



HAL
open science

Statistical physics of language evolution: the grammaticalization phenomenon

Quentin Feltgen

► **To cite this version:**

Quentin Feltgen. Statistical physics of language evolution: the grammaticalization phenomenon. Physics [physics]. Université Paris sciences et lettres, 2017. English. NNT: 2017PSLEE039 . tel-01753835v2

HAL Id: tel-01753835

<https://theses.hal.science/tel-01753835v2>

Submitted on 7 Jun 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE DE DOCTORAT

de l'Université de recherche Paris Sciences et Lettres
PSL Research University

Préparée à l'École Normale Supérieure

Statistical Physics of Language Evolution:
The Grammaticalization Phenomenon

**Physique statistique de l'évolution des langues :
le cas de la grammaticalisation**

Ecole doctorale n°564

PHYSIQUE EN ÎLE-DE-FRANCE

Spécialité PHYSIQUE DE LA COMPLEXITÉ

**Soutenue par
Quentin FELTGEN
le 11 octobre 2017**

Dirigée par **Jean-Pierre Nadal**
Co-encadrée par **Benjamin Fagard**



COMPOSITION DU JURY :

Mme HERNÁNDEZ Laura
LPTM, UCP, Présidente du jury

M. BLYTHE Richard
University of Edinburgh, Rapporteur

M. HILPERT Martin
Université de Neuchâtel, Rapporteur

Mme PRÉVOST Sophie
LaTTiCe, ENS, Examineur

M. TAILLEUR Julien
MSC, Université Paris Diderot,
Examineur

M. NADAL Jean-Pierre
LPS, ENS & CAMS, EHESS,
Directeur de thèse

M. FAGARD Benjamin
LaTTiCe, ENS, Co-encadrant

Résumé

Cette thèse se propose d'étudier la grammaticalisation, processus d'évolution linguistique par lequel les éléments fonctionnels de la langue se trouvent remplacés au cours du temps par des mots ou des constructions de contenu, c'est-à-dire servant à désigner des entités plus concrètes. La grammaticalisation est donc un cas particulier de remplacement sémantique. Or, la langue faisant l'objet d'un consensus social bien établi, il semble que le changement sémantique s'effectue à contre-courant de la bonne efficacité de la communication ; pourtant, il est attesté dans toutes les langues, toutes les époques et, comme le montre la grammaticalisation, toutes les catégories linguistiques. Dans cette thèse, nous étudions d'abord le phénomène de grammaticalisation d'un point de vue empirique, en analysant les fréquences d'usage de plusieurs centaines de constructions du langage connaissant une ou plusieurs grammaticalisations au cours de l'histoire de la langue française. Ces profils de fréquence sont extraits de la base de données de Frantext, qui permet de couvrir une période de sept siècles. L'augmentation de fréquence en courbe en S concomitante du remplacement sémantique, attestée dans la littérature, est confirmée, mais aussi complétée par l'observation d'une période de latence, une stagnation de la fréquence d'usage de la construction alors même que celle-ci manifeste déjà son nouveau sens. Les distributions statistiques des observables décrivant ces deux phénomènes sont obtenues et quantifiées. Un modèle de marche aléatoire est ensuite proposé reproduisant ces deux phénomènes. La latence s'y trouve expliquée comme un phénomène critique, au voisinage d'une bifurcation point-col. Une extension de ce modèle articulant l'organisation du réseau sémantique et les formes possibles de l'évolution est ensuite discutée.

Abstract

This work aims to study grammaticalization, the process by which the functional items of a language come to be replaced with time by content words or constructions, usually providing a more substantial meaning. Grammaticalization is therefore a particular type of semantic replacement. However, language emerges as a social consensus, so that it would seem that semantic change is at odds with the proper working of communication. Despite of this, the phenomenon is attested in all languages, at all times, and pervades all linguistic categories, as the very existence of grammaticalization shows. Why it would be so is somehow puzzling. In this thesis, we shall argue that the components on which lies the efficiency of linguistic communication are precisely those responsible for these semantic changes. To investigate this matter, we provide an empirical study of frequency profiles of a few hundreds of linguistic constructions undergoing one or several grammaticalizations throughout the French language history. These frequencies of use are extracted from the textual database Frantext, which covers a period of seven centuries. The S-shaped frequency rise co-occurring with semantic change, well attested in the existing literature, is confirmed. We moreover complement it by a latency part during which the frequency does not rise yet, though the construction is already used with its new meaning. The statistical distribution of the different observables related to these two phenomenal features are extracted. A random walk model is then proposed to account for this two-sided frequency pattern. The latency period appears as a critical phenomenon in the vicinity of a saddle-node bifurcation, and quantitatively matches its empirical counter-part. Finally, an extension of the model is sketched, in which the relationship between the structure of the semantic network and the outcome of the evolution could be discussed.

Acknowledgements

When I came up four years ago with the obsessive idea to work on language, I went to a ‘find your thesis supervisor’ event organized by the ENS and met Jean-Pierre Nadal. He was talking about his work on categories and neural networks, and at some point I interrupted him and said: “I would like very much to work on language. Would you agree to supervise a PhD on this topic?” He not only agreed, he also asked for two linguists to come, Benjamin Fagard and Thierry Poibeau, whom he knew were interested in empirical approaches on language. Fortunately, despite the vagueness of my wish, what Benjamin Fagard and Thierry Poibeau had in mind coincided exactly with what I really wanted, for it was the union, within the phenomenon of grammaticalization, of the two components I am really interested in: semantics, and historical change. Thus I wish to warmly thank Jean-Pierre Nadal for having so insightfully set the topic of my PhD. Yet he did much more along the years, and taught me a lot of things, including economy in scientific prose, tolerance for other hypotheses which we may not like but are worthy nonetheless, and diplomacy in the research world. Most wonderfully, he never ran out of patience, despite my frequent stubbornness, and my systematic tendency to incline a little bit too much towards the Humanities. I would also like to thank Benjamin Fagard, who was not officially my supervisor in title, but endorsed the role anyway as if he really were. We discussed a lot, and I wish I had found the time, the strength and the skill to explore all the possibilities he mentioned along the lines. There were numerous ideas I could not dig deeper in, or that were left aside without any further notice, and for this I am sorry. I learned from him many exciting and fascinating things about language and linguistics, and I have never ceased to be amazed on how open minded, tirelessly curious, and communicatively enthusiastic he may be.

I must also thank Mariana Estela Sánchez Daza, who closely followed this project, and provided me with an invaluable lot of advice, be it on the scientific or the personal planes, that I may not have always listened to, but which shaped anyway both the contents and the conduct of this work. I would have went on rampage in more than an occasion if she had not suggested me new perspectives and new ways to deal with the somewhat frustrating situations which abound in a PhD thesis; and discussing with her the concepts and the technical obstacles quite often helped me to accommodate the former and overcome the latter. To this long list I must also add that she proofread my thesis from start to end with great care and no less insight before I submitted it, which but deepens my abysmal indebtedness to her. I have also a most grateful thought for Tommaso Comparin, who have simply saved this thesis, and this is but an understatement. Any time I asked him a question on Python he answered me in full

detail, providing all possible help one could imagine; in truth I could not have done half of what I did without him. He convinced me that I should treat data from corpora in an automatized way, which ended up saving me months, if not many more. The scale on which the semantic change has been analyzed is thus due to the persistence, the dedication, and the profound generosity of Tommaso. My own mother Isabelle has also earned more than a mention, for she, too, proofread this thesis, hunting down with sharpness and application my grammatical mistakes. If the reader might find some that have escaped her notice, it is ascribable only to my own hiding them in convoluted formulations in a language which is not my own. It would also be unfair to mention my mother and forget my father Karl, who constantly pushed me forward when I was on the verge of giving up, providing me with the strength and the heart to face the occasional hardship of this work. Also, Physics is to me the most entertaining conceptual playground I can think of, and if I got to join the field, a choice I never had to regret, it is entirely thanks to my father. This never leaves my mind.

I also thank many researchers who devoted some of their time to have a word or a mail with me, either to help me deal with some issues of my own or to discuss further their own work: Christiane Marchello-Nizia, Tridib Sadhu, Sophie Prévost, Bernard Derrida, Michel Charolles, François Recanati, Ludo Melis, Eduardo Altmann, Igor Yanovich, Jérôme Michaud, Charlotte Danino, and the staff behind the Frantext database. In particular I thank Aleksandra Walczak, who tutored me since my first year of Master and up until the term of this PhD, and Thierry Poibeau, who always treated me with great kindness and respect, and very often sent me relevant and valuable papers or conference announcements. I am also in debt towards Vivien Stocker, for zealously providing me his irreplaceable insights on matters of Tolkienian scholarship, which he masters with the highest level of excellency, and Hélène Debrabandère, for all the help she abundantly offered me with the COHA and the study of *way too*, even though (and all the more so!) she disagreed frankly with my own preferred scenario of semantic change. I cannot forget either Michel Morange, who was my teacher in History and Philosophy of Science when I was still an undergraduate student. He showed his personal interest on the topic of this PhD and had the incredible kindness to listen carefully to the full presentation of my work. His advises and warm encouragements were precious to me, and they allowed me to go on at a time when my research was dismally stale. I also thank warmly all my jury members to have complied with the task of evaluating the present work, and especially my two referees, Richard Blythe and Martin Hilpert, who kindly and quickly agreed to come from afar to attend the defense.

My recognition go as well to the LPS, for hosting me in its halls. I have been provided with an excellent working environment, most pleasantly bathed in natural light, and a sumptuous desk that I could adorn with plants and book piles. I have also been equipped with a computer as well as a personal laptop, without which this research could not have been pursued with the same ease. I more personally thank the lab director, Jorge Kurchan, who always had an ear for our whimsy complaints, and the secretaries, for their dedicated help in all administrative matters, and most of all, their daily attention and kindness. I ought to mention in particular Annie Ribaudeau, for her trust, her care, her dedication, her constant willingness to make the lab a better place — and she certainly was successful at it. She supported me in all the ways she

could and my gratitude to her is of the keenest. I also thank the administrative staff of the EDPIF and the ENS, especially Jean-Marc Berroir, Jean-François Allemand, Vita Mikonovic, Stéphane Emery and Krisztina Hevér-Joly. They all contributed, with their personal touches, to the fulfillment of this work.

I have a warm thought for my companions of the E227d office: Tommaso Brotto, Marco Molari, Francesca Rizzato, Fabio Manca, Louis Brézin, Gaia Tavoni, Ivan Palaia, Caterina Mogno and, of course, Lorenzo Posani and his dramatic entrances. Anirudh Kulkarni has earned a special place, for we shared the same square meter or so, and it was a pleasing relief to just turn back and have a chat with him to ease the hard work, be it on books, society, politics, Carl Jung's personality test, mystics, astrological mysteries, or Wikipedia's fundraising campaign. I can hardly think of a better office mate. On top of this, he proofread the third part of this thesis (actually he did not proofread all of it only because I did not let him; he was so eager to help me that he offered to proofread the whole beast twice, which is eloquent of his amazing character). I must also mention those who helped me (or whom I helped) organizing the PhD and postdocs seminar of the lab, Manon Michel, Romain Lhermerout, Nariaki Sakai, and Samuel Poincloux.

On a more personal side, I have to end this long list of acknowledgements with a special mention of my students from the PSL undergraduate career. I could not dream of a better experience to start teaching. They were so incredibly smart, interested, agreeable, that I was always eagerly looking forward to my tutorial sessions with them. Thanks to Jean-Marc Berroir, I also had the great opportunity to teach tutorials in Statistical Physics during my last year, and with such a wonderful audience, it was truly a blast, and I enjoyed it immensely. Without a doubt, these were the best moments of my PhD experience, and I cannot be thankful enough to these two classes of astonishing and admirable students.

I acknowledge a three years grant from PSL Research University, which provided me financial support for the whole duration of this PhD thesis.

Contents

Résumé	i
Abstract	iii
Acknowledgements	v
Introduction	xv
I Grammaticalization in a phenomenological perspective	1
1 A linguistic framework: Construction Grammar	3
1.1 Definition	3
1.1.1 Ambiguity between constructions, paradigms, and collocation frequencies	4
1.1.2 Meaning of constructions	6
1.2 Constructions	7
1.3 Paradigms	9
1.3.1 Paradigm constraining	10
1.3.2 Paradigms at the intersection of competing constructional analyses	10
1.3.3 Speciation of paradigm members into independent constructions	11
1.3.4 External and internal paradigms	11
1.3.5 Paradigms and meaning	12
1.4 Collocations	14
1.4.1 Frequent collocations as actual constructions	14
1.4.2 Typology of collocations	14
1.4.3 Collocations and colligations	16
1.4.4 Asymmetry in collocations	17
1.4.5 Empirical dealing of collocations	18
1.4.6 Constraints on paradigms	20
1.4.7 Collocations and meaning	22
1.5 Meaning	24
1.5.1 Meaning in the semiotic view	24
1.5.2 Substantiating meaning	28
1.6 Summary	31

2	Grammaticalization as a phenomenon	33
2.1	Two examples of grammaticalization processes	33
2.1.1	Du coup	33
2.1.2	Way too	40
2.2	Theoretical discussion	49
2.2.1	Grammaticalization as a process	49
2.2.2	The dimensions of grammaticalization	51
2.2.3	The functional vault of language	55
2.2.4	Grammaticalization and construction grammar	64
2.3	The specificity of the grammaticalization phenomenon	75
2.3.1	Grammaticality as linguistic awareness	75
2.3.2	The mechanisms of grammaticalization	80
2.3.3	The scope of the speciation mechanism	87
2.3.4	A summary	89
II	Corpus study of grammaticalization	93
3	The S-curve: a signature of language change	95
3.1	First intuitions	95
3.2	S-curve in the sociolinguistic framework	98
3.2.1	The seminal paper by Weinreich, Labov, and Herzog	98
3.2.2	Lexical diffusion	101
3.2.3	Bailey's S-curve	104
3.2.4	S-curves based on synchronic evidence	105
3.2.5	S-curve and speakers' age	113
3.3	Empirical study of the S-curve	120
3.3.1	First corpus-based studies on the S-curve	120
3.3.2	The Constant Rate hypothesis	125
3.3.3	Further evidence from corpus data	132
3.4	The S-curve challenged	137
3.4.1	Theoretical challenging	137
3.4.2	Other templates for change	140
3.4.3	A Lotka-Volterra model of competition	143
3.4.4	Information encoded in the S-curve	144
3.5	Summary	145
4	Extracting patterns from frequency profiles	149
4.1	Extracting the S-curve from frequency data	150
4.1.1	Smoothing data	151
4.1.2	Identifying growth periods	152
4.1.3	Extracting the S-curve	152
4.1.4	Quality of the fit	154
4.1.5	Limitations and alternatives	157
4.1.6	Improving the extraction procedure	158
4.1.7	Summary	163

4.2	The Frantext corpus	164
4.2.1	Presentation of the database	164
4.2.2	Alternatives to Frantext	169
4.3	Four hundred instances of semantic change	174
4.3.1	Why focus on semantic change?	175
4.3.2	A note on French	182
4.3.3	How were the forms chosen?	183
4.4	Qualitative results	185
4.4.1	Case studies: <i>du coup</i> , <i>way too</i>	185
4.4.2	Cases of interest	188
4.4.3	S-curves in other linguistic domains	195
4.5	Statistical survey	199
4.5.1	How generic is the S-curve pattern?	199
4.5.2	Distributions	202
4.6	The full pattern of semantic change	216
4.6.1	A few examples of latency	216
4.6.2	Extracting latency	224
4.6.3	Distributions and correlations	224
4.6.4	A specificity of semantic change?	228
4.6.5	Robustness of the results	228
4.7	Summary	233
5	Paradigms and constructions	235
5.1	Co-evolution of forms within a paradigm	235
5.1.1	Detecting correlations	236
5.1.2	Preliminary test	237
5.1.3	A case study	240
5.1.4	Summary	253
5.2	Diachronic networks of co-evolution	255
5.2.1	Internal organization	256
5.2.2	Identifying competitors	268
5.3	The Constant Rate Hypothesis revisited	272
5.3.1	En plein	273
5.3.2	Idiosyncratic uses of <i>dans</i>	276
5.3.3	What shall we conclude?	279
III	Modeling semantic expansion	285
6	Previous works in language change modeling	287
6.1	Macroscopic models	288
6.1.1	The Abrams-Strogatz model of language competition	288
6.1.2	Macroscopic models in lexical diffusion	289
6.1.3	Internal and external factors of change	290
6.2	S-curve and learning	292
6.2.1	An early model of language change	292

6.2.2	Bayesian learning of language	293
6.3	Sociologic accounts of the S-curve	296
6.3.1	The Threshold Problem	296
6.3.2	Network structure	298
6.3.3	Social networks of influence	299
6.4	The Utterance Selection Model	300
6.4.1	The original model	301
6.4.2	Application to a linguistic case	303
6.4.3	The search for the S-curve	304
6.4.4	Another way to the S-curve	308
6.5	Another model of semantic change	310
6.5.1	Communication event	310
6.5.2	Semantic change of a lexical unit	311
6.5.3	Semantic replacement	311
6.6	Summary	313
7	Hypotheses of the model	315
7.1	Language change as a product of language use	315
7.2	Language as memory	318
7.3	The semantic territory	319
7.3.1	Non-overlap of meanings across a semantic replacement	319
7.3.2	An epistemologic perspective on the semantic territory	320
7.3.3	Empirical cartography of the semantic territory	322
7.3.4	Vectorial representation of semantic units	324
7.3.5	Summary	325
7.4	The semantic territory as a conceptual network	326
7.4.1	A context-based view of grammaticalization	327
7.4.2	Contexts and semantic units	329
7.5	The links of the network	332
7.5.1	Pragmatic inferences and implicatures	333
7.5.2	Metaphors	333
7.5.3	Analogy	334
7.5.4	Priming	336
7.6	Communication is efficient and successful	339
7.6.1	Communication errors	340
7.6.2	Mean field hypothesis	342
7.7	Summary	343
8	A model of semantic expansion	345
8.1	Presentation of the model	345
8.1.1	A random walk in frequency space	346
8.1.2	Deterministic approximation	351
8.1.3	Stochastic behavior in the critical region	357
8.2	Theoretical and numerical analysis of the model	360
8.2.1	First passage time	360
8.2.2	Trap region	365

8.2.3	Are the trapping time and the latency time really the same? . . .	371
8.2.4	Distributions	377
8.2.5	Summary	383
8.3	Model variants	384
8.3.1	Hearer perspective	384
8.3.2	Combined perspective	388
8.3.3	Different interpretations of the γ parameter	391
8.4	Discussion of the model	395
8.4.1	Understanding the phenomenon of semantic change	395
8.4.2	A near-critical behavior	399
8.4.3	Sociologic interpretation	403
8.4.4	Limitations	404
9	Semantic bleaching and other network effects	407
9.1	Pervading the network	408
9.1.1	Basic setup	409
9.1.2	A greater leap in frequency	411
9.1.3	A smoother S-curve	412
9.1.4	A theoretical approach	425
9.1.5	The growth distribution	431
9.2	Semantic bleaching	434
9.2.1	A conceptual picture of semantic change	434
9.2.2	The meaning of forms	436
9.2.3	Semantic bleaching and network propagation	441
9.3	Summary	447
	Conclusion	449
	A List of studied semantic expansions	457

Introduction

There is no true consensus regarding the emergence of language. In any case, language should be a trait acquired by our species through a series of random mutations, which happened to have been selected because it provided the speech community with somehow better surviving odds. The question that is subsequently asked regarding this matter is then: what possibly is the evolutive advantage associated with language? An interesting array of theories have been offered on this topic; to mention but one (Victorri, 2002), language was selected because it enabled to tell stories. This skill can occasionally come in handy: when two members of a tribe start fighting, threatening to plunge the whole community into a raging murderous feud, a wise, old member of the clan could remind everyone how disastrous it turned out to be the last time, and how much better it would be for everyone to just quiet down and work out their differences peacefully, without any unwarranted and fratricide bloodshed.

Though this little story offered by Victorri (2002) is certainly too much of a made-up, it conveys the crucial idea that the usefulness of language lies in its power of elicitation. In the same spirit, Chafe (1994), building upon a quote from Hockett, has defended the idea that language infuses our species with the ability ‘to capture and communicate thoughts’ which ‘has nothing to do with what is going around’ (Chafe, 1994, p.4). Actually, this evocating power of language even lay at the core of a decades-long scholastic debate regarding the true magical power of incantations. Nicole Oresme (c. 1320/1325 – 1382), famous for having set up the philosophical arguments in favor of a Copernician view of cosmos, rationally defended the idea of a natural, magical power of language, for it presents the ability to arouse imaginations in the human mind (Delaurenti, 2007). The long list of demented minds imputable to the reading of Alhazred’s treaty of witchcraft and black magic would also certainly speak in favor of this strong eliciting capacity of language.

If we do not have any firm theory regarding why language emerged and what it is actually used for (which is, to some extent, a similar topic), it is no wonder if there is no certainty either on the matter of language change. Not that the field has to deplore a lack of convincing ideas and theories on the matter, but none of them can pretend having reached a full and sound consensus. One can divide the different theories between those where change is language-external, and those where it is language-internal (Labov, 1994). A language-external account may be for instance that people have an incentive to change their way of speaking so as to mark their social identity (Croft, 2000), or that language changes because of an imperfect transmission from one generation to another (Niyogi, 2006). It may also be that social, political, and especially cultural changes call for language innovations, which perturb the system and

ripple through the entire linguistic organization (Meillet, 1906) — a possibility ruled out in Lass (1997). It has also been proposed that language changes because ‘there is in the mind of man a strong love for slight changes in all things’ (Darwin, 1871, p.60). On the language-external side, it has been proposed that language changes for the better, so that it keeps improving from one state of its evolution to the other (Müller, 1870; Farrar, 1870; Darwin, 1871). From what people casually say about language change, this is not the general impression of the speakers (Aitchison, 2013). Another idea would be that language is constantly repairing itself, because, as it is spoken, it gradually wears off: the evocative power of words is doomed to erode as they are uttered more and more. Used elements then have to be replaced (Müller, 1866; Geurts, 2000).

In this thesis, I will defend the idea that the cause of language change lies precisely in what language seems to be used for, that is, in its strong evocative power. For a representation to be efficiently summoned, it would be necessary for meaning not to be clear-cut or precise; it needs to be diffuse. Only through a rich cognitive structure of complex and numerous associations underlying the frame of the semantic territory can language achieve its narrative, imaginative purpose. One needs not, obviously, to believe that there has to be a ‘purpose’ of language (we strongly do not) to accept the idea that, because of this evocative power, the elements of language are likely to be able to elicit meanings, so as to inflate a representation on the basis of a limited number of components. There would therefore be a natural tendency for the meaning of a linguistic unit to expand whenever used. As we shall argue, this could be a sufficient mechanism to trigger occasionally a phenomenon of semantic change.

Yet, not all language changes amount to semantic change. Assuredly, phonetic change could not be explained by such a ‘diffusive meaning’ picture, although there could be similar phenomena in the phonemic space (Victorri, 2004). But most of all, syntactic changes, grammatical changes, arguably more crucial as they would impact the overall structure of a given language, do not fall into the category of semantic change. For instance, there is nothing in common between the process of semantic change which led the word *thing*, originally referring to an assembly in early Germanic and later Scandinavian societies within which judiciary and legal matters were settled, to its contemporary meaning of a loose pragmatic designation of any conceivable entity, and the process by which the paradigm of determiners emerged in French. One is an etymological curiosity, the other is a deep grammatical matter, affecting almost all possible utterances in the concerned language, and would certainly deal with the most intricate linguistic technicalities. The two are clearly better kept separate.

Or is that really so? Actually, this is where grammaticalization comes into the picture. Grammaticalization is a notion coined by Meillet (1912), who described it as one of the two processes by which the grammatical forms are renewed. The first one would be analogy (recruitment of new members in a grammatical paradigm by a calque of existing members’ constructional features), the second, grammaticalization, or how ‘an autonomous word endorses the role of a grammatical element’. Interestingly, the first example which is given by Meillet is that of *suis* in French (‘am’), which can be used as a content word (‘I am that I am’) or as an auxiliary. An English instance of it would be the [Be V_{ing}] construction, as in: ‘I am leaving no guard, because the vultures will not approach as long as anyone is near, and I do not wish them to feel

any constraint.’ (HOWARD, Robert, *A Witch Shall be Born*, 1934, Project Gutenberg Australia).

What this particular linguistic example shows is the continuity which exists between lexical and grammatical elements. This continuity is both synchronic (a polysemic word can have both lexical and grammatical meanings) and diachronic, since the etymological origin of a grammatical element can be a lexical one. For instance, *like*, as in ‘Thence up he flew, and on the tree of life, the middle tree and highest there that grew, sat like a cormorant;’ (MILTON, John, *Paradise Lost*, 1674), came from *lic*, ‘body’, through the intermediary of the (hypothetically reconstructed) construction *gelic*. Note that it has also developed a morphemic character through the reduced form *-ly* attached to adjectives to form adverbs, e.g. *fugitively*, a process repeated nowadays with nouns, especially proper ones, used to define a class according to its prototype (e.g. ‘doom-like game’, ‘Blair Witch-like movies’) — a construction which is also very productive in French, despite its English origin. Interestingly, grammatical elements can conversely be the source of lexical ones: the verb *like*, ‘to enjoy’, descends from the same origin.

Thus, grammatical elements can come from lexical elements, and semantic change has something to do with grammar, after all. But what about syntax? Syntactic change could be seen as a grammaticalization phenomenon, although this is not a consensual view in the study of this notion. This is especially the case if we consider the framework known as Construction Grammar, according to which there is no separation between the lexicon, i.e. the inventory of linguistic items, and syntax, i.e. the set of rules which allow to combine these items meaningfully to form a proper utterance. According to Construction Grammar, all language elements are constructions, from the more complex and schematic to the more atomic and simple ones, and belong to a single, unified, ‘construction’ (Hilpert, 2014). Therefore, syntax is just a set of argument structures constructions entrenched in language use, with specific behavioral and semantic features just as would be the case with any other constructions.

If we see semantic change as the broader process by which the features of a given construction are altered, then most language changes can be seen as instances of semantic change. Furthermore, we know that grammaticalization exists, that is, that an originally simple construction, as would be a lexical element, can become involved in more schematic ones, just as the lexical ‘am’, a lexical verb in its own due right, has come to be involved in the very grammatical [Be V_{ing}] construction. In that sense, the picture that was offered earlier, that language is prone to change because the meaning of its components can leak through the many links of a large and complex semantic network, could be a suitable frame for grammaticalization as well. We can ask, at the very least, whether grammaticalization would require any additional mechanism to occur compared to any regular semantic change.

Now, this explanation of change would fall within the class of the ‘Invisible Hand’ type of explanation (Keller, 1989), according to which the large-scale event of a language change would be the result of a collectivity of unwillingly concurring events which, isolated and independently, are not particularly related to the larger picture. In this case, it would be the succession of the events of communicative use of a linguistic form which would shape with time its semantic mutation. The problem with this kind of Invisible Hand explanation is that they are, basically, theoretical black

boxes. It assumes some process at the individual level (here the level of each utterance event), and then hopes that everything will conspire towards the intended result. As it happens, to ensure that the conspiracy is actually turning out the right way is precisely the aim of Statistical Physics.

The ultimate goal of this thesis is therefore to provide an analytical framework in which semantic change can be described, modeled, and discussed. Especially, if we want to eventually understand how a process such as grammaticalization can actually work, we first have to set the foundations of a more general model of semantic change, so as to ask which features of such a model could be exclusive to grammaticalization, or what is lacking to account for its occurrence. It may come up as a disappointment, but we shall not, in the end, provide a model of grammaticalization. We are merely offering a theoretical picture, grounded in a model through which we can test different sorts of hypotheses, which we hope will provide a way to discuss the phenomenon of grammaticalization with greater clarity and precision.

The model is a stochastic random walk on a network, describing a competition between two linguistic variants. As such, there cannot be a one-to-one comparison between a single outcome of the model, and a given instance of linguistic change. It can only answer two sorts of question: do the model outcomes present the same qualitative features as the actual changes? Do their statistical properties match quantitatively? If so, then the model successfully relates the proposed underlying mechanism to the intended outcome. This model can furthermore have predictive power if it comes with interesting regularities which could be investigated empirically, but it has none with respect to individual cases of language change.

The problem is that the qualitative features of a semantic change which an analytical model can describe are extremely few. It more or less amounts to the following: language change occurs according to an S-curve, which means that the rise of frequency of the new variant should obey a sigmoid function, or any other function the shape of which is sensibly similar. As for the statistical properties of language change, it would appear that the existing literature has mostly left them aside of its focus. Our work will thus follow two main steps: first, to establish empirical facts of semantic change; second, to describe, reproduce, and explain them through a mathematical model.

Before outlining the contents of this thesis, two major apologies are in order. The first one is that this is a PhD thesis in Physics, written by a student who received an academic formation in Physics and nothing else. Although it might seem at times, to the physicist reader, that this work deals more with Linguistics than Physics, it should be clear, for the linguist, that this is not the case. An example of this is that I use the words ‘qualitative’ and ‘quantitative’ in the physicist’s way, not the linguist’s one: a qualitative description refers to the mere appearance of things, while a quantitative one represents a step forward, as it provides a mathematical account of it — roughly speaking, it gets numbers. I am well aware that, in Linguistics, a qualitative analysis is a detailed one, while a quantitative one is less precise and polished, but usually encompasses a greater amount of cases.

As a result, and though I tried my best to acquaint myself with the rich literature on grammaticalization, diachronic construction grammar, corpus studies, and semantic change, it will be blatant that my knowledge on these matters is partial, in both senses of the term, for I have picked up only what I understood, and found convenient. I

nevertheless took the liberty to profess some of my personal views occasionally but it might be more of an enthusiastic chitchat than of a rigorous and scientific exposition of new theoretical ideas. I therefore pretend not that this work would bring any significant contribution to these fields, except hopefully by providing empirical facts and new tools of data analysis which may prove respectively enlightening and useful. At least, I made a point not to provide any example which would not be attested in actual instances of language use: there is not a single illustrative piece of language in this work which could be said to have been made up.

The second one is that I am fully aware that this thesis is far too long, and is about twice the size of a regular PhD thesis in Physics. The thing is, this work is addressed to two different audiences, two separate traditions of research, and I felt that I had to provide a decent amount of contents for both of them. The idea was, if one were to subtract everything which would not be clearly and undeniably related to Physics, one would still get a thesis of the regular size. Consequently, some chapters are more Linguistics-oriented, while some others are more exclusive to a physicists readership. Therefore I have hope that each audience will find a reasonable quantity of results to discuss, criticize, and ponder. I had yet to present my deepest apologies to the reader for the extent of this research account. As a partial remedy, the subsequent outline shall describe precisely what is to be found in each chapter, so that one is not bound to read them all to follow the line of reasoning pursued throughout this work.

Outline The first chapter presents a view of Construction Grammar, a linguistic framework that we shall adopt to discuss the different phenomena. As such, it is in no way crucial to the understanding of the remainder of this thesis. Construction Grammar states that language is made of constructions, and that an utterance (i.e. a string of linguistic items, attested in use, which combines the meaning of its components in a whole) is a hierarchical assembly of constructions of various degrees of schematicity. The simplest construction is an autonomous word (e.g. *donut*), a more schematic one would be the $\{N\}$ of $\{N\}$ possessive construction (e.g. ‘The History of the Duffle Coat’). This linguistic framework is extremely adaptive, and is driven by the linguistic evidence, which makes it efficient to account for diachronic semantic and behavioral changes of various linguistic elements. In this chapter, we discuss further the notion of paradigms, which are sets of constructions which can be substituted to one another. For instance, schematic constructions have slots to be filled; and not any construction is likely to enter these slots. The possessive construction we have just seen is extremely likely to be used with two nouns; therefore, it defines two related nominal paradigms (noted $\{N\}$). Not all pairs are equally expected either; for instance, ‘the lair of the beast’ is overwhelmingly more frequently attested than ‘the beast of the lair’. We argue that the apprehension of these paradigms is part of the knowledge of a language. The notion of collocations, i.e. the repeated co-occurrence of two neighboring constructions (as would be *fait* and *a fight* in ‘a fair fight’) will be discussed as well, including its relationship with the previous notion of paradigms. The crucial topic of meaning will be slightly touched upon, but not deeply explored.

The second chapter continues the overview of linguistic concepts with the notion of grammaticalization. We exemplify this phenomenon with two detailed examples. The first one is the French *du coup*, which became a common conjunction from the literal

meaning ‘as a result of the strike/the blow’. It can, for instance, replace the ubiquitous *donc* (‘so’) in some of its uses (e.g. ‘Tu vas faire quoi du coup?’, ‘So what are you going to do?’). The second one is the English intensifier *way too* as in ‘all too often, the Star Trek future seems way too rosy and sweet’. We shall then review different theoretical attempts to clarify the distinction between the lexical and the grammatical elements of language, which lies at the very core of the definition of grammaticalization. It will be proposed, as a working definition, that a linguistic element is grammatical as soon as it shows some sensibility to the other elements of the utterance, and acts upon them to constrain their meaning. In this perspective, adjectives should be seen as grammatical already. We will also discuss the specific character of the grammaticalization phenomenon, as it has often been argued that grammaticalization is an artificial subset of more general processes of language change, and that only its result is specific. We defend the opposite position, considering that grammaticalization is specific, and that, to occur, it may rely on several specific mechanisms. We suggest, however, that there might not be a single mechanism behind all grammaticalizations, but that different grammaticalizations could rely on different mechanisms. This would open the perspective towards a classification of grammaticalization cases.

As our objective is to deal empirically with language change, we consider next the most consensual qualitative signature of language change, including grammaticalization, which is the S-curve. Actually, there is no overall agreement to what the S-curve does actually represent. We shall review some of the positions held on this matter: that the S-curve describes the numbers of speakers having adopted a given linguistic variant over time; that it represents the evolution of the frequency of use of this variant; or that it represents the number of affected elements by an ongoing change (paradigmatically, words in a phonetic change). The S-curve needs not even be an S-curve in time, and can describe a population at a given moment of time, showing that the more peripheral members have adopted a variant more (or less) than the central members did, be it in terms of age, or in terms of spatial repartition. We will also review the actual attempts to empirically establish this pattern, the most important one being that of Kroch (1989b), who quantified the S-curve thanks to the logit transform of the data: assuming the data is described by a hyperbolic tangent going from 0 to 1, the logit transform would give a straight line as a result. Hence, a linear fit of the logit transform of the data allows to assess the relevance of the S-curve fit and to obtain its descriptive parameters. Kroch (1989b) formulated the Constant Rate Hypothesis, according to which the rate of the S-curve (the slope of the logit transform) should be, for a given change, identical in all contexts of change. We shall often make reference to this hypothesis, as well as to its rival, proposed by Bailey (1973), stating that, the later a context is affected by a change, the higher will be the associated rate. As this chapter is essentially a bibliographical review on the specific topic of the S-curve as an empirically attested pattern of change, it can be skipped entirely.

The fourth chapter constitutes one of the two main contributions of this work. We present an empirical, automatized procedure to extract S-curve patterns from a frequency profile. Then, we present a large-scale study of the S-curve, and the statistical distributions which can be derived from a few hundreds of instances of semantic change. These examples are drawn from French. The frequency data is obtained from the Frantext database created by the Atilf laboratory, and available

under subscription. We discuss at length the advantages and the limitations of this particular database, comparing it with other existing databases. We also ensured that our procedure could work on the data from other corpora, such as the COHA (Corpus of Historical American English). The distributions of both the growth times and the rates are not trivial, and are best fit by an Inverse Gaussian, among the different distributions we tried on. Furthermore, we provide evidence for the existence of a latency part of the pattern, preceding the S-curve: while the new meaning is already attested in the corpus, the frequency does not change, and will rise only after a period of time of random duration. This latency time actually proves to follow an Inverse Gaussian with a great precision, hinting that it is indeed a valuable and interesting phenomenon. We argue as a conclusion that this latency feature could be more specific to semantic change, and would constitute empirical evidence exclusive to language change, contrary to the S-curve which is found in various kinds of social, cultural and biological diffusions.

The fifth chapter also presents original research on the empirical characterization of language change. We especially tried to extract as much information as possible from the frequency profiles of different members of a given constructional paradigm. Specifically, we study the paradigms of quantifiers [un {N} de], the paradigm of pseudo-adverbials [par {N}] (*par surprise*, *par hasard*, etc.), the paradigm of the collocations of *en plein* ('in the midst of'), and the paradigm of idiosyncratic [dans {article} {N}] constructions. We investigate the correlations between different members of each paradigm and try to understand more closely the details of their individual evolutions. Moreover, we make use of these correlations to build networks of paradigm members, showing both correlations and anti-correlations between members. It appears that language change is made more complex by the fact that the element of language undergoing change may not be a simple linguistic form, but a cluster within a paradigm, for instance. Also, changes can be paradigm-internal (reorganization of the weights and the roles of the members) or paradigm-external (loss of paradigmatic features due to an external competition). Therefore, identifying the two members of a competition event (as semantic changes are assumed to be) is a most delicate matter. Finally, we also test the Constant Rate Hypothesis and its rival, the rate acceleration law. It appears that the Constant Rate Hypothesis can be useful to argue whether a change involves a paradigm as a whole (e.g. a schematic construction) or some of its individual members. We failed however to test efficiently the rate acceleration law, because the identification of 'contexts of change', in a semantic expansion process, is far from obvious. Indeed, there is an interplay between contexts and semantic features, so that the nature of the change itself is related to the contexts in which it occurs.

The last part of the thesis provides a usage-based model of language change, which is, more precisely, a random walk in the space of frequency of use. In chapter 6, we present a bibliographical overview of different physics-based models of the S-curve. We classify the different approaches into five different categories. The first one is that of macroscopic models; they usually rely on a simple, deterministic differential equation, which solution under the proper initial conditions is an S-curve. The second class of models explains language change as a result of language learning by subsequent generations. Most models of the S-curve, however, fall into the third class, which groups the agent-based models of social diffusion. It appears that the propagation of

a new variant crucially depends on the network structure of the speakers' community. Also, a bias in favor of the novelty is almost always necessary to account for its rise. Otherwise, it cannot challenge the entrenched convention. We then spend some more time on the Utterance Selection Model (Baxter et al., 2006), for several works in this line have specifically targeted the modeling of the S-curve. Once more, it appears that either a specific network structure or a bias in favor of the novelty are necessary to account for the propagation. One variant of the Utterance Selection Model proposes an original mechanism which does not rely on a preset bias, but on perceived momentum: speakers are sensitive to the fact that the use of some word has increased in the recent past as compared to the long-term past, and are expected to favor it accordingly. Therefore, an effective, non a priori bias induces the rise of the new variant. Finally, we present a model which does not address the S-curve, but nevertheless provides an enlightening view of semantic change. In this model, semantic change happens because the intended meaning is always diffuse, as is the meaning of any given word. As a result, different meaning domains overlap and evolve with time to adapt to the intended use of the words. Interestingly, it relies on the same cognitive hypotheses we already put forward in this introduction, and it is due to Victorri (2004), who professed the 'narrative'-based view of language origin we presented at the very beginning of this work.

In the next chapter, we present the cognitive and linguistic hypotheses on which our model is grounded. Notably, we discuss the complex and intricate interplay between constructions, contexts, and meanings, for the three are not formally distinct entities. If we want to model semantic change as a competition between two linguistic forms for a given context (or meaning), it must be clear that it is an idealization, the scope of which can nonetheless be quite large. We also present the idea that the semantic territory (the space of meaning) can be described as a small-world network, following numerous past works in that sense. Therefore, a semantic expansion could be described as the invasion by a given species' population into a new domain of the semantic network. The frequency of use of the species (the linguistic variant) can be directly related to its occupation of the network. We give it a cognitive interpretation, by stating that this network is actually entrenched in the memory of speakers. Since speakers have mnemonic limitations (they can only remember a finite number of occurrences), the size of the different semantic domains that a species can invade should be finite as well. This will prove crucial in our model. We also defend the hypothesis that communication between speakers could be approximated as perfect, and that a mean-field like, representative agent approach of language change, is legitimate. Therefore, our model is not agent-based.

Chapter 8 is the main contribution of this thesis. We propose a model of semantic expansion to retrieve the S-curve and latency pattern. Drawing inspiration from the Utterance Selection Model, and relying on attested mechanisms of change in the grammaticalization literature, we consider a minimal network of two sites, corresponding to two different meanings, asymmetrically related. Each site is populated by the same fixed number of occurrences. These occurrences are of two types, corresponding to two competing linguistic variants. At the beginning, each site is monospecific, i.e., it is populated by occurrences of only one of the two linguistic variants in competition. Moreover, one site is assumed to influence asymmetrically the other. At each time

step, one of the two sites (meanings) is chosen to be expressed. To express it, one of the two variants is retrieved from memory according to a probability function which depends on its effective frequency. This effective frequency replaces the true frequency of the variant (which would be the ratio between its associated occurrences and the memory size) and accounts for the contents of the influencing neighboring sites. That way, one of the two variants can start to invade the semantic dominion of the other. In a linguistic perspective, the inference leading from the source meaning to the target meaning is therefore a pathway for the semantic turf of a linguistic form to expand.

The model can be approximated by a dynamical system which shows a saddle-node bifurcation according to the value of the control parameter, here the strength of the inference/influence of the reservoir site on the invaded one. Below a critical threshold, the implantation of the new variant is only marginal; above, the new variant takes over the site and entirely evicts its former competitor. Furthermore, the non-linearity of the production probability function leads to an S-shape trajectory of growth in that latter case. What is more, in the vicinity of the threshold, the dynamics slows down infinitely and a latency part appears. If we now consider the full stochastic model, we can compute the statistical distributions of all relevant quantities associated with the frequency rise pattern: the duration of the S-curve, its rate, and the duration of the latency. It appears that the model closely captures the quantitative empirical features of the latency. Also, it predicts correctly the correlations between the different quantities. However, it fails at providing the right prediction regarding the growth time: the growth is too fast compared to the latency, and its variance too low.

We believe, however, that these shortcomings of our model can be overcome. The discrepancy between the large and variable empirical growth and the short, almost deterministic one that is described in our model, is certainly due to the fact that we only considered the domination over one site, whereas it is likely that the competition process would take place on a more extended and elaborated network. As such, the diffusion over the network would present a longer growth time, and a more variable one. In the last chapter of this thesis, we provide a few research perspectives in that sense, so we do not have any firm results to present. In this chapter, we also present an interesting extension of the model, adding a representation of meaning. This allows the model to account for the phenomenon of semantic bleaching, the process by which a linguistic form, as it undergoes a semantic expansion and all the more so if this expansion falls into the scope of grammaticalization, loses semantic focus and specificity, so that it becomes harder and harder to define and discriminate its actual meaning. What would be, for instance, the meaning of *at*? This interesting phenomenology of our model inclines us to think that it can indeed serve as a useful sandbox to discuss with greater clarity and efficiency the complex and intertwined processes at the core of grammaticalization.

Note on linguistic examples The linguistic examples have been extracted from various sources: Google queries, Google Books, COHA (Corpus of Historical American English), the TIME corpus of Mark Davies, Project Gutenberg, Frantext, and a few additional books which were available to me whenever needed. All translations of French examples into English are mine, unless stated otherwise.

Part I

Grammaticalization in a phenomenological perspective

Chapter 1

A linguistic framework: Construction Grammar

In this work, we adopt a linguistic framework to describe and to interpret the linguistic phenomena that we aim to model. The framework that seemed the most convenient to track semantic change through diachronic data is the one known as Construction Grammar, which has met an increasing success for the last thirteen years. The goal of this chapter is neither to provide an accurate bibliographical review nor to present a wide overview of Construction Grammar, but to lay down the few basic notions that shall be needed later on. We may eventually discuss some of the issues that they raise, yet it shall remain superficial, for it must be reminded that the aim of these discussions is not to contribute to the Construction Grammar theory, but to give depth to the definitions of the concepts that shall prove crucial to the understanding of change.

1.1 Definition

Quite naturally, Construction Grammar is first and foremost preoccupied with constructions. The notion of construction in the Construction Grammar framework is actually very specific. The term itself can be misleading, for it may be understood to refer to an elaborated building of some sort out of various constituents, while on the contrary, a construction is something entrenched, individualized, and seen as independent from the various materials it makes use of. It can be defined as follows:

Construction. *A construction is a conventionalized mapping from a cohesive form to some meaning domain, where conventionalized implies that it does not result from the processing of a given utterance in a specific context.*

A construction is not necessarily entirely fixed in form, but can be a flexible pattern as well (e.g. the French imperative comes from a particular pattern filled by verbs, so that the form associated with the ‘imperative’ meaning is not substantiated by anything but this pattern), or a mix between the two (e.g. the construction [in {X} way {ADJ/ADV}], where X belongs to the paradigm of determiners {a, this, that, no, any, every, some}, means that the adjective or adverb applies to what comes before with a certain modality X). It should be stressed that in this framework, all syntactic

patterns are seen as constructions as well. Words themselves can also be, and not be, constructions; the word [word] is a construction, while the word *soulless* is not, as it is made from two underlying constructions, [soul] and [$\{N\}_{\text{less}}$].

By ‘does not result from the processing of a given utterance’, we aim to highlight that a construction is processed as such, not as a computation from its components. This may be because it is non-compositional, or because speakers have memorized it as a unit. Throughout this work, we will represent by the sign { } an empty slot, to be filled by an appropriate linguistic form, while the brackets [] represent a construction (e.g. the genitive construction [$\{N\}$ ’s $\{N\}$]). A construction whose arguments are filled with specific constructions is called a construct (e.g. *soulless* is a construct, and ‘A beginning is the time for taking the most delicate care that the balances are correct.’, the opening sentence of Herbert’s *Dune*, is also a construct, of many different intricate constructions). Finally, the constructicon is the inventory of all constructions, be it in a random speaker’s mind, or in a grammarian description of a given language.

In this work, we would not go as far as to say that there is only one type of items in language, which is the construction (Hilpert, 2014), because there are also specific paradigms or categories which may be needed to state the specific array of uses of a given construction. We would therefore distinguish the following three components: constructions; paradigms (i.e. specific sets of constructions) or categories (what Traugott and Trousdale (2013) call “schemas”); and collocations (co-occurrence frequencies). Meaning should be considered as well, but we don’t consider it as a “constituent” of a language. By collocations, we mean that language users are accustomed to see certain groupings of words — e.g. you can end most sentences once you heard them out halfway. This kind of linguistic knowledge is part of what language *is* in the mind of language users, and so should be considered as a part of language as well. Of course, a high collocation score can lead to the entrenchment of a construction, so that there are no clear-cut boundaries between these objects; grammar stems at least largely from frequency effects, as Bybee theorized (Bybee and Thompson, 1997; Bybee, 2006, 2010). Paradigms are also likely to emerge from repeated collocations in specific constructions.

Paradigms and collocations correspond closely to the paradigmatic and syntagmatic axis of de Saussure (1995). It implies that a construction takes its meaning from both the paradigms it belongs to, and the other constructions it collocates with. This dual view is illustrated by the two semantic networks constitutive of the Semantic Atlas (Ploux et al., 2010): one way to build a network of semantic “cliques” is to make use of lexical synonymy, another is to use collocations.

1.1.1 Ambiguity between constructions, paradigms, and collocation frequencies

Contrary to Traugott and Trousdale (2013), we do not consider paradigms to be constructions; the nominal paradigm, for instance, has no clear form, and no clear meaning. Interestingly, constructions have two major kinds of relationships with paradigms: They can take them as arguments (e.g. French [$\{N_{\text{count}}\}$ après $\{N_{\text{count}}\}$]), which indicates a temporal series marked and rhythmized by a succession of N_{count} , like *jour après jour*:

Carotte après carotte, tomate après tomate, taro après taro, chou après chou, il a

rempli sa marmite.

Carrot after carrot, tomato after tomato, taro after taro, cabbage after cabbage, he filled up his cooking pot. GUIRAO, Patrice, *Crois-le !*, 2013 (Google Books)

Or they can as well belong to them (e.g. the previous construction belongs to some kind of “processual temporality” paradigm ; it can be replaced by *lentement*, *peu à peu*, etc.).

There can be also an ambiguity between a construction and a paradigm; for instance, we can propose a construction [$\{N_{\text{count}}\} \{\text{Prep}\} \{N_{\text{count}}\}$], with $\{\text{Prep}\}$ being the prepositional paradigm $\{\text{après, à, par}\}$, so as to account to a wider variety of attested constructs : *goutte à goutte*, *pas à pas*, *heure par heure*, *centimètre par centimètre*, *année après année*, etc. Yet we could also consider that all three constructions are independent (for instance, they are not compatible with the same sets of nouns) and posit instead a paradigm of constructions $\{\{\{N_{\text{count}}\} \text{par} \{N_{\text{count}}\}\}, \{\{N_{\text{count}}\} \text{à} \{N_{\text{count}}\}\}, \{\{N_{\text{count}}\} \text{après} \{N_{\text{count}}\}\}\}$. The only difference between the two cases is that in the former (construction of paradigms), all the constructions [$\{N_{\text{count}}\} \{\text{Prep}\} \{N_{\text{count}}\}$] share a common meaning, while in the latter (paradigm of constructions), each construction should be associated with a different, specific meaning.

However, this is rendered all the more complicated as the meaning of a construction is not a clearly delineated object. Indeed, most constructions are polysemic; in a paradigm of constructions, it is thus expected that the meanings of the different constructions belonging to the paradigm overlap to some extent. In chapter 5, we will discuss of a way to disentangle these two possibilities using an empirical approach, relying on the diachronic evolution of the different constructions or paradigms.

A similar confusion arises when it comes to construction and collocation. Is a frequent collocation a construction? For instance, the Late Middle French *tout de suite* meant ‘all in a row’, which can be seen as a collocation between the two constructions *tout* and *de suite*, as it would be in English. Yet, in Modern French, this construction came to be associated with a non-compositional meaning, ‘immediately, right now’. In this example, it appears that a frequent collocation has led to a construction. At a given state of any language, it is thus not always easy to distinguish between a construction proper and a frequent collocation of constructions. It has been proposed that frequent collocations are so entrenched in the mind of speakers that they should be considered as constructions as well (Goldberg, 2006), a claim which has been dismissed by others (Traugott and Trousdale, 2013, p.5). It can also be discussed whether a collocation becomes frequent because the collocated constructions have fused into a single constructional unit, or if a construction arises from the frequent collocation of its constitutive constructions. For instance, is the collocation ‘by the way’ frequent because it corresponds to a single, pragmatic construction, or did ‘by the way’ became such a pragmatic construction because of the frequent collocation of these three words in the flow of speech? Both can be true and a feedback loop of frequency and entrenchment is most conceivable; notwithstanding, the picture of change that we shall provide in chapter 8 inclines more towards the first of these two perspectives.

1.1.2 Meaning of constructions

Another question is the relationship between meaning and the constitutive paradigms of a construction. Let us take the widely discussed case of [be going to {V}]. It is well known that the paradigm of compatible verbs has expanded from {V_{action}}, verbs indicative of an action that the subject may be on its way to perform, e.g.:

“Where art thou going, Jack?” said the cat. “I am going to seek my fortune.”

CAMPBELL, Joseph, *Popular Tales of the West Highlands*, 1860 (Google Books)

to a more general {V} (e.g. *there’s going to be cookies*). In such a case, two competing interpretations are possible. Either we consider that the meaning of the construction has expanded, leading to a greater number of uses, or we can consider that only the paradigmatic scope has expanded, the semantic expansion being an incidental consequence of which. We will argue in chapter 8 that the first interpretation is to be favored; indeed, it explains such incongruities as:

I am going to set this right here. [<https://theprose.com/post/90933>, 2016]

where the locative complement indicates an absence of movement which contradicts the movement meaning, so that the meaning change is actualized even for the original {V_{action}} paradigm. This question is actually most relevant and a given example cannot set the matter straight, to be sure. We only want to hint at this point that there is a delicate interplay between contexts of use (hence collocational frequencies), meanings, and paradigmatic scope. We also want to express the idea that the paradigm appearing in a construction participates to its meaning, as well as it participates to its context of use.

This leads us to the last point we wanted to stress in this brief overview of the linguistic concepts and questions we are interested in throughout this thesis. What possibly is meaning? Is the meaning of the construction given by the frequency-weighted set of contexts it is used in (and hence derived from collocational frequencies)? This would be in line with the structuralist view professed early on by de Saussure (1995) and later by Harris (1968). Is meaning given by the paradigmatic array the construction belongs to?

In this work, we consider profitable to posit instead a space of meanings, and to assume the existence of ‘semes’, that is, individual and clear-cut semantic units. These units are to be distinguished from semantic features as they can encompass several features at once, of various kinds (e.g. pragmatic, modal, illocutionary, procedural, referential, etc.). Furthermore, those units are organized as a network, which we will describe in chapter 7. We will argue later that the meaning of a construction is also best seen as a time-changing distribution over these semantic units. However, it is clear that there exists a strong connection between meanings and contexts of use; to interpret the meaning of a construction encountered in a text written several centuries ago, one has to rely on its context of use (see also the difficulty posed by *hapax legomena*). Note also that a collocation can constrain the activation of the different semantic units covered by the construction, but does not have to reduce the meaning to a single semantic unit. A construction found in a full sentence in a complete discourse can still be polysemic in this very sentence, and despite of all the constraints posed by the context.

We shall now review in more detail the four main notions that will be of need in this work, that is Constructions, Paradigms, Collocations and Meaning.

1.2 Constructions

Constructions, as we defined them, can be anything from a morphological particle (e.g. [{V}-tion], in words like *destabilization*, *monetization*, *grammaticalization*) to an abstract pattern. For instance, the French construction:

[N(quantity of time)] de [N(activity)]

carries the meaning ‘a given amount of time filled by some activity’, as in *une minute de silence*, *un jour de pluie*, *une année de bonne administration*, etc., involving one metaphor (Lakoff and Johnson, 1980), time length as a container, and a second one, activity as an uncountable matter, for *silence* and *administration*, arguably also for *pluie*, which can be interpreted here not as the rain itself, but as the fact that it is raining.

However, it is not always clear what counts as a construction and what does not. Is the latter construction we presented a construction on its own, or more probably a metaphoric extension of the construction [N(quantity)] de [N(uncountable)] (*un baril d’hydromel*, *un monceau d’ordures*, *un tas de cailloux*, *un zeste de regret*)? One may also wonder if even the latter is a construction rather than a particular subtype of the construction [N(object)] de [N(material)], as in *une statue de porphyre*, *un verre de spins*, *un sac de jute*, but also *un bras de fer*, *un jeu de société*, *une mine de charbon*, which are all metaphorical extensions of the construction. In Spanish, similar kinds of construction can be encountered, and a broad debate stemmed from them. Some people were inclined to say *un vaso con agua* instead of *un vaso de agua*, because it was said that the glass contained water instead of being made of water. The Real Academia finally stated that *un vaso de agua* was correct as it corresponded to a specific [{quantity} of {something}] construction, instead of referring to the [{object} of {material}] one (glass being defined in the former construction as ‘the quantity of liquid that a glass can hold’). This anecdote usefully serves as a reminder that such kind of questions are not the appanage of theoretical linguistics alone.

In English, the situation is different and we can easily distinguish two different constructions, one being a compound [{material} {object}] (a silver ring, a spin glass, a wooden shield), expressing the material composition of the object, the other being the construction [{quantity} of {something}], e.g. the famous Dongshan’s koan *Three pounds of flax*, the brand *Tons of Tiles*, Leone’s movie *A Fistful of Dollars*, etc. Yet the construction [{object} of {material}] also exists, as in a ring of silver, a cup of porcelain, a shield of wood, etc.)

This kind of concerns pervades the theory of constructions. For instance, is a word like *question* not a construction in itself, because it is compositionally built out of the two constructions [quest] + [{V}-tion]? All Construction Grammar theorists would probably agree that this analysis is loose as best and that *question* counts as a construction, because [quest] as a verb does not exist anymore with the meaning ‘to ask’. But what about *action*? The way to derive action from the verb ‘to act’ and the [V-tion] construction is obvious and explicit. In French however, it seems clear that, however obvious the derivation might be, attested constructs like *faire une bonne action*, or worse, *agir dans l’action*, tend to prove that *action* has its own existence.

This has led Goldberg to postulate that being compositional does not prevent from being a construction; what counts is the degree of entrenchment of the form in the

mind of speakers. Yet, this view is not shared by all Construction Grammar tenants; one of the argument put forward is that this definition is not operable, as it seems impossible to properly define a threshold of entrenchment or frequency above which a string of words can be seen as a construction.

One (theoretical) way to settle this matter would be to focus on the definition of a construction as a form-meaning pair (the construction is not a specific form that can be filled with some meaning, it is a Saussurean sign). There have been satisfying attempts to model the meaning of linguistic units as vectors in some multidimensional semantic space — e.g. the *word2vec* family of semantic representation models (Mikolov et al., 2013). From a set of linguistic units, it computes individual meanings (i.e. vectors in the semantic space) for each unit, given the other units (it is thus a structuralist way of computing meaning). Let us say we have a string ‘X Y’ (e.g. ‘chercher des ennuis’) which we do not know whether it is a construction [X Y], or a construct of two constructions [X] and [Y]. We can compute, in a *word2vec* approach, the meaning for [X], [Y] and [XY]. The question would then be: does the co-occurrence of [X] and [Y] in language match the same meaning as the construction [XY]? To our knowledge, such an investigation has not been carried out, especially because the semantic space defined through this kind of models is a semantic space for units only, not for sentences. We will discuss this further in the section devoted to meaning.

As we adopt a diachronic perspective, the question of whether some construct is a constructional unit or not is of special relevance. Indeed, a construction usually arises from a repeated construct (e.g. [à défaut de {N}]), which means that at some point, a shift is happening from one way of processing the construct to another (as a composition of several units vs. as a single unit). Being able to define a stable criterion would be of great use. In the meantime, we are reduced to track down hints in corpora that a reanalysis has happened (Traugott and Trousdale, 2013). For instance, French [à force de {P}] displays a clear constructionalization, from the compositional uses:

Ruer par terre ces mastins et combatre A force d’armes pour du tout les abatre ;

To bring down those rascals and fight By the strength of (our) weapons so as to slay them all

DE LA VIGNE, André, *La Ressource de la chrestienté*, p. 140, 1494 (Frantext)

in which *à force de* introduces a physical means to achieve some goal, often military, to:

[L]e roy mua propos à force d’estre pressé du duc de Milan par lettres[.]

The King spoke otherwise as the duke of Milan was pressing him so much with letters.

COMMYNES, Philippe, *Mémoires*, p. 34, 1495-1498 (Frantext)

in which *à force de* does not introduce a mean anymore, but indicates that, if the King finally held a different position, it was because the Milan Duke repeatedly sent letters. In the latter occurrence, [à force de {action}] can be glossed by ‘as a result of the repeated iteration of action’, a meaning which cannot be inferred from a compositional parsing of the construct.

It is, however, important to note that these two occurrences are contemporaneous of each other: the compositional and the constructional meanings of a construct can coexist, which means that a single form can be at the same time a compositional construct and a construction proper. This alone should indicate that the focus put

on constructions may not be that enlightening, as a given form can be in some kind of a superposition of states, being a construction and not being a construction. It would suggest that the form in itself is a more stable and concrete entity than the construction it may serve to substantiate. It should be indeed once again stressed that, in Construction Grammar, a construction is not a form, with a meaning, but a non-compositional pairing between meaning and form; so that there is no theoretical contradiction to state that the pairing [[à force de {X}] ↔ [resulting from repeated instances of]] is a construction while [[à force de {X}] ↔ [relying on the physical means X]] is not, because the latter is a conventional pairing. This would imply that polysemy is at least partly treated as coexistence of different form meaning-pairs, where coincidence of forms is only seen as a diachronic accident. We argue instead that polysemy is to be seen as essential in the description of language.

This is the reason why, though we will make use of the Construction Grammar framework as it is rich and efficient to discuss many features of the diachronic changes we observe in corpora, especially as it highlights the existence of partially concrete, schematic forms, we will not always try to distinguish properly constructions and constructs and will speak more broadly and more vaguely of linguistic ‘forms’, ‘elements’ reserving the word ‘item’ for the components upon which complex constructions are built. In short, there is some unity in form, independently from any constructional considerations, and we rely on this unity to follow the changes undergone by that form.

1.3 Paradigms

Paradigms are sets of arguments entering into a free slot of a construction. They can be considered as prototypically organized (some arguments are more salient than others) ; for instance, in the French construction [de même {N}], the most frequent name in the paradigm, *nature*, accounts for 15% of the occurrences of the construction, and its close relatives (*espèce, ordre, type, sorte*) all belong to the ten most frequently encountered names in the paradigm. Other names like *couleur, sens, forme*, are semantically close to *nature*. Others, like *origine* and *famille*, are more atypical.

If we look now at the construction [d'égal {N}], which can be seen as another instance of one and the same [de {même, égal} {N}] construction, we find a rather different paradigm, dominated by quantitative characterizations (*valeur, épaisseur, hauteur, longueur, durée, intensité, pression, diamètre, indice, température, grandeur*). The second rank is occupied, quite surprisingly, by *importance*, which is not really quantitative, but it remains scalar, making it possible to speak of different degrees of *importance*. Note also that *valeur* is the only word which appear in both lists; this may be explained by the fact that *valeur* has both a qualitative reading and a quantitative one:

[I] est inutile de chercher à expliquer en quoi Merlin est un homme de valeur.

It is pointless to try to explain what makes Merlin a man of worth.

GUILLOUX, Louis, *Le Sang noir*, 1935 (Frantext)

[C]'est le temps nécessaire pour qu'une quantité quelconque de la substance considérée diminue de la moitié de sa valeur.

It is the time needed for any quantity of the studied substance to lower its value by half. CURIE, Marie, *L'isotopie et les éléments isotopiques*, 1924 (Frantext)

1.3.1 Paradigm constraining

If we consider the construction [de même, égal N], which we can rewrite [de {P₁} {P₂}], we see that {P₂} strongly depends on the value taken by P₁, so that it would be more accurate to write the construction as [de {P₁}{P₂}(P₁)]. We can, however, wonder whether these two paradigms {P₁} and {P₂} share the same status. Indeed, {P₁} and {P₂} seem like two very different objects. {P₁} is much more restricted (though we also find *pareil*, *semblable*, *différent*, *comparable*, *inégal*) than {P₂} (which can take more than a hundred different values), and {P₂} takes its values depending on P₁. This dependence can be seen as multiple levels of meaning, from the most schematic to the the most specified (e.g. [de {P₁} {P₂}] (two comparees share some P₂ to P₁ extent) → [de même {P₂}(même)] (two comparees share some attribute P₂ to the same extent) → *de même sens* (two comparees share the same meaning)).

What is more interesting is the fact that these paradigms evolve with time. The construction [de différent {N}] almost disappears at the twentieth century, while [de même {N}] reaches its peak of frequency in the mid eighteenth century. The paradigm P₂(même) also evolves; though the rank and domination of *nature* remains the same (about 11% and 16% of the total), nouns like *sentiment*, *avis*, *opinion*, are only prominent during the seventeenth century, while *de même couleur* (the second most frequent noun of the P₂(même) paradigm in the 19th and 21th centuries) appears only from the 18th.

A question worth investigating is the following. It is clear that the different terms of the {P₁} paradigm are competing among themselves across time. Then, what about the {P₂}’s? Is another competition taking place for each P(P₁) paradigm among its members? Or should we consider the P₁’s are competing for the P₂’s? In such a view, each [de P₁ {P₂}(P₁)] may be viewed as an environment for which the different P₂ can compete. Most probably, the two perspectives complement each other: the fact that the different members of {P₂} goes from one P₁ to another (e.g. *importance*, chiefly associated with *pareille* during the 16th and the 17th centuries, is taken by *égale* during the 19th) argues for the first view, while the changing frequencies of different P₂ members among a given {P₂}(P₁) paradigm speaks in favor of the second.

1.3.2 Paradigms at the intersection of competing constructional analyses

Another complication comes from the fact that a construct can be seen as an instance of two different constructions. For instance, we could postulate a construction [de {ADJ} importance] (*de première importance*, *de sublime importance*, *de maigre importance*, etc.) so that *même*, *pareille*, *égale*, *inégle* would only be instances of the {ADJ} paradigm. This paradigm is prototypically organized as well, with adjectives measuring the degree of the importance, specifically *de grande importance*.

There is, however, some specificity in the [de {même, pareil, égal, ...} {P₂}] construction, as it establishes a comparison between two terms (a grammatical category which can assume many shapes, as surveyed in Heine (1997)), while [de {ADJ} importance] typically qualifies one or several items as a group, without stating a comparison between its members.

Also, [de {ADJ} importance] has no reason to be considered as a construction; it is neither morphologically nor semantically specific, and should be considered as an instance of the more general [{NP₁} de {NP₂}] construction, where *de* NP₂ is an epithet of NP₁ and could be replaced by an adjective (e.g. *de couleur claire* > *clair*, *d'un âge avancé* > *vieux*). It cannot be excluded, however, that the comparative [de {même,...} {P₂}] construction is analogically strengthened by (and maybe also historically comes from) the epithet construction [de {NP₂}].

1.3.3 Speciation of paradigm members into independent constructions

The [de {P₁} {P₂}(P₁)] nominal comparative construction illustrates also one last point: some members of the paradigm can detach and exhibit a meaning shift. This is the case for [de pareil {P₂}(pareil)] and [de semblable {P₂}(semblable)], which were replaced by [de même {P₂}] and [d'égal {P₂}]; yet they did not disappear, but developed a new meaning, becoming nearly synonymous of [un tel {N}] ('such a {N}')

Depuis le ministère Clemenceau, les maisons de jeu n'avaient plus connu de semblable alerte.

Since the ministry of Clemenceau, the gaming houses had never known such a distress.

ARAGON, Louis, *Les Beaux quartiers*, 1936

[V]it-on jamais de pareil sapajou ?

Was such a sapajou ever to be seen?

RENAN, Ernest, *Drames philosophiques*, 1888 (Frantext)

As an amusing anecdote, the only other occurrence of 'de pareil sapajou' in Frantext is due to André Gide in 1932 who, in his personal diary, comment on the former occurrence from a drama by Ernest Renan, stating:

[I] me semble que le "de" n'est possible qu'avec sapajou(s) au pluriel.

It seems to me that the "de" fits only with sapajou(s) in plural form.

GIDE, André, *Journal : 1889-1939*, 1932 (Frantext)

This observation entails that the meaning of a construction can remain more or less constant, while the meaning of some members of its paradigm can shift away. In such a case, there is a phenomenon which is loosely analogous to speciation in biology: [de {pareil, semblable} {N}] has become a novel construction: a new, conventional, form-meaning pairing. This shows that the study of constructions cannot dismiss the importance of paradigms, which reveal a lot on the constructions they belong to. They remind us that language is chiefly a multi-level organization of mutual interactions and competitions (Melis and Desmet, 1998), between different kinds of constructions, between different items within a paradigm, even between different paradigmatic organizations.

1.3.4 External and internal paradigms

In a way, paradigms can even be seen as being part of the multiple features of the construction with respect to which they are defined. For instance, the paradigm of

nouns can be ascribed to the ‘Noun Phrase’ construction, which itself fills slots in various argument structure constructions (e.g. the transitive construction $[\{\text{Agent}\} \{\text{Verb Phrase}\} \{\text{Noun Phrase}\}]$). At a higher level, argument structure constructions can be seen as paradigm members of an ‘utterance’ construction. In that sense, the question of distinguishing a construction with a free slot from a ‘paradigm of constructions’ for each of which the slot would be filled (e.g. $[C_1 \{C_2\}]$ vs. $\{[C_1 C_2]\}_{C_2}$) could be settled by checking, in the latter case, if the paradigm can be seen as an argument of an overarching construction within which it would be properly defined. Note that this proposal amounts to consider that there are no paradigm of constructions as such, so that the first of these alternatives would be always true: every paradigm would count as a free slot within a construction, and the difference between $[C_1 \{C_2\}]$ and $\{[C_1 C_2]\}_{C_2}$ would be but the level of the paradigmatic variation in the embedding hierarchy of constructions.

Arises also the problem of shared paradigms. A paradigm, such as the paradigm of nouns, is obviously shared by a great number of constructions. However, as we just did, we can posit a ‘Noun Phrase’ construction, in which alone the full paradigm of nouns can unfold. All other constructions relying on the paradigm of nouns would then call forth the ‘Noun Phrase’ construction. For such a trick to be allowed, it is necessary that the ‘Noun Phrase’ construction can be ascribed to a meaning, which is certainly the case. A construct of the ‘Noun Phrase’ construction will be an ontologically determined entity which will provide content to the representation. Yet, the ‘Noun Phrase’ is itself a member of a paradigm $\{\text{Noun Phrase, Infinite Verb Phrase, Pronoun}\}$, which is shared itself among different argument structure constructions, and can even appear, with slight modifications, in different slots of some of those argument structure constructions (e.g. in the transitive argument structure construction, it can fill the ‘subject’ position as well as the ‘object’ one). On the other hand, in the transitive construction, the ‘object’ paradigm is much often constrained by the verb, which may indicate that the two paradigms, ‘subject’ and ‘object’, are separated. All those questions, no matter how fascinating, can nonetheless be left aside for the remaining of this work, for they are quite far from our primary focus, which is grammaticalization and its modeling. This topic has furthermore been discussed at great extent in the framework of Radical Construction Grammar (Croft, 2001).

1.3.5 Paradigms and meaning

As it has already been mentioned, meaning is related to paradigms. For instance, a list of synonyms may sometimes give a better idea of the actual meaning of a word, than a lexicographical definition from the dictionary (see for instance the Semantic Atlas for such a synonym-based approach on word meaning). However, a paradigm is not a list of synonyms; and some paradigms are so loose (e.g. the paradigm of verbs in the transitive argument structure construction), that they cannot be used to determine the meaning of its individual members. Paradigms, in that sense, are mainly guidelines of language use, recommending the use of items or group of items in certain parts of construction.

Yet, paradigms can inform the meaning of a construction, in at least two ways. First, the meaning of a construction is shaped by the paradigms which appear in it.

For instance, the Verb Phrase construction and the Noun Phrase construction share the same schematic structure, but strongly differ in terms of paradigm members (much more so in French than in English). Similarly, the French construction we mentioned earlier, the ‘nominal epithetical’ construction ($\{\{N_1\} \text{ de } \{N_2\}\}$), has the same schematic structure as the genitive construction $\{\{N_1\} \text{ de } \{N_2\}\}$ (whose meaning is ‘ N_2 ’s N_1 ’, a translation which is impossible for the first construction). These two constructions are schematically identical, yet they do not have the same meaning at all, albeit they share a common function of qualifying the N_1 argument. Their difference lies chiefly in the fact that they do not hold the same paradigms, be it for the qualifier N_2 or the qualifiee N_1 . This example shows that the meaning of a construction partly arises from the structure of its paradigmatic arguments.

The meaning of a construction is also construed by the paradigms to which the construction belongs. For instance, if a construction shows up in the second paradigm of the ‘nominal epithetical’ French construction, then the meaning of a construction is characterized by an ‘attribute’ trait, for only attributes can belong to this paradigm. Construction Grammarians have studied numerous examples of paradigmatic inscription meaning, the most famous being the use of verbs in unusual argument structures, eliciting a further meaning in agreement with the argument structure — see the famous examples of *bake* and *sneeze* in Goldberg (1995).

A consequence of this phenomenon is that, as soon as a linguistic item appears in several different paradigms — equivalently in different constructions — then it can be deemed as polysemous. The French word *voie*, for instance, is clearly polysemous in the sense that it appears both in the regular noun phrase construction — e.g. ‘La voie était momentanément libre.’, FÉREY Caryl, *Mapuche*, 2012 (Frantext) — and in the $[\text{en } \{\text{voie, cours, passe, train}\} \text{ de}]$ construction. Incidentally, the fact that a word appearing both in ‘subject’ and ‘object’ positions in, say, the transitive argument structure, does not necessarily imply that the word is polysemic, for instance with the quite monosemic word *screwdriver*:

[L]e tournevis dérape brusquement vers la gauche.

The screwdriver suddenly slides to the left.

BENOZIGLIO Jean-Luc, *Tableaux d'une ex*, 1989 (Frantext)

Passe-moi le tournevis.

Hand me the screwdriver.

PONTALIS Jean-Bertrand, *L'amour des commencements*, 1986 (Frantext)

This would speak in favor of the existence of a ‘Noun Phrase’ construction whose paradigm is consistent over different uses (different in that the ‘Noun Phrase’ construction can be used in different constructions).

Of course, a word can still be polysemic even if it belongs to one paradigm only. To take a rather extreme and humorous example, if one takes the word ‘paradigm’, it shows a remarkable array of meanings, even when used by one single author in one single work, at least according to Masterman (1970), which claims to have distinguished 21 different uses of the term in Kuhn’s *The Structure of Scientific Revolutions*. Polysemy is thus not due to different paradigm affiliation, though the latter is a clear sign of polysemy. There is indeed another usage-based component of meaning, which arises from collocations, with which we shall deal in the next section.

1.4 Collocations

There is an important aspect of constructions which must be stored somehow in speaker's minds, independently of constructions: it is the knowledge of collocations, that is, an overall idea of how frequently two constructions can collocate. This is to be distinguished from the knowledge of paradigms, because paradigms, as their names suggest, deal with paradigmatic alternations (which constructions can be substituted to each other in a given slot of the utterance), while the study of collocations aims at identifying recurring syntagmatic co-occurrences in the chain of speech. If a collocation is frequent enough, the two collocating constructions can even coalesce and lead to their reanalysis into a single, unified construction. For instance, in English, 'very' and 'much' frequently collocate, though it might not be fitted to consider that 'very much' is a construction in its own.

1.4.1 Frequent collocations as actual constructions

Actually, some proponents of Construction Grammar consider that frequent collocations of constructions indicate that such a construct, while fully compositional and still analyzable as such, can be stored nonetheless as a construction in speaker's knowledge of language. It is believed that a certain threshold of frequency must be passed above for such a storage to occur (Goldberg, 2006), yet this view is not consensually held (Traugott and Trousdale, 2013). It seems at least that the 'frequency threshold' is not absolute, but relative to the other members: if a collocate happens to be much more frequent than the others, then the collocation might act as a construction.

Let us take an example. The word *coup*, in French, is a substantive, atomic construction, whose literal meaning varies from 'blow' to 'shot'. Being very loose concerning the actual nature of the 'coup', it is often complemented through the genitive construction by another name. In Table 1.1, we show the four most frequent names immediately following *coup de*, with their frequency among all collocates. Remarkably, it appears that the present century has much more pacific concerns than the previous ones: it is all about glances and phone calls. Also, as it appears clearly, *coup de* shows a marked collocational preference for *œil* from the seventeenth century, in proportions which go on increasing. It is a strong sign that *coup d'œil* ('glance') emerged as an independent construction, through frequent, repeated collocation.

1.4.2 Typology of collocations

We can distinguish three kinds of collocations: collocation within a construction, extra-clausal collocation, and accidental collocation. Extra-clausal collocation happens when two constructions collocate in the utterance. For instance, in the following sentence:

Still, the Consul knew, to bring a ship as beautiful and vulnerable as the Yggdrasill
- one of only five of its kind - into a war zone was a terrible risk for the Templar
Brotherhood. SIMMONS Dan, *Hyperion*, 1989

Still, the Consul knew, the interpolation *one of only five of its kind*, are examples of extra-clausal collocations. On the contrary, we have a BE-attributive construction [X is Y] in which we can say that *was* and *a terrible risk* collocate. Within the Noun Phrase

XIV th (48)	lance (7)	baston (3)	coustel (3)	espee (2)
XV th (103)	espee (16)	lance (9)	fouet (3)	canon (2)
XVI th (304)	espée (26)	pied (12)	aventure (12)	fortune (11)
XVII th (1197)	œil (88)	foudre (50)	tonnerre (43)	pistolet (37)
XVIII th (2282)	œil (654)	foudre (110)	pistolet (61)	vent (55)
XIX th (10773)	œil (2371)	état (357)	vent (326)	poing (315)
XX th (17171)	œil (3264)	pied (702)	poing (560)	vent (518)
XXI th (2326)	œil (414)	téléphone (143)	fil (88)	pied (82)

Table 1.1: First four most frequent right-collocates of *coup de* for each century from the XIVth to the XXIth in the Frantext corpus. Each row corresponds to a different century. Number of occurrences of each collocate within the paradigm is indicated between parentheses; total number of occurrences of the form in the century is given between parentheses in the first column.

construction, *a terrible risk for the Templar Brotherhood*, *terrible* and *risk* collocate. The infinitive construction [to bring {something}] can also be said to collocate with *was* within the BE-attributive construction even if they are syntagmatically separated by a lot of additional information (the characterization of the ‘something’ and an additional precision -the location where the ‘something’ is brought). On the contrary, *a war zone* and *was* collocate only through a syntagmatic accident and not within a construction; hence it is an example of an accidental collocation. Accidental collocations also appear in our example *coup de*: for instance, *coup* collocates with *de toute sa force* (‘with all one’s might’):

[I] getta un coup de toute sa force[.]

He stroke with all his might.

Bérinus, 1350 (Frantext)

Of course, in this utterance, *toute sa force* does not collocate with *coup* within the genitive construction, as *coup d’épée* or *coup de tonnerre*, for *de toute sa force* applies not on *un coup* but on the whole ‘il getta un coup’. The construction *de toute sa force* collocate extra-clausally with the other members of the utterance, and collocate accidentally with *coup*.

None of these collocations must be neglected. As we shall see in the next chapter, even accidental collocations can be stored in speaker’s language knowledge and influence the course of language change. Most probably, they correspond to different psychological activities; collocation within construction correspond to the common knowledge of a language speaker about what usually go along; accidental collocations, when acknowledged, may invite to reanalysis and be a mechanism of constructionalization; extra-clausal collocations are probably not that significant, but they weave a network on the discourse level, leading to possible long-range co-occurrences of extra-clausal constructions — a very simple example of this would be the pair *on one hand* and *on the other hand*.

We already encountered collocations within constructions in the previous section; indeed, paradigms are abstracted out of frequent collocations. However, not all collocations may be bound to give rise to a paradigm. In that sense, Evert (2008) suggests

to distinguish between *empirical collocations*, the collocations evidenced through a statistical analysis on a corpus, and *lexical collocations*, corresponding to speaker's intuitions and ideally testable through psycholinguistic experiments. In which extent those definitions overlap is an open question; expectedly, the overlap would be greater if we restrict ourselves to collocations within constructions, albeit there is no safe way to distinguish constructions in a corpus, especially when new constructions arise, a topic with which we will be much concerned in this work.

1.4.3 Collocations and colligations

Another distinction which is made in the literature concerns collocations and colligations (see (Lehecka, 2015) for a review). The latter refers to the preference of certain grammatical categories manifested by lexical items. We shall see in the next chapter a particular verb construction, *give way*, which heavily favors intensified qualifiers over regular ones (e.g. *too soon* vs. *soon*). However, in Construction Grammar, there seems to be no need at first sight to distinguish collocations and colligations. We can indeed give a 'constructional' account of the difference between collocation and colligation: for instance, we can write the epithetical adjectival noun phrase construction in the following way, $[\{\text{adj}(\text{N})\} \{\text{N}\}]$, making explicit the fact that the adjective paradigm will depend on the evaluation of the name paradigm. It is then possible to claim that the adjective members of the $\{\text{adj}(\text{N})\}$ paradigm collocate with the name within this particular construction (e.g. *green* collocates with *grass* within the epithetical adjectival noun phrase construction, to invoke the most commonly referred to example of a collocation). A colligation would then occur when an evaluated member of the construction constrains the paradigm not in terms of arguments, but in terms of constructions. For instance, the use of the first person pronoun *I* might trigger a modal construction in the verb phrase within the transitive construction more often than the use of the third person. This would lessen the difference between colligation and collocation: in both cases one of the paradigm in the construction is constrained by the evaluation of another paradigm, but in the case of the (I > modal) colligation it is the whole verb phrase construction paradigm which is constrained instead of the atomic verb paradigm. Between collocation and colligation, there is only a difference of schematicity of the paradigm members, not a difference in nature: members of the 'Verb Phrase' paradigm in the transitive construction (such as the modal construction, the 'does not' negation construction, the simple atomic verb construction, etc.) are more schematic than members of the 'adjective' paradigm in the epithetical construction, but the mechanism is the same.

However, the examples provided by Sinclair (1998) show clearly that colligation is a different phenomenon. He cites the example of *budge*, which occurs only within a modal construction, or close to it. We can consider that, within the verb phrase construction, *budge* restricts some available paradigm to the modal construction; yet there is no construction within which the modal construction and the atomic *budge* construction are both members of a different paradigm. The order of things goes like $[\text{VP} = \{\text{modal construction, verbal negation, atomic verb construction, etc.}\}]$ (partial construct of the verb phrase into modal construction) > $[\{\text{modal}\} \{\text{atomic verb}\}]$, and not like $[\text{VP}]$ (partial construct into the *budge* construction) > $[\{\text{modal}$

X	<i>refuse to</i>	<i>might</i>
<i>begin</i>	4	265
Total	6079	289,520
Ratio	0.07	0.46

Table 1.2: Collocation frequencies between *begin* and two different modals, *might* and *refuse to*, showing a higher conditional probability of *begin* for *might* than for *refuse to*. Data from COHA.

construction, verbal negation, atomic verb construction, etc.}(*budge*) *budge*], the latter being clearly absurd, since the modal construction takes one argument, which would already be fulfilled in this case. While collocation is the phenomenon by which a linguistic item is associated with a distribution of co-occurring possibilities within a given construction, colligation is the phenomenon by which a linguistic item is likely to be found within particular constructions. At this point, to account for the idiosyncratic behavior of *budge*, we can either consider that *budge* is actually stored as a schematic construction, [{modal} *budge*], within the paradigm of the verbal phrase arguments, such as to directly encode its specificities, or we can consider that colligations are proper linguistic knowledge, alongside collocations (in the full sense encompassed above) and the constructicon.

1.4.4 Asymmetry in collocations

The existence of collocations raises several questions. First of all, how do we know which paradigm constrains the other, and not the other way around? In the epithetical adjective construction, why a given choice of adjective cannot be said to constrain the noun, a possibility all the more serious considering that, in English, the noun appears syntagmatically after the adjective? In a way, epithetical adjectives are used to modify a noun, so the noun is the starting point of the representation. It makes sense to add further detail in the representation by specifying the noun with an additional attribute; but it would be rather odd to start with some modifier, and then to ask what can be modified with this. Of course, as collocations are well-known by speakers, an adjective can create expectations. If a tale starts with: ‘In a peaceful clearing in the middle of the woods was a nice little house in which lived a young ...’, ‘maiden’ is more expected than ‘carpenter’, no matter how logical is the latter given the context.

The situation is all the more interesting with verbs and modals. In the modal construction, [{modal}, {atomic verb construction}], which of the two paradigms constrains the argument of the other? There are indeed such constraints; for instance, *begin* is more compatible with *might* than with *refuse to* (Table 1.2).

Actually, the {modal} paradigm is closed, so that it would be unexpected to further narrow a paradigm which is already limited. Yet, from a psychological point of view, it seems more reasonable to store, for each atomic verb construction, a prevalence towards a handful of modals, than the contrary. Indeed, it is easier to order and to feel differences for a small number of items, than for a huge number such as the entirety of English verbs. Thus, it might well be the atomic verb which constrains

the modal, than the other way around. Furthermore, the previous argument regarding the priority of content might still hold. However, this would imply a naive view of language production, in which the representation of the utterance is first built in the mind of the speaker and then expressed through linguistic means. Such a discussion would however lead us too far astray from the main focus of this work.

We saw in the section on paradigms that in constructions such as [de $P_1 = \{\text{pareil, égal, semblable, même, égal}\} \{N\}$], the less contentful, smallest paradigm was actually constraining the much more open noun paradigm. This might be because each of these paradigm members actually constitutes a construction on its own (the opposition between a paradigm of constructions and a paradigm within a construction); or we can consider that the paradigm is actually part of the definition of the construction, so that it cannot be evaluated by the noun which is a free slot (in the extent that it conforms to the rather loose ‘noun’ paradigm). If so, it would be the same for modals: the paradigm of modals cannot be constrained by the atomic verb, because the definition of the modal construction relies on the modals paradigm. A last possibility would be that it is the noun’s choice which constrains the paradigm member actually used in the construct. Indeed, speakers are more likely to pick the predicate of the comparison before choosing the right paradigm member to use in the construction. Yet, the choice of the P_1 paradigm member will be relevant not because of the noun, but because of the distributional features with which it is statistically associated. If *même* is more likely to be associated with a quality and *égal* with a quantity, it is not because each noun expressing a quality constrains the P_1 paradigm, but because some generality emerges from the nouns usually following each of the different members. Thus, the constraint is not associated with the actual production of the utterance, but is abstracted out the past linguistic experiences of the speakers.

1.4.5 Empirical dealing of collocations

This question of the asymmetry of collocations can also be investigated from an empirical perspective. If one looks for the collocates of *affreux*, using the Frantext query:

affreux &q

which will send back all occurrences of *affreux* and the word (including punctuation marks) following it, then one can find the first four nouns collocating with *affreux*, as displayed in Table 1.3.

If we now look for *malheur*, using the query:

&q malheur

which is the most frequent collocate of *affreux* in two of the observed centuries, we find that the collocation is indeed reciprocal, as evidenced on Table 1.4 (note that, contrary to most approaches, we do not yet divide the observed co-occurrences by the frequency of the collocate).

From this we can compute the asymmetric conditional probabilities of collocation for the three relevant centuries:

- XVIIIth century: $P(\text{affreux}|\text{malheur}) = 2.6 \cdot 10^{-3}$; $P(\text{malheur}|\text{affreux}) = 5.9 \cdot 10^{-3}$
- XIXth century: $P(\text{affreux}|\text{malheur}) = 7.9 \cdot 10^{-3}$; $P(\text{malheur}|\text{affreux}) = 1.9 \cdot 10^{-2}$
- XXth century: $P(\text{affreux}|\text{malheur}) = 2.1 \cdot 10^{-3}$; $P(\text{malheur}|\text{affreux}) = 5.8 \cdot 10^{-3}$

XVII th (573)	precipice (7)	spectacle (6)	périls (6)	supplice (4)
XVIII th (2710)	désespoir (31)	spectacle (19)	malheur (16)	supplice (14)
XIX th (3820)	malheur (71)	désespoir (28)	spectacle (26)	sourire (18)
XX th (2779)	malheur (16)	souvenir (13)	cauchemar (13)	courage (10)
XXI th (309)	accident (3)	docteur (3)	détails (2)	sentiment (2)

Table 1.3: First four most frequent right-collocates of *affreux* for each century from the XVIIth to the XXIth in the Frantext corpus. Each row corresponds to a different century. Number of occurrences of each collocate within the paradigm is indicated between parentheses; total number of occurrences of the form in the century is given between parentheses in the first column.

XVII th (3676)	grand (130)	dernier (28)	nouveau (18)	commun (16)
XVIII th (6189)	grand (182)	nouveau (59)	affreux (16)	petit (15)
XIX th (8932)	grand (352)	affreux (71)	nouveau (45)	dernier (20)
XX th (7611)	grand (206)	vrai (24)	nouveau (21)	petit (20)
XXI th (731)	grand (20)	nouveau (2)	gros (2)	—

Table 1.4: First four most frequent left-collocates of *malheur* for each century from the XVIIth to the XXIth in the Frantext corpus. Each row corresponds to a different century. Number of occurrences of each collocate within the paradigm is indicated between parentheses; total number of occurrences of the form in the century is given between parentheses in the first column. There are only three collocates of *malheur* in the XXIth section of the Frantext corpus .

Century	<i>affreux</i>	<i>malheur</i>
XVII th	0.042	0.056
XVIII th	0.034	0.046
XIX th	0.037	0.057
XX th	0.022	0.038
XXI th	0.016	0.033

Table 1.5: Conditional probabilities of the right-sided adjectival construction for the two words *affreux* and *malheur*, for each century from the XVIIth to the XXI^{20th}, as computed from the Frantext database.

Such numbers tell us that *malheur* occupies a more prominent place in the *affreux* paradigm than the contrary, which is not really a surprise; indeed, *affreux* is more specific than *malheur*. A greater still asymmetry is expected between *grand* and *malheur*, as *grand* is not specific at all and can combine with many collocates. It should be noted that these conditional probabilities correspond to the usual collocation score, though this is not the only way Corpus Linguists quantify collocations (Gries, 2013).

Another quantity to compute from those raw frequencies is the weight of the paradigm (truncated to its first five prominent members) among all occurrences of the form, as presented on table 1.5. Those results show that *malheur* exhibits a greater colligation for the epithetical adjectival construction than *affreux* (which indeed can easily occur in sentences such as *c'est/c'était affreux*, totalizing 601 occurrences in Frantext). Furthermore this colligation decreases with time for both terms. This decrease might be due to an increasing use of *malheur* in the left-sided adjectival construction (e.g. *un malheur terrible* instead of *un terrible malheur*).

1.4.6 Constraints on paradigms

Still the question which has occupied us so far remains without an answer: how can we know which one of two paradigms constrains the other within a construction displaying two paradigms? In other terms, in a collocation event, can we say that the two terms mutually collocate, or that one is collocating with the other rather than the contrary? At this point one can wonder why this is so crucial. To understand what really is at stake behind this issue, we must anticipate a problematic question with which we shall deal at length in the next chapter on grammaticalization. It concerns the organization of language along a cline, from lexical elements, to grammatical ones. Expectedly, modals are more ‘grammatical’ than regular verbs, for instance, because they are not providing actual content to the utterance, but help instead in processing it, or in shaping it into a particular representation, exerting thus a linguistic function which Chafe (2002) has called *orientation*. Yet, the distinction between the two elements is far from being clear, and while it lies on strong and widely held linguistic intuitions, no clear and precise consensus has emerged surrounding this question. Therefore, any empirical hint of the existence of that cline, such as the paradigm constraint in collocations which presently occupies us, is worth being digged into a bit further at least.

As will be argued in the next chapter, adjectives are more grammatical than nouns, just as modals are more grammatical than verbs. Also, in constructions such as [en {voie, cours, train, passe} de {N, infinitive verb}], nouns appearing in the closed, first paradigm of the construction are more grammatical than nouns appearing in the open, second paradigm (the very fact that a noun can assume such a grammatical status being one of the strongest features of the grammaticalization phenomenon). It would be nice, thus, to be able to compare the grammaticality of two paradigms, pairwise, by devising a rule stating, for instance: ‘If two paradigms are present in a construction, the most grammatical one constrain the other.’; or the other way around. As I tried to argue for constructions like [en {voie, cours, train, passe} de {N, infinitive verb}], it seems safer to consider that the smallest paradigm is constraining the biggest, in a loose way, meaning that regularities emerge from repeated collocations between members of the first paradigm and members of the second, so that semantic generalities start to profile themselves. The psychological argument invoked above does not hold, since the constraints on the paradigm are neither strict nor precise; they result from the perception of a distribution pattern over repeated exemplars.

In the case of constructions involving two noun paradigms, one being very narrow and the other being almost entirely open, it is clear which one is constraining the other. It is in no way meaningful to list, for each nouns of the open paradigm, which possibility of the restricted paradigm is the fittest. Interesting patterns come up when we discover unexpected preferences for each of the members of the closed paradigm, as Hilpert (2008) did with modals in Early Modern English. The conclusions drawn about specific constructions displaying two noun paradigms seem thus to be as relevant in the case of the modal construction. It would thus be very interesting, at this point, to establish that adjectives are constraining nouns in the epithetical adjectival construction. Such a feat would allow us to consider that there is a strong correlation between paradigmatic constraints in collocations within constructions, and grammatical status. Moreover, it could be extended at the level of more schematic constructions.

It would also have a small consequence for the definition of grammaticality. Indeed, what is grammatical in this view is not the linguistic item by itself, but the paradigm to which the item belongs. In that sense, we would have more lexical and more grammatical paradigms, instead of more lexical and more grammatical items or constructions. It follows immediately that grammaticalization, as it deals with the emergence of new paradigms and with the recruitment of new items within existing paradigms, is necessarily a usage-driven phenomenon, for paradigms result from repeated collocations, in ways which we shall unfortunately not address in this work.

In this picture, three phenomena are yet missing: colligation, accidental collocations, and extra-clausal collocations. Let us leave aside the first two ones, and focus on the third. Linguistic items occurring in extra-clausal collocations, especially discourse markers, belong to paradigms which do not collocate with other paradigms (a fact entailed by the very definition of an extra-clausal collocation). Hence, their grammatical status can never be compared with other paradigms according to the ‘constraints’ criterion. They may as well be ‘outside’ of the lexical-grammatical cline. Furthermore, there is apparently no graduality between extra-clausal collocation, and in-construction collocation, meaning that there could be a clear-cut gap between extra-clausal paradigms, and grammatical paradigms. This would go perfectly along a recent

proposal by Heine et al. (2014), minus the neurological claim.

To conclude this issue, it seems fitted to stress an interesting observation. While, in shaping an utterance, the speaker probably starts with the contentful elements and then adds an orientation, it would seem that the linguistic organization is completely upside down. There would be thus a sort of paradoxical duality in language: whereas language production implies to put content first and then to find fitting orienting elements (for instance a proper adjective, given the noun), language would be organized such that those orienting elements are fleshed out in their semantics by a reversal of the perspective. Knowledge of the language would thus be abstracted out of language use, but entrenched in the opposite way.

1.4.7 Collocations and meaning

We shall end this discussion on collocations by discussing their relationship with meaning. As we already implied, the meaning of a construction is certainly derived, partly at least, from its collocations, and from the paradigms it belongs to.

It was actually the project of Harris (1968) to describe language through collocation frequencies, with the hope that meaning of words can emerge in a structuralist way. This program has been finally gotten through with the rise of computational power in the two first decades of this century, as exemplified by very many different approaches, the most representative being perhaps that of Mikolov et al. (2013), which makes use of learning theory to build a vectorial representation of words meaning through their collocations in training corpora. This representation is high dimensional (ranging from 20 to 1000 dimensions, 640 for the model giving rise to the results that they present), and the vector representation is an array of numbers, none of them having any significance by themselves. In a purely structuralist fashion, those vectors make sense through the relations they have with each other. For instance, if one performs the operation *sushi - Japan + USA*, one gets *pizza* (Mikolov et al., 2013). The representation may not be perfect (e.g. *Berlusconi - Silvio + Putin* gives back *Medvedev* instead of the expected *Vladimir*), yet it is a remarkable feat of computational semantics. It shows that there is indeed a structural organization of meaning emerging from the mere collocation relations between words.

A different approach is taken in the Semantic Atlas (Ji et al., 2003). The idea is not to represent all words together, but to provide a semantic map of the meanings of a given word. Building on corpus-evaluated co-occurrences, they consider a network of words, mutually linked if they co-occur. They then restrict to the subgraph of all the contexonyms of the word under study (i.e. the words with which it co-occur). From this network, they compute ‘cliques’, i.e., complete subgraphs not included in a bigger complete subgraph. Those cliques are then located in a multi-dimensional space, where each dimension corresponds to one of the contexonyms. To get a suitable, graphical representation of the meaning of the word under study, this procedure is complemented by a principal component analysis, giving rise to a two-dimensional schematic representation of the word’s meaning. The two dimensions of such a representation have no straightforward interpretation, but they are less opaque, and more meaningful, than the dimensions in word vector approaches, as one can visualize which cliques (and the associated contexonyms) are attracted to which axis.

The spatial representations of the Semantic Atlas, contrary to those of word vector approaches, unfolds the meaning of a single word through a geometric structure. This is obviously of high interest, and it gives a vivid picture of the semantic complexity of language items. However, the meaning of a word is still given in relation with other words (as in a dictionary, there is no ‘entry point’ in the language). Furthermore, the ‘cliques’ are not items to which we are accustomed, and the precise understanding of their dispersion pattern is not always easy to get at. Still, it shows that co-occurrences (a wider term than collocations, for two words are said to co-occur as long as they belong to the same sentence) encode deep and interesting information about the semantic of the language elements.

Yet, those approaches share a strong limitation, at least from the perspective of Construction Grammar. Indeed, they focus on single words, not on constructions; hence, the meaning of most constructions lies beyond the scope of these empirical representations. This is all the more regrettable as the issue of the meaning of constructions, especially the most schematic ones, remains a debated topic, and doubt is cast upon this idea even by self-proclaimed tenants of Construction Grammar:

[T]he present-day English transitive construction is ostensibly void of all meaning
(though no construction grammarian could accept its meaninglessness)

(Noël, 2007)

Even atomic constructions can be problematic; especially discourse markers. For instance, the ten first contonyms of *cependant* offered by the Semantic Atlas are: *dont, peut, entre, était, leurs, soit, celle, jusque, aucun*. None of them are entirely relevant, especially considering that they all are quite frequent words. This kind of data, in any case, is of very little help to figure out the meaning of *cependant*.

Another aspect of the relationship between meaning and collocations which is left aside in these approaches is the fact that collocations not only build meaning, it also constrains it. For instance, the word *worm* can elicit a wide variety of representations, from *Dune*’s majestic sandworm to the little worm of remorse gnawing inside of you. The adjective to which *worm* can possibly collocate constrains drastically this representation; there is indeed a striking difference between Poe’s ‘Conqueror Worm’:

And, over each quivering form,
The curtain, a funeral pall,
Comes down with the rush of a storm,
While the angels, all pallid and wan,
Uprising, unveiling, affirm
That the play is the tragedy, “Man,”
And its hero, the Conqueror Worm.

POE, Edgar A., *The Conqueror Worm*, 1843

and the more common ‘tiny worm’:

A tiny worm one day was found
Within a timber crawling round ;
The shipwright thought a little worm
Like that could surely do no harm,
The plank is in the vessel’s side,
The vessel o’er the waves does ride;
Freighted with many hopes and fears,

She battles with the storms for years;
 But when most needed it is found
 One of her timbers is unsound
 A fearful storm the vessel racks,
 Her planks are strained, the timber cracks;
 Down goes the vessel in the storm,
 And all for just that little worm.

CICELY Cousin, *The Budget*, 1854 (Google Books)

Interestingly, both of these quotes show a close association of worms with death; something that the collocates of *worm* do not reveal (with the possible exception of *canker*: the five first collocates of *worm*, according to the COHA, are, left-side: *silk*, *canker*, *poor*, *laidly*, *white*; right-side: *turn*, *hole*, *dieth*, *snake*, *gnawing* ('Worm Fence' has been ignored and the pair (*turned*, *turns*) has been combined into its lemma).

This constraining of the meaning could be described through Bayesian learning. As the Semantic Atlas shows, the meaning of a construction, no matter how simple, is geometrically extended in some abstract semantic space, and presents several semantic domains. When presented alone, a word will probably weakly elicit all those different meanings, so that its meaning can be described as a uniform prior distribution. Then, as the context of the utterance is given, this distribution will peak around the most relevant meaning domain of the semantic map associated with the word. Collocation, and further co-occurrence, is thus a good way to reduce the polysemy of words within a particular utterance. However, it may not be that those 'peak-meanings' are grasped by the language users, as the meaning of each individual construction will probably be blended into a single representation corresponding to the whole utterance.

The contribution of collocation to construction's meaning is thus real, but limited: one cannot reduce the meaning of a construction to its collocates. Yet the knowledge of the collocates of a given construction, and especially so for atomic and substantive constructions, informs us of its meaning and, importantly, of the structure of this meaning. However, the meaning of some constructions (especially the complex and the schematic ones, but also substantive and atomic constructions carrying a pragmatic meaning) completely eludes a characterization through collocates. Finally, collocates not only contribute to the meaning of a construction, it also constrains and specifies this meaning within a compositional construct.

1.5 Meaning

We shall end this short presentation of Construction Grammar with the question of meaning — an essential question, as a construction is chiefly defined by a semiotic association between a form (possibly schematic) and a meaning. Though we already partly addressed the question of meaning in the previous sections on paradigms and collocates, there remains a few points which seems worth addressing.

1.5.1 Meaning in the semiotic view

The semiotic nature of the construction is at the core of Construction Grammar:

The basic tenet of Construction Grammar [...] is that traditional constructions — i.e., form-meaning correspondences — are the basic units of language.

(Goldberg, 1995, p.6)

Grammatical constructions in construction grammar, like the lexicon in other syntactic theories, consist of pairings of form and meaning that are at least partially arbitrary [...]. Thus, constructions are fundamentally symbolic units [...].

(Croft and Cruse, 2004, p.257)

All levels of grammatical analysis involve constructions: learned pairings of form with semantic or discourse function, including morphemes or words, idioms, partially lexically filled and fully general phrasal patterns . (Goldberg, 2006, p.5)

The basic unit of grammar is the construction, which is a conventionalized pairing of form and meaning (Traugott and Trousdale, 2013, p.3)

and more so, expectedly, in Sign-Based construction grammar:

Sign-Based Construction Grammar [...] assume [...] that grammar is an inventory of *signs* — complexes of linguistic information that contain constraints on form, meaning, and use — and that constructions are the means by which simpler signs are combined into more complex signs.

(Michaelis, 2013), emphasis original

This pairing is further evidenced in Traugott and Trousdale (2013), for they represent the construction through the following scheme:

[[Form] ↔ [Meaning]] (Traugott and Trousdale, 2013, p.7)

There are, however, two issues with this semiotic view of constructions. The first one is that it hides a theoretical asymmetry between form and meaning in the concrete definitions of constructions. The second is that it virtually ignores polysemy, despite the pervasiveness of this phenomenon (Ploux, 2011). We shall address them successively.

Though form and meaning should be addressed on equal ground in this semiotic stance, the construction is, most often, defined chiefly by its form, rather than its meaning. For instance, in her seminal work, Goldberg (1995) devotes a whole chapter to the *way* construction — as in ‘King Richard the Lion Heart, bankrupt of wealth and glory, is plundering his way back to England after ten years on his crusade.’ (from the opening prologue of Ridley Scott’s 2010 movie *Robin Hood*) — whose structure is stated in the very first paragraph; the form is not an issue, it is given by the corpus evidence available (the real issue with which Goldberg deals rather lies in the constructional status of this particular structure). However, the meaning of the construction is not immediately clear, and indeed give rise to a precise and detailed discussion, from which it appears that the *way* construction has in fact two meanings: a ‘means’ meaning, which specifies the means of the movement:

She stretches her thin neck forward, plows her way through a mass of stalled travelers and covers thirty more feet of the Moon.

GILBOW S. L., *Who Brought Tulips to the Moon?*, 2007 (COHA)

and a ‘manner’ meaning:

He craves to lie down at night without these apparitions that have wormed their way under his fingernails and can’t corkscrew themselves back out.

LIVERS, Paulette, *Soldier’s Joy*, 2012 (COCA)

The example of King Richard pillaging on his road back from the Third Crusade would illustrate yet a third meaning of the construction (Israel, 1996).

At this point of Goldberg's reasoning seems to appear a duality between two views. In a section named 'Constructional polysemy', where she investigates the relationship between the two different meanings, she states at some point that 'the constructions are related by polysemy inheritance' (Goldberg, 1995, p.210), implying that two constructions of same form but two different meanings should be distinguished (quite consistently with the semiotic form-meaning pairing at the core of the definition of a construction). There would be thus the *way* construction, 'means' meaning, and the *way* construction, 'manner' meaning. However, this interpretation is never made explicit, and the conclusion of the chapter indicates otherwise:

The *way* construction demonstrates the need to recognize *constructional polysemy*, parallel to the polysemy often posited for lexical items and grammatical morphemes. (Goldberg, 1995, p.217)

The work of Traugott and Trousdale (2013) also suggests that form and meaning are not equal. Indeed, they establish the crucial distinction (to which we shall come back in the next chapter) between *constructional changes* and *constructionalizations*, the former referring to changes of features (e.g. phonetic) within a construction, the latter to the emergence of a new construction, 'the creation of a new form_{new}-meaning_{new} pair' (Traugott and Trousdale, 2013, p.22). However, there might be some problems with this characterization. First, semantic expansion can happen, so that the meaning of the construction will be altered, at a varying extent. As this involves a change of the semantic features of the construction, it is described as a constructional change. Similarly, phonetic attrition of the form is a possibility (they give the example *a lot of > allota*) but this won't be described as the creation of a new construction either, since only the phonetic features of the construction are changing. Yet, semantic expansion can occur concomitantly with phonetic attrition, leading, technically, to a form-meaning pairing where both the form and the meaning are new. And still, none of the changes involved can be assimilated to a constructionalization.

How can we explain this? A first proposal would be that constructionalization must imply big, simultaneous shifts of form and meaning; gradual semantic expansion, and slow, continuous phonetic reduction may not be sufficient to speak about 'new' form and 'new' meaning. In this perspective, the difference between constructional changes and constructionalization would just be a matter of degree. A simpler explanation is at our disposal though. The constructionalization giving rise to *a lot of* as a modifier is described as follows:

$$[[N_i \text{ [of } N_j]] \leftrightarrow [\text{part}_i - \text{whole}_j]] > [[N \text{ of } N_j]] \leftrightarrow [\text{large quant} - \text{entity}_j]$$

(Traugott and Trousdale, 2013, p.25)

It seems clear that what is crucial in the constructionalization is the reanalysis from $[N_i \text{ [of } N_j]]$ to $[[N \text{ of } N_j]]$, that is, a change in the internal cohesion of the form. Semantic change happens throughout this long diachronic evolution, to what (Traugott and Trousdale, 2013, p.28) pre-constructionalization constructional changes and post-constructionalization constructional changes, respectively coined 'PreCxzn Ccs' and 'PostCxzn Ccs'. The specific semantic change happening during the reanalysis step is no different from all other semantic changes occurring throughout the long process, except maybe in degree, and what really matters is this internal, formal reanalysis.

Albeit Traugott and Trousdale (2013) always put a great care at stating the ‘meaning’ part of a construction in this double brackets notation of theirs, it appears from their work that constructionalization require formal analysis, while meaning change can be associated with a mere constructional change only. Hence, the form and the meaning do not seem to be treated equally. Usually, in Construction Grammar, the construction is apprehended through its form, not its meaning. Also, it is the form which is available in corpora, not the meaning, which may explain, from a methodological point of view, this prevalence of form over meaning in the study of constructions.

I would like to argue that this asymmetry, though it may be a little inconsistent with the definition of construction usually provided, is not an issue, and rather is to be expected. Indeed, and this will lead to our second point, it seems a bit odd to look for ‘the’ meaning of a construction, while its meaning is more accurately described as polysemic, and semantically extended. For instance, (Traugott and Trousdale, 2013, p.153) study the construction $[[V_{TRi-able}]_{Aj} \leftrightarrow [[\text{‘can be SEM}_i\text{-ed’}]_{PROPERTY}]_j]$, where TR stands for ‘transitive’. In doing so, they come to distinguish examples of ‘full sanction’ of the pattern, as *squeezable* and *fixable*, from ‘partial sanction’, such as *despicable* or *drinkable*, the latter adding an ‘esthetically pleasing’ value which is not derived from *-able* alone (though we can say that it is motivated through euphemism).

In French, from which *-able* comes, we find the same diversity, between constructs such as *opérable*, *finissable*, *mesurable*, etc., whose meaning is compositionally derived from the construction, constructs such as *rentable*, *incommensurable* and *potable*, *miserable*, entirely opaque nowadays as the corresponding verbs only survive in this construction, or others such as *jouable*, whose meaning is partly non-compositional (it cannot be translated by ‘playable’, despite the translation *jouer* \leftrightarrow *play*, because it refers to a very specific meaning of *jouer*), *pendable* (which has nothing to do with *pendre*, ‘to hang’ as to execute), *vénérable*, *agréable*, *semblable*, *probable*, *coupable*.

Many of these constructs can be seen as instances of speciation, that is, they have become independent constructions. There are however some nuances of the construction; if it usually indicates the property that ‘one can V_{TR} whatever is characterized by this property’, some uses of the construction show a different meaning. This is the case of *épouvantable*; if something has the property of being *épouvantable* (‘frightening’), it means that this thing can *épouvanter* (‘frighten’), not that it can be *épouvanquée* (‘frightened’). This parallel use of the construction is even compatible with intransitive verbs, such as *convenir* (‘to suit’), as in *convenable*. Other examples show a discrepancy with the main pattern while still being related to it, such as *profitable*, indicative of the property that ‘one can profit *from* whatever is characterized by this property’, and also *délectable*, or yet another variant *intraitable*, ‘with which one cannot deal’. Another family of *-able* constructs, which takes nouns as arguments, shows yet another meaning of the construction, ‘worthy of N’, ‘full of N’, as exemplified by *véritable*, *pitoyable*, *raisonnable*.

At this point, we can either posit a full range of very closely related *-able* constructions, or consider that there is only one single construction (and some speciated, independent offsprings), covering a range of meanings. To do this, we must slightly alter its formal characterization, and instead posit a less specified paradigm, whose definition can only be given by the actual uses of the construction, in the spirit of Radical Construction Grammar, according to which the categories of language are

only defined with respect to the arguments appearing in different constructions (Croft, 2001). In any case, it seems profitable to posit that the meaning of the construction is not a precise and discrete entity, but that it is extended, and correspond to a domain in some semantic space, rather than to a closed and specific definition.

1.5.2 Substantiating meaning

We will now shortly deal with the substance of meaning. This is a vast question, that we cannot explore nor survey; we shall be contended here with a brief discussion about what can occasionally be referred to by ‘meaning’ in this work. A more detailed view of the representation of meaning we subscribe to in the modeling of language change will be the object of chapter 7, and the present reflections should be held as complementary and tangential to the somehow more pragmatic point of view exposed there.

As has been argued so far, the meaning of a construction is not easily captured by the sole datum of its collocations and paradigmatic alternations. Both show something of a construction’s meaning, but none is able to give a description getting close to exhaust its complexity. To deal with this question of semantic complexity, Wierzbicka (1991) takes an interesting approach which we shall briefly explain. She criticizes three currently held positions on meaning: that meaning cannot be analyzed so that linguistics shall not deal with it; that meaning is by nature fuzzy, so that defining a construction would be pointless and vain; that meaning can be explained with simple words, especially the infamous definition of ‘to kill’ by ‘to cause to die’. However, her approach may look like, at first glance, to this latter proposal.

Over the years, she has established a list of ‘semantic primes’, ‘simple’ concepts which, first, cannot be defined by any other words, and second, are attested in all known languages. Such semantic primes include things such as ‘I’, ‘you’, ‘say’, ‘want’ ‘something’, ‘good’, ‘bad’, etc. She then claims that the meaning of any construction can be characterized by using these semantic primes. An interesting, yet mildly convincing example, would be her definition of *fire*:

at some times, something happens in some places
 at the same times, one can see things in these places because of this
 at the same times, one can feel something in these places because of this
 after this, something in these places is not the same
 at some times, people do something in some places because they want this to
 happen (Wierzbicka, 1991, p.86)

Her definition of *cat*, actually more interesting, was too long alas to be reported here.

What is most striking in those definitions is the way they combine the semantic primes, so as to describe how people are actually experiencing the characterized notion. This is of course particularly valuable, because when asked to explain some meaning, people are prone to set a situation of use (by linguistic means most of the time) in which the meaning is apparent. Though she does not state this explicitly, Wierzbicka rationalizes, through the use of semantic primes, this somehow primal of way of providing substance to meaning. The arguable success of her enterprise highlights that meaning cannot be captured by the sole linguistic means, only conveyed by them; collocations, alternations, help in delineating a territory, but are mute as for its contents. This failure of language-based description of meaning can be understood and expected once acknowledged that meaning refers to the entire range of human experiences in

their own inner and external environments. Semantic primes can conversely be used to describe the meaning of a construction, but then it cannot be derived a posteriori from the sum of all the utterances in which the construction appears.

The specific method of Wierzbicka, i.e. resorting to semantic primes, might not be optimal, and has no vocation to be definitive. It has shown its use in making apparent the nuances between similar words in different languages (Wierzbicka, 1997), yet it suffers from some limits. To begin with, it separates linguistic items into two separated classes: the semantic primes in the first, all others in the second. If this dichotomy were established, it could be conjectured that one could see traces of it, — be it psychological, for instance. Such a dichotomy could also be reflected in language change, which does not seem to be the case. The deictic *this* and the conjunction *because* appear both in the list of semantic primes, despite being attested loci of change. Semantic primes such as *I* and *you* have not equal importance in all languages — especially in those where the flexion of the verb carries this information, such as Latin or Spanish. Their dominion exhibits thus a high degree of variation between languages, so that we have no way to compile a list of semantic primes which will present the same range of uses in all languages. However, this might be a matter of detail.

Another limitation is that it is not clear whether all elements of language can find a definition in terms of semantic primes. This would chiefly be the case for grammatical elements, and, most of all, discourse markers, the most elusive linguistic items to any attempt at definition, circumscription, or even translation, for they have seldom cross-linguistic satisfying equivalents. It might also be that some languages have internalized new semantic primes, required to describe the meanings of some of their words, but not attested in all languages. Temporal semantic primes, for instance, seem rather hard to be passed over for the description of most Indo-European languages. This would open up a new line of investigation, for phylogenetically related languages would share the same primes, assuming they are indeed inherited.

A last point concerns the definitions themselves. Indeed, they not only make use of semantic primes, they also combine them syntactically, or at least representationally. Are there also syntactic primes, as there are semantic primes? Such a question is especially relevant in the framework of Construction Grammar, in which there is an ontological continuum between complex schematic constructions such as argument structure, and atomic, substantive constructions, like morphemes and words. Could the transitive construction be for instance such a universal syntactic prime?

Actually, Goldberg (1995) has criticized Wierzbicka's work, but not on these grounds:

She then proposes that the entire semantics of any lexical item can be captured by paraphrases involving these atomic semantic primitives combined in determinate ways. We have taken rather the opposite approach to semantics, arguing that lexical items are associated with rich frame-semantic or encyclopedic knowledge, and that decomposition into atomic elements is impossible.

(Goldberg, 1995, p.224)

The Frame Semantics is superficially defined in the following way:

Many researchers have argued that words are not exhaustively decomposable into atomic primitives. However, it is not necessary to conclude that meanings have no internal structure. Instead, it has been argued that meanings are typically defined relative to some particular background frame or scene, which itself may be highly

structured. I use these terms in the sense of Fillmore (1975, 1977b) to designate an idealization of a “coherent individuatable perception, memory, experience, action, or object”.
(Goldberg, 1995, p.25)

Yet, we saw that Wierzbicka’s definitions can be said, in a sense, designed in such a way that they set one or several frames in which the meaning is apparent. Furthermore, the definition of *cat* she provides clearly shows that meaning is informed by and filled with encyclopedic knowledge. Indeed, her definition lists characteristic behaviors of cats, characteristic interactions people have with cats, characteristic physical features of cats, all of those items being expressed in the perspective of an experiential frame.

What we might argue is that Wierzbicka’s approach leaves polysemy aside; words, possibly constructions, have only ‘one’ meaning. As she devoted many efforts to explain the semantics of culturally salient words, including emotions, relationships, even objects, all sorts of linguistic items which indeed have a very strong and delineated meaning, this focus on a monosemic approach makes perfectly sense. However, in this work, we are concerned with semantic expansion, a phenomenon which evidences the multi-faceted meaning of a given linguistic item. It can even be that one of these semantic facets is ruled out in a given utterance, as in:

Est-ce que je vais rester là planté comme une idole [?], se dit à lui-même le bretteur impatienté de ses propres tergiversations [...].

Am I going to stay right here standing still like an idol?, the swordsman said for himself, losing patience with his own shilly-shallying.

GAUTIER, Théophile, *Le Capitaine Fracasse*, 1863 (Frantext)

Here the standing still of the character suppresses all idea of movement implied by the lexical root of the *aller* future construction (the French equivalent of *be going to*).

Although the relevance of Wierzbicka’s framework to our concerns is limited, it is important to remind that the semantic units that a linguistic item encompass in its polymorphous meaning can be given true substance, as Wierzbicka’s enlightening definitions strongly illustrate. Furthermore, she also investigates (Wierzbicka, 1991) the relationship between meaning of an atomic construction, and meaning of a construct of several atomic constructions, concluding that her metalanguage based on semantic primes is also operative in this latter case. This is a point of uttermost importance, as it shows that meaning of constructs, and from this, meaning of utterances, are contiguous to meaning of atomic constructions: they can be described in the same semantic framework, or semantic space. This has far-reaching consequences about language, as it suggests that lexical items are already ‘small’ utterances in their own. For instance, uttering *Attention !* (‘Beware!’) is a small utterance, and it conveys a meaning roughly rendered as: ‘There is something which is dangerous for you in your immediate surroundings that can result in serious harm, and of which you seem to be unaware, but you can still probably avoid it by appropriately reacting in the briefest possible delay.’ However, not all linguistic items (even fully lexical) may be seen as a minimal utterance, and most of them might require other linguistic items to get a proper meaning. In this perspective, we would need different meaning spaces, one for utterances, and one for the components of the utterance, that is, the constructions.

1.6 Summary

In this chapter, we discussed the main notion of Construction Grammar, which is the construction. We professed that constructions alone might not suffice to describe linguistic knowledge, and that they must be complemented with paradigms, i.e. idiosyncratic sets of constructions (which can, however, be part of the definition of the constructions they appear in), and with an apprehension of collocational frequencies (which constructions suit which), that can only be learned through a long exposition to the language. Our goal was not to contribute to a theoretical debate within Construction Grammar, but to lay down some useful notions which will enable us to describe specific instances of language change as they are manifested in corpora. The choice of Construction Grammar may be criticized on the basis that it is not a unified framework, and that it is instead declined into multiple variants, still in mutation. However, Construction Grammar is a flexible usage-based framework, which describes utterances as they appear and only as such, adapting to the linguistic data and not constraining it theoretically. As such, it makes for an efficient analysis tool to investigate empirical data in a diachronic perspective, where notions such as acceptability or grammaticality (in the sense of ‘well-built’) are always changing and thus cannot be tracked down.

Especially, we recalled and discussed a few notions which will be frequently called forth in the subsequent chapters:

- Constructions are the basic items of language, reflecting the cognitive knowledge of the language users. They are to be contrasted with constructs, which are the speech realization of a specific combination of constructions. Constructions can be either simple or schematic, and present in the latter case a free slot in their definition, that we represent with curly brackets {.
} (e.g. the English construction [because of {NP}]).
- These slots are not entirely free and are subject to strong constraints regarding the constructions that they can host. The set of constructions appearing in a given slot of a schematic construction is what we called a paradigm. This notion will be extensively discussed in chapter 5.
- We also argued that speakers have some knowledge of frequency, since the different constructions are not stored in their memory with the same weight. This knowledge accounts for both the frequency of occurrence (some constructions are more frequent than others) and the frequency of cooccurrences (once a given construction has been uttered, speakers have some expectations regarding which constructions are likely to follow, e.g. *safely*, *late* and *early* commonly comes after *arrive*).
- Crucially, the two notions of collocation and that of colligation have to be contrasted with each other: the first one refers to a syntagmatic cooccurrence, the second one to the specific attraction that a construction can manifest towards the hosting paradigm of another schematic construction.

- Finally, we advocated that the meaning of a construction cannot be reduced to the datum of its collocations and colligations, though a change in these frequency features does imply a semantic change.
- Collocations, colligations, paradigm contents and organization, and semantic changes, have clear empirical counterparts in corpus data. However, there is no definitive criterion to distinguish constructions from mere constructs. The quality of the extracted corpus data thus relies on the successful identification of these constructions.

Before turning to actual corpus data, we first have to discuss one of the most intriguing and yet pervasive phenomena of language change, namely grammaticalization. Indeed, our aim in this work is to seek after empirical specificities of this phenomenon, and to understand by which mechanisms of change they arise. Having set the frame of Construction Grammar shall be of great convenience in the discussion of these matters.

Chapter 2

Grammaticalization as a phenomenon

In the present chapter we focus on the notion of grammaticalization, which can be crudely defined as follows:

Grammaticalization. *A grammaticalization is a process by which a given construction acquires with time a more grammatical meaning.*

This definition immediately raises several important questions: Is this process specific? Is it always the same in all instances of grammaticalization? What are its specific signatures? Where does it take place, what does it involve, and which forces are shaping it? What are the typical timescales of grammaticalization? And how can one assess how grammatical a meaning really is?

Before addressing a few of these questions, we will first provide two original illustrations of grammaticalization processes, so as to ground the following discussions on concrete instances of this phenomenon. After reviewing some of the main issues surrounding the theoretical notion of grammaticalization, we shall end this chapter by a section presenting our own views on the matter.

2.1 Two examples of grammaticalization processes

2.1.1 Du coup

French *coup* (or its older spelling variant *cop*) has long been a highly polysemic word, at least since Middle French (Marchello-Nizia, 2006). As such, it has been used in numerous constructions, either lexical ones (*coup d'état*) or grammatical ones, among which *du coup*, on account of its pervasive use in contemporary French, is probably the most striking example. Representative of this vivid polysemy is the list of possible translations proposed by the Semantic Atlas :

Translation of *coup* in English:

bang, bell, blow, bow, jerk, job, knock, move, punch, shock, shot, slap, splash,
stab, stroke, throw, time, trick, tug (Semantic Atlas)

As a result, *coup* appears in numerous grammatical constructions, such as *sur le coup* ('at the moment'), *pour le coup* (no satisfying English translation for that matter; interestingly, in the examples of translation given by the Linguee search engine, *pour le coup* is not translated in the majority of cases), *beaucoup* ('a lot'), *coup sur coup* ('in quick succession' according to the Collins dictionary), *après coup* ('right afterwards'), *au coup par coup* (progressively, as things present themselves individually), *tout à coup* ('all of a sudden'), and of course, *du coup*, to which the Collins dictionary gives the limited translation 'as a result'.

The grammatical meaning of *du coup* as a connective has already been studied in a synchronic perspective by Rossari and Jayez (2000), yet there is no study of its diachronic emergence in the literature that we know of. This development has, however, already been described as an instance of grammaticalization (Rossari, 1998), though this view may not be fully consensual. Indeed, the change undergone by *du coup* could be seen as an instance of pragmatization, a process which can be considered as distinct from grammaticalization, as shall be discussed in the next section. We nonetheless believe that the development of *du coup* presents all the most relevant marks of a grammaticalization and therefore consider it to be a clear and illustrative example of this phenomenon.

Earlier occurrences of *du coup*

The composite *du coup* had no independent existence prior to its grammaticalization. It arose as a syntagmatic, fully compositional sequence, of *de* (meaning 'of' or 'from' according to the context) and *coup*, whose oldest example in Frantext dates back from the twelfth century:

Nicholas fu iriez **du cop** qu'a receü;
Pour un seul petitet qu'il ne l'a abatu;

Nicholas was angry **for the blow** he had received, which but failed to slain him.
DE PARIS Alexandre, *Roman d'Alexandre*, 1180, p. 36 (Frantext)

Very quickly, it appears that *du coup*, or its longer variant *de la force du coup*, may convey a further meaning of result, or consequence. It is reflected by an atypical syntactic position, and unexpected repetition:

Sy l'assena en l'escu sy qu'il le fendy en deux, et le brach senestre ly / endormy de le forche du cop. Et **de le forche du cop** ly entra ly bastons bien demye paume dedens le cuisse, et fu ly rois sy angousseux qu'il jetta ung grant cry.

And so he [=Henry] stroke on the shield that it broke apart in two, and the left arm got numb by the strength of the blow. And **by the strength of the blow** his [=Henry's] staff thrust into his [=the King's] thigh for at least half a palm, and the King was so distressed that he fiercely cried out.

Ysaïe le Triste, 1400, p. 304 (Frantext)

Of course, the lexical reading is still pregnant, and there is a clear blow in this context to which *coup* refers, but still, it is unusual to repeat twice in a row a construction as substantially heavy as *de le forche du cop*, which shows a commonality of this construction (especially in this particular novel, whose author is anonymous). Other examples of a possible 'as a result' reading are numerous, such as:

Aussi tout que Roland eu finé sa parouille, son oncle l'empereur, moult indignez contre luy, a grant melancolie de son gant destre [...] va donner a travers le visaige

de Roland et l'attaint tellement sur le nés que le sang en vint habondamment **du coup**.

And as soon as Roland ended his speech, his uncle the Emperor, furious against him, with great sadness came to knock the face of Roland with his left glove, and he hit him so strongly on the nose that it bled abundantly **from the blow**.

BAGNYON Jean, *L'Histoire de Charlemagne*, 1465-1470, p. 32 (Frantext)

Interestingly, *du coup*, by the sixteenth century, can also appear in an interpolated clause (which may be motivated by poetic metrics):

La Rivière au contraire, en frappant ce grand Roy
Mangelard, d'un caillou le meit en tel desroy,
Que **du coup**, le cerveau par les naseaux sortoit :
Et la terre du sang toute couverte estoit.

La Rivière, on the contrary, striking this great king Mangelard with a stone, put him in such a disarray that **du coup** the brain went out through the nostrils, and the ground was all stained with blood.

MACAULT Antoine, *Le grand combat des ratz et des grenouilles*, 1540, p. 8 (Frantext)

In this sentence, an appropriate English translation could be 'because of the blow', indicative that the inference of a result, or a consequence, can already be strongly invited by the utterance. Such an interpolated clause, with a strong 'consequential' reading, is still highly related with the concrete, lexical meaning of 'du coup', which can even be fully specified:

Et Altoviti, **du coup d'espée qu'il avoit reçu**, demeura mort sur la place.

And Altoviti, **because of the sword blow he had received**, remained dead on the square.

DE L'ESTOILE Pierre, *Registre-journal du regne de Henri III : t. 5 (1585-1587)*, 1587, p. 186 (Frantext)

This utterance is especially interesting, since *du coup* appeared so far immediately or closely after the actual depiction of a blow; here the blow, though it also refers to an event described a few lines before in text, is nonetheless described within the interpolation. The whole excerpt, not fully quoted here, is all the more remarkable as it makes use four times of *du coup* (with the variant *dudit coup*) in only two sentences. A common trait of all those occurrences is that the result is always a violent and often deadly consequence of the mentioned blow. In the same register, we find, in the seventeenth century, an occurrence of *du coup* where the result is somehow indirect:

Ayans repris de nouvelles lances, ils recoururent pour la seconde fois avec plus de fureur et de force que la premiere, les rives de la Tamise resonnerent **du coup**, et leurs claires ondes en furent troublées.

Having equipped new spears, they dashed for the second time with more strength and fury than the first, the shores of the Thames resounded **from the blow**, and their clear waters were troubled by it.

D'AUDIGUIER Vital, *Histoire trage-comique de nostre temps, sous les noms de Lysandre et de Caliste*, 1615, p. 369 (Frantext)

In the second half of the seventeenth century, and throughout the eighteenth century, *du coup* no longer seems to be used idiosyncratically, and most of its uses are purely compositional. The implicit ‘result’ reading of *du coup* is seldom carried by *du coup* alone and must be complemented by more explicit constructions:

Le maure saisit cet instant pour lui décharger sur le cou un revers de son épée, dont il connaissait la trempe ; **et telle fut la force du coup**, que l’animal en tomba roide mort sur la place, au grand étonnement de tous les spectateurs.

The Moor seized this opportunity to deliver him on the neck a revers of his sword, of which he knew the sturdiness; **and such was the strength of the blow**, that the beast fell stone dead on the square, to the great amazement of all the witnesses.

LESAGE Alain-René, *Histoire de Guzman d’Alfarache*, 1732, p. 165 (Frantext)

A few occurrences akin to *mort du coup* are nonetheless encountered now and then.

Onset of grammaticalization

This situation would change in the nineteenth century. Quite early, we encounter an occurrence where *du coup* has its contemporary meaning: no actual blow is mentioned in the context, and the result is not physical either (as were stunned, dead, fell down, etc.).

On voulut l’éloigner, mais il était déjà trop tard : on avait touché l’homme, porté ses bagages, **et du coup**, tous les habitants des grottes succombèrent au terrible fléau.

They wanted to take him [= a plague victim] away, but it was too late: they had touched the man, carried his belongings and, **as a result**, all inhabitants of the caves fell to the terrible plague.

POTOCKI Jean, *Manuscrit trouvé à Saragosse*, 1815, p. 600 (Frantext)

Interestingly, in the same book, an occurrence of *du coup* in its previous use is also encountered:

ensuite, il me donna un coup de pied, de ceux qu’on appelle crocs-en-jambe, et me fit tomber le nez dans le ruisseau. **Étourdi du coup**, je me relevai couvert de boue

Then, he gave me a kick, of those we name ‘croche-pied’, and he made me fell down into the river. **Stunned from the blow**, I got up, covered with mud.

Idem, p.504 (Frantext)

However, we can certainly rule out the former occurrence. Indeed, the *Manuscrit trouvé à Saragosse* had a complex publishing story. Originally written in French by a Polish writer, the manuscript was not published at first, and parts of it were lost. When it was finally published in France in 1989, the missing parts were replaced by a contemporary translation into French from a Polish translation of the original French version, made in 1847. Since the digitalized text on Frantext comes from this partly retranslated version, we cannot know for sure whether this particular excerpt dates from the early nineteenth century, or the late twentieth, but the latter is most probable, as the next occurrence of *du coup* with this meaning appears only about thirty years later:

Eh bien, capitaine, voilà un petit contre-temps. **Du coup**, nous n’arriverons pas avant minuit

Well, captain, this shall delay us a little. **As a result**, we will not make it before midnight.
 HUGO Victor, *Le Rhin : lettres à un ami*, 1842, p. 283

This occurrence is remarkable for three reasons at least. The first one is that it is extracted from a special novel, in the sense that it is made of fictional letters, partly drawn from a travel diary, partly made-up as a writing process, partly copied from real letters. It is thus deliberately written in a much less formal tone than the common lettered or academic style. The second one is that *du coup* appears in a dialogue, in what Christiane Marchello-Nizia has coined ‘represented orality’ (Marchello-Nizia, 2012). It is thus a literary device aiming to reflect language as it was spoken. Very often, the first occurrences of a novelty in textual databases occur in this kind of represented orality. Here is no exception.

The third feature, and the most striking to be sure, is the complete abruptness of the change. For centuries, we encountered contexts in which *du coup* could be at least partially read with the meaning ‘as a result’. However, the contexts in which this reading could happen were very restricted; an actual, physical blow was always mentioned in the frame of the utterance; the consequence of the blow was also direct and physical. Even if this use showed a timorous expansion by the sixteenth century, it then retracted and almost disappeared in the eighteenth, leaving only occurrences with an explicit blow and an explicit physical result, as we already found six hundred years earlier. It thus seems that no further conditions were met for the grammaticalization to occur; which means, it was probably possible as early as the thirteenth century, and yet waited centuries to happen.

Actually, a possible explanation for this unexpected and long delayed actuation would be that, at this time, *coup* itself is undergoing a semantic expansion. This can be seen from the diversification of the collocates paradigm (Table 1.1), and the frequency profiles of various uses of *coup* (*un coup*, *coup de*) show as well a large increase of frequency at this time. It might thus be that this overall semantic expansion would have blurred the meaning of *coup*, allowing for *du coup*, which already elicited a resultative meaning, to be fully reanalyzed in that sense.

Entrenchment and further grammaticalization

In the same year 1842 as this ‘first’ occurrence in our corpus, another book is published in which we can count at least two unambiguous uses of *du coup* as a conjunction expressing a causality link between two events. As soon as 1845, this conjunctive causality meaning is further extended to a more pragmatic meaning, once more in a context of represented orality:

Du coup, reprit Farrabesche, je compris M.. Bonnet, il n’eut pas de grandes paroles à me dire pour m’expliquer ma besogne.

‘So’, Farrabesche resumed, ‘I understood Mr. Bonnet, and it did not take a long speech for him to explain my work to me.’

BALZAC Honoré de, *Le Curé de village*, 1845, p. 779 (Frantext)

In this context, there is no event of which ‘I understood Mr. Bonnet’ is a result. The speaking character has been interrupted in his speech, and so he uses *du coup* to mean ‘Well, I will not carry on with what I was saying to you, the important thing

is that, I understood Mr. Bonnet, so...? The timescale of the meaning expansion which led from a causality conjunction to a pragmatic marker is of three years, to be compared with the centuries required for the first step of the grammaticalization to occur. This example hints at the fact that the grammaticalization involves an important step of meaning to pass from a lexical use to a grammatical use, while no such step is encountered when the grammatical meaning extends towards new, more grammatical uses.

As usual in grammaticalization, we can observe in this case the phenomenon of semantic retention: though the meaning of the form has evolved to new innovative uses, its former meaning is still present in the language:

Il partit — toutes les dames crièrent de surprise — s' élança au plafond, et retomba sur une cloche à fromage, qui se mit à vibrer **du coup**.

It popped up — all ladies shouted with surprise — sprang up to the ceiling, and fell down on the cheese cover, which came to vibrate **from the blow**.

FLAUBERT Gustave, *La Première éducation sentimentale*, 1845, p. 259 (Frantext)

At this time, the new use of *du coup* remains sparse. However, the construction seems to explore new semantic possibilities, as in the following, isolated occurrence:

Mais voyez donc, notre maîtresse, comme il est devenu? Je m' étonne bien comment vous l' avez acconnu **tout du coup**.

But do you see, mistress, how he became? I wonder how you recognized him **right away**. SAND George, *François le Champi*, 1850, p. 344 (Frantext)

This meaning of *du coup* as introducing something which is immediately subsequent in time, but not causally related to a physical blow, appears in other occurrences, without *tout*:

[E]t, comme mes yeux restaient toujours méchants, elle avait enfin poussé son cri des grands abois, ce miaou particulier et sinistre des chats qui se sentent en passe de mort. Toute ma colère **tomba du coup** ;

[A]nd, as my gaze remained harsh, she finally cried aloud, as if at bay, this peculiar and sinister miaow of the cats who feel on the verge of death. All my anger **fell down immediately**.

LOTI Pierre, *Le Livre de la pitié et de la mort*, 1891, p. 85

Occurrences such as *étourdi du coup*, *mort du coup*, disappear at the turn of the twentieth century. *Du coup* as a conjunction, introducing a result of a former event, becomes more and more pervasive, and can appear in various syntactic positions:

Bravay, l'agent du pacha, ayant acheté deux glaces de Saint-Gobain 250. 000 francs, les a revendues 1. 500. 000 au pacha qui, **du coup**, n'a plus voulu rien acheter de Saint-Gobain.

Bravay, the Pasha's agent, having bought two ice creams from Saint-Gobain for 250.000 francs, sold them for 1.500.000 to the Pasha who, **therefore**, never again consented to buy anything from Saint-Gobain.

GONCOURT Edmond de et GONCOURT Jules de, *Journal : mémoires de la vie littéraire, t. 2: 1864-1878*, 1878, p. 21(Frantext)

The occurrence of *du coup* at the end of the sentence, syntagmatically appearing after the 'result' event, apparently comes up a bit later:

Ne fais pas ça, c'est stupide, quand tu devrais gagner dix mille francs. Tu te fermes l'avenir **du coup**.

Don't do this, it's stupid, when you could earn at least ten thousand francs. It would shut down your future **then**.

MAUPASSANT Guy de, *Bel-Ami*, 1885, p. 9 (Frantext)

However, the position at the start of the utterance is highly favored, and remains so nowadays, contrasting with the former use of *du coup* according to which it usually followed the stated result of the blow.

By the end of the nineteenth century, we can thus consider that *du coup* has fully developed its grammatical use, both on the syntactic and semantic planes. Its current use may have been pragmatically extended, showing a sign of what is known as 'semantic bleaching', a phenomenon by which the meaning of a linguistic item becomes blurred as the scope of its uses widens. For instance, in the following occurrence, the meaning of *du coup* no longer carries any idea of consequence or result:

Océane le voyait aussi, ses yeux grands ouverts ne bougeaient plus, je ne savais pas si elle allait hurler ou vomir ; elle non plus ne le savait pas. **Du coup**, elle ne dit rien.

Océane was seeing it as well, her eyes wide open were no longer moving, I did not know if she was about to scream or throw up ; and neither did she. **In the end**, she did not say anything.

JENNI Alexis, *L'Art français de la guerre*, 2011, p. 124 (Frantext)

Summary

To sum up this long story, we can point out that *du coup*, literally 'from the blow', has almost always invited a further interpretation implying the idea of 'resulting from the blow'. This interpretation seemed to be at its peak in the sixteenth century, and yet, this meaning of result does not come to be encoded in the construction before the first half of the nineteenth century. As we have observed, there were apparently no new semantic or syntactic conditions at this time allowing for this grammaticalization to occur, which could suggest that the grammaticalization could have happened as soon as the thirteenth century. Yet, it did not. Triggering the grammaticalization may have required additional conditions, such as the constitution of a paradigm in the French language upon which the analogical reanalysis leading to the new grammatical uses of *du coup* could have been based. This hypothesis could be explored further through a compared study of other connectives.

Another striking feature of this grammaticalization is how suddenly it appears in the corpus. True, this may be due to a mismatch between the spoken language and its representation through written documents; yet, we saw that *du coup* first appeared in a context of represented orality anyway, suggesting that the gap between spoken and written language might not be as important as one can suspect in this case. This discontinuity is further marked by the observation that, as soon as the grammaticalization has taken place, further semantic expansions towards more grammatical meanings became immediately possible. The suddenness of the process, however, should not hide the fact that the occurrence of a novelty is far from its entrenchment in use, which can take decades, or more. Furthermore, to anticipate the theoretical discussion to come,

‘grammaticalization can be semantically instantaneous but formally gradual’ (Nicolle, 1998), so that it still makes sense to regard grammaticalization as a process instead of a sudden switch such as the one we witnessed. In chapters 4 and 5, we shall indeed deal with this gradual character, notably in terms of frequency of use, whose change is far from being as abrupt and discontinuous as the semantic reanalysis. It will give us an opportunity to investigate the grammaticalization of *du coup* from an empirical perspective, and to assess how much it concurs with the occurrence-based story that we have just outlined.

2.1.2 Way too

As *coup*, *way* is remarkably polysemic, and enters into multiple grammatical constructions, including the famous WAY construction which is a paradigmatic example in Construction Grammar (Goldberg, 1995). By comparison, its close French equivalent *voie*, with which it shares a common Indo-European root, has a very narrow range of grammatical uses. Among these grammatical constructions to which *way* belongs, *way too* may be among those who have developed the most recently. In Present Day English, *way too* serves as an intensifier of adjectives, strengthening *too* alone, so that *way* can be said to intensify *too*, which itself plays its usual role. This duality of interpretations (*way too* as a fully cohesive construction or as compositionally built up of *way* and *too*) has led to uses of *way* as an intensifier all by itself, but it should not conceal the fact that, diachronically, it emerged as a construction, and that the intensifier function of *way too* was attached to the construction as a whole and did not come from a compositional reading of its two components. Actually, *way too* in its contemporary meaning has emerged quite suddenly, just like *du coup*.

In the following, we shall provide a short historical enquiry to dig up the occurrences of *way too* prior to the grammaticalization, in order to understand how the syntagmatic co-occurrence of *way* and *too* could have led to such a constructionalization. We are not aware of any prior work describing the development of this construction. As for *du coup*, we assume nevertheless that it counts as a convincing case of grammaticalization, especially since the ‘intensifier’ target meaning of the evolution is consensually considered as grammatical, as attested by numerous examples of it throughout the lexicon of grammaticalization compiled by Heine and Kuteva (2002).

Way too as an accidental collocation

In the Corpus of Historical American English (COHA), *way too* appears as soon as 1820 (the first texts of the corpus are dated of year 1813):

If the opinions recorded in scripture relative to a future state of existence are to be relied on, as being dictated by God himself, and **in a way too**, that was not mistaken;

BALLOU Hosea, *A Series of Letters in Defence of Divine Revelation*, 1820 (COHA)

However, this occurrence is of no interest to us as *too* does not serve to intensify an adjective. A quite similar example shows nonetheless a more interesting sequence:

A disturber of the peace has only to run beyond a certain street to put himself out of the reach of his pursuers; for though the constable or watchman beyond his own

district might exercise the right belonging to every citizen to arrest an offender and detain him till the arrival of the proper officer, yet the process becomes **in this way too inconvenient and complicated** to be always attempted.

New Englander and Yale Review: July 1844: 453-470, 1844 (COHA)

In this occurrence, *in this way* applies to *the process becomes*, as does *too inconvenient and complicated*, unlike *way too* which applies to the adjective in the *way too* construction. Nonetheless, the construction [in {determiner} ({qualifier}) way] is one of the three most frequent contexts in which the string ‘way too’ appears.

The second one is due to the idiomatic *have one’s way* construction, as in:

she **has a way too** of knocking a smile on the head

SANDERSON John, *The American in Paris*, 1838 (COHA)

Here, *too* applies to the whole ‘she has a way’ part of speech, and it is not followed by an adjective (*too* being used with the ‘also’ meaning), so we can reject the occurrence right away. However, this particular context is compatible with an adjective:

I am afraid I have suffered Maude **to have her own way too much**, and for the future I must be more strict with her.

HOLMES Mary Jane, *Cousin Maud and Rosamond*, 1860 (COHA)

Here, *too much* applies to *Maude [has] her own way*, so that *way* and *too + adj* are less syntactically distant than they were in the previous context.

The third syntagmatic context in which *way too* appears most builds on the *give way* construction:

“Firing hastily,” said he, “the trigger may **give way too soon**.”

TUCKER Nathaniel, *George Balcombe, Volume 2*, 1836 (COHA)

the interpretation of which is close to the previous occurrence.

None of these contexts particularly invite a reanalysis, even if some occurrences make it slightly possible, as in:

I see plainly, Harry, that you have some scruples, and I caution you against **giving way too much to them**.

COOPER Susan Fennimore, *Elinor Wyllys, Volume 2*, 1846 (COHA)

But this ‘reading’ is, most probably, an anachronic projection of the contemporary use of the construction.

A fourth context may seem more promising, in that it already shows the *way too* construction, however in a more substantial form, *every way too*:

For, sir, I do not and can not like you! I do not say but that you are far too good and wise, **and every way too worthy** for such a girl as I am – and that you do me the very greatest honor by your preference, but still no one can account for tastes – and, sir, I can not like you – pray, pardon me!

SOUTHWORTH Emma Dorothy Eliza Nevitte, *The Missing Bride*, 1855 (COHA)

The alternation with *far too* in this particular occurrence is especially interesting, and revealing of the analogical attractions within the paradigm of modifiers. This *every way too* construction can appear with different adjectives:

You need not fear my falling into what you call the Pantheism of the moralists; it is **every way too cold** for my hot blood.

MITCHELL Donald Grant, *Doctor Johns: Being a Narrative of Certain Events in the Life of an Orthodox Minister of Connecticut*, 1866 (COHA)

However, there is a strong reason to assume that this *every way too* syntagmatic construct does not lead to the contemporary *way too* construction. In the COHA, there are three occurrences of such a construct, respectively located in the 1840-1849, 1850-1859 and 1860-1869 decades, while the *way too* construction only took off around the 1930s, as we shall see. Also, *every way*, attrition of *in every way*, is much more frequent alone than in combination with *too*, so that the occurrence of *every way too* is in no way indicative of an ‘every way too’ construction. It seems thus highly implausible to posit an attrition of *every way* to *way* which then came to recruit *too* within his construction, before losing it again in the course of its evolution. There are at last strong behavioral differences between the two constructions: [every way {adj}], where the adjective can be further intensified with *too*, is indeed not the same as the [way too {adj}] construction in which it is *way too* as a whole which serves as an intensifier. Of course, a reanalysis is always possible, but the scenario leading from the former construction to the latter would be, as we pointed out, convoluted at best.

Constructionalization of *way too*

The first clear occurrence of *way too* + adj is not related to any of those contexts:

‘Pass over them documents for Cherokee Hall, an’ don’t hold out nothin’ onto us.
We-alls is’ **way too peevis** to stand any offishul gaieties to-day.’

LEWIS Alfred Henry, *Wolfville*, 1892 (COHA)

This occurrence is certainly remarkable, as it is yet another example of represented orality. Furthermore, it aims to reflect a Southern way of speaking, hinting that *way too* may have originated from there. However, the author hailed from Ohio, yet he spent one year in the Southern regions (New Mexico, Arizona and Texas), the year before he wrote his *Wolfville* book, as indicated by its obituary notice in the New York Times. That his prose is a faithful reflection of the Southern dialect is thus a possibility, but no firm conclusion can be drawn.

This occurrence is isolated and the next one is only to be found almost fifty years later, once again in a context of represented orality:

“My dear fellow, your ears are **way too keen** for a gentleman’s gentleman,”
remarked Terrier. QUADLAND Preston, *Setter and Terrier*, 1939,(COHA)

The absence of occurrences in-between may be due to the scarcity of texts in the COHA, but from the 1940s they are to be found on a regular basis:

And such an explosion, he felt sure, would certainly blow everything for miles around into the smitheriest kind of smithereens. The activity of the vortex stayed high,’ **way too high**. The tiny control room of the flitter grew hotter and hotter.

SMITH Edward Edwin, *The Vortex Blaster*, 1941 (COHA)

It’s slavery all right, and especially for the married men. A married man that brings his family down here is stuck proper because if he ever had enough money to pay his family’s fare and to pay for his clearance he’d never come in the first place, probably, and by the time he sees what he’s up against he’s **way too deep** in to get out.

CLYDE Brion Davis, *Jeremy Bell*, 1947 (COHA)

A first explanation: give way as a catalyst context

At this point, we can venture to explain what happened. A first possibility would be that one of the contexts in which *way too* was occurring spontaneously, in the syntagmatic sequence of speech, gave rise to a reanalysis of some sort. The best candidate for this to have happened is the *give way* context. Indeed, *give way* is commonly used in a construction [give way to + {recipient}], which can also be used in a ditransitive form [give {recipient} way]:

Reason should confirm Our hearts' emotions, ere we **give them way**.

BOKER George Henry, *The Betrothal*, 1883 (COHA)

A qualifier intensified with *too* can occur next to *give way*, as in 'give way too much' and 'give way too soon', both attested in the COHA, while 'give way much' and 'give way soon' are not. Then, if the two are combined, we are left with a situation very akin to what we are looking for: [give way too {adj} to {recipient}] becomes, in the ditransitive construction, [give {recipient} way too {adj}]. The fact that the recipient is inserted between *give* and *way* interferes with the reading of *give way* as a unit, as it activates the reading of *give* alone and blocks the parsing of *way* as being part of the *give way* substantive construction. Then, *way too* + {adj} can be reanalyzed as a whole part of speech serving to intensify [give {recipient}].

Three problems are met with this scenario. First of all, *way too* starts being used sporadically in the 1930s, a time during which *give way* is in the decline: according to the COHA, this latter construction is used almost three times less than it was at its peak in the 1860s (frequencies per million words of 9.97 vs 3.98). Second, *give way* is seldom used in a ditransitive way, and if so, it is never intensified with *too* + {adj} in the COHA. The huge database Google Ngram does not show any occurrence of this type either. The third argument is that such a scenario would predict that the first uses of the *way too* construction should show traces of this 'give way too much' presumed origin. However, the first four occurrences all present a stative verb (either *be* or *stay*), and we have to wait the 1950s to find otherwise:

"It happens all the time. Either they use **way too much** or **way too little**."

MONSARRAT Nicolas, *Cruel Sea*, 1951 (COHA)

In another corpus, the TIME corpus of Mark Davies, the first five occurrences all show the same prevalence of stative verbs, with the possible exception of *start*:

The question was whether he had started **way too late**, and whether his enthusiastic amateurs would be much of a match for the old crowd.

The Contenders, 1957

But sex, at the rate O'Hara burns, is **way too hot** for the average U.S. movie exhibitor to handle;

The New Pictures, 1958

Figure it out again. You're **way too low**. I think it will be over \$100 million.

Nickel Counter, 1959

Both estimates are probably **way too high**.

Lawless Terror, 1961

Kennedy ordered the second strike reinstated, but it was too little and **way too late**.

Bay of Pigs Revisited, 1963

Way too as a member of the [way + {preposition}] paradigm

Another explanation is thus required. There is, actually, another possible source for *way too*, which is based on analogy. Indeed, *way* could be used with many prepositions (which all present adverbial uses), and some adjectives, such as *way above*, *way back*, *way before*, *way behind*, *way better*, *way more*, *way off*, *way out*, *way up*. Not all of these share the same path of formation, although most probably originated from the construction [a {adj} way {preposition}], *a little way* being especially frequent. For instance, it is the case for *way above*:

and then, in the dead hour of the night, when the moon was **but a little way above** the horizon, and divided her quiet empire with Mike's own nose, he crept forth silently upon his destined exploit.

KENNEDY John Peddleton, *Swallow Barn; or, A Sojourn in the Old Dominion*,
Volume 2, 1832 (COHA)

“How can he care for these dirty, dull-witted fellows that can't spell their own names, when he is so smart and **such a long, long way above** them?”

WHITNEY Adeline Dutton Train, *Faith Gartney's Girlhood*, 1863 (COHA)

And Cousin Kate, excuse the expression, but your reception to them is **way above** par.
DE MILLE Henry Churchill, *Men and Women*, 1879 (COHA)

It is also the case for *way back*:

Look **a little way back** in memory, and see if no image will arise to remind you that then[.]

FOSTER John, *An essay on the importance of considering the subject of religion*,
1828 (COHA)

But, O dear, I cant find out half so much about it now, here in this great city of Portland, where all the Governors live, as I could six months ago among the bear traps and log houses in our town, **way back** in the woods.

SMITH Seba, *The Life and Writings of Major Jack Downing of Downingville*,
Away Down East in the State of Maine, 1833 (COHA)

But you must remember that this is a very extensive subject. It would lead us a **long way back**, to Kant and even to Plato.

A Plain Discussion with a Transcendentalist, New Englander and Yale Review:
October 1843: 502-517, 1843 (COHA)

As well as for *way before*, which, however, occurs alone significantly after the other previous two:

A little way before them a footman held two horses, one caparisoned for a lady, the other a noble warhorse in military harness.

INGRAHAM Joseph Holt, *Burton; or The Sieges Volume 2*, 1838 (COHA)

After all, it is **a great way before** we come to the home of the Esquimaux, and the desert of ice where Sir John Franklin perished.

ANDREWS Christopher Columbus, *Minnesota and Dacotah*, 1857 (COHA)

I haven't been over it since **way before** the rains.

STEINBECK John, *Wayward Bus*, 1947 (COHA)

The same construction can also account for the emergence of *way behind*:

Nanamakee informed him of his dreaming, and told him that his two brothers remained **a little way behind**.

BLACK HAWK, *Autobiography of Ma-ka-tai-me-she-kia-kiak, or Black Hawk*, 1834 (COHA)

We dwell upon this point the more particularly, because our strong prevailing impression is, that the mass of the Fourierites in this country, are falling **a great way behind** the mark...

Fourierism, *New Englander and Yale Review*, 1846 (COHA)

"And while he's about it," said Alice, "you might ask him to make a little list of some of the new music. I've got **way behind** the times, being without a piano so long."

FREDERIC Harold, *The Damnation of Theron Ware*, 1896 (COHA)

To this list can be added *way out*:

He has come **a great way out** of the country, from the very fields where I was born...

MATHEWS Cornelius, *The Career of Puffer Hopkins*, 1842 (COHA)

Mr. Booth presents his compliments to Mr. Rae, and is sorry to inform him that he finds himself so extremely ill from the agitation he has suffered during this last week, that it is totally out of his power to perform this evening, and that he is gone **a little way out of town** to restore his health.

FORD Thomas, *A peep behind the curtain*, 1846 (COHA)

"An' that' minds me," said the white man, "that I heard at the same time, that Walter Greer, who used to own the plantation afore yer Marse Desmit bought it, died sometime lately,' **way out** in Texas. It's quare, ain't it, that they should both go nigh about the same time. ..."

TOURGÉE Albion Winegar, *Bricks Without Straw*, 1880 (COHA)

And *way off* also:

But when he was yet **a great way off**, his father saw him, and had compassion, and ran, and fell on his neck, and kissed him.

COX Francis Augustus, *Female Scripture Biographies, Volume I*, 1817 (COHA)

I left him as I said this, and had got **a considerable way off**, when I heard a shriek – a scream – a short brief cry that pierced my heart like a knife.

NEAL John, *Authorship: A Tale*, 1822 (COHA)

A wicked man was coming to take little Harry away from his mother, and carry him' **way off** in the dark;

STOWE Harriet Beecher, *Uncle Tom's Cabin*, 1852 (COHA)

The participate *past* having clear propositional uses, it can combine well with *way* in this construction, though we find, a bit surprisingly, *way past* alone before members of the fuller [a {adj} way past] construction:

“It’ll be **way past** dinner time when you and I get there, Emily,” she said.

PRENTISS Elizabeth, *Stepping Heavenward*, 1869 (COHA)

A little way past the market, come, within the space of a few blocks, the leading shops of the city.

PEIRCE Zina Fay, *The Externals of Washington*, Atlantic Monthly, December 1873 (COHA)

He was a man **a good way past** middle life, but still full of vigor and quick of wits.

ROYCE Josiah, *Reflections after a Wandering Life in Australasia*, Atlantic Monthly, June 1889 (COHA)

Finally, *way up*, unsurprisingly, follows the same schema:

They had about half a mile to go to the burial ground, which was **a little way up** the side of a hill fronting the southeast, not far from Turtle Creek.

MCHENRY James, *The Wilderness; or, Braddock’s Times.*, 1823 (COHA)

“Yes, yes,” said a tar, “I know that to my sorrow. I was up the Straits last v’y’ge,’ **way up** to Smyrna and Zante, after reasons, and we ketch’d one of these thundering Levanters, and was druv’ way to h – ll, away up the Gulf of Venus ...”

AMES Nathaniel, *An Old Sailor’s Yarns*, 1835 (COHA)

Look up, and see, **way, way up**, where the water first springs over the rock: there you must go.

SEDGWICK Catharine Maria, *The Boy of Mount Rhigi*, 1848 (COHA)

From this lengthy review, we can conclude that an attrition of the construction [a {adj} way {preposition}] occurred, leading to the simpler construction [way {preposition}]. All members of the paradigm behave more or less in the same fashion, and the attrition happen at comparable times, given the uncertainties inherent to the corpus study.

The possible role of *way better* and *way more*

However, *better* and *more* are not prepositions. We can thus expect that they arose from a different means. They more probably came from the construction [in {any, every, some, no, this, that, etc.} way], and especially ‘in every way’, which carries an intensifier meaning and behaves as an independent construction, leading to a possible loss of the *in* preposition, [every way {adj}], as we already encountered.

Yes, with a little attention, perhaps, we might discover many a hidden fountain, far preferable to any we have yet tasted, flowing with pleasures more copious, more constant, and **in every way more valuable**.

W.H., *Happiness More Dependent on Ourselves than our Condition*, 1834 (COHA)

The farmers, too, have found a better market for their surplus produce, and are **every way more prosperous**.

REED Andrew & MATHESON James, *A Narrative of the Visit to the American Churches by the Deputation from the Congregational*, 1835 (COHA)

Yet, *way more* only appears in the corpora (TIME and COHA) much later:

When there were no more holds, she dropped, using her space training for the best landing she could make. It was not a light one. She would bear **way more** than one bruise.
NORTON Andre, *Ice Crown*, 1970 (COHA)

That is **way more** than enough to handle the increased deficit.

BYRON Christopher, *The Great Deficit Dilemma*, 1982 (TIME Corpus)

It should be stressed that there is an obvious discrepancy between [every way more {adj}] and [way more]; in the first one, *every way more* never occurs without an adjective, while *way more* often does.

The closely related *way better* follows, as expected, the same evolutionary scheme:

As nothing like even a fair specimen of the speeches of the two houses could be given, without unduly occupying the pages of the Register, we conceive it **every way better** [...] to confine the account of congressional proceedings to the history of their progress through the stages of legislation [...].

American Annual Register, 1830 (COHA)

“That decided it,” she said. “I saw how well I did in swimming, and I knew I could do **way better**.”

MITCHELL Michele, *Downers S. swimmer seeks thumbs-up finish*, Chicago Tribune, 1989 (COHA)

While the sound quality of the Sony’s external-speaker system is **way better** than the Aiwa’s for, say, sitting quietly and alone on your bed, it’s too mellow for car trips and howling Quittners.

QUITTNER Joshua, *Tearjerkers to Go*, 2000 (TIME Corpus)

It should be noted that *every way better* is not to be found in the COHA after the 1900s, while *way better* appears in its modern use only in the late 1980s... It must be concluded, then, that *way better* and *way more* may have more probably come as a side product of the evolution [way too + {adj}] > [way + {adj}], a possibility which would be more consistent in terms of times of first occurrence.

Summary

The question is then: has *way too* appeared as a paradigm member of [way {above, back, off, past, up, etc.}]? Several reasons speak against such a scenario: in the latter construction, *way* intensifies a preposition, while in *way too*, it intensifies an intensifier. Furthermore, *way too* spread in the language much later than the [way {preposition}] construction, even if some members of the paradigm, such as *way past*, took off in the same time in terms of frequency, whereas they show the same kind of occurrences as the others. But, most of all, we do not find occurrences of constructs such as ‘a little way too soon’ or ‘a long way too much’ in the COHA. This might just be an effect of the small size of the corpus. In Google Books, which is the biggest corpus available, there is only one occurrence of ‘long way too much’, but it must be rejected for obvious reasons:

Too much, by the way, does not indicate anything excessive. It is merely the simple superlative. Thus, if a [Polynesian] native is asked the distance to a certain village, his answer will be one of these four: “Close-up”; “long way little bit”; “long way big bit”; or “**long way too much.**”

LONDON Jack, *The cruise of the Snark*, 1911

However, there are many occurrences of ‘in every way too’ prior or contemporary to the first occurrences of *way too* (albeit most of them, being found in British English textual documents, are perhaps not all that relevant to highlight a process which seems to have its roots in American English):

Thinking that my information was too limited, my health too much impaired, and my circumstancers **every way too unfavourable**, to enable me to write such historical and explanatory notes on the following poem, as would be full and satisfactory;

HESTON Jacob Franklin, *Moral & Political Truth*, 1811 (Google Books)

The change is **in every way too great.**

Pennsylvania School Journal, Volume 54, 1905 (Google Books)

We already discussed the possibility, with reservations, that ‘(in) every way too’ could be the ancestor construction of *way too*, and these observations would tend to conform the latter scenario.

There is nevertheless an other feature which is better be stressed, as it applies to all constructions we have investigated so far: why would the intensifier *way* be associated with *too*? The *every way* construction, for instance, is much more often encountered with an adjective alone rather than with *too*, with which it would seem a bit redundant. Then, why is the construction which emerges during the 1930s [way too {adj}] instead of a simpler [way {adj}]? Admittedly, it might be due to the analogical influence of the [way {preposition}] paradigm that we reviewed.

As it is time to conclude, I will pick, among all the failed hypotheses reviewed so far, one which may seem a little more appropriate than the others: the *give way* hypothesis. First of all, *give way* shows an intriguing preference for *too*: ‘give way too much’, ‘give way too soon’, are attested occurrences in the COHA while ‘give way much’ and ‘give way soon’ are not; though they are not impossible, as Google Books shows. This prevalence is all the more remarkable if we remember that *too soon* covers only 2 % of all uses of *soon*, *too far* 3 % of uses of *far*, and *too much* 9% of those of *much*. Second, we pointed out that *give way* knew a decrease of frequency by the time *way too* emerged. This situation, however, may have favored the reanalysis, as the meaning of *give way* was presumably becoming less salient (the fact that *give way* remained used without a qualifier for most of its occurrences weakens nonetheless this claim). Third, we saw with *du coup* that the semantic expansion posterior to the reanalysis step of the grammaticalization could be extremely fast. Thus, it is not impossible for *way too* to have reached, and prospered in, semantic and syntactic environments which were not primarily associated with its uses prior to the reanalysis. The initial colligation with stative verbs, particularly, may have served as a switch from adverbs (more fitted in the *give way* construction) to adjectives. This scenario, while compatible with the evidence we detailed in this study, remains certainly speculative, but might at least be possible.

Of course, a complete account of this grammaticalization should make use of the fact that, by that time, there were several factors favoring a reanalysis: both the existence of an extended paradigm [way {preposition}] and the use of *way* in an intensifier construction, [in every way {adj}], must have analogically motivated the reading of *way too* as a unit. The first one probably favored the extraction of *way too* instead of *way* alone, while the second may have strengthened the intensifying semantics. This example illustrates thus, if needed, that a grammaticalization is not an isolated process, but depends on many linguistic factors which guide, catalyze and constrain the process in many ways.

2.2 Theoretical discussion

Now that we have settled the notion of grammaticalization on a more concrete ground, we can come back to its definition and try to flesh it up a little. In the following, we shall first discuss why it matters to consider grammaticalization as a process, and not only as a result. Then we will present the traditional, parameter-based definition of grammaticalization, and see which problems it raises. A longer discussion shall ensue, devoted to the crucial point behind the definition of grammaticalization: the nature of the functional contents which allows to consider a linguistic element as grammatical. The uncertainty surrounding this definitory distinction is indeed one of the greatest weaknesses of the study of grammaticalization, as it prevents to reach a consensual agreement regarding the actual scope of this phenomenon. We will end with a brief review of some of the interesting ties between grammaticalization and Construction Grammar, asking whether grammaticalization still has its place in a Construction Grammar framework.

2.2.1 Grammaticalization as a process

Saying that grammaticalization is a process may sound like a truism. However, we could see grammaticalization as the result of a process instead of a process as such, and this has been, in fact, a long-lasting issue in the literature. It also stresses that the steps of the change are as important as the difference between the starting point and the end point. Otherwise, we could content ourselves with accounts of grammaticalization of the following kind:

Au tournant de (at the turn of):

Grammatical meaning (spatial preposition) → More grammatical meaning (temporal preposition)

Spatial preposition:

puis, doublant le pas, il disparut au tournant du sentier

then, doubling his pace, he disappeared at the turn of the road

TOEPFFER Rodolphe, *Nouvelles genevoises*, 1839, p. 372 (Frantext)

Temporal preposition:

on lui prédisait au tournant de la quarantaine une bedaine formidable

he was told he would get a formidable belly by the time of his forties

LAFON Marie-Hélène, *Les pays*, 2012, p. 26 (Frantext)

This kind of account is for instance the one followed in the *World lexicon of grammaticalization* (Heine and Kuteva, 2002) which lists several hundreds of grammaticalization processes, grouping them by parallels in their target/source association, across many different languages. It does however not imply that the authors do not regard grammaticalization as a process; quite the contrary: the word ‘process’ appears 523 times in their book.

Then why is it so important to consider grammaticalization as a process? There are several reasons for this. First of all, it emphasizes the path between the source and target meanings of grammaticalization. Saying that lexical words and a grammatical morphology system have components in common is one thing; highlighting the path between the two domains is another. It stresses that there is no clear-cut boundary between grammar and lexicon, and that the tools of the former are recycled from parts, chunks, and fuzzy assemblages of the latter. It also implies a strong methodological focus on diachronic inquiry through past textual or recorded evidence, rather than on a plausible reconstruction, contrasting, for instance, with phonetic changes as they were theorized in the second half of the nineteenth century.

The goal of such an inquiry in time is not only to find the lexical origin of some grammatical form, to give its etymology in some sense, but also to understand precisely how such a change has happened. This latter interrogation encompasses all relevant aspects of the process, and particularly, the motivations of the speaker, for grammaticalization studies have for long been interested in the cognitive aspects of the change. As early as (Givón, 1979), one finds reflections on what grammaticalization processes tell us about the organization of cognition and the pragmatic, cognitive-based factors of the process have been discussed since at least the eighties (Traugott, 1988). This cognitive aspect of the change has been a long lasting trope of grammaticalization studies.

Seeing grammaticalization as a process is thus a way to stress the very real incarnation of this process in speakers’ cognition, contrasting thus with parameter-based accounts of language change, which hold dictionary and grammar as two separate entities (a dichotomy magnified by Chomsky’s hugely famous ‘Colorless green ideas sleep furiously.’), while the two are intrinsically merged and entangled in grammaticalization.

The diachronic perspective often adopted in grammaticalization is not only interested in the motivations of the speakers, but also in the specific ‘substance’ of the

process. Indeed, if grammaticalization is conceived of as a process, then it is possible to dissect it and to work out new haruspical insights from the contemplation of its entrails. As such, an account of an instance of grammaticalization will be typically devoted to tracking down, through texts and contexts, the studied linguistic form as it comes to develop in turn new uses, new meanings, and a new syntactic behavior. Thus, grammaticalization is best seen as a gradual process (Lichtenberk, 1991), which follows some characteristic steps (Heine, 2002), wherein lies its identity and specificity, besides its defining association between a grammatical target meaning and a less grammatical source meaning.

2.2.2 The dimensions of grammaticalization

More accurately, grammaticalization is seen as moving along different dimensions. For each of these dimensions, a certain kind of change is expected; but the overall progress of the change is far from being predictable on a step-wise basis. A four-dimensional picture of grammaticalization has been proposed by Heine and Kuteva (2002), in the spirit of Lehmann's criteria 1995, which I shall not present here. Indeed, whereas the six criteria of Lehmann (1995) may picture more closely the grammaticalization, some of these dimensions make only sense in the case of forms which are already grammatical, such as structural scope. They also tend to limit severely the notion of grammaticalization, a theoretical reduction which is profitable in some contexts, but which is of no relevance to us in this work. In (Heine and Kuteva, 2002), the dimensions along which the grammaticalization process acts on the form are the following:

- **Semantics:** The meaning of the form is expected to become more and more blurred and imprecise as grammaticalization happens. This is the phenomenon known as 'bleaching' (Givon). For instance, *down*, which has undergone multiple grammaticalizations throughout its history, has completely lost the concrete, representational meaning of Old English *dun*, meaning 'hill'. We shall, however, discuss further this notion of bleaching.
- **Usage:** The number of contexts of use of the form undergoing a grammaticalisation process usually increases. It is chiefly related to the trivial fact that more meanings mean more uses, but also, to the less trivial fact that functional meaning is less constrained to the utterance situation than a representational, lexical meaning. For instance, in the grammaticalization of French *côté* (approximately translated by *side* for most of its uses, which also shares an anatomical origin), the present-day prepositional meaning can be used in a much wider array of uses than the original lexical source meaning, 'flank' (see the closely related *côte*, 'rib').
- **Category:** The form becomes harder to categorize as the grammaticalization proceeds. Usually stemming from one of the major lexical categories (noun, verb, adjective, adverb) associated with a more or less rigid syntactic pattern and a rich morphology (agreement marking, case marking, tense marking, etc.), the form loses this specific behavior and ceases to be attached to a clear-cut category. It can become a preposition, a pronoun, a conjunction, an adverbial, a discourse

marker... Those categories are important as well, but they are not associated with a distinct, codified syntactic behavior, and they form paradigms rather than broad, unlimited categories such as the four mentioned. The form also presents a smaller variety of formal aspects, a limited agreement marking flexibility, when it is not completely frozen in one of its particular former manifestations. For instance, the French modal *falloir* has a limited verbal morphology, since it can be used only with the third person singular. Its source form, the verb *faillir* ('to lack of'), did not exhibit such restrictions. An example of category loss is to be found with the grammaticalization of the construction [needless to say]; needless is an adjective, and yet [needless to say] can be used with much greater syntactic freedom and does not need to qualify any noun to be used (e.g. 'It would be needless to dwell upon his misery as this conviction forced itself upon his mind.' vs 'Well, needless to say, the rest of the shift passed under a definite pall.'). It belongs, in fact, to the adverbial category.

- Phonetics: A phonetic reduction can be concomitant to the grammaticalization process. The latter instance gives us an example: The adverbial 'needless to say' is a reduced form of 'It is needless to say'. This fourth dimension of grammaticalization has raised a lot of controversy. Some people consider that the phonetic reduction is a mere consequence of the rise of frequency due to the context expansion mentioned above (Bybee and Thompson, 1997; Diessel, 2007), and therefore dismiss the theoretical importance of this dimension (they do, however, consider it as a valuable hint that a grammaticalization process is occurring). Others consider that this phonetic reduction is central to grammaticalization (Geurts, 2000). When the phonetic flesh of a form wears off, it has to be replaced with a new one, usually by putting together two or more existing forms. As the newly created form has more salience and can be understood better than the previous, diminished form, it usually takes over. Phonetic reduction of existing functional forms would just act as a sink drawing into existence new grammatical forms from what lies available in the lexical domain.

It should be immediately pointed out that none of these movements is specific of grammaticalization (Melis and Desmet, 1998). Semantic bleaching happens very frequently, without involving any grammaticalization. The English *key* for instance, is semantically bleached compared to its initial, restricted meaning of 'metal piece that works a lock' (OED). The following use of the word shows that its semantics have been considerably bleached, and only the vague idea of 'opening' remains:

Voyage tape fed into the controls of the ship had taken the men, and, when rewound, had – by a miracle – returned them to Terra with a cargo of similar tapes found in a building on a world which might have been the central capital for a government comprised not of countries or of worlds but of solar systems. Tapes – each one the key to another planet.

NORTON, Andre, *The Defiant Agents*, 1962 (COHA)

Key presents also, as an adjective, a slight amount of decategorization (but it remains in the four main lexical categories of English), which considerably increases its array of uses. Colombian Spanish *ojo*, on the contrary, has really undergone such

a decategorization; a lexical item meaning ‘eye’, it can now be used as a warning (see the close English equivalent *Watch out*), thus decategorizing from noun to interjection. Extension of context of uses is also pervasive in lexical evolution. For instance, the French *portée* was originally a measure for liquids. When it was extended to the meaning of *scope*, for projectiles especially, it opened to many metaphorical extensions.

A ces paroles ez vos .II. Sarrazins ;
 Une portee apportoient de vin,
 Sus el palés en voloient servir ;
 Mes quant il virent les ruistes cops ferir,
 En fuie torment, si lessent tot chair.

At these words came two Saracens;
 They were bringing a *portée* of wine,
 Which they intended to serve at the palace;
 Yet when they saw how fierce the blows were struck,
 They turned away, dropping everything on the ground.

La prise d'Orange, c.1200, p.110 (Frantext)

Parfois, ils avaient la consigne de renvoyer les décisions aux AG de Jussieu. [...] Mais ils n’y parlaient pas, à moins qu’un événement d’une portée symbolique certaine soit intervenu dans leur fief et qu’on les invite à le raconter - mise en scène, quasiment, de la « parole du peuple ».

Sometimes, they were instructed to postpone the decisions to the general meetings at Jussieu [the nickname of the university Pierre et Marie Curie]. Yet they were not involved in the discussion, unless some event of some symbolic *portée* had occurred in their dominion and they were invited to speak about it — almost theatrically representing the voice of the people.

VIENNOT Éliane, *État des lieux*, 2012, p. 67 (Frantext)

However, the extension of the contextual uses of *portée* is limited. Indeed, almost all instances of *portée* stick to the pattern — as defined in Hunston (2000) — [det portée de N], e.g. ‘la portée de ton acte’ (Frantext). A better example would be the French verb *se rappeler* (‘to remember’), which use has extended from a mere transitive pattern ([se rappeler + N]) to a propositional one ([se rappeler que + Proposition]):

Ah ! dans ses yeux confus je lis ses perfidies ;
 Et son trouble, appuyant la foi de vos discours,
 De tous ses attentats me rappelle le cours.

Ah ! In his troubled eyes I read his treachery,
 And his confusion, supporting the truth of your speech,
 Remind me the course of all his assailments.

RACINE Jean, *Esther : tragédie tirée de l'Écriture sainte*, 1697 (Frantext)

[U]n heureux hasard me rappelle qu’il y a dans la maison que j’habitais, un caveau dont il m’ étoit permis de disposer: c’est-là que j’ensevelis, pour ainsi dire, ce misérable jeune homme que je plainois encore [...].

By a fortunate coincidence I remember that in the house I lived in there was a vault that I could make use of: It is there that I buried, or so to speak, this miserable young man that I am still pitying.

DE BACULARD D'ARNAUD François-Thomas-Marie, Les Époux malheureux ou Histoire de Monsieur et Madame de ***, 1745, p. 211 (Frantext)

In this pair of examples, it appears that *se rappeler* has benefited from an extension which lies in its use rather than in its semantics; in both examples, the verb has more or less the same meaning, however, in the second example, the complementive *que* allows the use of this verb with a much wider range of arguments, including complex events. Thus, *se rappeler* has not grammaticalized, and yet it has known a very clear expansion, increasing greatly the number of contexts in which it can be used.

I have not mentioned phonetic reduction, but clearly, this phenomenon is not bound to grammaticalization processes. Phonetic reduction instances are plenty, and it should suffice to quote but a very few ones. In French: cinématographe > cinéma > ciné; cafétéria > cafèt; je suis > chui; Comment ça va? > Comment va?; il ne faut pas > faut pas. In American English: What's up? > 'sup?; see you > see ya; congratulations > congrats; and > 'n'; probably > prolly. If some of them affect grammaticalized forms, they also affect completely lexical forms, or frequent language segments such as *je suis*.

Thus, not a single one of these transformations along the four dimensions is truly characteristic of grammaticalization; yet, we can consider that their conjoint work is what makes grammaticalization phenomenologically specific (Prévost, 2003). However, is it always true that all four of these movements are found in every single grammaticalization processes? We have already seen several examples of linguistic changes presenting several of these moves, and yet, they could not be seriously considered as instances of grammaticalization. Additionally, grammaticalization can leave one of these four dimensions untouched; for instance, phonetic reduction. An increase in syntactic freedom is not necessary, either. The grammaticalization of the French construction [de {N_{time}} en {N_{time}}], for instance, does not display a significant change in its syntactic behavior. Also, grammaticalization can lead to a reduction of syntactic freedom, for instance in the case of the French determiners, syntactically bound to the beginning of the noun phrase they belong to.

I would thus tend to disagree with the idea that grammaticalization is a specific phenomenon, insofar it is a specific conjunction of non-specific changes happening in different dimensions. Let us recall here a point made in (Prévost, 2003):

[L]a grammaticalisation n'est pas « définie » par un ensemble de mécanismes, mais par un mouvement : celui qui conduit une forme linguistique à acquérir un statut plus grammatical [...].

Grammaticalization is not “defined” by a conjunction of mechanisms, but by a movement: the one which leads a linguistic form to acquire a more grammatical status. (Prévost, 2003)

Phenomenologically, it seems thus that there is no difference between grammaticalization and a broader phenomenon known as constructionalization, i.e. the emergence of a new construction, or in our terms, a new linguistic cohesive form. We can also consider some grammaticalization processes as cases of semantic expansion, especially the instances of further grammaticalization, or secondary grammaticalization in the sense of Waltereit (2011), when an already grammatical form becomes still more so due to some ‘semantic widening’ (e.g. *cependant*, in French, which evolved from the meaning ‘meanwhile’ to a concessive marker, semantically closer then to ‘however’ (Marchello-Nizia, 2009), the two uses still remaining nowadays). The first process implies that a

new fixed constructional form becomes entrenched in language, the second that an existing form gains a new meaning. The latter change is encompassed by what Traugott and Trousdale (2013) call a ‘constructional change’ (defined as a change in the features of a construction, including semantic features, as in the case of semantic expansion).

The fact that primary (i.e. with the emergence of a new form) and secondary grammaticalizations correspond to two distinct processes of change is problematic, to be sure. Should we see constructionalization and semantic expansion as two subprocesses possibly occurring in the course of a grammaticalization? This is an interesting possibility, as grammaticalization is thought to follow distinct steps (see for instance (Heine, 2002), which is discussed further in chapter 7), a few of which correspond to constructionalization and semantic expansion. Yet, it suggests that the macro-process of grammaticalization can proceed in different ways, relying on different mechanisms of linguistic change. What individualizes grammaticalization as a distinct process would then be the sole fact that its output is more grammatical than its input.

Of course, this raises the question that we have eluded so far: What could possibly mean that a form is grammatical?

2.2.3 The functional vault of language

By definition, a linguistic form undergoes a grammaticalization if, and only if, it comes to be used with a more grammatical meaning. This is the only criterion which can decide whether or not a linguistic change counts as a grammaticalization. It lacks, however, a proper characterization of what ‘grammatical’ means, and it should be expected that disagreement could stem from this imprecision.

First of all, there are often two views on this matter, which are but superficially opposite to each other. According to the first one, there exists a cline between lexicon and grammar (Hopper and Traugott, 1993), a ‘continuum’ along which we can find all status from the purely lexical to purely grammatical (Langacker, 1987; Haspelmath, 1999), and everything in between. However, those two poles remain to be identified. This is what the second approach tackles with: The dichotomic distinction between two natures, the lexical one and the grammatical one. Interestingly, some linguistic changes happen to be neither grammaticalizations nor degrammaticalizations (the process by which a word comes to be less grammatical), for they remain purely lexical. It would thus seem that there is a boundary beyond which things start to become grammatical, a plateau before the start of the cline, hence justifying the idea of a dichotomic separation between the two. Furthermore, this dichotomy can well be prototypical, so that there are prototypically grammatical constructions and prototypically lexical ones, with words such as the French preposition *à* on one side and words such as *petrichor* on the other. A linguistic form or construction would thus be grammatical as soon as it enters the cline, but can become arbitrarily more grammatical from that point. In this view, primary grammaticalization would occur on the onset of the cline, and secondary grammaticalization would describe anything which comes next.

In the following we will thus present several dichotomies which have been proposed to account for the idea that not all components of an utterance play the same role in its processing.

The conceptual/procedural distinction

One of the most famous bipartition of linguistic elements is drawn from Relevance Theory and focuses on pragmatic inferences. The linguistic load of an utterance is seen as made of two contributions: a ‘conceptual’ one and a ‘procedural’ one (Blakemore, 1987), which helps to process the conceptual part and constrains its interpretation. This theory especially accounts for the role of pragmatic markers, or discourse markers (Hansen, 1998), discussing for instance the different constraints brought by the two markers *after all* and *so* (Blakemore, 2002). This distinction has arisen in a very specific context which has little to do with grammaticalization, and grammar, in this context, means something entirely different:

This distinction — between the process of decoding messages and the process of making inferences from evidence — is the basis of their [=Sperber & Wilson’s] distinction between semantics and pragmatics. The first kind of process is performed by an autonomous linguistic system — the grammar — which is dedicated to the performance of mappings between a linguistic stimulus (utterance) and a semantic representation for that utterance. The other kind of process — the inferential process — integrates the output of the decoding process with contextual information in order to deliver a hypothesis about the speaker’s informative intention. (Blakemore, 2002, p.60)

Grammar is thus seen here as a cognitive device which helps to build a semantic representation out of a linguistic input; therefore this input cannot be seen as made of ‘lexical’ and ‘grammatical’ components. ‘Lexical’ rather seems to be pairing with ‘syntactic’ (Blakemore, 2002, p.180), in line with the dictionary/syntax dichotomy of Generative Grammar (Hilpert, 2014), a distinction which is devoid of all relevance in grammaticalization theory.

Why, then, is this distinction useful at all for our purpose? Actually, it is only through a bleaching of these two terms that they came to be used to describe the lexical/grammatical opposition. The first to have tied together grammaticalization and the conceptual/procedural dichotomy appears to be LaPolla (1997), who argues that all human communication is based on inference. In an inference-based theory of communication, meaning is built by the hearer through a reasoning about the speaker’s communicational intentions. In this view, people assume that they are spoken to for a reason. The meaning of the utterance which is addressed to them and the reason for which they are addressed are pretty much the same thing. Then, a large part of the utterance is ‘procedural’ in the sense that it serves to support this pragmatic reasoning about the utterance’s ‘raison d’être’ (which underlies the active role of the speaker in the guiding of the utterance’s reception by the hearer).

Apart from discourse markers, procedural items, in this enlarged sense, also provide ‘contextual cues’, of which it is said that ‘while [they] convey information, the information is not like that of lexical items, as it is independent on the propositional meaning, and conveyed only in the interactive communicative process, and so cannot be discussed outside of that process’. A dramatic metaphor of my own would be to compare conceptual and procedural information to the difference between the bare script of a drama piece and the way it is played on an actual stage. Then, grammatical elements, as they actually ‘set the stage’ for the representation of meaning, can be

seen as a procedural input about the utterance:

The fixing of repeated patterns into grammar is nothing more than the development of conventionalized forms that restrict interpretation. (LaPolla, 1997)

Having bridged the gap between the procedural information of Relevance Theory and the lexical/grammatical distinction, LaPolla can now re-tell grammaticalization in those new terms:

Lexical items would not grammaticalize into so-called procedural information if their conceptual information was what was important. The conceptual/procedural distinction then is not a simple either/or situation, but more of a privative opposition (with gradations): all lexical items can have a procedural function, while grammatical markers generally only have a procedural function. That is, some elements involve only so-called procedural information, but lexical items are not used exclusively for conceptual information; they often have a procedural function, and it is having this function that allows them to become grammaticalized into items with only procedural information. (LaPolla, 1997)

Interestingly, in this quote, we find the two ideas expressed earlier, that there exists a continuum between the lexical and the grammatical pole, as well as a clear-cut boundary: A linguistic item becomes grammatical as soon as it ceases to convey conceptual information. Contrary to what I presented before, the cline is more on the lexical side of the limit and the plateau on the grammatical one. Linguistic items do not start falling down the cline by ‘gaining’ some grammaticality, they stop shifting along the continuum once they have ‘lost’ all conceptual content.

Similar ideas are to be found in (Nicolle, 1998). In this work, Nicolle strives to reconcile the gradualness of grammaticalization, with the dichotomic character of the conceptual/procedural distinction. His solution is similar to the one of LaPolla (1997) to the extent that he recognizes the possibility for a linguistic form to be conceptual and procedural at the same time:

If an expression can [...] encode both conceptual information and procedural information, then grammaticalization (the development of a procedural semantics) needs not result in the loss of lexical (conceptual) semantic features.

(Nicolle, 1998, p.15)

However, there is a major difference with (LaPolla, 1997): Grammaticalization is no longer ‘the loss of conceptual features’ (LaPolla, 1997), but ‘the development of a procedural semantics’ (Nicolle, 1998). With (LaPolla, 1997), grammar starts when conceptual content ends. With (Nicolle, 1998), the lexical side ends when procedural content starts. In both cases, grammaticalization is semantically abrupt (see Fig. 2.1 for a schematic picture of this divergence). Actually, while LaPolla (1997) focuses on the end point of grammaticalization (the creation of grammar in a language through the fixing of inferential constraints), Nicolle (1998) is more concerned with the onset of grammaticalization (how grammaticalization comes to happen). The latter position is actually more consensual: as Prévost (2003) states, ‘Il semble peu pertinent de parler de formes plus ou moins lexicales.’ (‘It does not seem much relevant to speak of more or less lexical forms.’) Usually, it is the grammatical quality which is deemed as gradual,

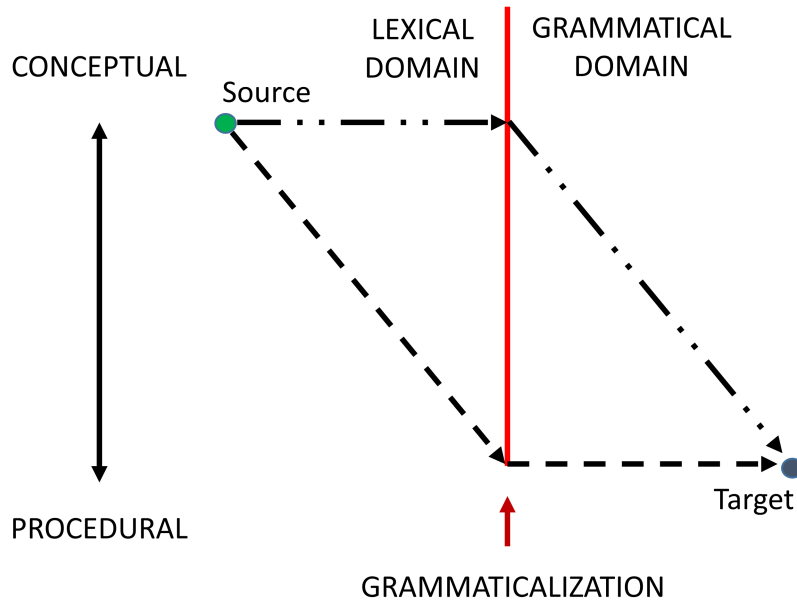


Figure 2.1: A schematic view of the difference perspectives of (LaPolla, 1997), represented in dashed line, and (Nicolle, 1998), in semi-dotted line.

not the lexical one, and definitions of grammaticalization almost always leave way for a grammatical > more grammatical evolution. Anyway, the two authors agree, in the end, that grammaticalization can be seen as the conventionalization of conversational implicatures.

There is another difference between these two authors. While LaPolla (1997) fully equates procedural and grammatical, (Nicolle, 1998) does not. In his own terms:

[G]rammaticalization involves a shift from conceptual encoding to procedural encoding in a single expression over time. This is not to say that all exponents of procedural encoding are grams; discourse connectives [characterized] as encoding procedural information, are not grammatical markers. (Nicolle, 1998, p.6)

Then, having a grammatical status is not the same as conveying procedural content; grammatical markers are a subclass of procedural markers. The difference between discourse markers and grammatical markers is also made explicit:

[Discourse markers] constrain the processing of fully propositional conceptual representations with respect to the implicatures which a speaker intends an addressee to compute, whereas grams, such as modality, tense, aspect, and case markers, constrain the construction or identification of propositional conceptual representations. (Nicolle, 1998, p.6)

We thus end up with a two-dimensional view of the organization of language, encompassing a conceptual/procedural separation as well as a discourse/grammar one (Fig. 2.2). If we accept this distinction, then grammaticalization is but one of two processes — with pragmatization, the rise of a discourse marker — having both in common a result, the addition of procedural content in the semantic load of a linguistic

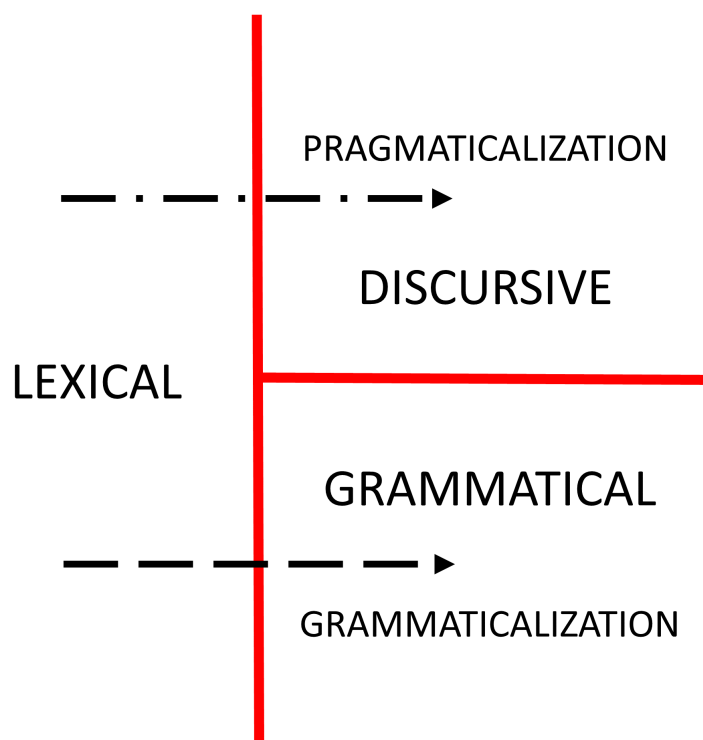


Figure 2.2: A two-dimensional view of the organization of language, distinguishing two kinds of change, pragmaticalization (from lexical to discursive) and grammaticalization (from lexical to grammatical). There is no transition from discursive to grammatical in this view.

expression, and a mechanism, the conventionalization of a pragmatic inference. This raises many questions, which are not addressed in Nicolle (1998). Can there be a vertical movement (discourse \leftrightarrow grammar) in semantic change as well as a horizontal one (conceptual \leftrightarrow procedural)? If so, should we expect directionality as well, as it seems we have in the (conceptual \leftrightarrow procedural) dimension?

In present day, the (discourse \leftrightarrow grammar) differentiation is still heavily debated. Heine (2013), for instance, considers that they correspond to two completely separated cognitive functions, and correspond to the two halves of the brain. In this view, pragmaticalization and grammaticalization are of course distinct, and there can be no movement along the (discourse \leftrightarrow grammar) dimension.

Other distinctions

The conceptual/procedural distinction has been further bleached out and can sometimes be considered as a convenient metaphoric way of speaking about the lexico-grammatical polarity:

The dimension of type of concept concerns whether a construction is contentful ('lexical') or procedural ('grammatical'). 'Contentful' material can be used referentially [...]. 'Procedural' material has abstract meaning that signals linguistic

relations, perspectives and deictic orientation [...]. [[Footnote] The term ‘procedural’ was originally suggested by Blakemore (1997); we adopt it without intending any theoretical connection with Relevance Theory. Another useful metaphor highlighting the role of grammatical items was [...] ‘a sort of functional glue tying together lexical concepts.’] In Terkourafi’s words, linguistic expressions encode procedural meaning when they ‘contribute information about how to combine [...] concepts into a conceptual representation’ (2011: 358-359).

(Traugott and Trousdale, 2013, p.12)

Bybee (2002) also adapts the notion of procedural meaning to suit her own purposes. She opposes ‘procedural’ knowledge to ‘propositional knowledge’:

Propositional knowledge is ‘knowing that’ or knowing facts [...]. Procedural knowledge is ‘knowing how’ and includes knowing how to tie shoelaces and how to drive a car. (Bybee, 2002, p.111)

As Sperber and Wilson (1986) did for conceptual and procedural information, Bybee argues that propositional knowledge is mainly conscious while procedural knowledge is unconscious. She concludes with a claim similar to LaPolla (1997): ‘lexical items involve at least some propositional knowledge, while grammatical constructions are largely procedural’(Bybee, 2002, p.111).

Bybee’s dichotomy stresses that the distinction between lexical material and grammar is chiefly based on a cognitive ground. It is not defined formally or structurally, within language, but from an external point of view, in reference with human cognition in a much broader way. Though this cognitive stance is interesting and enlightening, it lacks some operative criterion. Especially, when facing diachronic data, we have no access to the cognition of the speakers except through their linguistic output. Since grammaticalization is a diachronic process, this proposal is of little help in deciding whether or not some process is an instance of grammaticalization.

The idea that linguistic items are split in two categories, lexical and grammatical, the latter being operative on the former, is anyway well accepted. These two terms can receive many names in addition to ‘lexical’ and ‘grammatical’ — e.g. ‘contentful’ and ‘functional’ as proposed by Hopper and Traugott (1993) — yet the basic intuition behind this distinction remains the same. Talmy (2000), in his *magnus opus*, illustrates it with a simple example, the sentence ‘A rustler lassoed his steers’, listing three lexical items, and eleven grammatical ones (Talmy, 2000, pp.33-34). The problem with this kind of listing is that it would be near impossible to perform on actual sentences. It may be insightful on a made-up example, but it relies only on the linguist’s intuition. Also, the actual role of discourse markers, which we have seen to be problematic, is not discussed in this work. The definition of grammar underlying this separation effort is given in the following terms:

[W]e take a sentence (or other portion of discourse) to evoke in the listener a particular kind of experiential complex, here termed a **cognitive representation** or CR. The grammatical and lexical subsystems in a sentence seem generally to specify different portions of a CR. Together, the grammatical elements of a sentence determine the majority of the *structure* of the CR, while the lexical elements together contribute the majority of its *content*. The grammatical specifications in

a sentence, thus, provide a conceptual framework or, imagistically, a skeletal structure or scaffolding for the conceptual material that is lexically specified.

(Talmy, 2000, p.21), emphasis original.

Chafe (2002) offers an idea along the same lines: ‘It is in fact what one would expect if one views grammaticalization as a shift, not from a word or construction to a *grammatical* element, but from an *idea* to an *orientation*.’ (Chafe, 2002, p.409, emphasis original). The dichotomy is placed here between ideas and orientation, orientation being a new concept to deal with. Its characterization is close to Talmy’s ‘structure’:

Ideas are located within our thoughts in a variety of ways [...]. It is often observed that language is a social phenomenon, that our ideas are not verbalized in a vacuum but to a large extent in order to communicate them to others. One result is that it is necessary to locate our ideas so a listener will know where to place them within his or her own store of knowledge. (Chafe, 2002, p.401)

This cognitive ‘location’ of ideas is precisely the role of orientation. Interestingly, discourse markers are considered to provide an orientation as they ‘orient events with respect to the ongoing discourse, but also with respect to the ongoing interaction’ (p.401), so that in this framework, pragmaticalization would count as grammaticalization. As in Talmy (2000), an emphasis is put on the communicational nature of language; it is through the necessity to communicate with others that grammatical elements comes to be important. This idea agrees with the core claim of the conceptual/procedural distinction, in which procedural elements serve to guide the inferential process of the hearer in communication (LaPolla, 1997).

An alternative definition of ‘grammatical meaning’, in the context of defining precisely what grammaticalization really is, is provided by Idiatov (2008) in terms of ‘obligatoriness’:

After all, the term at stake is grammaticalization. The notion of grammatical meaning is best defined via the notion of obligatoriness: a meaning is grammatical in a given language if the speaker cannot choose to leave it unexpressed. [...] An important consequence of this is that a given meaning is grammatical or non-grammatical only with respect to a particular linguistic system. It cannot be grammatical a priori, universally. [...] It is also important to notice that the criterion of obligatoriness does not necessarily imply that the border between the domains of grammatical and non-grammatical meanings is always strict and clear. (Idiatov, 2008, p.155)

Obligatoriness is indeed an intriguing aspect of language, and quite pervasive. For instance, in French, the determiner of a noun is grammatical in the sense that a noun cannot be uttered without a determiner. Similarly, the subject pronoun is grammatical as a verb cannot be used without an explicit subject (contrary to other Romance languages). Verbal tense, in English, can be seen as grammatical, as one cannot use a verb without specifying a tense. However, sentences such as ‘What do we eat tomorrow?’, largely attested, show that tense may also be specified by lexical means instead of being carried by the verb — which illustrates the fact that the ‘border ... is [not] always strict and clear’.

Idiatov is aware that this very specific definition ‘excludes from the scope of grammaticalization’ (p. 166) many instances of language change which were considered

as such, e.g. ‘the development of derivational affixes’ (p.166). It goes without saying that discourse markers are excluded as well from this definition, while the grammatical status of prepositions is not obvious. The notion of obligatoriness, though very limiting and not entirely clear (one could say, a bit caricaturally, that most functional items are obligatory in the context of use of the function they serve and optional otherwise; locative prepositions, for instance, are frequently required to refine the utterance with a spatial adverbial complement, which is itself non-mandatory), is at least insightful in that it stresses the fact that not all languages constrain the same features in their grammar — a fact largely acknowledged by other authors, especially LaPolla (1997). It also raises the question of how obligatoriness comes to be conventionalized through language use; we shall briefly address this point in the next section.

Prévost (2006) also addresses this question to distinguish grammaticalization from lexicalization. Much like Talmy (2000, p.22), according to whom ‘[t]he distinction between the two is made formally — that is, without reference to meaning — in terms of the traditional linguistic distinction between “open-class” and “closed-class”’, she notes:

Se pose évidemment la question de la frontière entre catégories lexicales et grammaticales, ou, formulée en d’autres termes, entre catégories majeures et mineures, entre classes ouvertes et fermées (même si les trois oppositions ne se recouvrent pas exactement).

This naturally raises the issue of the boundary between lexical categories and grammatical ones or, in other terms, between minor and major categories, between open classes and closed classes (albeit those three dualities do not exactly overlap).
(Prévost, 2006, p.133)

However, if there are clear examples of the two kind of categories, such as nouns and adjectives in French and English for the lexical pole, and conjunctions for the grammatical one, a lot of linguistic items lie in between — especially because of grammaticalization and the resulting porosity between the two poles. Auxiliaries, for instance, are problematic; they both belong to the open class of verbs, and to their own closed class. Especially concerning is the case of extra-clausal items, such as most discourse markers, whose autonomy seems at odd with what we usually expect from grammatical items.

A frequent remedy to the difficulty of establishing clearly the demarcation line between the two categories consists in a ‘prototypical’ approach, in which the lexical and grammatical categories are not defined by a separating boundary, but by prototypical, expected traits or features:

On peut aussi envisager la question dans le cadre d’une approche prototypique : les items d’une classe affichent plus ou moins des traits définitoires de cette classe, traits qui permettent de distinguer classes lexicales et grammaticales (en particulier pour ce qui est du caractère concret/abstrait).

One can also consider the issue in the framework of prototypicality: items of a given class present a small or large number of definitive features of this class, features which allow to distinguish lexical and grammatical classes (especially concerning the concrete/abstract trait).
(Prévost, 2006, p.134)

This kind of reasoning perfectly agrees with the well-established idea that there is not, in any way, a clear-cut distinction between lexical and grammatical items. Thus, we can easily consider that some items are clearly grammatical, that some others are obviously grammatical, so that they can serve as references for the categories in which they stand and which they do represent. However, I do not believe this approach to be entirely satisfying. Indeed, grammaticalization precisely lies in the middle zone; furthermore, it is chiefly defined by the fact that it goes from lexical to grammatical or from grammatical to even more grammatical. If, among the items on which it acts upon, the distinction between lexical and grammatical components becomes blurred and unclear, then it seems that we are left with very few theoretical insights to decide whether a language change is a grammaticalization or not. And indeed, a lot of disagreements are still going on, especially concerning the problematic issue of discourse markers (Heine et al., 2014; Degand and Evers-Vermeul, 2015).

This small and non-exhaustive review will be concluded by the discussion of yet another distinction, whose chief characteristic is that it precedes all others by at least a handful of decades, having been developed by the German linguist Karl Bühler (1879-1963), presented here through the work of Marthelot (2012). This distinction stemmed from a context which had nothing in common with grammaticalization, and yet, as pointed out by Melis and Desmet (1998), it remains of relevance in this diachronic context. Bühler develops a distinction between signs with field value, and signs with symbolic value. He is motivated, in this theoretical proposal, by a rejection of a former distinction due to Anton Marty, a Swiss linguist, between autosemantic (*selbstbedeutend*) and synsemantic (*mitbedeutend*) signs (Marthelot, 2012), which respectively correspond to signs able to carry their own meaning, and signs which cannot mean anything by themselves, but need to rely on other signs to elicit meaning. This distinction is furthermore strengthened by the idea that those two families of linguistic signs correspond to two different kinds of psychological abilities.

However, Bühler does not accept the terms of this bipartition. He accepts a dichotomy, but he does not consider that linguistic signs can either have a meaning on their own or in relation with other signs. On the contrary, signs are arranged on a polar continuum between autosemantism and synsemantism:

What remains instead of the living sentence in the syntactic scheme is a sequence of isolated signs (most of the time “linking” signs, as copulas, prepositions, etc.) and a more or less important number of union signs (such as suffixes and prefixes, or vocalic flexional endings on roots). All these isolated signs and union signs behave as a whole as a lattice of slots or, so to say, as an unfilled blank piece of paper. Everything which fills the empty slots immediately receives a meaning, fully specified by the whole assembly of relations given within the syntactic scheme. [...] Those signs and parts of signs create, in their surroundings, a field for the concrete content of the representation. Values of this field are especially determined by the choice of the signs which compose the scheme. We thus see, on one hand, that all signs, from the substantives to the nouns, create around them at least optional slots for other signs; on the other hand, that the filling of those slots may be realized at various degrees. There are thus the most diverse intermediary degrees between the autosemantic and the synsemantic signs. Yet for not a single one of them will it be fitting to say that it is entirely autosemantic, or inversely, that it does not mean anything by itself.

quoted in Marthelot (2012, pp.175-176),
directly translated into English from the French translation

How close this view is to contemporary Construction Grammar is quite striking: atomic items need to be part of more complex constructions for their meaning to be fully specified, and complex, schematic constructions, still remain meaningful and print their own semantic orientation on the specific evaluation of its filling slots in a given construct. However, it is clear, from the beginning of this quote, that one can flesh out a ‘living’ sentence, so that only the syntactic scheme remains, which is itself made of signs. Therefore, while the distinction between synsemantic and autosemantic fail to split the lexicon into two separate categories, another dichotomy is profiled. In Marthelot’s terms:

The sharp divide between parts of signs utterly devoid of meaning and signs which suffice to themselves in the building up of their meaning does not hold anymore, first because of the impossibility, for a sign, to be autosemantic: every sign needs a field to specify its meaning. [...] All language signs are therefore, in a sense, synsemantic signs, in that they open up, around them, empty slots, which must be filled for a meaning to emerge. However, all signs are not synsemantic to the same degree. The issue of the intermediary states is therefore of the uttermost importance, since it allows to distinguish a language sign (as a noun) from a field sign (*Feldzeichen*) as a preposition, which directly contributes to determine the order of the field.

(Marthelot, 2012, p.177),
directly translated into English from the French translation

Though this distinction between field value and symbolic value of a sign is not particularly original with respect to the other distinctions which we already reviewed, it dates back to 1928 (Marthelot, 2012, p.178), which shows that linguists have been preoccupied pretty early on with the lexical/grammatical dichotomy. Unfortunately, Bühler’s work is not interested in diachronic aspects of language, and the question of how a linguistic element can exert a field on other items is left entirely aside. Anyway, even if Bühler may have not exerted much influence in subsequent works in Linguistics, its insights resonate closely to more contemporary concerns. The proposal by Chafe (2002) of an ‘orientation’ operated by grammatical elements is for instance much reminiscent of this idea of a semantic field. Pessimistically, we can also consider that no much progress has been done in this direction, and the richness of the terminology developed across the years to describe the lexical/grammatical distinction may well reflect how long-lasting has been the puzzlement on this particular subject.

2.2.4 Grammaticalization and construction grammar

Before presenting our own views on grammaticalization, it seems mandatory to shortly investigate this notion from the viewpoint of Construction Grammar, which has been presented in the previous chapter. Indeed, it would not be appropriate to promote a framework in order to understand and investigate language change, and yet dismiss it entirely when it comes to discuss grammaticalization. As a specific phenomenon of language change, it should be possible to address it through a Construction Grammar perspective.

A long-lasting antagonism

This proposal, of course, is far from being new. An interesting review can be found in Noël (2007). However, though it seems inviting to think grammaticalization in the Construction Grammar framework, as the two share many interesting similarities (Diewald, 2006), a wariness still remains between the grammaticalization theorists and the proponents of Construction Grammar. One of the main issues explaining this distrust would be that, in Construction Grammar, grammaticalization has no *raison d'être*. Indeed, in Construction Grammar, the grammar of a given language is entirely described by an inventory of its entrenched constructions, the constructicon (Hilpert, 2014), so that there seems to be no room left for an opposition between lexical categories and grammatical categories. This point has been raised several times (Noël, 2007).

The thing is, grammaticalization theory has stated more than often that the opposition between lexical and grammatical is not clear-cut or absolute. The crossing of this elusive boundary, which is the very core of a grammaticalization process, shows that they are part of a ‘continuum’ (Haspelmath, 1999), so that, because of the acting of the grammaticalization process, there is no absolute separation between the lexical and the grammatical material of the language. The only opposition between lexical and grammatical categories is that of a polarity. Analogically, you do not need an Equator to go either North or South since both are largely independent from any absolute separation line on the globe. Hence, a holistic view of language stems from grammaticalization, whose very occurrence entails a global unity, in their nature, of linguistic items. That they hold different properties, allowing for a possible typology, is expected; and so is the case with constructions. As pointed out by Diewald (2006), grammaticalization and Construction Grammar can go along perfectly well.

Construction Grammar also provides interesting new insights on grammaticalization processes. Traugott and Trousdale (2013) make clear that grammaticalization happens simultaneously on the formal and the semantic planes, to which Diessel and Hilpert (2016) concur:

Although research on grammaticalization has focused on individual grammatical items, it must be emphasized that grammaticalization generally concerns strings of linguistic expressions rather than isolated words (e.g., *be going to*, *in front of*). Grammaticalization is a complex phenomenon involving both formal and semantic changes. (Diessel and Hilpert, 2016)

Hence, Construction Grammar, as it focuses on signs made of form-meaning pairings, would be a privileged framework to study grammaticalization in a more holistic, semiologic way.

Why, then, are there any difficulties to reconcile the two approaches? The main stumbling block lies in the question of specificity. Grammaticalization theorists claim very firmly that grammaticalization is a specific process; construction grammarians state that it is only one among many constructional changes, which all rely on the same mechanisms. Symptomatic of this opposition is the work of Traugott. As one of the leading figures in grammaticalization studies — notably, she co-authored the reference textbook on the subject (Hopper and Traugott, 1993) — she now adheres to Construction Grammar and works on what has been called ‘Diachronic Construction

Grammar’ (Construction Grammar applied to the understanding and the description of language change). In a recent book on the subject which she co-authored, she wrote that ‘grammatical constructionalization is the outcome of changes, not a process’ (Traugott and Trousdale, 2013, p.147). Similarly, Trousdale (2014) wrote that ‘constructionalization cannot apply to grammatical categories alone’. It would thus seem that one cannot endorse Construction Grammar without rejecting the specificity of grammaticalization.

A pragmatic solution of coexistence has therefore emerged to remedy to this antagonism, assuming that the two approaches are interested in different aspects of the change (Noël, 2007). This stance recently culminated in a somehow aporetic statement by Heine et al. (2016):

Both frameworks search for regularities in grammatical change, but whereas Construction Grammar has a focus on constructional change, that is, change in the development of constructions, the central question asked by students of grammaticalization is how and why, e.g., lexical categories give rise to grammatical (or functional) categories. (Heine et al., 2016)

This would perhaps imply that construction grammar aims at describing one particular change, while grammaticalization theory is focusing on cross-linguistic recurring lexical - grammatical associations; yet it more probably reads as a confession that no agreement has been made, so that it would be better to keep the two lines of work separate.

A second point at stake in the opposition would indeed be the methodology. Construction Grammar has been proposed as a synchronic description of a language, and Diachronic Construction Grammar was developed later on. On the contrary, grammaticalization has always been diachronically oriented. And, as it happens, what we have a hold on in diachronic corpora is the linguistic form — a word, or a recurring string of words. Constructions are more abstract; words can be constructs of constructions (e.g. *unusual*) and constructions can be schematic. They do not constitute readily available data. Hence, for empirical (and not theoretical) reasons, grammaticalization may have focused more on individual forms rather than on constructions proper. This is not to say that grammaticalization has virtually ignored the possibility for more complex items to grammaticalize; the crown jewel example of grammaticalization, for instance, the *be going to* future, is clearly constructional in nature, be it in the traditional sense of construction or in the Construction Grammar sense.

A corollary of this focus is that constructions have been held as ‘contexts’ in which the grammaticalization occurs. This is particularly clear in the work of Diewald (2006), but also in (Himmelmann, 2004):

Strictly speaking, it is never just the grammaticizing element that undergoes grammaticization. Instead, it is the grammaticizing element *in its syntagmatic context* which is grammaticized. That is, the unit to which grammaticization properly applies are *constructions*, not isolated lexical items.

(Himmelmann, 2004), emphasis original

Traugott (2003), in a book chapter written before her full adhesion to Construction Grammar, offers a similar view on this matter, and concludes her work by defining grammaticalization as:

The process whereby lexical material in highly constrained pragmatic and morphosyntactic contexts is assigned grammatical function, and once grammatical, is assigned increasingly grammatical, operator-like function.

(Traugott, 2003, p.645)

Hence, grammaticalization theorists have maintained a distinction between the forms and the constructions, taken not in the Construction Grammar sense, but in a so-called ‘pre-theoretical’ sense, which excludes, for instance, the possibility to regard simple words as constructions. In Construction Grammar, this vision of constructions amounts to consider only the ‘complex constructions’, leaving aside the ‘atomic’ ones. This incomplete acceptance of Construction Grammar has probably contributed to keep the two parties apart.

Grammatical contrast in Diachronic Construction Grammar

Construction grammarians had shown a tendency to reject grammaticalization as a process in its own right on the ground that there is no such specificity as lexical and grammatical material in Construction Grammar, making grammaticalization irrelevant compared to more general processes of change such as constructional change and constructionalization.

Yet, there are different kinds of constructions. Aside from the difference between atomic and complex constructions, other important differences exist. Traugott and Trousdale (2013) admit that there are lexical (or contentful) constructions alongside grammatical (or procedural) ones, a distinction which is directly inherited from the grammaticalization literature. More conventionally, construction grammarians distinguish substantive and schematic constructions (Noël, 2007), schematic constructions having usually more slots to be filled. If we draw a square representing those two dimensions, we would have, at the (atomic, substantive) corner, lexical words, such as *druid*, at the (atomic, schematic) corner, morphological markers, such as [ADJ *-ly*] which turns an adjective into an adverb (noun-modifier to verb-modifier indicative of manner), whereas at the (complex, substantive) corner we would have idiomatic expressions (e.g. ‘The worm has turned.’), and finally, at the (complex, schematic) corner, we find argument structures, such as the passive construction, where all slots have to be filled in a construct of this construction.

As those dimensions naturally emerge from a Construction Grammar perspective, it could be tempting to rephrase grammaticalization along these lines. This is what Noël (2007) has tried to do, distinguishing schematicization (the emergence of a new construction) from grammaticalization (a change in the semantic features of an existing construction towards a more grammatical meaning). Also, grammaticalization, understood as the development of a grammatical meaning of an existing construction, can only happen subsequently to a prior schematicization; indeed, “a grammaticalizing construction is never fully substantive” (Noël, 2007). According to this viewpoint, schematicization would be more on the formal side and grammaticalization, on the semantic one.

This is to be contrasted with the proposal of Traugott and Trousdale (2013), which distinguishes between constructionalization (the emergence of a new construction), and constructional change (a change in the features of a construction). Grammaticalization,

in this new dichotomy, would encompass both: in a grammaticalization process, constructional changes prepare the constructionalization to occur, which is itself followed by further constructional changes. As has been stressed before, grammaticalization happens on the formal and on the semantic planes at the same time.

In this line of thought, Traugott and Trousdale (2013) propose some kind of a criterion to distinguish constructionalizations whose output is procedural (roughly equals to grammatical) and those whose output is contentful (i.e. lexical):

The constructionalization of schemas always results from a succession of micro-steps and is therefore gradual. New micro-constructions may likewise be created gradually, but they may also be instantaneous. Gradually created micro-constructions tend to be procedural, and instantaneously created micro-constructions tend to be contentful. (Traugott and Trousdale, 2013, p.22)

In other terms, grammaticalizations, understood as grammatical constructionalizations, are likely to require more constructional changes to occur. They must be, in a way, more elaborately ‘prepared’. As the schematic constructions also involve a gradualness, we can understand that they are likely to be the result of grammaticalizations as well. It can already be argued that this necessary long constructional genesis of grammaticalization hints that something more may be at work in this process, a point which I will develop in the next subsection.

We are now facing a dilemma. Grammaticalization can be powered by two different kinds of Diachronic Construction Grammar mechanisms: constructionalization (the emergence of a new construction) and constructional changes (change of features of an existing construction). Several stances are thus possible:

- the sequence: constructional changes > constructionalization > further constructional changes can be seen as constitutive of the grammaticalization process, which would then call for this sequentiality. Hence, grammaticalization would be a specific macro-process made of a recognizable sequence of constructional micro-processes. At this point, two stances are possible. Either this happen only by chance, as these micro-processes can combine freely, so that grammaticalization is but a remarkable output, with no phenomenological coherence. Or it can be that the micro-processes responsible for grammaticalization have a reason to follow this particular sequence. This would suggest that an overall guiding mechanism exerts some drift over the micro-changes, and this mechanism would ensure the phenomenological specificity of grammaticalization.
- grammaticalization can be seen as sometimes caused by constructionalization, other times by mere constructional changes, and an intricate mixture of both in yet other instances. We have already argued that some grammaticalizations present several, separate and independent steps, which can be held as individual grammaticalizations each, and only one of them usually involves a constructionalization. A grammaticalization can thus well be, as Noël (2007) suggests, the effect of constructional changes alone. Again, we are faced with an alternative. This plurality of mechanisms can suggest that grammaticalization is but a linguist’s reconstruction over a diversity of independently motivated mechanisms which have worked along separate ways. It would thus be a diachronic illusion

of no explanatory value, and of no phenomenological status. But we can also consider the fact that, if no proper constructional mechanism is able to explain grammaticalization, then there must be something else at work.

It should be clear, at this point, that regarding grammaticalization as a phenomenon or not is a matter of choice, as there is no empirical evidence in favor of either of these two possibilities; nor is there a criterion to specify which empirical evidence would be relevant to settle this question. Admittedly, the impressive catalogue of concordant grammaticalizations gathered by Heine and Kuteva (2002) speaks strongly in favor of the reality of grammaticalization as a linguistic phenomenon. However, it may only reflect strong, repeated, directional tendencies in the metaphorical working of the human mind, but this would still stand as a relevant and remarkable phenomenon. Whether this phenomenon is reflected in the structure of language, be it described by Construction Grammar or not, is another matter entirely. To reiterate, I would maintain that, as long as no empirical criterion is available to assess of the specificity of grammaticalization *as a process of change*, it might just be a fascinating and enlightening picture of the main dynamical lines behind the organization of human cognition.

Bleaching and obligatoriness in a Construction Grammar perspective

From the preceding, non-exhaustive review of contact points between grammaticalization and Construction Grammar, it would appear that tying the two together only leads to an embarrassing and probably unnecessary theoretical discussion. However, Construction Grammar provides interesting insights on several issues encountered in the theory of grammaticalization.

The first one would be that of semantic bleaching. Though the idea was already discussed earlier on, actually as soon as the XIXth century (Hopper and Traugott, 1993), the term *semantic bleaching* has been first coined by Givón (1979), who used it to describe a pathway of semantic change often observed in grammaticalization process: ‘the process of *semantic bleaching* by which spatial concepts develop into temporal concepts but never vice-versa, and temporal ones into expression of existence-identity but never vice-versa’ (p.316). Lehmann has theorized the concept further in the devise of his famous ‘criteria’:

We now turn to the semantic integrity or semanticity of a word. For the sake of simplicity, I will assume that the semantic representation of a sign consists of a set of propositions taken from some semantic metalanguage commonly called semantic components or features, and that those propositions which are conjoined (rather than disjoined) contribute to the semantic complexity or semanticity of the sign [...]. Desemantization, or semantic depletion (Weinreich 1963:180f) or bleaching, is then the decrease in semanticity by the loss of such propositions. As said above, the last proposition is commonly lost at the moment where the last rest of the significans also disappears; but as we shall see, either one can continue, at a submorphemic level, without the other. (Lehmann, 1995, p.114)

Sweetser (1988) offers a more contrasted perspective on the question of bleaching; she says that meanings are metaphorical schemas enriched by the domain in which they

are used. During a grammaticalization, the form undergoes such a domain change, so that the meaning-schema loses some of the features brought forth by the source domain and are replaced by those characteristic of the target domain. However, she adds:

the meaning shifts involved in grammaticalization are necessarily shifts towards a relatively abstract and topological domain of meaning, since those are the meanings that we use in grammatical systems. This being the case, there will be less “fleshing out” of the transferred image-schematic topology when the transfer is into a domain which centrally refers to the topological aspects of meaning, rather than to some of the other aspects of rich lexical meaning.

(Sweetser, 1988, p.401)

On the question of semantic bleaching, Construction Grammar might provide new insights. An atomic, substantive construction comes to be enrolled in a schematic construction: for instance, *provided* in the construction:

[[Proposition₁], provided [Proposition₂]]

with the approximate meaning ‘[P₁], as long as [P₂] holds’, e.g.:

Senior Level residents like to look at young people, provided they aren’t real young people, who are, you know, too noisy.

CHET, Arthur, *The Trouble with heaven*, 2013 (COCA)

In this construction, *provided* participates in a more schematic construction, so that its original substantive meaning is ‘bleached’. Bleaching can thus result from a schematization of the form.

However, this counts as bleaching only because we consider the wrong item. That the lexical meaning of *provide* is bleached in the PROVIDED construction is only indicative of the fact that, in the latter, *provided* does not act alone as a sign, for it is the construction which means something, as a whole, not through its individual components. We can then debate whether the meaning of the schematic PROVIDED construction is richer or poorer than the meaning of the substantive construction *provide*, yet it is clear they both have strong semantic features.

This being said, the notion of schematicity might not be able to entirely settle the issue raised by the notion of semantic bleaching. For instance, an increase in complexity with no or little increase in schematicity can lead to bleaching as well. This would be the case for most discourse markers (e.g. *par contre*, which is not schematic, yet is more complex than *par* and *contre* taken as isolated items). It is also the case for the French *d’emblée*, with the meaning of ‘straight away’. It has stemmed from a now extinct verb *embler*, as in :

Soiées sur vostre garde, car pour certain il i a asés priès de chi une grant quantité de Gantois, car je les ai veus et oïs, et portent eschielles, et embleront Audenarde...

Be on your guard, as for sure there is, not far from here, a great deal of Gantese, for I have heard and seen them, and they carry ladders, and they will take over Audenarde...
FROISSART, *Chroniques*, XI, c.1375-1400, p.138 (Frantext)

The participle *emblée* has then lexicalized into the construction *d’emblée*, with the reduced meaning of *off guard*, to speak exclusively of the taken over a city:

Et neantmoins fist sejourner son arriere garde à l'entour et près de Beauvais, affin de la prandre d'emblée s'il veoit que les gens d'armes en fussent hors.

And yet he kept his rear guard in the neighborhood of Beauvais so as to take it offguard were he to see its troops sortieing it.

LE CLERC Jean, *Interpolations et variantes de la Chronique scandaleuse*, 1502, p.299 (Frantext)

A semantic expansion occurs about three hundred years later, after which *d'emblée* has the meaning of 'right away':

Parmi celles que l'abbé coucha de ce temps-là en joue, je me souviens d'une grisette assez jolie pour qu'il l'eût amenée d'emblée, si elle n'avoit résisté à toutes ses épreuves.

Among all those the abbot spotted on in these times, I remember a maid pretty enough for he to bring her right away [over the prince], if she had not resisted all his trials.

VARENNE Jacques de, *Mémoires du chevalier de Ravanne*, 1740 (Frantext)

The first step of the evolution is difficult to account for since data is scarce, so that the *d'emblée* construction seems to arise out of the blue. However, the second step displays a clear semantic bleaching, while no increase of schematicity of the construction is to be reported in this instance. This example also illustrates the fact that semantic bleaching and semantic expansion are two concurrent notions, in the very etymological meaning of the word, as the construction loses its semantic specificity at the same as it gains a wider array of uses. This parallelism between bleaching and expansion will be modeled in chapter 9.

Semantic bleaching can probably be better understood through collocational analysis. The more an item shows collocational preferences, the more it displays a clear, overt and salient meaning. If an item can be used with pretty much anything, then its meaning is not quite substantial. Here again, we meet some limitations of the theoretical analysis, especially due to discourse markers. Discourse markers, conjunctions, being extra-clausal, have very low collocational preferences. In fact, it is quite a wonder how users of a language can properly understand them, since their meaning is extremely opaque most of the time; for instance, what could be the meaning of *décidément* in the following utterance?

Mais voilà, il n'est pas religieux, et dans le trouble provoqué par une telle scène, il n'a pensé qu'à offrir ses services. *Décidément*, la complaisance ne suffit pas toujours; il faudrait la conviction.

But there, he is not religious, and in the confusion that such as scene arouse, he thought but to offer his help. *Décidément*, kindness not always is enough; there would need conviction.

ARTIÈRES Philippe, *Vie et mort de Paul Gény*, 2013, p. 72 (Frantext)

Yet, speakers know perfectly how to use them and in which circumstances. Once more, the special status of discourse markers is evidenced; to understand them, and to be able to account for their own semantic bleaching, it may be that further descriptive tools are needed.

Another issue of grammaticalization which dissolves in a Construction Grammar perspective is the idea of obligatoriness. As we saw, obligatoriness of a form in a given context had been seen as a hint for its grammaticality (Idiatov, 2008). Indeed, the fact that some forms are mandatory is curious; French determiners, for instance, are required in a Noun Phrase construction. This fact receives a straightforward explanation in Construction Grammar. Some items are obligatory, because they are part of a fixed, conventionalized construction. The Noun Phrase construction, in French, is minimally made of a determiner and a noun (and can be enriched through the genitive construction, the adjectival construction, the relative proposition construction, etc.). The [det N] noun phrase construction has competed against the [N] over the history, and finally won the game, so that determiners have become obligatory. Yet, this ‘obligatoriness’ is a side effect of the evicting of the formerly entrenched noun phrase construction by a new one, just as two words can compete to express one meaning.

The processual view

We shall now present one last approach, due to Himmelmann (2004). Though most authors would agree on a schematic definition of grammaticalization of the kind (primary: lexical > grammatical; secondary: grammatical > more grammatical), Himmelmann denies the relevance of the lexical/grammatical distinction, which he calls the ‘box metaphor’ (according to which the items of language can be dispatched between two separate boxes respectively labeled ‘lexical’ and ‘grammatical’). The chief weakness of this ‘box metaphor’ lies precisely in the uncertainty surrounding the lexical/grammatical distinction, so that ‘for quite a number of items it is not clear in which box (grammar or lexicon) they belong’ (Himmelmann, 2004, p.25). He proposes then a reversal of perspective: instead of calling ‘grammaticalization’ what leads from a lexical to a grammatical status, one can view grammaticalization as a process with distinct phenomenal features, so that a grammaticalization can be identified as such without any reference to the box. This shall provide in turn a characterization of what is grammar and what is lexicon: respectively, what results from a grammaticalization process on the one hand, and what results from a lexicalization process on the other hand (Himmelmann, 2004, p.25). In this perspective, the distinction does not lie in the results of the processes, lexical or grammatical, but in the processes themselves, lexicalization or grammaticalization.

Himmelmann insists, in particular, on the importance of constructions in a grammaticalization process. It is not a single, isolated item which can be said to grammaticalize, but the item in a particular construction. Grammaticalization is thus associated with three kinds of expansions: host-class expansion, syntactic context expansion and semantic expansion (the latter including pragmatics). Let us consider these three processes in turn, in the light of the notions developed in chapter 1.

The host-class expansion directly refers to the fact that the grammaticalization happens in the context of a construction. To follow the notations of Himmelmann (2004), we shall note $[\{A\} B]$ (the exact order and number of the elements in the construction is not relevant), the host-class expansion corresponds to a widening of the paradigm $\{A\}$ in the context of the construction. The example given to illustrate this phenomenon is the grammaticalization path *demonstrative* > *article*: an item B which follows this path within a $[\{Noun\} B]$ construction will become compatible with

a greater number of nouns. We can provide other instances of host-class expansions; the French construction [une sorte de {N}] underwent a host-class expansion of its noun paradigm. First, *une sorte de* was used to refer to a particular subspecies of the species given by the noun N, as in:

Marius avoit souverainement aymé une sorte de gobeletz que l'on appelle Nerytum[.]

Marius had sovereignly treasured a kind of goblets that is called Nerytum.

DE SAINT-JULIEN, Pierre, *De non se courroucer*, 1546, p. 158 (Frantext)

Then, *une sorte de* came to express a mere similarity with the noun N (similarly to the English *a kind of*), which no longer needs to be associated with different subspecies. As such, the noun N can be instantiated by a proper noun, e.g.:

[La gravure] représente une sorte de Pierrot lunaire, vague ovale d'un visage, petites fentes pour les yeux et la bouche !

The carving displays a kind of lunar Pierrot, hazy oval of a face, small slits for the eyes and the mouth!

PICQUET, Pauline, *Sans illustration*, 2013, p. 72 (Frantext)

Syntactic context expansion is the process by which the construction can be used in a greater number of different positions in the utterance. This is the case for the French quantifier [plein de {N}]: since *plein* is originally an adjective, it could at first only be used to complement a noun ('Chevalier plein de foi', 'a knight filled with faith', *La Queste del Saint Graal*, 1220, Frantext) or as a subject attribute ('il sont venu plein de joie', 'they came full of joy', idem). As a quantifier, it is no longer bound to this syntactic constraint (e.g. 'Les femmes pour la plupart avaient encore plein de cheveux.', 'Women had still plenty of hair for the most.', AKERMAN, Chantal, *Ma mère rit*, 2013, Frantext). According to the terminology we presented in chapter 1, this would correspond to an increase of the range of colligations of the construction undergoing a grammaticalization.

Semantic context expansion, deemed as the 'most important' of the three by Himmelmann (2004), is the equivalent of the phenomenon of 'semantic bleaching': it corresponds to a widening of the collocates of the construction. We already met numerous instances of this phenomenon, for instance the semantic context expansion of *coup* within the genitive construction ('un coup d'épée' > 'un coup de téléphone'). We can also think of *d'emblée* ('at once'), originally referring to the military taking of a city by means of surprise, which can now be used with all sorts of collocates (e.g. 'j'aimai d'emblée tout ce vert et ce brun').

Therefore, the three processes put forward by Himmelmann could be seen as a widening and loosening of the colligations, of the collocations, and of the internal paradigm of a construction. None of these three processes is specific of grammaticalization (as shown by our examples above), and a given grammaticalization process is not bound to present all three. The emergence of discourse markers, such as *du coup*, typically does not involve a host-class, modals are seldom associated with an expansion of syntactic contexts (on the contrary, they tend to be increasingly bound to a specific syntactic position), while semantic context expansion is likely to take place in all cases.

Himmelmann is the first to admit that it would be interesting to investigate to which extent these three phenomena are necessary for a process to be classified as

a grammaticalization, and how they would interact for grammaticalization to occur. Yet, at least two of these expansions are optional. We may be left with a characterization of grammaticalization as a combination of a semantic expansion and either a colligation or internal collocation paradigm widening (or maybe both), and a bunch of new questions. For instance, do these three possibilities correspond to three subtypes of grammaticalization? Does the order with which the subprocesses occur matter? Or must they work in parallel?

We would have two major issues regarding Himmelmann's account of grammaticalization. First, grammaticalization most often involves a constructionalization step (we posited that such was the case for primary grammaticalizations). However, host-class expansion and colligation bleaching are both defined with reference to a given construction. As the construction of interest may not be the same before and after the grammaticalization step, it is problematic to define these two expansion subprocesses with respect to the whole process. Second, there is no reason why the collocation between two constructions, and the collocation of a construction within another one, would not similarly contribute to shape the meaning of these constructions. Therefore, we can consider that all three processes are associated with a phenomenon of semantic expansion. This is especially the case if we recall that syntactic contexts, in a Construction Grammar perspective, are nothing more than constructional contexts, just like the two others. Since an expansion of the constructional contexts of use of a construction amounts to a semantic expansion, we can argue that Himmelmann's proposal leads to equate grammaticalization with semantic expansion, in a wide, fully constructional sense.

Truly enough, there may be differences between syntagmatic bleaching (a construction tends to loosen its ties with all its neighboring correlates) and paradigmatic bleaching (a construction comes to be used in a wider array of other constructions, or hosts a larger and less focused paradigm). The latter, in particular, will be crucial regarding the categorization of an item, as we discussed in the previous chapter. Since grammaticalization is sometimes seen as the movement towards tighter and more closed categories, the paradigmatic axis of the bleaching may indeed be specific, or at least symptomatic, of grammaticalization. However, paradigmatic and syntagmatic bleaching can both be associated with semantic expansions which are consensually not regarded as grammaticalizations, they cannot serve as a sure criterion. For instance, *rappeler* {N} > *rappeler que* is an instance of a colligation expansion, and *attraper* + something > *attraper* + {A = froid, chaud, mal}, where A is a closed paradigm of adjectival attributes usually associated with bleached and frequent verbs such as *avoir*, *prendre*, and *faire*, is an instance of host-class expansion. Yet they can hardly be considered as grammaticalization processes.

One last remark is in order. Himmelmann not only defines grammaticalization, but lexicalization as well. He proposes to see lexicalization as the individuation of a member of a paradigm within an existing construction, e.g. [{A = A₁, ... A₁, ... } B] > [A₁ B], where the construction [A₁ B], which was a possible construct of the former construction [{A} B], becomes now a construction in its own. This closely corresponds to what we have called a 'speciation' event. Once [A₁ B] becomes independent, it can acquire new features, which will bring it further apart from its originating construction. Lexicalization, seen as the addition of a new item in the repertory of linguistic forms, is

not relevant as such in this work: since all linguistic forms are constructions, the forging of a new one amounts to a constructionalization. Therefore, we rejoin Himmelmann's particular definition of lexicalization.

To summarize, the proposal of Himmelmann (2004) fully embraces the view of grammaticalization as a process. Yet, it seems that the particular characterization of this process offered therein fails to achieve the expected level of specificity which is required to account for grammaticalization. We have indeed argued that his proposal amounts to a semantic expansion in a broad sense of the term. It appears, however, that this work could serve as a basis for a typology of semantic expansions.

2.3 The specificity of the grammaticalization phenomenon

In this section, I will expose the position on grammaticalization which shall be adopted in this thesis. As we want to investigate it empirically, it is important for this position not to be too much restrictive. If it appears that some language change processes in the subclass behave differently, it will then be an argument to further focus the definition of grammaticalization. This particular purpose rules out restrictive definitions such as the six criteria proposed by Lehmann (1995). First of all, we want to argue that grammaticalization is best seen as a semantic expansion process. It can also involve a constructionalization, which we claim to be mandatory in instances of primary grammaticalization. Here we partly follow Himmelmann (2004), when he states that the locus of grammaticalization is not a single item, but an item within a construction. As constructionalization involves a new form-meaning pairing, it is automatically associated with a semantic expansion.

What is crucial, however, in a grammaticalization process, is that the source meaning is less grammatical than the target meaning — we acknowledge therefore that the following position is tied to the 'box approach' criticized by Himmelmann (2004). This stance requires to address two main issues. The first one is the lexical/grammatical distinction. The second one is more complex: given that a grammaticalization is a semantic expansion process, how can the fact the target meaning is more grammatical than the source meaning entail a specificity of the process itself? In other terms, are there specific mechanisms which would lead to an increase of the grammaticality of a linguistic form?

2.3.1 Grammaticality as linguistic awareness

To summarize the non-exhaustive review we gave in the previous section, it is clear that scholars share a common, intuitive idea of the lexical/grammatical distinction. Grammatical items act on an operative, meta-discursive level. They serve to constrain the mental arrangements that the addressee can perform on the contentful, lexical items. However, even this view is not without problems. For instance, Melis and Desmet (1998) point out that the word *truc*, in French (suitably translated by 'thing' in many instances, with the important difference that 'thing' appears in paradigms alongside grammatical items, e.g. the paradigm $P = \{\text{thing, way, where, how}\}$ in the [somep] construction, while 'truc' does not) formally behaves as a very prototypical

noun, but carries a clear pragmatic meaning and is almost devoid of content. It can be nonetheless noticed that even if *truc* is contentless for some its uses, it does not perform any operations on the other components of the utterance, nor it provides major clues for the hearer to interpret the utterance. Also, prefixes such as *un-* ('unnatural') or *re-* ('reconsider') can be said to perform some operation on the meaning of the lexical elements they associate with, while they are usually not considered as properly grammatical. The distinction, therefore, is not clear, and the list of ambiguous items would be nearly endless.

The lexical/grammatical distinction relates to psycholinguistic processing

It is more or less a certainty that the difference between grammatical and lexical items chiefly lies in the psycholinguistic processing of these items. Albeit the field is in dire need of more advanced psycholinguistic results on this question, the overall intuition has been laid down for a few decades now, starting with Relevance Theory. However, no matter how extensive the progresses would be in that sense, they would be pointless in a diachronic perspective: as we stated numerous times, historical linguistic data cannot provide access to the psychological processes happening within the speakers' minds of older times. Should we conclude that, since the lexical/grammatical distinction is chiefly a matter of cognitive processing, we have no way to discriminate the two for older states of languages?

Obviously this is not the case and the diachronician expertise is often able to recognize some forms as clearly grammatical, or clearly lexical. Actually, we can hypothesize that the way linguistic items are processed will be reflected in the structural features which characterize these items, for language is the product of its cognitive use, and is shaped by it. Consequently, there should be traces, that could be empirically evidenced, of this difference in psycholinguistic processing. As for which these structural features could be, it is of course the key question, and answering it would be a crowning achievement. In the following, we will therefore content ourselves with a mere speculative proposal, which cannot lead yet to empirical measurements, and only aims at stating the view adopted in this thesis.

A tentative definition

We shall consider that a construction is grammatical as soon as it comes to express information concerning other linguistic items — in other, vaguer terms, as soon as it 'feels' the utterance. Grammatical items illustrate the fact that the utterance presents a dimension of reflexivity; it speaks about itself (this is exactly what is covered by the idea of a 'procedural' meaning component). This statement is rather consensual, and does not solve in any way the issues that have been raised above, in that it is of little help, at least as such, in distinguishing lexical from grammatical items from the linguistic data alone. At least, it leaves no possible ambiguity regarding discourse markers: insofar as they are extrapredicative and yet introduce or orient the remainder of the utterance, they have to be considered as grammatical, for their meaning is highly 'aware' of the fact that they are part of an utterance. Hence, they conform to the reflexivity criterion.

Allow us at this point a short terminological note. Reflexivity, in linguistics, is associated with many different notions and using it now would only add to the confusion. Therefore, we will favor the rather metaphorical term of ‘linguistic awareness’ to express this idea that the meaning of some language units makes explicit reference to the fact that they are linguistic units, partaking in an utterance. The idea behind this terminological choice is to stress the fact that we aim to transpose the idea of a ‘procedural’ meaning in terms of specific structural features, so that we make complete abstraction of the speakers, and adopt, so to say, the viewpoint of the forms themselves — roughly, nouns do not ‘see’ the other linguistics forms, while a conjunction will be most ‘aware’ of them.

As stressed above, we nonetheless believe that these differences, hereby seen from a strictly language internalist perspective, have their roots in the cognitive organization of language within individual speaker’s minds, of which we aspire to find a reflection in the datum of recorded utterances. In no way the use of the term ‘awareness’ should imply that words are conscious entities; but they do encode in some way the cognitive activity of the speakers when manipulating language (Chafe, 1994). In that sense, the ‘awareness’ of grammatical constructions would be the dim trace of the fact that language users are themselves aware of the linguistic nature of their language, which would be a requisite condition for grammar to emerge (i.e. language would no longer be limited to interacting with the world so as to elicit a given behavior from other language users, but would become something that can itself be interacted with) . In the following, when I say that a given construction is ‘aware’, this term must therefore be understood as a metaphorical shortcut to convey the idea that the description of this construction’s meaning cannot avoid to make at least partially reference to its linguistic status.

A crucial question which is immediately raised by this proposal is: how far this theoretical metaphor coincides with schematicity? Indeed, we could consider that a construction, as soon as it is schematic, is not autonomous and therefore aware of its constructional status. This would make all schematic constructions grammatical. A grammaticalization would then be a constructionalization involving the emergence of a paradigm. However, some grammaticalizations are categorial change, that is, one linguistic construction comes to be recruited in a grammatical paradigm (or: a repeated collocation emerges out as a construction belonging to a grammatical paradigm). This would be the case for *du coup*, which grammaticalized by joining on the paradigm of discourse markers, not by developing an internal paradigm. One could argue that this is specific to discourse markers, yet other examples of this fact are to be found, e.g. with complex adverbials (*peu à peu*) or temporal markers (*soudain*). On the other hand, some schematic constructions are hardly grammatical (e.g. the one we already mentioned, [{avoir, faire, prendre} {froid, chaud, peur, pitié, part, mal, peine, ...}]).

This is why I would like to go as far as to argue that the grammatical status of a construction is not formal, but semantic above all. True, schematicity is usually a very good hint of grammaticality, but it is neither sufficient nor necessary. The specificity of the grammatical status would be this ‘linguistic awareness’ of grammatical meaning. Unfortunately, I have no empirical criterion to offer so as to support it. For now, the ‘awareness’ of a construction could only appear in the lexicographers’ account of its meaning. Especially, if an approach such as the one of Wierzbicka (1996) were

developed for any construction, then to describe the use and meaning of a grammatical construction, one would need propositions with explicit reference to the fact that this construction either acts upon other parts of the utterance as arguments, or is part of an utterance.

Applying the distinction

According to this toy definition, conjunctions and prepositions are grammatical, but also syntactic constructions such as the ditransitive one, and the majority of schematic constructions. Importantly, there are two lines between which grammaticalization unfolds, depending whether the linguistic ‘awareness’ of the form is rather external or rather internal. Discourse markers are ‘externally aware’, as they are blind to their own status, but they are aware of the utterance and even of the discourse situation in which the utterance occurs. Syntactic constructions are internally aware in that they perform linguistic operations upon their arguments. Basic argument structure constructions are the closest to the ‘internal awareness’ pole of the awakening. Modifiers (be it morphological or propositional) are less ‘internally aware’, in that they are applied to another element, but they are still far from the external pole. Conjunctions, on the other hand, lie closer to the external side than to the internal. Prepositions would be in-between. This picture is summarized on Figure 2.3. Of course, it is expected that different languages would present a different location of its different elements: conjunctions, prepositions, morphological markers, argument structure, discourse markers, are to be understood as Standard Average European features (Haspelmath, 2001), rather than universal categories.

The status of morphological items, such as the English and French construction Verb-to-Noun construction [V_{tion}], is nonetheless debatable. It is used to mark a change of category; in that sense, the linguistic categories of both the source and the target item become salient. As linguistic categories only make sense with respect to the utterance, it can be safely considered that such kind of category-shifting or category-attachment markers are indeed grammatical, as their role shows some awareness of their taking place in an utterance. Similarly, the English -ISH construction (Traugott and Trousdale, 2013), insofar as it involves a category marking, would count as grammatical, and so would the ordinal marker -th. More complicated is the case of tense marking (e.g. -ed in English, which interestingly presents the same historical origin as -th). Does it show some awareness of the fact that it belongs to an utterance?

I would like to venture into arguing that as soon as an item operates on another item, it already shows some linguistic reflexivity, equated to grammatical status in this attempt of a definition. This would imply that adjectives, at least in French and in English, and especially when used epithetically, are already partly grammatical. To quote a famous philologist:

The incarnate mind, the tongue, and the tale are in our world coeval. The human mind, endowed with the powers of generalization and abstraction, sees not only *green-grass*, discriminating it from other things (and finding it fair to look upon), but sees that it is *green* as well as being *grass*. But how powerful, how stimulating to the very faculty that produced it, was the invention of the adjective: no spell or incantation in Faërie is more potent. (Tolkien, 2014, p.41), emphasis original

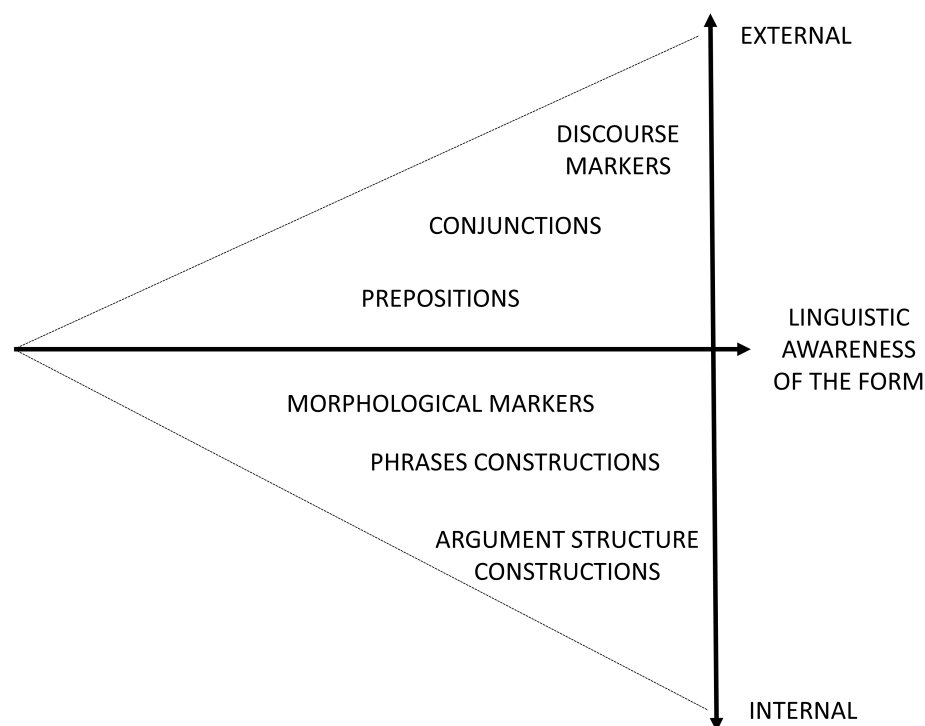


Figure 2.3: An ordering of French and English grammatical categories according to the external/internal polarity of linguistic awareness.

Of course, adjectives are contentful items, which would locate them on the lexical side of language. However, it is by no means forbidden for a grammatical item to have some contents. English modals, which are consensually considered as grammatical, are distinguished from each other because they have a different lexical load. Admittedly, they form a much closed class than the seemingly open-class of adjectives; but they are also much more grammatical in that they are characterized by a way greater linguistic reflexivity. Also, in some languages, adjectives arguably do not exist as a proper category (Thompson, 1989); there is no necessity of an adjective class as it exists in most European languages, and properties can be attributed to entities through a wide array of linguistic means (Dixon, 2004). Yet a cross-linguistic perspective is not necessary to see that the adjective class has some peculiarities which, as we argue, hint at its slightly grammatical character.

In both English and French, numeral cardinal adjectives can be used as determiners, determiners being a closed class, consensually held as grammatical. Other observations along this line have even led to posit an ‘adjective-determiner continuum’ (Haspelmath, 2000a). In French and English, past participles, such as *perdu*, *bouilli*, *écrabouillé*, *stolen*, *frightened*, *splintered*, although built out of a grammatical construction, can frequently be used as adjectives and are probably stored as such cognitively. These two examples show a clear porosity between adjectives and more traditional grammatical categories; a porosity which is not as pervasive for nouns and verbs, the two other ‘major’ or lexical categories. Another argument for the ‘awareness’ of the

adjective is that in English, the compound construction of two names, such as *a monster outfit*, *a bone ship*, *a light day*, though historically related to a genitive, is clearly analogically related to the epithetical adjectival construction, so that the first noun, in these compounds, plays a role similar to an adjective. Interestingly, a historical corpus study has shown that those compounds were relying both on noun + noun compounding and adjective + noun compounding (Gavranović, 2015), though in the latter case, a fusion has occurred which has lexicalized the association between the two parts of the compound, hence blurring the adjectival construction. In any case, the compound construction relies on a grammatical operation between words to perform its function of attributing properties to a noun. It thus seems legitimate to assume that the epithetical adjectival construction involves a grammatical operation as well, or, in other terms, exhibit some awareness of the status and ordering of the words it takes as arguments.

To summarize, there can be a lexico-grammatical continuum based on the continuous character of the ‘awareness’ quality. However, it seems best to complement the lexical/grammatical polarity by a second one, the internal/external one, which unfolds as linguistic items evolve further towards the grammatical pole. Grammaticalization will thus correspond with an increase in semantic awareness, while pragmaticalization, as opposed to schematicization, will combine with the grammaticalization process to increase this awareness in the external or in the internal direction, respectively.

2.3.2 The mechanisms of grammaticalization

Having stated what I consider as ‘grammatical’, a new, arranged definition of grammaticalization follows (which is a mere metaphoric rephrasal of ‘semantic expansion towards more grammaticality’ in the light of our own viewpoint on grammaticality):

Grammaticalization. *Grammaticalization is an awakening of a linguistic form to its being part of an utterance, or to its own status of linguistic form.*

However, here we focus on the specificity of the result of the process, and not of the process itself. The idea is therefore to identify special, specific mechanisms, which would ensure this ‘awakening’ of a form. Indeed, such an awakening is not implied in a semantic expansion: an additional, specific mechanism seems required. Also, the possibility of degrammaticalization becomes all the more puzzling: How can a linguistic form become oblivious of its surroundings, or of its own operative structure? Unfortunately, this crucial question will not be addressed in that work. Yet, as shall be clear by the end of this chapter, I believe that different mechanisms should be necessary for such a process to occur; especially, the entrenchment of a specific construct of a grammatical, schematic construction as a construction of its own (which corresponds to (Himmelmann, 2004)’s lexicalization) can lead to such a reduction in linguistic awareness. As an overall distinct process, it will be left aside for the remaining of this thesis.

Four mechanisms behind grammaticalization

The specificity of this ‘awakening’ does not rule out the possibility for other, non-specific mechanisms to act towards a grammaticalization process. Here we shall present

four of them which appear to play a role in many instances of grammaticalization.

1. **Semantic shift.** A construction can be subject to a semantic shift and therefore become more grammatical. In such a process, the construction associated with the source meaning is the same than the one associated with the target meaning. Formally, it is likely to be observed through a change of the collocates' paradigm, so that the most prominent collocates of the paradigm would have changed by the end of the process. It is equivalent to a semantic bleaching, possibly followed by a semantic tightening. In the framework of a semantic network we shall develop later, it consists in invading new semantic contexts, while losing those which had been formerly acquired. A theoretical representation of this process will be sketched in chapter 9. We hypothesized this process to be chiefly associated with secondary grammaticalization (e.g. the source meaning is already grammatical). Semantic shift is for instance responsible for the frequently attested 'temporal marker' > 'causality marker' change (e.g. English *since*), listed in Heine and Kuteva (2002, p.291).
2. **Constructionalization.** A new construction emerges out of repeated collocation patterns, through inductive abstraction and/or reanalysis. The new construction can host a paradigm (e.g. the [en {N = vertu, prime, droit, état, place, vue, faveur...} de] construction in French), but not necessarily (typically, *du coup*). If the new construction has a grammatical meaning, then it counts as a grammaticalization. The difference with semantic shift is that the construction associated with the source meaning is not the same as the one carrying the target meaning. This specific way to obtain a grammaticalization would coincide with what we consider to be a 'primary' grammaticalization, following Traugott and Trousdale (2013).
3. **Colligation, or paradigmatic growth.** Once a construction with a paradigmatic slot has formed, it can recruit constructions within that paradigm. For instance, the complex preposition construction [à {NP} de] (*à la lumière de, à la vue de, à l'issue de, à la perspective de, à la charge de...*) can recruit new members. These members are then said to colligate with the schematic construction. This increases the abstractness of the hosting construction and leads to its further semantic bleaching (now a paradigmatic semantic bleaching, by contrast with what happens in a semantic shift, where the bleaching is syntagmatic instead). However, the hosted constructions (*lumière, issue, sortie, ...*) do not become grammatical. No grammaticalization is therefore entailed by the colligation of new items in an already grammatical construction. An isolated complex preposition would then be considered as grammatical only insofar as it is a construct of a grammatical construction, but this we consider to be an improper metonymic expansion of the term grammaticalization. Furthermore, this colligation mechanism can be the source of new grammaticalization through the next mechanism.
4. **Speciation.** This mechanism is the same as Himmelmann's lexicalization. A specific member of a paradigm becomes a construction in its own right. Usually, if the construction from which the new member is speciated is grammatical,

with an ‘internal’ orientation of its ‘linguistic awareness’, then the construction is likely to be lexical (e.g. [coup d’œil], speciated from the grammatical genitive construction [{N} de {N}], is lexical). Conversely, if the originating construction was ‘externally’ grammatical, then the speciated construction has good chances to be grammatical itself (such as [en train de] from the [en {N} de] construction), making it an instance of grammaticalization. One should note that the speciation mechanism can be considered as a constructionalization (it involves a new form-meaning mapping), although it is a very specific one, and is best to be distinguished from the previous use of the term. Due to this speciation mechanism, complex schematic constructions can act as a template from which new grammatical constructions are to be forged, through the intermediary of repeated colligation.

Among the usual mechanisms invoked in a grammaticalization process (reanalysis, analogy, inference), analogy is especially active in colligation (new members are recruited by analogy with existing members), reanalysis in constructionalization (contrary to speciation, the new construction is not inferred within the boundaries of an existing construction, but cross over the boundaries of several existing constructions), and inference is always crucial save for colligation, since it allows for the use of a form to be mapped onto a new meaning.

A dichotomic view of the lexical/grammatical opposition

What is most remarkable with grammaticalization is that a word which is purely lexical, or contentful, or referential, that is, which provides a conceptual image, is enrolled to serve an operative function in the process of shaping an utterance meaning. Also striking is that the gradual, continuous character of grammaticalization, is seldom to be found in data, where the use of a given linguistic form is instead prone to exhibit abrupt shifts: at some point, seemingly without warning, a new use starts to appear in texts or speech, and reconstructing the process linking the old use to the new one is, most of the time, quite a puzzling challenge. What is gradual is the entrenchment of this new use through a long and comparatively slow increase of frequency. In line with Traugott and Trousdale (2013), we can consider that grammaticalization is a three-step process: small constructional changes, abrupt constructionalization, which requires a reanalysis (which they renamed ‘neoanalysis’) to be achieved, and further constructional changes. Note that this account focuses on ‘primary’ grammaticalization, since semantic shift and speciation do not require a reanalysis, though they may be equally abrupt. The timescale of the whole grammaticalization can thus be of the order of a century, whereas the crucial shifting step happens in less than a decade; at least it appears so according to corpus data.

The crucial question is thus: how can this semantic transition take place? In a speciation process, the grammatical character of the new construction is almost mechanically inherited from the construction it originates from; and we already suggested that this inheritance depends on the sort of grammaticality of the source construction. In semantic shift, there is room for further argument. The source meaning can, for instance, already be grammatical, but the process is a (secondary) grammaticalization insofar as the target meaning is more grammatical still. There may not be the need

for a special mechanism to be involved here, since once a construction reaches the grammatical domain, almost all semantic expansions should take it further away from its originating point, hence further away from the lexical domain. There is thus a good probability for a semantic change within grammar to lead to further grammaticality. As for semantic shifts starting within the lexical domain, we already argued that they cannot lead, by themselves, to the grammatical domain. We are left therefore to deal with the second process, that is, the creation of entirely new constructions of grammatical status from lexical elements, resulting from a constructionalization step.

As will be argued at length in chapter 7, I propose that constructionalization can be partly understood as a semantic expansion, so that grammaticalization is a semantic expansion no matter what, with a meaning gradually expanding. But at some point, there is a gap between two meanings of a very different nature — the lexical and the grammatical one, the image and the function. This is what has been metaphorically described as some sort of an ‘awakening’. Unless we deny the existence and the specificity of this semantic gap, there might be little to no reason to think of grammaticalization as a distinct and relevant phenomenon. This entails to throw away the prototypical, polar view of the lexical/grammatical opposition, and to assess a clear chasm between the two (the ‘equatorial’ view). Both parts can be associated with meanings, as Construction Grammar has striven to show for the fully schematic constructions (Goldberg, 1995), but these meanings are, somehow, further away from lexical meanings than the typical distance between meanings. If we adopt a view of the meaning space so that different meanings can be represented as discrete units localized in the meaning space, then it might be that a lexical meaning L_1 can be further away from another lexical meaning L_2 than from a closer grammatical meaning G . Yet, it is to be expected that there will always be an other lexical meaning L_3 closer to L_1 than G is. In other terms, the meaning space is not homogeneously filled and there will be gaps between the areas corresponding to the lexical domain, and those corresponding to the grammatical one.

I would like to go further still in this dichotomy. To stick with the spatial representation metaphor, it might be that lexical meanings and grammatical meanings belong to different planes (or more generally speaking, different non-intersecting subspaces). Of course, there might be a succession of planes, and I am inclined to believe so, but let us assume, for the sake of the discussion, that there are chiefly two of them. For a linguistic form to go from a semantic node of one plane, to a semantic node of another plane, it must pass through a semantic expansion (or perform a semantic leap); but this semantic expansion will not be so straightforward as the one required to go from one node to another of the same plane. Something more must be at work — something which would explain why primary grammaticalization requires constructionalization.

Transfer hypothesis

One possible solution has already been hinted at: the lexical linguistic form has to be hauled to the grammatical plane by another, already grammatical form. Once there, the lexical linguistic form will not have a life of its own, but only as part of the construction it now belongs to, pairing up with the grammatical element which dragged it up. Without this tutoring of the grammatical element, which somehow transfers its

grammatical status to the lexical form, there is no way this form can grow out a grammatical meaning of its own. Note that this hauling is not necessarily an instance of paradigmatic growth (which would not count as a grammaticalization). The pairing can indeed happen within the context of a new, emerging construction, thanks to a reanalysis, and would be therefore classified as a constructionalization. This requires to define what would be grammatical forms or elements, by contrast with grammatical constructions; a possibility is that they are separate and individuated forms which have no lexical uses (e.g. *that, at, to*), forms that appear in grammatical constructions, but never as constructions of their own.

Thus, according to this transfer mechanism, grammaticalization could be seen as the process by which a grammatical item transfers its grammatical status to a lexical item within an emergent construction. In these terms, grammaticalization is a specific kind of constructionalization, one that involves, among its components, both an already grammatical item and a lexical element. It is, therefore, a distinguishable phenomenon.

Of course, this view presents its own load of flaws. First of all, it might not hold for all languages. It works at least for the Standard European matrix of languages (Haspelmath, 2001), but languages such as Mandarin Chinese might rely on different grammaticalizing mechanisms. Second, what are those grammatical elements? In French, they are items such as *de, à, par, sur, pour, ce, en*, etc. This list of items is not fixed in time as coalescence can occur among the members of a construction and lead to new grammatical items, such as *dans* which emerged out of the coalescence of *de* and *intus* (Fagard and Combettes, 2013), and can now serve for the dragging up of the lexical items into partly functional constructions: *dans la mesure où* ('insofar as'), *dans l'idée de* ('with the idea of'), possibly and recently *dans quoi* ('where'), etc. In English, those items would be *of, to, in, out, from, for, by, so, as, all, on*, etc.

Third, it might not be always easily to distinguish such processes from instances of speciation. It has been argued that some placeholders have the property to attract lexical items to grammaticalize them (Bisang, 1998), especially in the formation of complex prepositions such as [in N of] (Hoffmann, 2004), giving rise to *in view of, in terms of, in place of, in regard of*, etc. Such colligations can indeed lead to the grammaticalization of some of the constructs if a further step of individuation is involved (as have most clearly happened with *instead of*, for instance). This illustrates only the productivity of the [in {N} of] complex preposition construction, and most of its constructs cannot be properly considered as resulting from a grammaticalization. Some of our previous examples are ambiguous as well in this regard; *dans la mesure où* can be seen as a construct from the standard use of the [dans {N}] prepositional construction. Yet, its structure ([dans la mesure où {Proposition}]) is distinct from the former and is not right away inherited. An alternate view on this matter could be to distinguish templates such as [in N of] from constructions, templates being devoid of any meaning, but serving as productive patterns to generate independent constructions, these constructions belonging then to the same formal paradigm. We, however, do not favor this latter view.

Fourth, the proposal that grammatical items have to haul lexical items to the grammatical plane is a bit at odds with Construction Grammar, as it states that those little grammatical items, which have no existence out of the constructions they figure in (they are not atomic substantive constructions), still remain somehow inde-

pendent. This suggests that even entrenched constructions, whose meaning is clearly non-compositional, still remain analyzable by language users in terms of constitutive items. This would imply that the construction does not contain all language knowledge of the speakers, as they still perceive the unity and independence of certain items, which, contrary to words, do not stand as constructions of their own. This also has consequences for the understanding of grammaticalization: a lexical item can be said to grammaticalize, even if the grammatical meaning is only carried by the construction it belongs to, as the lexical item keeps some sort of individual existence even within the construction (e.g. *mesure* in *dans la mesure où*).

A grammaticalization case at odds with the transfer hypothesis

There might be, of course, exceptions to this process, that is, lexical items which grammaticalize without the help of grammatical ones, or grammatical constructionalizations which do not involve any grammatical item. This would be the case of English modals, for instance. Other instances of grammaticalization do not go well with this picture either, such as the emergence of modal particles (English *I think* (Aijmer, 1997), French *j'imagine*). Also, *beaucoup*, in French, stemmed from [*beau + coup*] which are two lexical items. However, exactly as English *a lot of*, it came from a [*beau, grant coup de*] construction (Marchello-Nizia, 2006), with *de* as a grammatical item. This is clear in such a sentence:

la poïssiez veoir a l'assembler meint biau coup de lance et meint bon chevalier a la terre verser, et meint bon cheval corre tout estraié parmi le champ, qu'il n'estoit qui les retenist[.]

you could see there altogether many a fierce blow of the spear and many a good knight bite the dust, and many a good horse running in wander across the field, as there was nothing to hold them back.

La mort le roi Artu, c.1230, p. 232 (Frantext)

Here, the former marker of quantity, *meint*, is used, and *biau coup de lance* does not mean 'many spears', but 'a fierce blow of the spear'. However, *beaucoup* also exists as an intensifier in uses without *de*, such as:

Nos engins getoient au leur et les leurs aus nostres, mes onques n'oy dire que les nostres feissent biau cop.

Our machines were shooting at theirs and theirs at ours, yet I never heard that ours did much. (Marchello-Nizia, 2006, p.145)

[Q]ui ne va randonnant ja ne ferra biau coup

He who dashes not will never do much.

Les enfances de Doon de Mayence, c.1250, f. 21v (Frantext)

Assuming that *beaucoup* as an adverbial intensifier and *beaucoup de* as the expression of a large quantity have grammaticalized separately, and can still be considered as two separate constructions, it might be that *beau*, at this time, had a grammatical status sufficient to operate the grammaticalization of *coup*. It is true that *beau* is used as an adverbial, at least as soon as the fifteenth century:

[E]t si a beau crier, il n' est ame de nulz sens qui le puist oyr

And though he may scream, there is no sensing soul that can hear him.

Les Cent Nouvelles Nouvelles, 1456, p. 186 (Frantext)

and even earlier occurrences might be interpreted along this line as well:

[B]iau m'est que or t'en voy souprise.

It marvels me that I see you surprised nevertheless.

Le roman d'Eneas, c.1160, p. 518 (Frantext)

However, *grant* can be used with *coup* in intensification contexts as well (Marchello-Nizia, 2000). Furthermore, it would not solve the issue that we are concerned with, for it would remain to explain how *beau* has grammaticalized. The same can be said of *very* in English, which comes from *verray* meaning 'true' or 'truly'. It should be stressed that, at the time of these transformations, the available data is limited, so that we can easily miss the crucial steps of the grammaticalization. This is also true for semantic expansions: *tromper*, for example, which means 'to trick', or 'to cheat' and has been used as such, is attested as early as the fourteenth century, but its origin is unclear. The assumed path from a former attested meaning of *tromper*, 'to play the trump', is not fully convincing and lexicographers have not fully elucidated this puzzling etymology (Rey et al., 2010). In the case of *beaucoup*, one can also consider that the path of change which led to *beaucoup* as an intensifier includes a former step *faire beau coup* (literally 'to perform a great strike', with the meaning of 'do much'), but it would still include lexical elements only.

I will nonetheless favor this last proposal. Indeed, the adverbial slot seems to be a grammatical attractor; hence, lexical items can grammaticalize as they come to be attracted by this atomic, schematic construction. This would especially be the case for intensifiers. A reanalysis *verb complement* > *adverb* is often observed, attesting the attraction of the schematic adverbial construction. An example of this can be found in the much more recent emergence of another intensifier in French, *masse* (which actually exists in many variants: *des masses*, *des masses de*, *masse de*, *masses de*, etc.). Just like *beaucoup*, it can be used as a quantifier in the [des {P} de] construction (*des masses de*, *des tas de*, *des monceaux de*, *des tonnes de*, etc.; see chapter 5 for a broader study of this paradigm), where it gets a loosely grammatical status, as in:

Dans ma vie, je n'en avais pas rencontré des masses de type capables de faire ça.

In my life, I hadn't met many people able to do such a thing.

TORRÈS Tereska, *Une Française libre : journal 1939-1945*, 2000, p. 161
(Frantext)

It can also be used as an intensifier, without the final *de*, distinguishing it from other members of the paradigm, as there is no quantity implied, only a degree of quality:

Ça n'a pas l'air de tourner des masses, ton business...

It doesn't look to work that much, your business...

FÉREY Caryl, *Mapuche*, 2012, p. 226 (Frantext)

This use of *des masses* may have arisen out of the reanalysis outlined above. Indeed, in utterances such as the following:

Au moment où il vendit sa maison de commerce au fils Beauvisage, il possédait une forte partie de cotons achetés en pleine hausse, tandis que de Lisbonne, on en introduisait des masses dans l'Empire à six sous le kilogramme [...].

When he sold his trading house to the Beauvisage son, he owned a major part of the cotton bought when the prices were at their peak, whereas from Lisbon, lots of it were introduced into the Empire at six guineas per kilogram.

BALZAC Honoré de, *Les Paysans*, 1850, p. 187 (Frantext)

In this occurrence, *des masses* can be read compositionally as a complement, or can be reanalysed as an intensifier, reading 'it was massively introduced' instead of 'lots of it were introduced'. Regarding the contemporary use of *masse* alone as a quantifier, as in:

En même temps cette année j'ai séché 300h de cours (j'ai eu masse de problèmes), je trouve que je m'en suis pas mal sorti

In a way, this year I missed three hundred class hours (I had a lot of problems), I think it didn't went so bad. [<http://www.jeuxvideo.com>, 2011]

it is probably due to an attrition of substance of *des masses de*. Other examples of this adverbial attraction can be found and discussed. In vernacular French, *grave* has such a use:

Un pote me saoule grave

A friend bothers me really much [<http://www.jeuxvideo.com>, 2017]

This use can have resulted from a reanalysis, for instance in a sentence such as:

l'un d'eux mourra d'ici peu, deux autres sont en danger grave.

One of them should die soon, two others are in serious danger / definitely in danger. ARTIÈRES Philippe, *Vie et mort de Paul Gény*, 2013, p. 71 (Frantext)

Other lexical items used adverbially, such as *limite* in the meaning of 'almost', may have emerged out of the same adverbial attraction mechanism, or by attrition of the substance of a more complex construction (as might be the case for the now pervasively used *genre*, which plays the same pragmatic role as the English *like*). For all those examples, a complete and thorough diachronic investigation would be necessary to settle the matter. The point is, it might be possible to posit the existence of a near-atomic, schematic adverbial construction which could haul up lexical items to the grammatical plane. Of course, this mechanism, as it relies on a construction, which is a living linguistic object subject to change and evanescence, is specific of a given historical period of a given language. The grammaticalizing mechanisms, as they rely on the combination of a lexical item with special grammatical items, are thus both language-specific and period-specific.

2.3.3 The scope of the speciation mechanism

There would be no transfer in such a case, for a lexical item would just colligate with an entirely schematic construction. This raises an important question regarding speciation. Are the formation of *can*, of *beaucoup*, due to their colligation with the fully schematic 'modal' and 'adverbial' constructions, or do they involve a step of constructionalization? In the former case, the lexical item becomes grammatical in

the sense that it colligates with a grammatical construction; in the latter, it is not the item which is more grammatical, but the new construction itself (hence, Noël's proposal that grammaticalization is only a subsequent semantic expansion of a new construction). Before settling down this matter, let us discuss further the first scenario, according to which a lexical item can grammaticalize through its colligation in a purely schematic construction.

This would entail that colligation itself is a way to grammaticalize an item. Indeed, if we describe, say, the grammaticalization of *des masses* as a speciation scenario, it would process along the following steps: construct of the transitive construction [V des masses] R> construct of the adverbial construction [U des masses] S> independent construction [U des masses], where 'U' stands for 'utterance', 'R>' for 'reanalysis' and 'S>' for speciation. The increase of awareness typical of grammaticality is achieved during the first reanalysis step, as the adverbial construction allows to specify the orientation of the whole utterance. In this case, the grammatical status of the independent *des masses* construction would have been infused into the form by its temporary status of a construct of the much grammatical schematic adverbial construction.

An interesting, related question would be to discuss whether the speciation step is instantaneous or not, and how much it really relies on frequency to happen, frequency being the trigger of this independence according to Goldberg (2006). Actually, this line of questioning is akin to the difference between a paradigm of analogically related constructions and a construction with an empty slot to be filled by a proper member of a given paradigm (e.g. [N preposition N] vs [N preposition N], denoting respectively the fact that the paradigm is external and internal with respect to the construction). Diachronically, in the first case, constructions are created directly by analogy with other members of the external paradigm (so that the speciation is instantaneous and only the reanalysis step occurs), while in the second case, constructs are produced by analogy with other members of the internal paradigm, and only then, through speciation, they can become independent constructions (giving rise, most probably, to an external paradigm).

If we accept that, for adverbials, the first scenario describes the situation better, then they are all cases of spontaneous constructionalizations based on a lexical item, towards a more grammatical meaning. On the contrary, if the second case is to prevail, then it would mean that a construction can serve as a template to smuggle lexical items into the grammatical plane. We would have thus to understand which mechanism allows the construct to become an independent construction, and what is the timescale of such a speciation. In chapter 5, we shall see that such constructs have a frequency evolution which can be independent of the frequency of the construction as a whole, hinting that they may become independent before showing any significant increase of frequency (assuming separate evolutions indicate that two items are treated as different in the minds of speakers).

In such a case, postulating a 'construct' step is still possible, but then the second, speciation step, would be near-instantaneous; therefore, we are left with the first 'speciation' scenario no matter what. This raises two problems; first, the 'adverbial' construction would only serve, in this case, to produce new constructions on a given, specified pattern, a claim which is incompatible with the semiotic character of a construction in Construction Grammar; second, this theoretical proposal would be

gratuitous, and there would be no possibility to test it empirically, as the middle step (lexical item recruited in the internal paradigm of the adverbial construction in a given construct) would not be extended in time, and would not leave any trace.

A last difficulty emerges. Many words colligate with schematic constructions, and they do not become more grammatical as such. Lexical items colligating with the genitive $[\{N\} \text{ de } \{N\}]$ are in no way grammatical (e.g. ‘pou’ and ‘poule’ in ‘le pou de la poule’). More crucially, almost all lexical items colligate with argument structure constructions, and that does not make them grammatical. What makes the items more grammatical, in our example, is the reanalysis step; and reanalysis is not a prototypical feature of the colligation + speciation mechanism of grammaticalization.

The role of reanalysis

What allowed *beaucoup* and similar adverbials to achieve grammaticality is the following reanalytic mapping:

$$[S V O]_{\text{transitive}} > [[S V]_{\text{intransitive}} A]_{\text{adverbial}}$$

An individual argument of the Object paradigm of the transitive construction is reanalyzed as an independent adverbial construction applied on the whole remaining of the utterance.

In that case, the reanalysis is made possible by, on the one hand, the bleached semantics of *coup* (Marchello-Nizia, 2006; Diewald, 2006), on the other hand, the solid entrenchment of the $[[S V] \text{ adverbial}]$ pattern in xvth century French, which might have favored the expectation of an adverb consecutively to a verb. The blurry semantics of verbs likely to have co-occurred with *beaucoup*, such as *faire*, might also have favored such a reanalysis (due to the colligation of *faire* with the intransitive construction). In this case, the grammatical awakening would be achieved through the parsing of a semantically ambiguous item in the light of the common linguistic knowledge. The conclusion of which is that grammaticality, in this case, would be inherited from a specific cognitive processing of the item.

Thus, it seems safer to assume that the ‘transfer’ mechanism of grammaticalization might not be the only one able to grammaticalize a lexical item through a constructionalization process. Another mechanism, which would be a simple constructional reanalysis, e.g. from part of a construct in a transitive construction to an adverbial independent construction, seems to be able to grammaticalize a lexical item without requiring any transfer of grammatical status from a grammatical item. This opens up the question of the unity of the grammaticalization phenomenon: can there be different ways to achieve grammaticalizations? If so, can we map the mechanisms to the outcomes? Indeed, it would seem that this second mechanism, which is an analogically motivated reanalysis of a chunk of a construct into a construction of its own, is especially encountered in the rise of intensifiers, and some discourse markers.

2.3.4 A summary

In any case, and this will serve as a conclusion, it seems both productive and relevant to assume that grammaticalization is a specific phenomenon, and that the changes leading a lexical item to take on functional uses are not as trivial as they may seem with regard

to usual constructional changes and constructionalizations. It might indeed require an additional explanation, a stronger mechanism.

More specifically, I tried to argue that it might require some pulling force from the grammatical side, either exerted by strongly grammatical items within the context of a specific construction, or by a paradigm of existing constructions serving as a productive template for new constructions to emerge through reanalysis. In short, grammaticalization needs grammar in order to occur.

If, however, we deny all relevance to a typology of constructions, then grammaticalization has no place in a Construction Grammar view of language change; but if we consider that there are differences of nature among constructions, then the phenomenon through which those separate categories of constructions diachronically interact with each other is worth being considered as such, and may give new insights to the mechanisms by which constructions evolve with time in the course of language use.

Furthermore, grammaticalization might involve different mechanisms, each one possibly leading to a different part of grammar, or a different kind of function. True, in such a view, grammaticalization would cease to be a unique, precise phenomenon, but its specificity would lie elsewhere, that is, in the mapping between specific mechanisms, and specific changes, or equivalently, specific families of functions. Since not all languages share the same kind of constructional organization, it would thus be expected that, in the details, grammaticalization would not rely on the exact same mechanisms.

Grammaticalization would then be much more than a phenomenon; it would entail that the description of a language grammar (i.e. the typology of its constructions) can and should be understood through the detailed study of its diachronic genesis (Givón, 2015). Grammaticalization would then be the constant feedback between the typological organization of a language, and its constant, continuous renewal.

To sum up, I would propose such a hypothetical picture of grammaticalization:

- There exists a strong dichotomy between lexical and grammatical constructions. Of course, this crude typology might and most probably must be refined much further; but there must exist some sort of a severance between the different categories.
- Grammatical constructions are specifically characterized in that they manifest, through their meaning, a linguistic awareness of either their being part of an utterance or their own internal structure. This internal-external dimension unfolds as one moves towards the grammatical pole. Syntax and pragmatics lie at the two extremes of this new cline.
- Given the lexical and grammatical categories are separate, specific mechanisms are required to pass from one to another. Different grammatical categories may require different mechanisms (e.g. compounding to create adjectives; formation of complex constructions involving a grammatical item leading to prepositions).
- Most (primary) grammaticalizations are semantic expansions of a lexical item accompanied by a change in hosting construction. They can therefore be described as constructionalization.

- Further secondary grammaticalizations can occur through mere semantic expansion of a given grammatical construction (e.g. *still*, from temporal marker to concessive marker), with the possible involvement of new constructionalizations (e.g. French *côté*, from spatial preposition to topicalization marker). The latter might be the mark of a categorial change.
- Speciation of an individual member of a schematic construction can also lead to the grammaticalization of this member. In that case, the new construction can inherit the grammatical status of the construction it stems from. Such grammaticalizations are bound to move away from the ‘internal’ pole of the awareness cline.
- Since none of these processes is a specific mechanism, it follows that the specificity of grammaticalization should lie in the special semantic link between the source and target meanings. Specific constructionalization processes are therefore required to build up such a cognitive channel between the source and the target.
- Two frequent mechanisms of grammatical constructionalization in the standard European languages are the transfer mechanism (grammatical status is transferred from a grammatical item within the context of a newly reanalyzed construction) or category-based reanalysis (a lexical construction is reanalyzed, in specific contexts, as a member of a more grammatical category, e.g. from noun to adverb).
- All grammaticalizations, either primary or secondary, gradually cause a renewal of the categories: the typology is not fixed in time, but altered by the successive grammaticalizations.
- Different languages, being characterized by different typologies structuring the constructicon, may rely on different mechanisms of grammaticalization.
- Empirically, it follows that a (primary) grammaticalization must at the very least be characterized by the phenomenological signature of a semantic expansion and that of a constructionalization.

The next part will deal with the determination of such empirical criteria.

Part II

Corpus study of grammaticalization

Chapter 3

The S-curve: a signature of language change

Grammaticalization is, first of all, a language change. If we are to look for an empirical manifestation of the specificity of the grammaticalization phenomenon, then it is mandatory to first assess the empirical characteristics of language change in general, in order to position grammaticalization within this broader picture. The empirical characterization of change will thus be our primary concern in the two next chapters to come.

One of the most well known and most widely accepted regularity in language change, is the fact that a language change follows an S-curve over time. This claim, however, is an understatement of some sort, for it doesn't say what exactly is changing. We shall see that there is no clear consensus on this matter, and that different answers have been provided to this question.

3.1 First intuitions

The S-curve has long been used to describe social phenomena. One of the first example to be found is the work of Lehfeldt (1916) in which, drawing on several datasets, it is argued that the amount of British trade evolved with time according to an S-curve, mathematically identified as a cumulative normal distribution. This behavior is assumed to be generic of any kind of cultural change and caused by the work of some “social forces”:

[P]rogress must consist in a transition from one stationary state to another, and this transition may be expected to show some uniform characteristics. It must be produced by the uprising of a group of social forces, modifying a condition in which they have been long latent: time is needed for the forces to develop their strength and spread their effects through the community. There is a period of acceleration in the movement: then one of steady change: then a period in which these social forces lose their efficacy, or die away, leaving a new stationary condition of society.
(Lehfeldt, 1916)

Pearl and Reed (1920) try to show that the population growth in the United States of America should follow an S-curve. Observing that the population has grown expo-

nentially since 1790, they state that it is unrealistic to assume that this growth can continue much further, and propose to fit the population curve with an S-curve, bound to saturate at some point. They conclude that the concavity of the growth should have switched around 1914, so that the growth would slow down, and the population size reach an asymptotic size around 197M of inhabitants. They next come to discuss the difficulties of such a population size in terms of available resources, summarizing in the following way their growing concerns :

Altogether, we believe it will be the part of wisdom for anyone disposed to criticise out asymptotic value of a hundred and ninety-seven and a quarter millions because it is thought too small, to look further into all the relevant facts.

(Pearl and Reed, 1920, p.286)

To mathematically describe the S-curve, they make use of the logistic function, noting with some interest that this function is solution of the differential equation describing autocatalytic reactions in Chemistry. These ideas are further applied by Chapin (1928) in a book entitled *Cultural change*, which shows that cultural quantities grow as an S-shaped curve, as do the number of cities adopting manager plan, or the institution size in various cities. Interestingly, he distinguishes three phases of the S-curve, (a slow start, a fast subsequent growth, and a slow end), providing for each a suitable explanation, reminiscent of the one sketched by Leffeldt (1916):

First, the period of slow growth in the growth curve corresponds to the first phase of the societal reaction pattern at which mores are enforced and few innovations (additions) allowed or attempted. The middle period of the growth curve, when rapid increases in the complexity of the structure are observed, corresponds to the second phase of the societal reaction pattern in which special legislation is the order of the day. It is in this period of the institutional cycle that growth is most rapid because the law-making body is experimenting with all sorts of expedients. Finally the third period of diminished growth corresponds to the third phase of the societal reaction pattern in which general laws are passed and the structure consolidated and simplified. At this time the varied experiments are sifted and the real contributions are integrated into a new pattern of institutional structure.

(Chapin, 1928, p.384)

This is most different from the previous work, where the process was seen as a whole and explained as such. We shall see that this three-part division, and the necessity to explain each part separately, will be most persistent in Linguistics works.

In a very nice paper, Pemberton (1936) provides three very striking examples of S-curves, respectively for the number of countries adopting the stamp postage system, the number of states adopting a tax limitation law, and finally the number of states in which school is compulsory (studying separately Northern/Western and Southern states). He claims thus:

The series of cases presented here offers evidence that within any given culture area the diffusion of a culture trait tends to occur at a rate which may be described by the cumulative curve of a normal frequency distribution. [...] distribution. The curve of diffusion is simply the cumulative expression of this symmetrical binomial distribution. [...]

(Pemberton, 1936)

This work is interesting for at least three reasons; it provides convincing illustrative cases displaying an S-curve; it tries to explain the S-curve as a whole, like (Pearl and Reed, 1920); it states very clearly that any specific mathematical function could fit the data well, but that the choice should be made according to explanatory criteria :

Goodness of fit, however, was not regarded as the fundamental criterion for deciding whether one of these three curves or some other would be the most adequate for describing the typical curve of culture diffusion. It was assumed to be more important that the mathematical equation used should be constructed on the basis of the specific theory of culture interaction which could be considered the most acceptable explanation of the curve of culture diffusion. That is to say, the a priori conditions of the mathematical equation should have their theoretical counterpart in the conditions of culture interaction producing each sequence of diffusion.

(Pemberton, 1936)

Pemberton goes on with explaining why he favors the cumulative curve of a normal distribution, saying that the time of adoption of the new trait by a “population unit” (individual, state, country) is determined “by the interplay of an infinitely large number of elements in the social milieu”, so that it is expected to follow a normal distribution. The main weakness of this explanation is that we don’t understand why the individuals would be bound to adopt the new trait eventually. Interestingly, in this view, first adopters are not precursors imitated by others, they just happen to be early compared to some predetermined course of change. Roughly speaking, Pemberton explains the S-curve, but leaves the very explanation of the change entirely left over.

The idea of an S-curve governing social diffusion was thus well established, but it took two more decades to find its way into Linguistics. Let us mention an unexpected reference to the S-curve by Reid (1944), stating that the expected Zipf straight line is not be found for French literature: he finds an S-curve instead. However surprising this reference to the S-curve may be, it is clearly out of place in this review.

The first mention of the S-curve in this field seems thus to go back to 1954, within the *Psycholinguistics* book by Osgood and Sebeok. Addressing the question of how change takes place within the speech community, they note that an S-curve is the most likely way for a change to proceed:

Social change. Language change in a community will be gradual and cumulative, representing a continuous changing proportion of individuals who do or do not hear and produce a particular feature or set of features. The process of change in the community would most probably be represented by an S-curve. The rate of change would probably be slow at first, appearing in the speech of innovators, or more likely young children; become relatively rapid as these young people become the agents of differential reinforcement; and taper off as fewer and fewer older and more marginal individuals remain to continue the old forms.

(Osgood and Sebeok, 1954, p.155)

This claim is grounded on nothing but likeliness, as shown by the repeated use of “probably”; the authors do not provide any actual data to assess it -contrary to all previous works we surveyed. However, they suggest an interesting, innovative explanation, whose influence will last up to the twenty-first century: that the S-curve is not so much due to societal diffusion, but to the passing of generations. As Lehfelddt

(1916) and Chapin (1928), they explain the three phases of the S-curve separately. First, innovations are produced in a situation of “stress”, so that “errors” are likely to occur, such a situation of stress being language learning. In language learning, most of these errors will be censored by the “agents of differential reinforcement” (the adults at that time) but some changes may nonetheless escape this censorship (e.g. if they are not detrimental to the understanding of the utterance). Then, the children will grow up, so that they would become enforcers of their new trait: the change will be able to spread, which explains the “fast” phase. However, change will not affect elder people, while it is adopted by the newcomers. Only with the slow departure of the elder will the change achieved dominance: this is the slow final part of the curve. This kind of explanation brings forward the problem of simultaneous, similar innovations ; Osgood and Sebeok partly address the question by saying that change is constrained, especially by the balance between reduction and efficiency already discussed by Zipf.

In this work, the authors also propose interesting hypotheses on how change is adopted by individuals, proposing three possibilities : 1 - change is gradual within the individuals ; 2 - change is abrupt within the individuals; 3 - change does not occur within the individuals: they learn the language in a new way at each generation, so that individual changes, while theirs idiolects do not — the latter one being bound to occupy a dominant position for the remaining of the century:

The nature of language change within the individual is a difficult question — some linguists feel that this is an all-or-nothing matter akin to mutation, whereas most psychologists feel that there should be a period, at least, of oscillation between competing forms. Perhaps, in a manner akin to imprinting in birds, individuals never change in the features they hear and produce after early childhood experiences, language changes being purely a matter of sociological shift in the composition of the group. (Osgood and Sebeok, 1954, p.155)

This work by Osgood and Sebeok is important for at least two reasons: It is the first work bringing the S-curve pattern of community diffusion into Linguistics, and it introduces the idea that this pattern may be related to the age stratification of the community. These two ideas will soon bear their fruits in the Sociolinguistic framework.

3.2 S-curve in the sociolinguistic framework

The idea that the S-curve is relevant to describe language change is brought further alight by the important work of Weinreich, Labov, and Herzog, *Empirical foundations of language change*, published in 1968. In this work, the mention of an S-curve appears quite incidentally, in the middle of a lengthy, careful discussion of Hermann Paul’s theory of change (which he himself developed in his *Prinzipien der Sprachgeschichte*, published in 1880).

3.2.1 The seminal paper by Weinreich, Labov, and Herzog

In Paul’s theory, changes in the idiolect of an individual arise both from a desire to conform to the speech community, and spontaneously from within. In the latter case, Weinreich, Labov, and Herzog note that if the changes were random, individual errors,

they would scatter and never produce a shift in language of the community. Chance in general, unlike Pemberton's view, is not seen as a suitable explanation:

[I]f the beginnings of changes were random processes, occasional losses of balance would alternate with restorations of balance, and beginnings of infinitesimal change would alternate with cessations of infinitesimal change. Thus, chance is here invoked illegitimately, since we are out to explain a specific, not a random process. (Weinreich et al., 1968, p.112)

There should thus be some favored direction of change; new variants must have some sort of an intrinsic advantage to be produced repeatedly, at least according to Paul. The question of change becomes thus the question of *actuation*: if the change is bound to happen because it fits speech better, why does it happen at a certain time, within certain idiolects? Why then, and not at any other time?

The authors then explain that Paul has in mind "what we might call the 'avalanche mechanism' " (Weinreich et al., 1968, p.112), as the following quote from Paul's work indeed tend to prove:

Once a definitive shift in the kinesthesia [or any other idiolect feature] has taken place through the elimination of the inhibitions exercised by communication [i.e., speakers' desire to conform to their interlocutors' idiolects], a further small shift is made possible by the continuing effect of the tendency. Meanwhile, however, a whole minority is swept by the movement. The very factors which prevent the minority from getting too far ahead of the general custom also prevent it from remaining significantly behind the progress of the majority ... The movement proceeds in such small distances that a salient opposition never arises among individuals standing in close intercourse with each other. (p.62)

quoted in Weinreich et al. (1968, p.113)

From this, Weinreich, Labov, and Herzog understand, having Osgood and Sebeok (1954) in mind, "that the progress of a language change through a community follows a lawful course, an S-curve from minority to majority to totality" (p.113). They even pursue the very same line of thought that the two previous authors unwinded, briefly evoking how and on which conditions the "S-curved social trajectory of a change [...] can be correlated with the universal differentiation of speech communities by age." (p.114) Yet, no matter how enriched by Osgood and Sebeok (1954)'s ideas, this reference to the S-curve is but the summary of Paul's description of the spreading process. Though this S-curve proposal has been largely referenced in subsequent literature, it is only the rendering of Paul's own theoretical insights, not the claim of the authors themselves. The latter even go on to conclude their presentation of Paul's *Prinzipien* by pointing to a major inconsistency regarding this "avalanche" mechanism leading to the S-curve:

An idiolect or dialect may also change by "borrowing" forms from other idiolects or dialects. Such borrowing is selective, but no explanation is offered for particular selections. Opportunity to borrow from other idiolect' depends on exposure to them: however, both borrowing and nonborrowing are attributed to conformity — either with the innovators or the conservers. (Weinreich et al., 1968, p.119)

That is, they criticize the explanation which supports the S-curve on the ground that it relies to a single mechanism — conformity — which acts in opposite, conflicting ways along the process: there is a resistance to change because speakers to conform to the conservers, and there is a speeding up of the process because they conform to the innovators. As such, the authors point out that the ‘avalanche’ picture lacks some crucial elements to satisfyingly explain the process (e.g. something akin to what is referred to by the authors as an ‘evaluation ‘mechanism’, as presented below).

The interest of this work thus does not lie in its theoretical contribution in the understanding of the S-curve. In the list of seven main stylized fact that a theory of language change should account for (pp. 187-188), no mention is made of the S-curve, for the authors are actually more interested in the geographical and sociological diffusion of a change, rather than its diffusion over time — which is probably due to an implicit assumption that the S-shaped time diffusion of a language change only mirrors the time course of this sociological diffusion.

However, they provide a list of problems which a theory of language change needs to solve: 1- the Constraints Problem, which copes with the constraints posed on language variation, and consists in finding which changes can happen and which cannot; 2- the Actuation Problem: why the change happens at some arbitrary time; 3- the Transmission Problem: how an individual can adopt in his own idiolect a variant he is in contact with. This process is seen as gradual:

Change takes place (1) as a speaker learns an alternate form, (2) during the time that the two forms exist in contact within his competences, and (3) when one of the forms becomes obsolete. (Weinreich et al., 1968, p.184)

4- the Evaluation Problem: Sociolinguistic description of language change puts the emphasis on how the different communities a speaker belongs to shape his knowledge of language and influence the different variants he will use in the array of various social circumstances he is accustomed to. The Evaluation Problem thus deals with the building up of the speaker’s knowledge of this heterogeneity, i.e. the knowledge of how he should select appropriately the different variants at his disposal to deal with these different situations. 5- the Embedding Problem: how the change spreads after its Actuation, through Transmission from speaker from speaker. Most interestingly, this problem is split into two subproblems: embedding in the social structure, and embedding in linguistic structure. It means that diffusion takes place on the sociological plane as well as on the linguistic one, synchronously. There is thus mutual influence between the two processes of diffusion.

At this point, the S-curve of language becomes suddenly more complicated. It is not only a diffusion among what Pemberton (1936) called ‘population units’, but also a diffusion within language, where no ‘population units’ are easily defined. One might furthermore state that the diffusion can also occur within the idiolects, for the depiction of the Transmission Process can also very well be described by an S-curve. Obviously, this entangling of planes and levels complicates the matter tremendously.

Thus, even if this paper does not discuss the S-curve thoroughly, it lays down important concerns and questions which are of immediate relevance for the understanding of the S-curve. This uttermost importance was actually quickly recognized, as is shown by this quote from (Bailey, 1971), which reviews the history of linguistic paradigms:

The paper by Weinreich, Labov, and Herzog (1968) remains one of the fundamental theoretical writings of the new paradigm. (Bailey, 1971, p.320)

However, its direct influence on the S-curve question was only mild, or so it would seem. It will require new theoretical considerations for the importance of the S-curve to be fully highlighted.

3.2.2 Lexical diffusion

Alongside this Sociolinguistics framework, another view of language change and of the S-curve progressively emerged, putting the emphasis on what is called “lexical diffusion”, thus putting the emphasis anew on the mechanisms tied with the emergence of the S-curve. The term itself seems to have been coined by Wang (1969), in the context of phonological change. The notion of “lexical diffusion” is summarized in the following way:

Phonological change may be implemented in a manner that is phonetically abrupt but lexically gradual. As the change diffuses across the lexicon, it may not reach all the morphemes to which it is applicable. If there is another change competing for part of the lexicon, residue may result. (Wang, 1969, p.9)

This view goes against the Neogrammarian view of phonetic change, according to which a sound change should affect all relevant items straightforwardly. Though this paper has been published shortly after the paper by Weinreich et al. (1968), it acknowledges fully the Sociolinguistics ideas expressed therein, but goes slightly further as it considers that the diffusion is bi-dimensional: there is of course a societal diffusion, across speakers, but there is also a diffusion across words, within language itself, the so-called “lexical diffusion”:

Ideally, before the change, all speakers will use sound X in all relevant morphemes; after the change, all speakers will use sound Y in the same set of morphemes. The dimension of time may be studied in each of three relatively independent parameters: (1) phonetic, i.e. from sound X to sound Y; (2) lexical, i.e. from morpheme to morpheme in the relevant part of an individual’s vocabulary; and (3) social, i.e. from speaker to speaker in the same dialect. (Wang, 1969, p.13)

However, the paper does not provide further details on this diffusion “from morpheme to morpheme”; also, it is never stated that this diffusion should obey an S-curve as well. Not a long time would be needed to elapse before the S-curve could surface again. In the next paper on lexical diffusion in 1970 (yet published in 1977), Wang and Cheng, providing data from Chinese, claim that an S-curve should conveniently describe the lexical diffusion as well. Interestingly, contrasting with Osgood and Sebeok (1954) view according to which the S-curve was somehow granted, because of its generic character in cultural diffusion, Wang and Cheng (1977) consider this shape to be motivated by empirical considerations:

An extreme interpretation of the notion of lexical diffusion may be represented by the graph in Figure 2. What the graph indicates is that the diffusion of the change across the relevant sector of the lexicon proceeds at a relatively uniform rate. The percentage of the affected lexicon, in other words, increases linearly with time (for the time being, we are not concerned with the exact slope of the

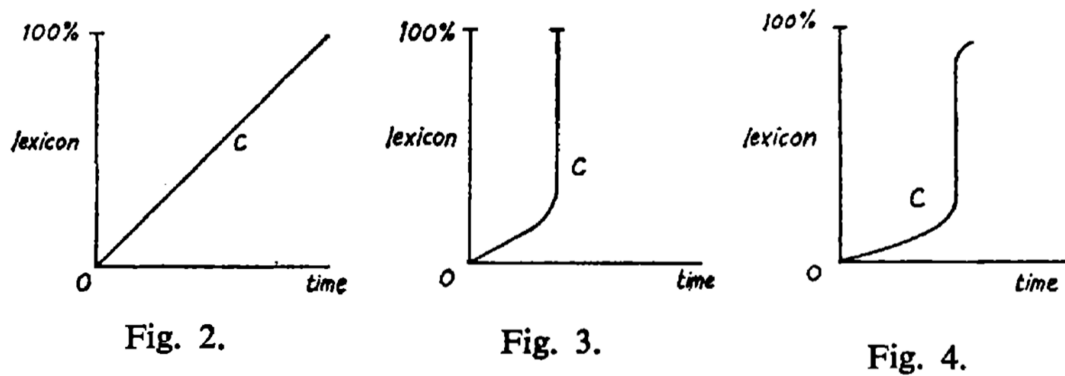


Figure 3.1: Figures 2, 3 and 4 of (Wang and Cheng, 1977).

change C). There are some empirical considerations which may be thought of as negative evidence against the picture of phonological change implied in Figure 2. If Figure 2 were true it would mean that whenever we are able to catch a change in progress we should have as good a chance of catching it around the middle of its time span as of toward either end of the time span. Yet our experience with the incomplete changes in the Chinese dialects has been that either a large majority of the words (say, over 80 percent) have changed or a large majority of the words have not changed. (Wang and Cheng, 1977, pp.151-152)

They nonetheless provide a now traditional step by step explanation for the curve:

There is also a theoretical argument against Figure 2 that goes something like this. When the change first enters the language as a minor rule the number of words it affects may be too small for the rule to serve as a basis for extrapolation. As the change gradually diffuses across the lexicon, however, there comes a point when the rule becomes "felt" and it is generalized to many other words. During this span of time, when the minor rule becomes a major rule, we would expect diffusion to be much more rapid. The graph that portrays such a conception of the chronology of change may be seen in Figure 3. (Wang and Cheng, 1977, p.152)

The "slowing" part of the S-curve follows in due time:

To go one step further, we might consider cases where the change tapers off before it completes its course, leaving a handful of words unaffected. Such a situation would appear as the graph in Figure 4, where a second point of inflection marks the time where the change tapers off. If these residual forms persist, then an unconditioned phonemic split would be the result.

(Wang and Cheng, 1977, p.153)

All figures mentioned in these quotes have been reported on Fig. 3.1.

Apart from the fact that this is the first instance of a theoretical proposal to describe lexical diffusion as an S-curve (even if it is not named as such), several interesting details are worth to be highlighted. First of all, the three different mechanisms responsible for the three different parts of the curve (initiation, acceleration, and inflexion) are not viewed as necessary. According to the latter quote, it is plain that the "tapering off" of the curve is not bound to happen every time. However, the three

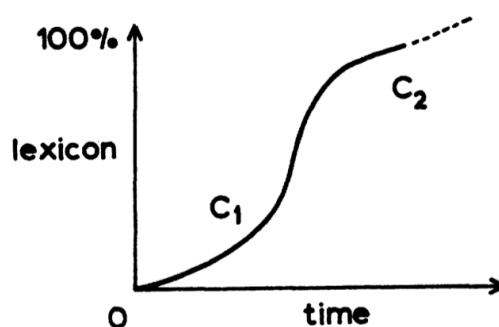


Fig. 1.

Figure 3.2: Figure 1 of (Chen, 1972).

mechanisms remain rather mysterious, apart from the second one, which is akin to analogical regularization.

The authors claim in their paper that data seems to be best explained by the S-curve. It should be stressed at this point that their data is not of a diachronic nature. The S-curve never appears as such, but only through several hints pointing towards its validity: the fact that ongoing changes seldom show a *statu quo* between two competitors; and the “residues” in phonetic change. One might argue that the existence of residue does not require an inflexion point to be explained, but a saturation mechanism; it could as well happen in an abrupt way. Also, an inflexion point does not prevent to reach completion eventually. Then, the S-curve is neither necessary nor sufficient to explain the empirical synchronic data proposed by the authors, and is but a theoretical extrapolation.

The same ideas are re-assessed in a much similar way by Chen (1972). Discussing the data from Wang and Cheng (1977), he writes:

The chronological profile of sound change we think the Shuang-feng case is suggesting in effect may be portrayed by an S-curve such as the one plotted in Figure 1. (Chen, 1972, p.475), figure reported on Fig.3.2

However and most surprisingly, the author, while considering another set of data, makes the exact inverse move than Wang and Cheng (1977) and conclude that a linear diffusion may be more appropriate after all:

[W]e suggested that the chronological profile of the lexical extension of a phonological change resembles an S-curve [...]. The increasing cases of lexical diffusion ‘caught in mid-stream’ seem to suggest that in some cases the progression of a sound change along the lexical dimension can be compared rather to a straight diagonal line [...]. (Chen, 1972, p.489), figure reported on Fig.3.2

Interestingly, the reason put forward in favor of the linear curve is consistent with the reason that served to justify the S-curve in the previous work. Furthermore, the evidence under discussion consists once more in the observation that few speakers are using the two variants in equal proportions of contexts, the S-curve being then regarded as the most plausible underlying dynamical process responsible for the observed regularities of this data.

3.2.3 Bailey's S-curve

In the wake of the foundational paper by Weinreich et al. (1968), one work in particular should be mentioned: *Variation and Linguistic Theory*, by Bailey (1973). This book happened to have become one of the most authoritative ones in the S-curve literature, though it touches the matter only incidently. As an illustration of this lasting influence, the values of 20% and 80% proposed by Bailey for the frequencies thresholds at which the curve 'picks up momentum' and then 'tails off' will be retained by a majority of the subsequent authors (e.g. (Chambers, 1992; Preston, 1996, p.17) and (Rickford, 2002, p.162), who mistakenly refer to them as 'inflection points'). Though later authors justify these values by a due quotation to Bailey (1973), he himself, however casual he may sound while identifying these values, had nonetheless some reason to do so, for the sigmoid function indeed takes its two extrema of curvature as it respectively reaches both 21% and 79% in ordinates.

This work presents the "wave-based model" of language change, putting the emphasis on societal diffusion as waves among the communities of speakers. In this context, he presents the S-curve as representative of this diffusion process (the so-called "Principle 17"):

A given change begins quite gradually; after reaching a certain point (say, twenty per cent), it picks up momentum and proceeds at a much faster rate; and finally tails off slowly before reaching completion. The result is an S-curve: the statistical differences among isolects in the middle relative times of the change will be greater than the statistical differences among the early and late isolects.

(Bailey, 1973, p.84)

In his terminology, the notion of isolect is akin to the one of dialect, but is more fine-grained. How exactly Bailey relates the S-curve to the wave model is not quite obvious. First of all, it must be noted that the S-curve does not represent change through time, but through relative time, relative being 'defined on minimal (isolectal) changes'. Second, it is not clear whether the S-curve diffusion describes the change within a given isolect, or among the ensemble of all isolects, and this needs further clarification.

1. According to Principle 18 (Bailey, 1973, p.85), the S-curve applies to all parts of the affected lexicon, separately, similarly, and at different times.
2. All individuals of the isolect are assumed to speak in a similar way (this defines the isolect), so that the percentages of change within an isolect can only mean that individuals use both the old and the new variant with different probabilities, those probabilities shifting in favor of the new variant with time according to the S-curve.
3. The change spreads from one isolect to another, faster than each isolect is changing gradually towards the new variant (geographical diffusion is faster than diffusion in use within an idiolect).

We can summarize these three remarks by postulating a diffusion function of the following kind, assuming that the change spreads according to a circular wave originating from some isolect geographically situated at $x = 0$, x denoting the spread radius, t

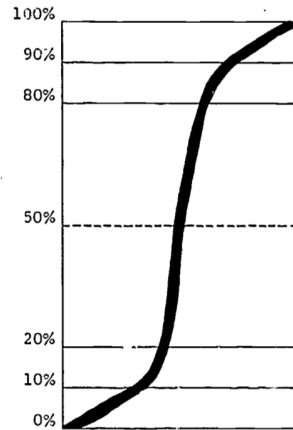


Fig. 6. The \int -curve generated by Principle 17.

Figure 3.3: Figure 6 of (Bailey, 1973).

the time, i a subpart of the lexicon (an ‘environment’ in Bailey’s words), and $\phi_i(x, t)$ the likeliness, for speakers of the x isolect, to use the new variant in this environment i at time t :

$$\phi_i(x, t) = \frac{1}{2} [1 + \tanh(\alpha(t - t_i) - \beta x)] . \quad (3.1)$$

However, as Bailey is considering time on the X-axis, the S-curve he displays (reported on figure 3.3) would represent the passing of the wave through a given isolect, say situated in $x = x_0$ (hence $\phi_i(x_0, t)$). The S-curve according to Bailey (1973) — i.e. the change in frequency of use across time for a given isolect and a given part of the lexicon — thus closely corresponds to the S-curve we will model in chapter 8.

It should be stressed that Bailey did not propose the S-curve because of some empirical evidence at his disposal, but as a theoretically convincing picture of the process of change. The data he provides (which results from previous works by other researchers) is only synchronic data, from which he deduces that a S-curve shaped wave diffusion is going on, a stance which is implicit in the ‘Principle 17’. In his work, time is only a conceptual reconstruction. As we shall see, this is characteristic of, but not entailed by, the Sociolinguistics framework.

3.2.4 S-curves based on synchronic evidence

The wave-based model of Bailey increased the legitimacy to seek the evidence for the S-curve in synchronic data. The first tenants of lexical diffusion were already using synchronic data to establish the validity of this pattern (by stating that the S-curve was hinted by the low probability to catch the change at mid-process, so that the latter was faster in the middle than at both the beginning and the end); however, Bailey’s insights allowed to go still further in this direction.

In the wave-based view of language change, synchronic evidence is not only relevant as it allows us to find traces of the dynamical process which is going on; it permits

to fully reconstruct the S-curve. Indeed, to find an S-curve, we can either look at one point in space (one isolect) through time, or look at one time, in different isolects. By surveying the language of speakers from different, adjacent isolects, it thus becomes possible to find an S-curve. Another difference from the lexical diffusion framework is that, in the latter, the only relevant quantity is the number of items in the lexicon affected by the change, while in Bailey's framework, the frequency of use of the new variant in a given lexical environment is already sufficient. That is, the 'percentages' do not indicate the same thing: in lexical diffusion, they are the percentage of affected words or lexical contexts, as for Bailey (1973), they refer to the frequency of use of a variant compared to its competitor, and make sense even in a single lexical context.

These two differences are fully exploited by Bickerton (1975). When he offers an idealized view of the S-curve, reported here on figure 3.4, time is no longer used on the X-axis: it has been replaced by individual speakers, presumably representative of different isolects, sorted by their adoption of the novelty. This choice is clarified, and theoretically supported, in the following quote:

In the classic S-curve, when the percentages of a variable are assessed for a number of speakers, [...] the line jointing the speakers' positions on the graph will resemble a flattened S. Such a curve may be taken as representing a change passing fairly rapidly through a population. The low percentages are those of speakers who are in the process of acquiring a new feature; the high percentages are those of speakers who have almost fully acquired this new feature and are in process of getting rid of its antecedent; these two groups account for the bulk of speakers involved in the change. The reason for relatively few speakers being found among the middle percentages is that, at a certain stage of development, grammatical reinterpretation takes place; some speakers have a grammar that rewrites a given feature 'a, with b as a possible variant', others have one that rewrites the same feature as 'b, with a as a possible variant', and while performance factors tend to blur the distinction between the two groups, few if any speakers are found to have the two variables in balance. (Bickerton, 1975, p.65)

It should be noted that, while the wave idea is overtly present, the explanation of the curve features is not wave-based, but rule-based, which seems somehow at odds with the idea of the 'passing through' of a wave. Yet Bailey (1973) himself was speaking of a 'rule' change, indicative that the change, within individuals, is paradoxically seen as simultaneously abrupt and continuous.

However, when this theoretical view comes to be supported by empirical findings, hereby reported on figure 3.5, it turns out that the obtained curve is not as S-shaped as one might have expected. Still, this is no reason to doubt the S-curve, and indeed Bickerton provides an elaborated explanation for this deviation, stating that there are actually two changes involved in the adoption of *doz*, so that 'the 100% mark for *doz* is therefore in reality the 50% mark in Fig. 3.2.' (Bickerton, 1975, p.65)

It is thus interesting to see that, at this point, the S-curve is assumed to be theoretically valid and well established, while no convincing empirical support has ever been provided in its favor. Fortunately, in another case, the empirical evidence finally leads to a nice S-curve (Fig. 3.6a). Yet, whereas one could be much satisfied with this beautiful pattern, Bickerton is unsatisfied by the fact that he finds speakers 'among the middle percentages', which he assumed to be theoretically impossible, for there is no rule 'a, and also b'. Thus, he strives to explain why the two speakers labeled 165 and

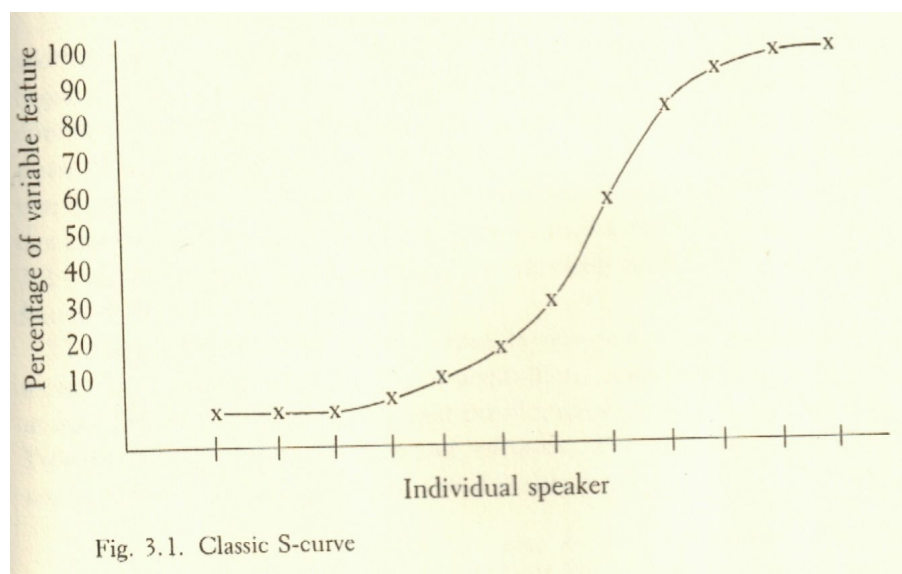


Figure 3.4: Figure 3.1 of (Bickerton, 1975, p.65).

196 show those unwanted percentages:

165 is not an individual, but a cover-term for participants in a bar-room conversation who could not be individually identified (and thus, though members of the same primary group, probably had differing outputs) and 196, as we have seen, was under the pull of contrary social influences. (Bickerton, 1975, p.77)

This justification of a feature which is entirely expected for an S-curve propagation, and yet unwanted according to the theoretical explanation of this very phenomenon, magnifies the existing hiatus between the wave-based, continuous perspective, and the rule-based, abrupt view on this matter.

A third example (Fig. 3.6b) contradicts the S-curve pattern once again and shows a more linear distribution of the use percentages over the speakers. Facing this result, Bickerton still accounts for it in terms of the S-curve, in the most surprising way, saying that 'the familiar S-curve is almost reversed'. By pointing out this 'reversal', Bickerton certainly means that we find more speakers in the middle percentages than at the low and high ones. Instead of concluding that the S-curve may not be fundamental in the spatial diffusion of a language change, Bickerton invokes then the fact that 'a change which is phonologically vague may be sharpened by social factors' so as to save the theoretical relevance of the S-curve.

Later, along the line of Bailey's proposal that the change follows a different S-curve in each lexical environment, Bickerton provides, for three separate 'grammatical environments' (Bickerton, 1975, p.154), both raw data from fourteen speakers (Fig. 3.7) and a theoretical abstract view of the process (Fig. 3.8). The three different curves are interpreted as three different changes:

[T]he data in Fig. 4.4 conform more closely to the S-curve pattern already shown to be characteristic of linguistic change. [...] A very slight idealization of Fig. 4.4 would give us the pattern of three consecutive S-curves (indicating three consecutive changes) shown in Fig. 4.5. (pp.155-156) (Bickerton, 1975, pp.155-156)

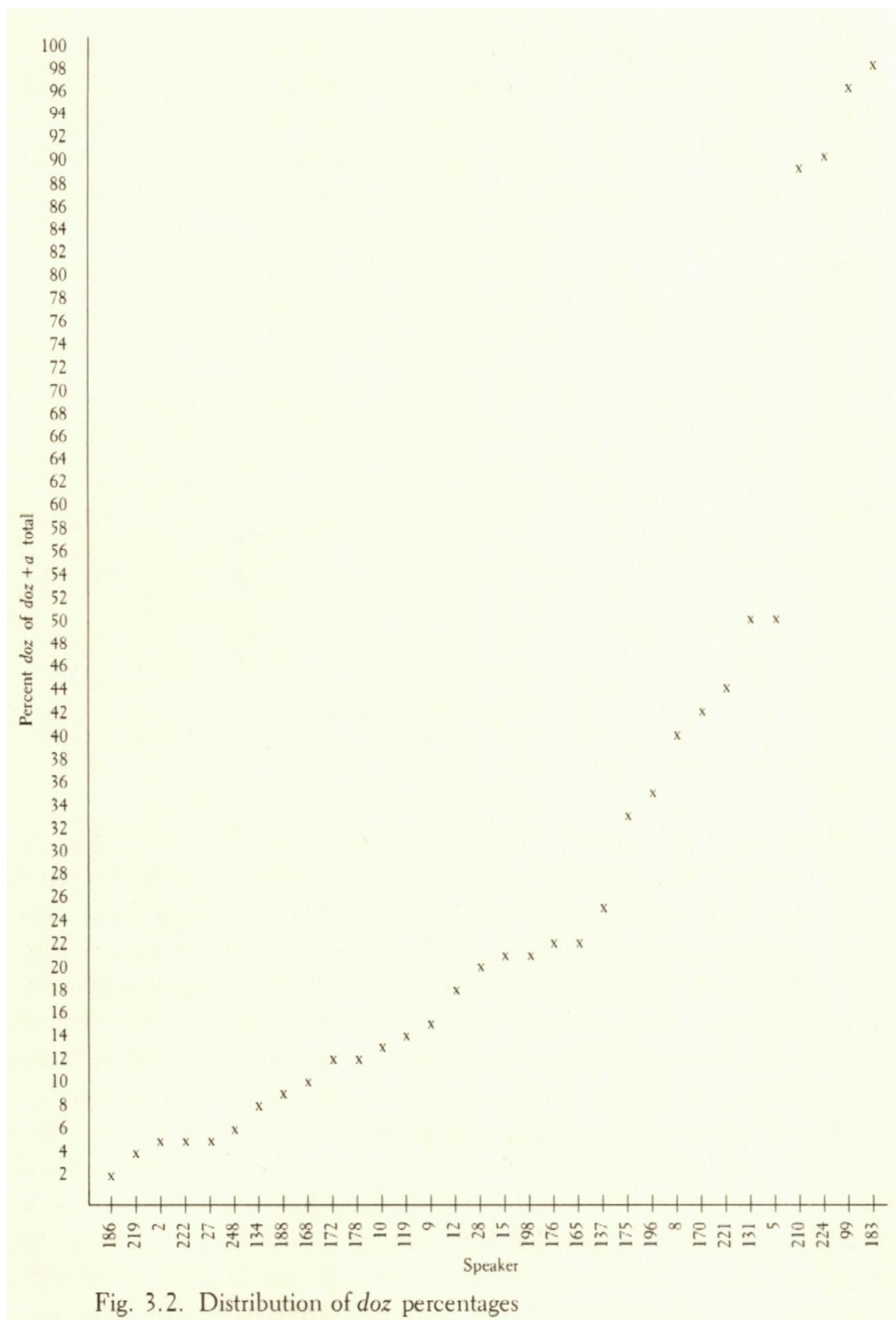


Fig. 3.2. Distribution of *doz* percentages

Figure 3.5: Figure 3.2 of (Bickerton, 1975, p.66).

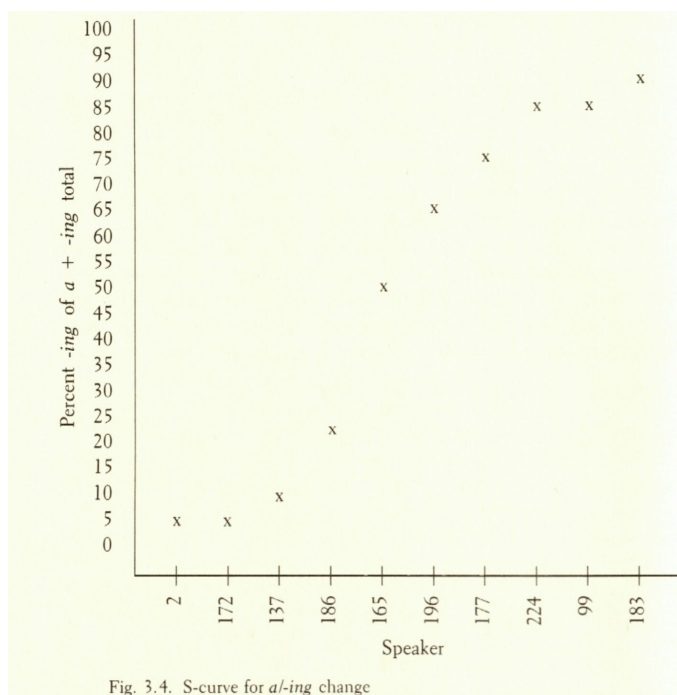


Fig. 3.4. S-curve for a/-ing change

(a)

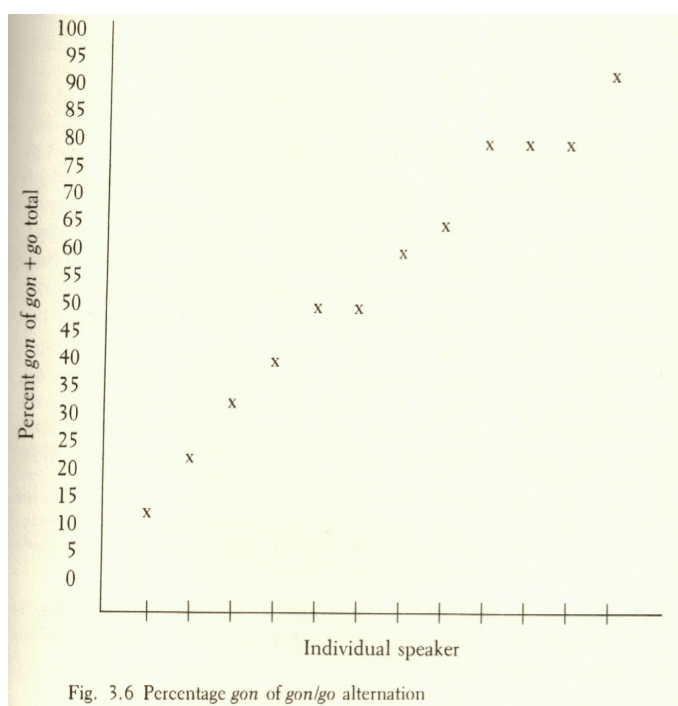


Fig. 3.6 Percentage gon of gon/go alternation

(b)

Figure 3.6: (a) Figure 3.4 (Bickerton, 1975, p.78) (b) Figure 3.6 of (Bickerton, 1975, p.89)

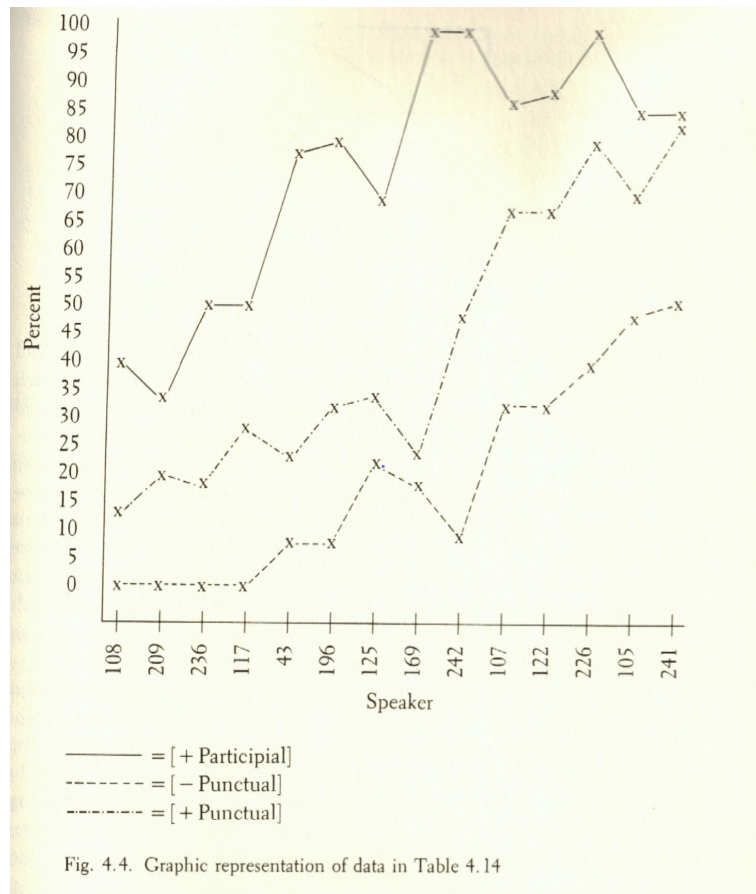


Figure 3.7: Figure 4.4 of (Bickerton, 1975, pp.155).

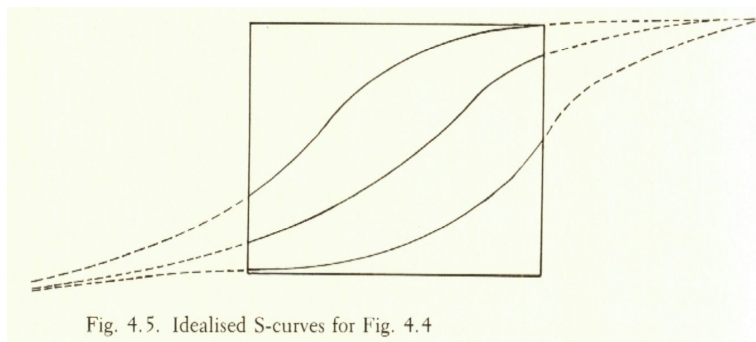


Figure 3.8: Figure 4.5 of (Bickerton, 1975, pp.155).

The shift from the lexical diffusion view of the S-curve is therefore fully acted.

However, the theoretical understanding brought forth by the lexical diffusion school remains prominent. In a textbook on dialectology, Chambers and Trudgill (1980) offer indeed the exact same picture as the one put forward by (Wang and Cheng, 1977). They first recognize the importance of lexical diffusion :

For dialectologists, the theory of lexical diffusion has instant credibility because it gives theoretical status to the kind of variety that is a commonplace of dialect surveys. [...] [The] hypothesis that change is 'lexically gradual' fits the data, since it predicts that in any ongoing change some words will undergo the innovation before others. In other words, lexical diffusion accommodates the kind of heterogeneity that exists in transition zones. (Chambers and Trudgill, 1980, p.160)

They provide next (Fig. 3.9) the stylized S-curve already present in Chen (1972), similarly justifying the validity of the S-curve by the low probability of finding a change halfway to the end:

If lexical diffusion can be studied at various stages of its progress, we should be able to determine how it proceeds through the lexicon. Here, the simplest hypothesis would be that diffusion occurs at a uniform rate. This situation is represented graphically in Fig. 10-5, which shows the percentage of lexical items that have undergone the change along the ordinate and the time interval on the abscissa. The progress of diffusion is uniform, taking in 25 per cent of the lexicon in each interval n . It follows from this hypothesis that any case study of a change in progress should in principle have an equal probability of 'catching' it at any point in its progress. However, it happens that changes almost never are found in the middle of their time span - around 50 per cent - and are most often found at one of the two extremes - above 80 per cent or below 20 per cent. This remarkable fact is already such a common observation in variation studies that it effectively refutes an assumption of uniform rate of change. In its place, we assume a rate of change that is quite rapid in the middle stages and slower at its beginning and end. This assumption is called the S-CURVE model of diffusion, after the figure that represents it, shown in Fig. 10-6. (Chambers and Trudgill, 1980, pp.162-163)

The lexical diffusion view of language change thus seems undisputed. Yet, when it comes to providing data to substantiate these theoretical claims (Fig. 3.10), Chambers and Trudgill (1980) make implicit use of the wave-based S-curve diffusion of Bailey. For thirteen speakers, they measure how many items in their lexicon are affected by the phonetic change, and they then sort the speakers according to this percentage of affected lexicon, in order to draw the spatial pattern of the S-curve.

The perspective shift from a diachronic diffusion to a synchronic variation symptomatic of this line of work is most clearly stated and draws on the notion of 'apparent time':

One significant difference is seen by contrasting [Fig. 10-7] with Fig. 10-6. While the ordinates in both figures represent the same value, the percentage of the lexicon that has undergone change, the abscissae represent different values, the former being time and the latter different speakers in the transition zone. [...] Fig. 10-7 is really just an apparent-time representation of Fig. 10-6, which is based on real time. In other words Fig. 10-6 could describe a single speaker in a transition zone who is recorded on several different occasions in real time [...]; in that case the S-curve would describe that speaker's personal progress through a linguistic

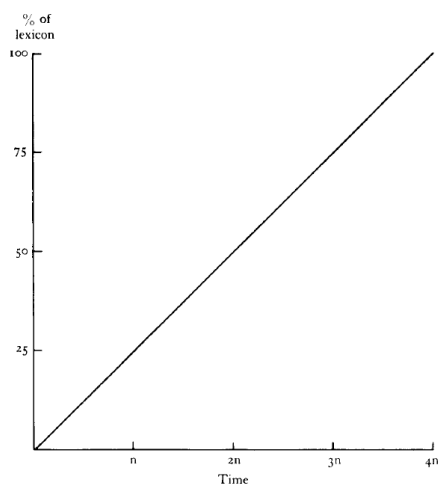


Fig. 10-5. Progress of lexical diffusion on the assumption that diffusion proceeds through the lexicon at a uniform rate

(a)

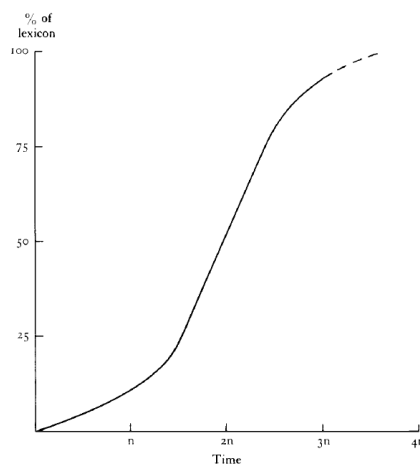


Fig. 10-6. Progress of lexical diffusion in the S-curve model, based on the assumption that diffusion is more rapid in the middle stages than at initiation and completion

(b)

Figure 3.9: Figures 10-5 (a) and 10-6 (b) of (Chambers and Trudgill, 1980, pp.162-163).

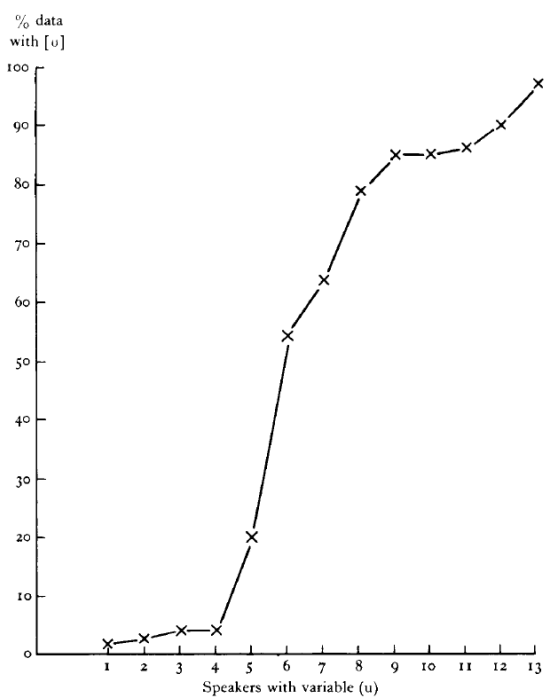


Fig. 10-7. Speakers in the transition zone for variable (u)

Figure 3.10: Figures 10-7 of (Chambers and Trudgill, 1980, pp.164).

change. However, [...] real-time data for any speaker is not available. Instead, the progress of the change is represented inferentially by describing contemporaries in the transition zone. Some people in the transition zone are further ahead in the change than others. Assuming that those people with low percentages will eventually move into the middle areas and then into the high percentages in the course of time, then the progress of this change can be viewed in apparent time by looking at the cross-section of speakers at different stages.

(Chambers and Trudgill, 1980, pp.163-165)

The position held by the authors is thus quite different from that of Bailey (1973) and Bickerton (1975), who consider that the change proceeds independently in each linguistic environment, so that the choice of the individuals between the old and the new variant, given a linguistic environment, is not binary, and makes room for variation (even in Bickerton's rule-based view); a different S-curve is thus associated with each environment. Here, the S-curve takes place over all environments, and the spread of a new variant corresponds to the number of environments affected. Yet, the wave idea that spatial diffusion is a reflect of time diffusion is fully endorsed.

Thus, as we have seen, the main ideas of lexical diffusion, combined with Bailey's theoretical contribution, shortly became an essential part of the Sociolinguists' view on the S-curve. It brings forth the idea that individuals are not making binary choices between two variants, but use them both in different lexical or sociolinguistic context, possibly with different proportions within each of these contexts. However, beyond this wave-based model of spatial diffusion, the Sociolinguists took quite a specific turn on the S-curve question, shifting away from the initial ideas of Bailey, by inferring the S-curve pattern from a new kind of synchronic evidence, no longer based on spatial repartition, but on the speakers' ages.

3.2.5 S-curve and speakers' age

In the same tendency to probe the ongoing change through a synchronic snapshot of language variation, various attempts have been made to capture the S-curve through the differences in language use between speakers as a function of their age. One of the first is Chambers (1992), who tracks the occurrence of change over the speech of six different Canadian children and adolescent speakers. He provides curves in the apparent time of speakers' ages, such as Fig. 3.11, whose grey histogram can indeed suggest an S-curve shape.

However, Chambers (1992) does not refer to their shape to interpret these curves as S-curves; rather, he considers, in line with the lexical diffusion school, and adding some of Bailey's new insights, that the S-curve pattern is suggested by the very low probability to find speakers in the mid-stream of the change:

[T]he responses of the Canadian youngsters to the three best-developed phonological processes conform to a pattern which is proving remarkably robust in studies of Lexical Diffusion. The typical pattern [...] is the S-curve, with phonological changes occurring slowly for the first 20% or so of possible instances and then rising rapidly to about 80% before tailing off toward categoricity. The empirical basis underlying the S-curve is the sparsity of speakers caught in the middle three fifths, 20-80%, at any given time, in contrast to the clusters of speakers found at either end. These figures are taken to signify that speakers must sporadically acquire new pronunciations for about 20% of the available instances as the basis for

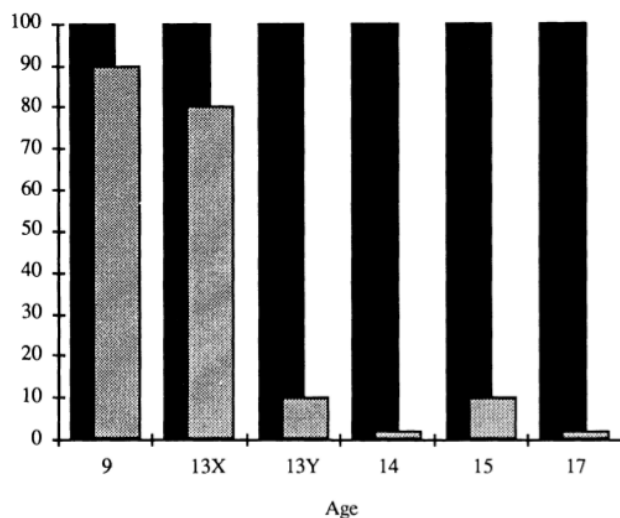


FIGURE 7. Absence of Low Vowel Merger in the speech of 6 Canadians, from youngest to oldest. In background, the scores of their English age/gender mates.

Figure 3.11: Figure 7 of (Chambers, 1992).

generalizing a rule, and that, once the process becomes rule-governed, about 80% of the instances will be affected immediately, with some portion of the remaining instances [...] resisting change and perhaps remaining as residue.

(Chambers, 1992, p.695)

In this theoretical picture, the age of the speakers plays thus absolutely no role: what matters is that none of them happens to be ‘caught in the middle’ of the change. That such a neat correspondance holds on Fig. 3.11 between the speakers’ age and their propensity to lack low vowel merger seems to be a completely incidental feature and escapes proper notice. Thus, if it were not for the age-sorting of the speakers on the X-axis of the graph, this work would clearly belong to the previous tendency.

Labov (1994, p.67) seems to be the first to reconstruct an S-curve by taking into account speakers’ age (Fig.3.12). At the core of this method lies the idea that speakers of different ages represent different time stages of the language — the elder speakers representing older states of language — so that it would be theoretically possible to have a diachronic picture of language based on the synchronic variation between speakers of different ages. However, the S-curve is obtained through a more sophisticated procedure. First, changes are sorted according to their degree of completion (which corresponds to the frequency of use of the Y-axis). Then, things go somehow deeper as the derivative of all these changes (instantiated by a little arrow on the figure) is also computed, and corresponds to a quantity called ‘age coefficient’ which is obtained by measuring the effect of age on the adoption of a feature, using regression over data obtained from a group of 180 speakers (Labov, 1994, pp.57-58). The overall picture is a bit complex, since each phonetical change seems to be adopted independently through time (apparent, as it is reconstructed from the speakers’ ages), yet they all commonly follow a bigger pattern which is the S-curve proper. To which change this S-curve corresponds is a little bit unclear, but it might be the phonetical ‘fronting’

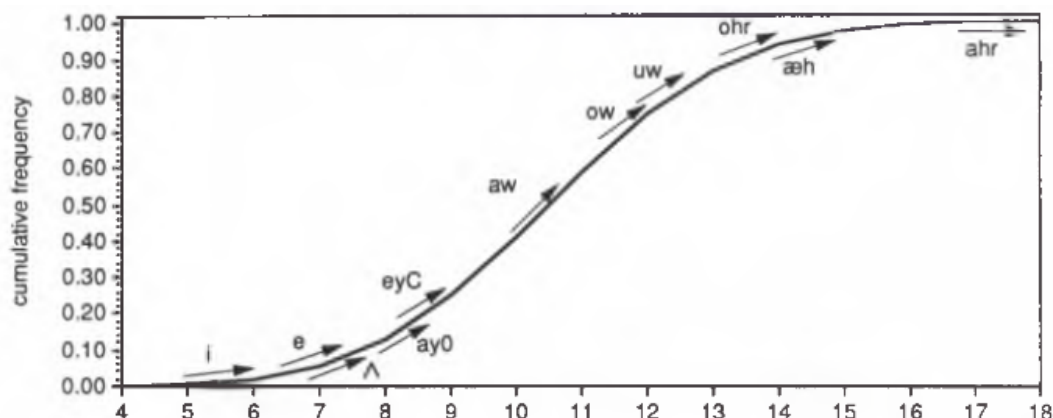


Figure 3.11. Placement of Philadelphia sound changes on S-shaped curve.
Slope of arrow (y/x) = (age coefficient of figure 3.9)/100

Figure 3.12: Figure 7 of (Labov, 1994, p.67).

in the pronunciation of a wide set of phonemes. In any case, Labov proposes that the S-curve emerges as the result of the competition of two linguistic forms, where the newer one is ‘easier to produce or easier to understand’ (Labov, 1994, p.66).

Chambers (1995) followed along these lines, focusing on an instance of a lexical replacement (*chesterfield* being replaced by *couch* among Canadian speakers) — which may be the one of the first instances, in the Sociolinguistics literature, of a replacement which is not phonological —, plotting the proportion of use of each variant for eight age classes (Fig. 3.13). In this graph, the speakers’ age is decreasing, so as to make the correspondence between apparent and real times plainer.

However clear and convincing this result may be, one might wonder whether language really becomes frozen at the end of the early childhood, so as to make people, pretty much like rocks recording Earth magnetism at the times of their formation, reliable testimonies of past stages of the language entity. Also, the modalities of this recording deserve to be questioned. For instance, do the quadragenary people really speak a frozen idiolect in which they use both variants with equal proportions? Or should we consider that each variant is used unilaterally by each speaker, but that the proportion of users within an age bracket varies from bracket to bracket? Chambers states that 70 out of 935 Canadian speakers provided more than one answer when asked “about the name they used for the long upholstered seat that holds three or four people”, so that it would seem that the data inclines towards the second hypothesis. What is more, this leaves entirely apart the actuation problem —how some speakers have started to learn language in a different way than others in the first place.

Chambers (1995) does not identify the replacement curve with an S-curve, though he used the term in previous works. It was but a matter of time: in (Chambers, 2002), discussing data published in 1998, he explicitly states that the obtained curve is an S-curve, presenting the studied replacement as a ‘well-behaved change in progress’ (Chambers, 2002, p.360). Interestingly, on this curve, the starting point and the end point are not the expected 0% and 100%. The endpoint at 90 % is explained by the

Figure 1: Use of *couch* and *chesterfield* by age groups.

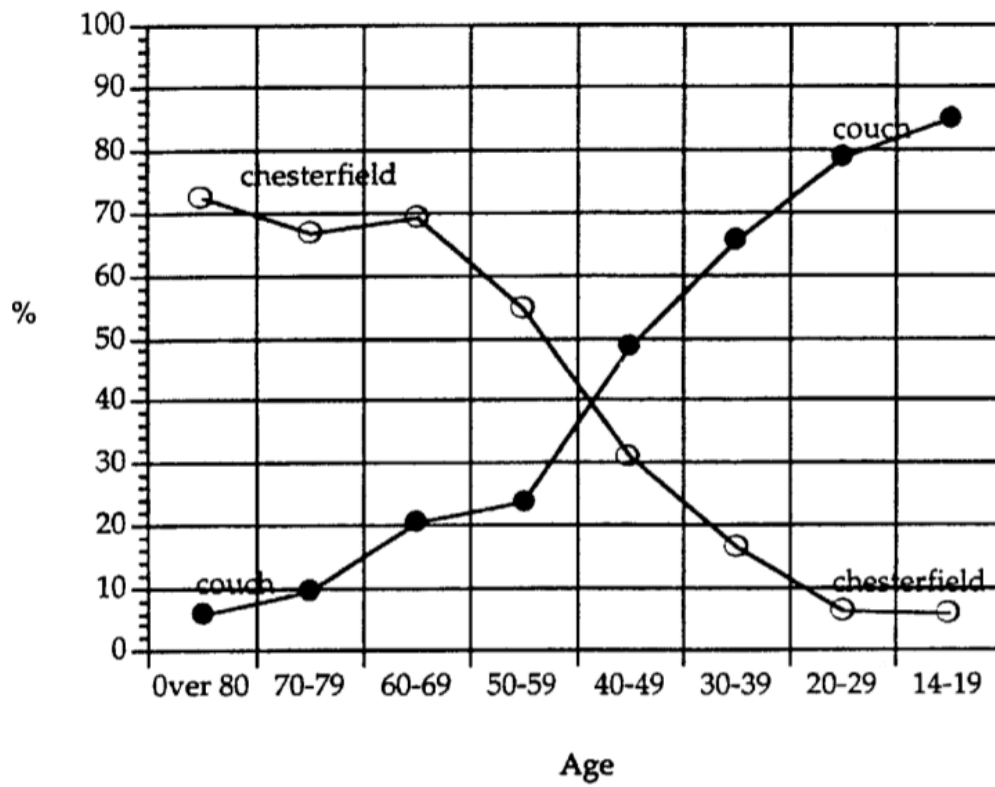


Figure 3.13: Figure 1 of (Chambers, 1995).

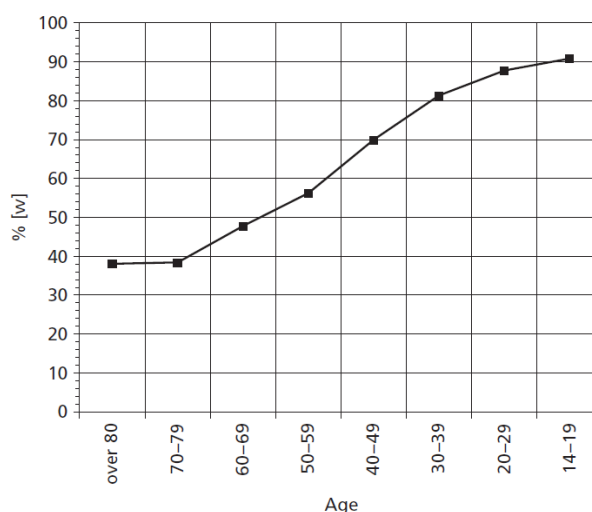


Figure 14.5 Percentage of speakers with [w] not [hw] in words like *which* and *whine* in central Canada by age

Figure 3.14: Figure 14.5 of (Chambers, 2002, p.360).

fact that the apparent time representation offers a limited time window on the ongoing process, and ‘happens to have caught the change as it nears completion’ (Chambers, 2002, p.361). The ‘initial stasis’ at 40 % is not explained; it would seem that the language was allowing for a stable alternation between the two competing phonetic variants.

This picture is interesting, as it may be one of the first S-curves represented in the literature where the starting point is different from 0 (or a very low value). We shall see later on, in chapter 4, that this is not an uncommon feature in language change. It speaks in favor of a possible stable coexistence between competitors — yet a quite unexpected one, as for no given reason, as soon as the new variant went above that seemingly arbitrary threshold of 40%, the wiping out of the older variant was inevitable:

[W]e discover a time when (wh) was stable, and the [hw] variant was the more frequent. That situation was disrupted by the 60-year-olds, whose usage changed the norms of the people older than them, so that, for them, the two variants were about equal. From that point forward, in the speech of people under 60, the [hw] variant was doomed. (Chambers, 2002, p.361)

In dynamical systems terminology, this would mean that the 40%/60% stage was an unstable equilibrium, and once perturbed, the system had to move towards its only stable equilibrium, the complete triumph of the variant (wh). Why it would be so, and why there would be such an equilibrium at 40%/60%, is however a complete mystery.

Chambers (2002) attributes the fatal eviction of [hw] to the notion of ‘critical mass’: once the frequency of the new form reaches a critical mass (which it seems to be bound to reach anyway, since it keeps increasing towards it), then the bell tolls for the old variant, for the novelty is revealed to public unconscious knowledge, and will be adopted thereafter:

Just as the tailing-off period is a recurring pattern in linguistic change, so are the initial stability and the sudden rise. Before a change takes hold, there is a

gradual, almost imperceptible, rise in frequency until the new form attains some kind of critical mass. At the earliest stage, the change apparently affects too small a population to serve as a model, but at some point it becomes perceptible, though usually beneath consciousness, and spreads through the community. No one has been able to establish the point of critical mass as an absolute value, and it appears to be different for each change, subject, as are all social developments, to countless possible influences. Once that point is attained, however, the change accelerates relatively rapidly toward the tailing-off point. (Chambers, 2002, p.361)

We shall see, in our modeling of the phenomenon, a similar idea (there is some low threshold of frequency which has to be overcome for the S-curve to unfold).

Finally, the generic, established and consensual character of the S-curve as the recurring signature of language change is asserted:

The combination of these three stages - initial stasis, rapid rise, and tailing off - gives a characteristic shape in graphic representations that is known as an S-curve. The significance of the S-curve pattern for linguistic change was introduced by Wang and his associates (esp. Wang and Cheng 1970, Chen 1972) as an adjunct of lexical diffusion, a type of change in which lexical items undergo a sound change one at a time, so to speak, and extrapolated in a variationist context by Bailey (1973: 77). The S-curve has since been observed in diffusions of all kinds (Chambers and Trudgill 1998: 162-4), and is now established as a kind of template for change. (Chambers, 2002, p.361)

The S-curve remains largely unexplained (as an example, why would it tail off before completion whereas the old variant is bound to disappear quickly once the critical mass is attained does not receive any discussion). One might also note that the reference to Chambers and Trudgill (1980) covers only one example of an apparent time S-curve based on the behavior of thirteen speakers, as was shown earlier, and relates only to phonetical change. This hardly counts as “diffusions of all kinds” and certainly does not make the S-curve an established template of language change.

A little more ought to be said concerning the notion of ‘critical mass’, for it leads to an interesting paradox. As long as the frequency of the new form lies below the threshold corresponding to the ‘critical mass’, the propagation of the novelty cannot rely on social imitation, and unless we posit another spreading mechanism, the frequency of the new form has no reason to increase. But if the frequency remains constant, how can the threshold ever be reached? This so-called “Threshold Problem” (Nettle, 1999) was actually solved by assuming an inherent fitness of the variants, able to sustain the novelty while it is not yet embraced by enough speakers:

In real languages rare variants get adopted and spread through entire communities in a rising S-shaped curve. Even given the existence of sources of variation, it is not entirely clear how this can happen. [...] [T]he solution may well lie in the fact that the learner does [not] weight all the sources to which he is exposed equally but uses a biased learning strategy. If the learner were at least sometimes biased towards new variants for some reason or other, then those variants would have a chance of overcoming the threshold of rarity. There are two possible sources of such a bias. One is social; the learner may favour the speech of some individuals more than others, and so, if socially influential people are from time to time the bearers of new variants, transmit those variants. The other is linguistic or functional; certain linguistic variants may have some functional attribute which makes them easy to acquire or use which favours their adoption over their competitor variants.

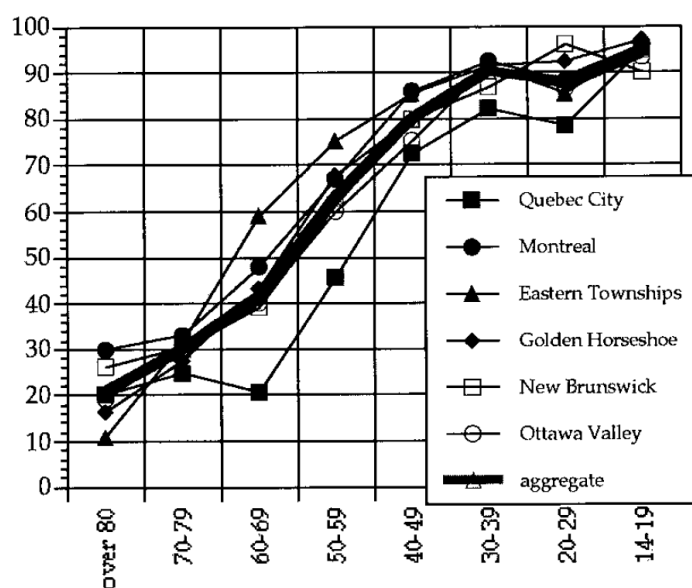


FIGURE 2

Percentage of *snuck* in the speech of different age groups in six Canadian regions, with aggregate score superimposed to illustrate the linearity of change

Figure 3.15: Figure 2 of (Chambers, 2007, p.31).

(Nettle, 1999, p.99)

This shift from geographical to age variation is made all the more explicit in a later work of Chambers (2007), as he shows (Fig. 3.15) the on-going S-curve for speakers of different ages, for different locations. According to his dataset, it would seem that the geographical diffusion is near instantaneous, so that the age of speakers matters most, as they all record the language state as it were by the end of their “critical learning period”, which is the end of adolescence. The wave-based model of Bailey is no longer of relevance to explain these data:

The S-curve pattern that we found in the Golden Horseshoe essentially repeats itself in all the Dialect Topography regions, as Figure 2 shows. Notwithstanding the relative autonomy of the regions, each with their own central places and self-contained communication networks, the replacement of *sneaked* by *snuck* follows much the same trajectory in a similar time frame in all of them.

(Chambers, 2007, p.30)

This approach, called the “apparent time” construction (time is given by the age of the speakers, because of this ‘freezing’ mechanism at the end of the learning period of the speakers, just like successive soil layers captured the state of the environment when they were on the surface, so that a synchronic view of the different strata gives a reconstructed view of the diachronic change of the environment), has thus completely evicted the spatial reconstruction of the diffusion pattern.

3.3 Empirical study of the S-curve

As we have seen, the S-curve remained for long a theoretical belief, which gradually came to be supported by limited synchronic empirical data under the assumption that a spatial S-curve would correspond to a temporal one, due to the wave-based nature of the diffusion. Only with the development of corpus-based studies could the diachronic pattern of the S-curve truly emerge. Interestingly, we will see that this “corpus-based” tradition will not especially focus on phonetic change (not easily traceable in corpora), but on all other kinds of language changes: lexical, morphological, syntactic...

3.3.1 First corpus-based studies on the S-curve

The oldest of these studies is probably a Russian 1974 paper by Piotrovskaja & Piotrovskij, which unfortunately I could not find, yet of which we dispose an account in (Altmann et al., 1983). Piotrovskaja & Piotrovskij considered two sets of data (collected previously by other researchers): the appearance of new, simplified forms for the expression of the genitive in Russian, comparing the two variants on a period lasting from 1881 to 1910, with a one-year time window (hence based on thirty data points); and the proportion of Arabic loanwords, on a period dating back from 963 to 1200, with only five data points. In the first case, there is a clear competition between two well identified variants, and their relative frequency is measured, while in the second one, the proportion of the novelty is computed from the raw frequency of tokens in the corpus.

The authors did not only make use of diachronic data; they also tried to fit them with an actual mathematical function. In that regard, their work is all the more innovative and may be the first instance of a mathematical characterization of the S-curve in Linguistics. For this fit, they use the Arcus Tangent function:

$$p = \frac{1}{\pi} \arctan \mu(t - t_I) + 0.5 \quad (\text{Altmann et al., 1983, p.106}), \quad (3.2)$$

μ and t_I being two parameters to be fitted.

Altmann et al. (1983) criticize this choice, on the ground that:

[I]t is not easily possible to justify [this function] from a model of the transition process [...], since the arc tan function does not appear as the solution of a simple differential equation. (Altmann et al., 1983, p.106)

They thus offer to make use of the hyperbolic tangent instead:

$$p = \frac{1}{2} \tanh w(t - t_I) + \frac{1}{2} \quad (\text{Altmann et al., 1983, p.108}), \quad (3.3)$$

as it derives from a simple and well-know dynamics of the change (just as Pemberton (1936) stated), explaining then how to fit the data, using the logit transformation and then a simple linear regression.

Interestingly, if the S-curve fits very well the Russian genitive data (Fig. 3.16a), it is not so obvious for the Arabic loanwords in Persian (Fig. 3.16b). Piotrovskaja & Piotrovskij proposed to fit them with a linear curve, but Altmann et al. (1983) insist

that the S-curve is still relevant in this case, however partial it may appear. They do not specify, however, if the pattern stops because the process itself came to an end, or because the data do not go further.

From an empirical point of view, this short paper is of course of great interest and has been duly recognized as such. Furthermore, it also represents some sort of a theoretical shift compared to the previous work, sociolinguistically oriented. Crucially, the S-curve no longer appears as a pattern of diffusion, but as resulting from a living, ecological competition between populations of “language entities”:

Language, like a living organism, is subject to (synchronic) variability and to permanent (diachronic) alteration. From the historical point of view new entities come in, others die out, change their frequency of occurrence or change their form. The majority of these alterations can be conceived as growth or decay of classes of language entities. [...] These classes do not remain intact during history but continuously grow or diminish by either creation or complete loss of elements or by transition of elements from one class to another. It is this process of transition which is studied here. (Altmann et al., 1983, p.104)

Also, the authors provide a new perspective for future research, based on the characterization of the two parameters constitutive of the S-curve:

More empirical data are needed to verify the proposed theory [that language changes according to a hyperbolic tangent]. In case it should prove to be valid, the important problem of historical linguistic concerning the transition of elements from one class to another has been reduced to the interpretation of the two quantities, t_I and w , which may turn out to be either universal constants or special constants depending on different language units or on different languages. (Altmann et al., 1983, p.112)

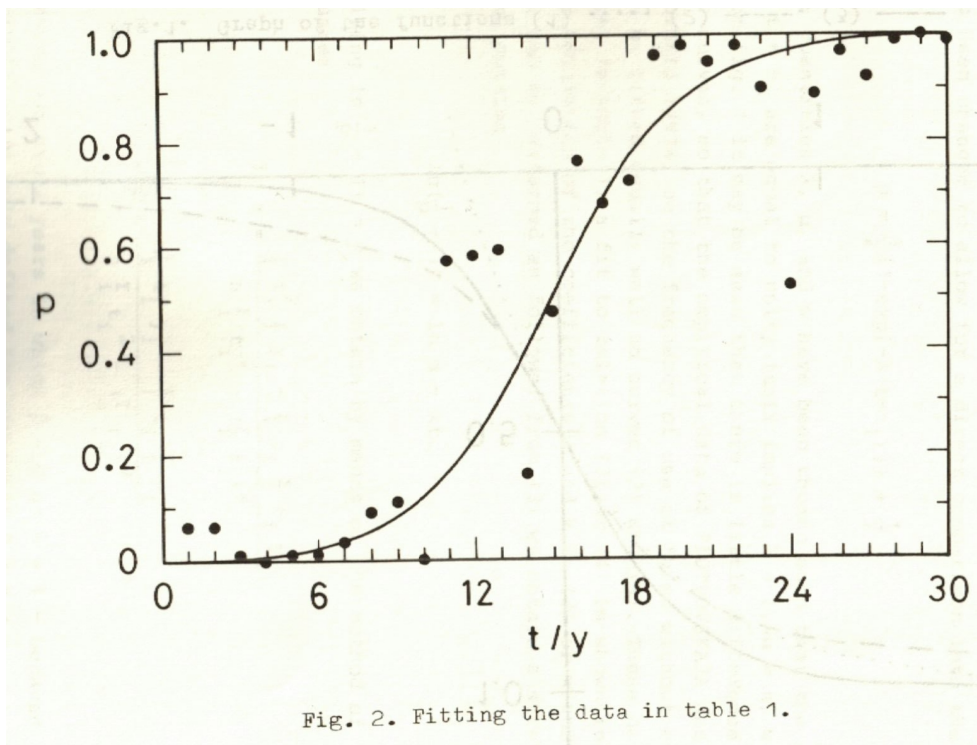
The perspective shift of this work is not to be dismissed. In this view, the speakers seem not to play any relevant part — which is at the exact opposite of the Sociolinguistics position. Language is seen as an autonomous ecosystem, whose change is not explained, but held as a natural course of things.

A same line of thought is pursued in (Kroch, 1989a). After reporting data on the rise of periphrastic *do* in English in various lexical environments, initially gathered by Ellegård, Kroch (1989a) proposes that the “roughly S-shaped growth curve” (p.113) obtained this way could be modeled by a logistic curve of equation:

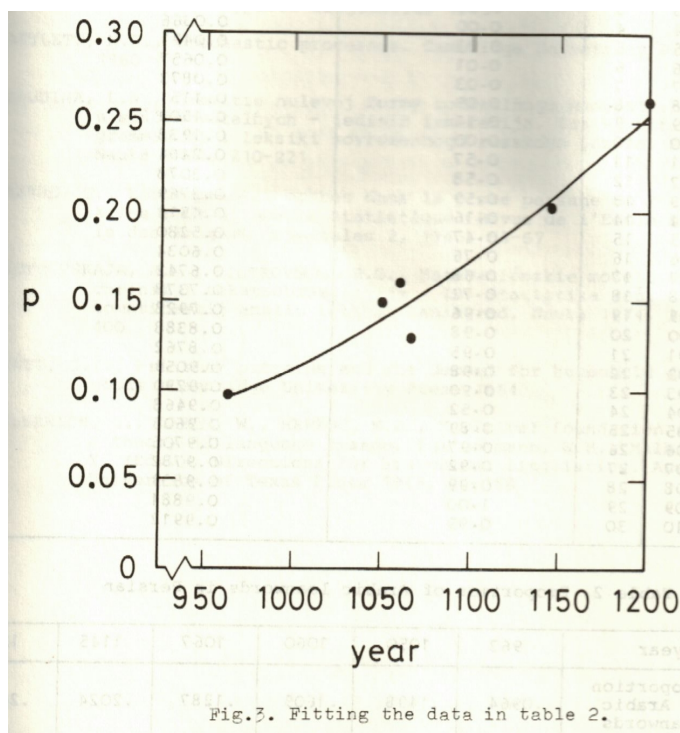
$$p = \frac{1}{1 + \left(\frac{1-p_0}{p_0}\right) e^{st}} . \quad (3.4)$$

To justify this modeling, he brings forward interesting considerations:

The logistic function expresses a number of basic growth relationships in population biology and genetics. Of most interest to us is the fact that it expresses the rate of replacement of one species for another in a context where the two compete with differential reproductive success for the same resources [...], a situation that is exactly analogous to the replacement of one linguistic form by another when the two are unequally likely to reinforce their own future use. The value of assuming that the curve of replacement of one linguistic form by another is logistic is that one can then use sampling data to estimate the two parameters of the



(a)



(b)

Figure 3.16: (a) Figure 2 of (Altmann et al., 1983, p.110). (b) Figure 3 of (Altmann et al., 1983, p.113).

logistic equation, s and p_0 . Of these parameters, s represents the advantage of one form over the other and p_0 the initial relative frequency of the favored form. By comparing the values of these parameters across linguistic environments, one can obtain a quantitative description of the time course of a change [...].

(Kroch, 1989a, pp. 111-112)

Surprisingly, this line of thought is close to the one already pursued by Altmann et al. (1983). Once again, the individual speakers of language fade away and are replaced by the idea of an ecological competition between two linguistic entities. The speakers are but an environment to which the linguistic forms adapt -with differentiated success. In this perspective, the two parameters of the S-curve receive a straightforward interpretation. Also, in line with lexical diffusion ideas, he considers that the growth of the new form obeys an S-curve in each linguistic environment separately.

It thus becomes possible to see if a given form is more or less adapted to different environments by computing, from corpus data, the value of the s parameter of each corresponding S-curve. However, the difference between the various linguistic environments can manifest itself in a different, opposite way:

The second model for the change might be called the ‘mediated influence model.’ Under this model the influence of processing effects on changes in the frequency of syntactic alternants in different environments is constrained by the way that the forms are categorized by the grammar. If a form (here *do*) is introduced into a number of different environments by a single rule, then it must change its frequency of use in all environments. Psycholinguistic effects that cause one environment to favor *do* more than another will be reflected in a higher frequency of use of *do* in the favored environments, but change in frequency of occurrence affects all environments concurrently. Such a model implies that the rate of increase in the use of *do* should be the same in all environments and that the degree to which each environment favors or disfavors the use of *do* should be the same at every point in time.

(Kroch, 1989a, p.155)

Thus, we can distinguish two scenarios: either the affinity between a form and an environment leads to a higher rate of change in this environment, or to a higher initial frequency. In the first hypothesis, each environment is characterized by the same value of the p_0 parameter but a different value of s , while the reverse holds in the second hypothesis. Actually, Kroch (1989a) shows that, in the case of *do*, a logit transform of the data for the four different environments speaks in favor of the second hypothesis (Fig. 3.17).

As he understands it, this is the consequence of a rule-based change — and, by the way, a serious challenge to the lexical diffusion phenomenon:

[This] model implies that the speech community shares a norm for the overall rate of use of the *do* form, which is defined as single form by the fact that a single grammatical rule governs its distribution.

(Kroch, 1989a, p.156)

One limitation of this view is that a rule-based change seems fitted for a syntactic change in a generative grammar framework; yet it is not so for other kind of changes, such as the one observed in (Altmann et al., 1983), which concerns the borrowing of Arabic words in Persian language. However, a particularly interesting idea is put forward to explain how the rate of change could be constant in all syntactic contexts,

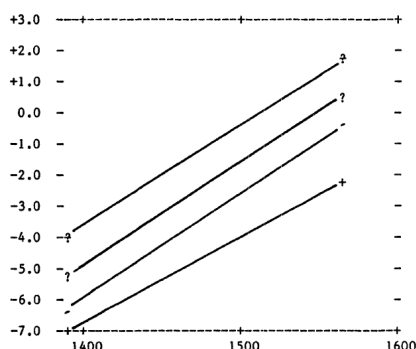


Figure 5. Actual regression lined fitted to the transformed frequency data for the rise of *do*.

Figure 3.17: Figure 5 of (Kroch, 1989a, p.161).

making explicit reference to psycholinguistic considerations: “these processes would be constrained to influence only the overall rate of use of *do*, because this is the only frequency that language learners would be tracking and matching”.

Also, the interpretation of parameter p_0 , especially in the second model where it depends on how fitted an environment is for the novelty, is quite puzzling. Indeed, one might wonder how the frequency of a new form could suddenly jump in such a discontinuous way. While data are probably compatible with such a jump, it is certainly because the time window is quite large (about 25 years), so that there is a jump of frequency when the novelty appear (as much as there is between two adjacent data points). We can note, in the detail of the data, that the frequency of the novel variant jumps from zero to non-zero at different times for different environments, so that the notion of ‘initial frequency’ does not make a lot of sense. All the worse, without a proper redefinition of time (here given by the different years), p_0 would be the frequency of the term for year $t = 0$, while the process starts around the fourteenth century. This initial frequency p_0 thus corresponds to a frequency of the word at a time where Middle English did not even exist as such. There is actually some confusion regarding this parameter, since Kroch identifies it as the intercept of the logit transform of the S-curve, while this intercept b is related to p_0 by:

$$p_0 = \frac{1}{1 + e^{-b}}. \quad (3.5)$$

In the case of Ellegård’s data, the p_0 value obtained for the four different environments is of course negligible.

Furthermore, this interpretation of the p_0 parameter is not the only possible one. One might as well consider that this p_0 is actually the equivalent of a time t_0 :

$$t_0 = \frac{1}{s} \log \left(\frac{p_0}{1 - p_0} \right), \quad (3.6)$$

so that different values of this time would only mean that the S-curves are temporally shifted from one another, not that they start with different initial frequencies at some

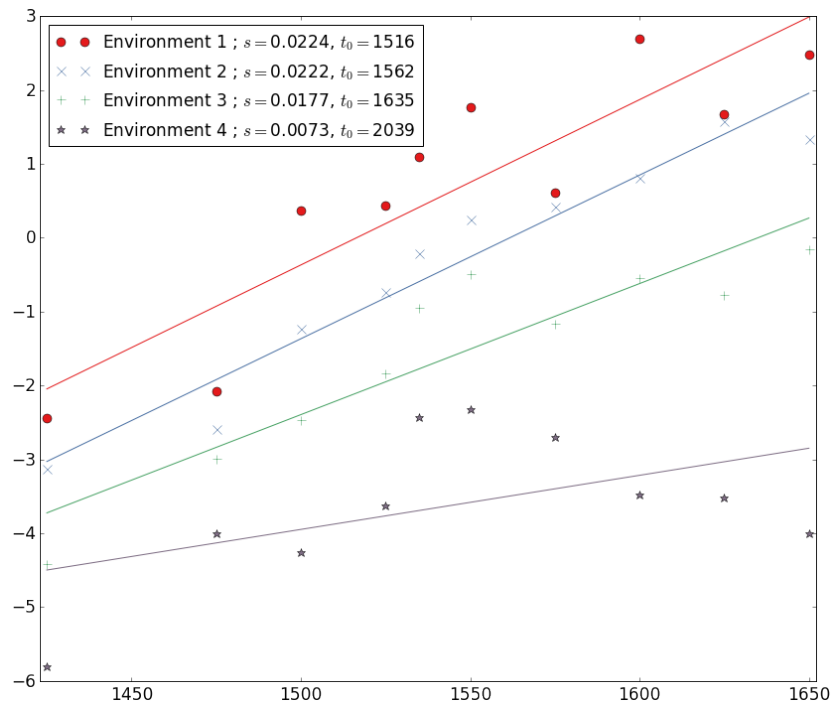


Figure 3.18: Computation of the logits' slopes from Table 1 in (Kroch, 1989a, p.135).

arbitrary origin of time. In that sense, different values for the intercept would simply mean that some contexts are affected later than some others. We can note, from our own reappraisal of Ellegård's data (Fig. 3.18), that environments which differ in terms of rate also differ in terms of intercept. Actually, environments affected later (with a greater value of t_0) show a slower rate s ; this conforms exactly to Bailey's hypothesis —“What is quantitatively less is slower and later; what is more is earlier and faster.” (Bailey, 1973, p.82)— which Kroch aims to contradict. However, with a sample of three environments (the fourth environment does not display a proper sigmoid and should have been dismissed), one cannot conclude anything firmly from all this.

Aside from these flaws, this paper was clearly innovative, both theoretically and methodologically. The S-curve was not only taken as a phenomenological signature of the change, but also as an empirical tool to compare two theoretical hypotheses on language change. It is now empirically grounded, theoretically motivated, and conceptually productive.

3.3.2 The Constant Rate hypothesis

Though the former work already contains all the most important and pioneering ideas of Kroch, his most well-known contribution to the subject is a paper entitled *Reflexes of grammar in patterns of language change*, published the same year (but certainly posterior in its writing), which extends the former one and grounds its theoretical claims on several other examples from corpus data. This paper (Kroch, 1989b) was actually so important on the topic of the S-curve that it has sometimes been regarded

as having introduced this concept (Marchello-Nizia, 2006, p.42).

The goal of this paper is twofold. On one hand, it tries to assess the validity of the fact that when a change proceeds, it does so at the same rate within all linguistic environments affected — an idea he calls the ‘Constant Rate hypothesis’. This, according to Kroch (1989b), is the sign that change results from the actuation of a single change of a grammatical rule. On the other, he tries to understand more closely the nature of the change which is going on, relying on the Constant Rate Hypothesis. Indeed, a same grammatical change can take multiple ‘surface’ forms (which makes sense both in Generative Grammar framework, of which Kroch makes use, and in the Construction Grammar framework, as discussed at length in chapter 1). The idea of Kroch is that if two changes exhibit the same rate, they might well be two instances of the same change, in two different linguistic environments.

So far, in the literature around the S-curve, it was always safely assumed that the ongoing language change was clearly identified; this, probably, is due to the historical influence of sound change on the S-curve literature, as we have surveyed. However, this is seldom the case for other domains of language, and the question of *what* exactly is changing during a language change is a very delicate one, as we tried to show in chapter 1). Kroch’s line of research is thus quite innovative in that sense. Once again, the fact that change follows a logistic S-curve is taken for granted, so as to be used to draw new insights on language change, in this case, to help identify the linguistic locus of a given change:

We are now in a position to tackle the basic question posed jointly by our grammatical analysis of the change in the English auxiliary system and our hypothesis on the relationship of changes across contexts. Can the rise in use of periphrastic *do* be related quantitatively to any other reflex of the loss of V-to-I raising? Under our hypothesis, we would expect any other such reflex to change at the same rate as the use of *do*. If we find such a result, our constant rate hypothesis will be considerably strengthened, since the changing forms will not exhibit any superficial relationship to one another. (Kroch, 1989b, p.225)

We can however note that there is a curious bootstrap between the two purposes of Kroch: The Constant Rate Hypothesis is used to show that two linguistic changes are in fact one and the same because they proceed at the same rate, as well as this very same fact is used to corroborate the said hypothesis.

Kroch (1989b) proposes four examples on which to substantiate his claim. The first one deals with the replacement of *have* by *have got* in British English in two different semantic contexts, one in which the possession is ‘temporarily bounded’, the other where it is deemed ‘permanent’. The provided corpus data come from an earlier work by Noble, and only three data points are provided for each of these two semantic contexts, covering a time period ranging from 1750 to 1930. Though three data points and two contexts would be most insufficient to assess the relevance of the Constant Rate Hypothesis, the evidence supports it (Fig. 3.19). Kroch, however, does not compute the slope and the intercept of the logit transform of these data as done on Figure 3.19, but instead compares how the occurrences of the new variant are shared out among the different contexts, and shows that this distribution remains constant. He legitimately deduces from it that the spread of the new variant proceeds at the same rate in both environments.

The second one regards the rise of the definite article before possessives in Por-

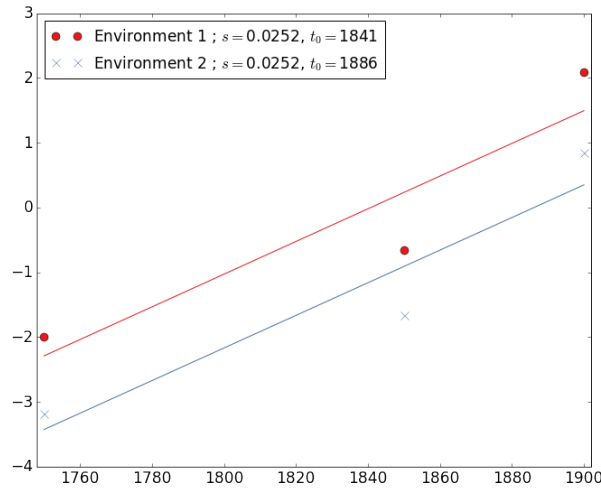


Figure 3.19: Computation of the logits' slopes from Table 1 of (Kroch, 1989b, p.209).

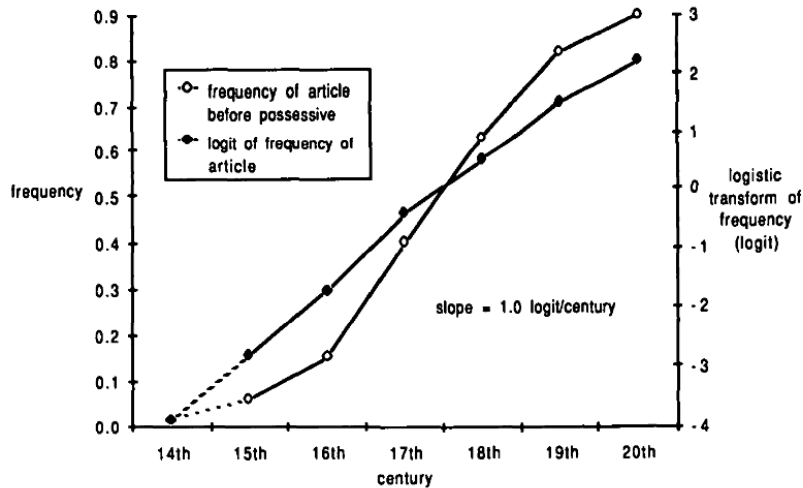


FIGURE 1: Rise in the use of the definite article before possessives in Continental Portuguese (adapted from Oliveira e Silva, 1982).

Figure 3.20: Figure 1 of (Kroch, 1989b, p.209).

tuguese. The data reported by Kroch from a thesis manuscript by Oliveira e Silva shows a beautiful sigmoid which spans over seven centuries (Fig. 3.20).

Here again, the Constant Rate hypothesis is not supported by direct empirical evidence, but by the fact that all contextual effects on this rise are constant (according to a procedure performed by Oliveira e Silva). In a third example, focusing on word order in Middle French, slopes of the logit transform of the change in three different environments are displayed, but the empirical data from which they are computed (taken from a work by Fontaine) is not provided. The three logit transforms indeed seem to be parallel lines, supporting the Constant Rate hypothesis accordingly. In this

TABLE 4. Slope and intercept parameters of logistic regressions on data in Table 3 (slope is measured in logit units per century)

Negative Declaratives		Negative Questions		Affirmative Transitive Adverbial & Yes/No Questions		Affirmative Intransitive Adverbial & Yes/No Questions		Affirmative <i>wh</i> - Object Questions	
Slope	Intercept	Slope	Intercept	Slope	Intercept	Slope	Intercept	Slope	Intercept
3.74	-8.33	3.45	-5.57	3.62	-6.58	3.77	-8.08	4.01	-9.26

Figure 3.21: Table 4 in (Kroch, 1989b, p.225).

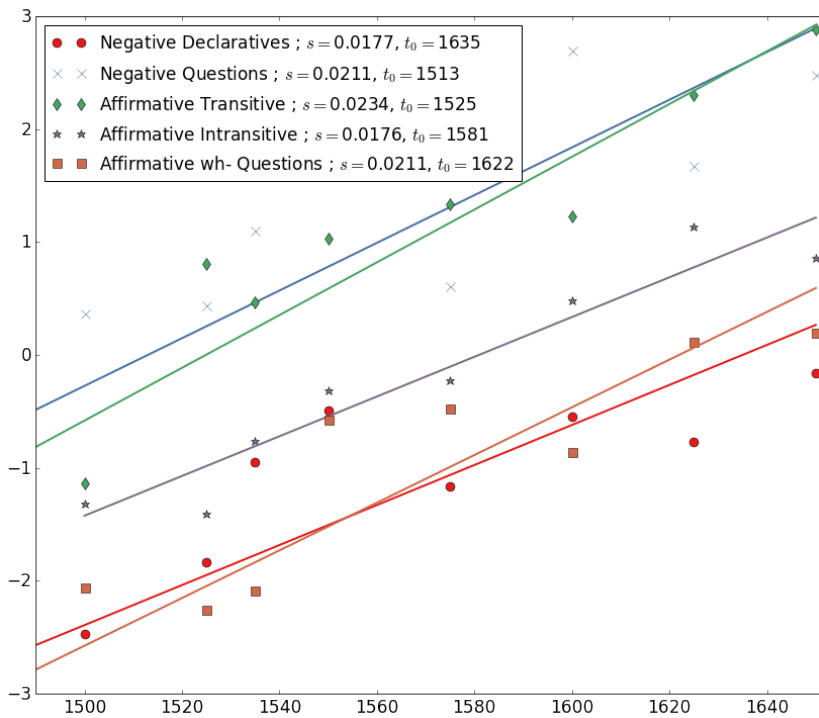


Figure 3.22: Logits computed from Table 3 of (Kroch, 1989b, p.224).

case, Kroch also considers another change in Middle French (using data from Priestley) and finds the same slope as the three previous lines, from which he deduces that this other syntactic change is actually another instance of one and the same grammatical rule change.

The fourth and final example is once more the rise of periphrastic *do*. The date, still adapted from Ellegård, is now spread out between five environments (two of the same being the same as in the previous work). For some reason, we were not able to reproduce the results of Kroch and do not find the same slopes (Fig. 3.22) as he does (Fig. 3.21, and both sets are not proportional one to another. This, however, is a mere detail, and his conclusion that the changes indeed follow S-curves holds no matter which slopes are considered, his or ours.

Now that we exposed most of the empirical content relevant to the S-curve, we turn towards his interpretation of the two parameters. As we saw, the situation of the parameter p_0 was far from receiving a clear understanding in his previous work. Indeed, if the constant rate hypothesis holds, then one has to explain why this change looks differently in different linguistic environments:

[W]hen one grammatical option replaces another with which it is in competition across a set of linguistic contexts, the rate of replacement, properly measured, is the same in all of them. The contexts generally differ from one another at each period in the degree to which they favor the spreading form, but they do not differ in the rate at which the form spreads. (Kroch, 1989b, p.200)

Hence, the second parameter of the logistic becomes of crucial importance, since it is responsible for these differences. The problem is, this parameter is equivalent to a shift in time of the S-curves. Kroch now recognizes that this is the proper interpretation of the logit transform intercept:

$$p = \frac{1}{1 + e^{k+st}} \quad (\text{Kroch, 1989b, p.204}). \quad (3.7)$$

The constant s is the slope of the function and hence represents the rate of replacement of the new form by the old, whereas k , the intercept parameter, measures the frequency of the new form at the fixed point in time $t = 0$. For a given value of s , the curve has a fixed form. Changing the value of k merely slides the curve along the time axis. Conversely, changing the time point to which the value $t = 0$ is assigned alters the value of k , a fact that will become important later in our discussion. (Kroch, 1989b, p.204)

However, Kroch cannot accept this interpretation:

In principle, actuation might occur in three different ways across the various contexts in which a new form appears. First and most obviously, it might occur sequentially, with the new form appearing at the start in the most favoring context and then successively in less and less favorable contexts. [...] On the other hand, actuation might instead occur simultaneously in all contexts. Then two further possibilities present themselves: either the initial frequency of the new form will be the same in all contexts or it will vary by context. Under the former scenario, at the point of actuation there will be no distinction among more and less favoring contexts [...]. However, fitting data to the logistic [...] will not allow us to choose between the latter two options. (Kroch, 1989b, p.205)

Indeed, he understands the Constant Rate hypothesis as a manifestation of a rule change. As an acknowledgement of this new rule, the change must proceed in all environments in exactly the same fashion; but also at the same time. Otherwise, it would mean that the rule is successively activated as the change worms its way from environment to environment. It would imply that a syntactic rule can apply or not in relevant environments, depending on it how it has spread diachronically over language; but a rule whose application is variable as such is not a rule proper (at least in the Generative Grammar sense of a grammatical rule). Then, Kroch strives to see, in this diversity of intercepts, the expression of different ‘initial frequencies’, which he defines in the following way:

Note that [...] the logistic, like other functions used in statistics, idealizes the empirical situation. Under the model, there is no time t for which $p = 0$, nor any for which $p = 1$. Of course, actual linguistic changes have starting and ending points, so the model can only approximate real data, and this approximation falsifies the change process precisely at the beginnings and ends of changes. In particular, at the beginning of a change p jumps from zero to some small positive value in a temporal discontinuity [...]. (Kroch, 1989b, pp.204-205)

This idea is not empirically aberrant. In the case of the periphrastic *do*, there seems indeed to be such a discontinuity. After a sudden increase from zero to a higher value (typically between ten and twenty percents), the change slows down again, displaying the long start of the sigmoid. However, there is a much more appropriate way to capture this behavior, replacing the expression of the sigmoid by:

$$p(t) = p_0 + \frac{1 - p_0}{1 + e^{-s(t-t_0)}}. \quad (3.8)$$

By defining p_0 as the lowest non-zero frequency for a set of frequency values of the novelty in a given environment, we can adjust the logit transform accordingly to still get a linear curve:

$$\log\left(\frac{p_t - p_0}{1 - p_0}\right) = s(t - t_0). \quad (3.9)$$

In that case, the Constant Rate hypothesis does not hold anymore (Fig. 3.23). Empirically, this overall picture proposed by Kroch is thus dubious — a judgment which clearly does not aim at overshadowing the praiseworthy methodological advance of this work. Kroch proposes the Constant Rate hypothesis alongside other hypotheses; and he tries to devise possible ways to empirically distinguish from them, discussing which criteria can be drawn from corpus data analysis and which cannot.

Theoretically, we also find interesting insights on the S-curve topic. For instance, Kroch provides an interesting reflection on the possibility of a coexistence between two syntactic rules:

It is obvious that [...] the process of language change is not a fact of grammar but a fact of language use and so must be studied with tools appropriate to that domain. The study of language use is the study of the choices that people make among alternative forms in their repertoire of grammatical knowledge in formulating utterances. [...] [V]ariation often reflects choices that are not categorically determined by linguistic principles at any level but instead are only probabilistically influenced by features of context and situation. (Kroch, 1989b, p.202)

The idea that language change is rooted in language use, which is akin to the psycholinguistic considerations proposed in (Kroch, 1989a), is certainly not new, but contrasts interestingly with the two dominant theories, attributing language change to sociological influences and identification mechanisms, or to imperfect learning by new generations. As was suggested by Altmann et al. (1983), language change is not held as an occasional and external deviation from the normal course of language.

This raises the question of which mechanism is responsible for the S-curve pattern. Contrary to Altmann et al. (1983), who stated that the advantage of the S-curve was

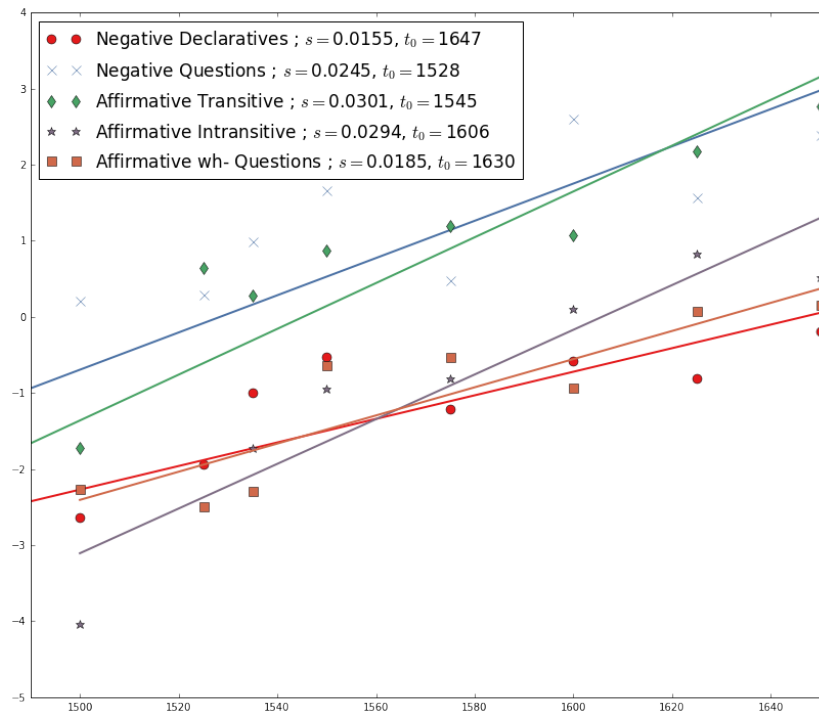


Figure 3.23: Modified logits computed from Table 3 of (Kroch, 1989b, p.224).

its association with a proper dynamical model of the change, Kroch (1989b) is not so optimistic on this matter, and favors the S-curve for more practical reasons:

Thus, given the mathematical simplicity and widespread use of the logistic, its use in the study of language change seems justified, even though, unlike in the population genetic case, no mechanism of change has yet been proposed from which the logistic form can be deduced. (Kroch, 1989b, p.204)

This idea is very interesting if we consider it at the light of the historical retrospective we sketched in this chapter. At the very beginning of this research tradition, we saw, with Osgood and Sebeok (1954), that the S-curve was taken for granted as an apt description of social diffusion of language change, based on analogies with other cultural diffusion phenomena. Then, the S-curve, considered to be theoretically consensual, was looked at in different perspectives, gradually shifting further away from the idea of social diffusion proper; on one side, there was lexical diffusion, on the other side, sociolinguistic differences (areal or age-based). In both cases, the S-curve was (supposedly) reconstructed from synchronic evidence. With this new empirical tradition, empirical data is finally provided to support the diachronic S-curve ; but at the same time it completely loses its initial theoretical support as a signature of social diffusion. Hence, we are left without any explanation for a pattern whose robustness is empirically established in the same time. In stating this, Kroch also stresses that the S-curve in language change is not another example of a simple diffusion of some sort (at least the terms of the analogy are not straightforward). For the first time, the S-curve ceases to be theoretically obvious.

X	Slope 1	Slope 2
Remove	1.45 ± 0.53	0.91 ± 0.24
Assign (10^{-2})	1.80 ± 0.33	1.80 ± 0.43
Assign (10^{-3})	2.41 ± 0.48	2.29 ± 0.65

Table 3.1: Different values of the slopes of the two changes studied in (Santorini, 1993), according to the way the extreme values 0.0 and 1.0, producing divergence in the logit transformation, are dealt with.

3.3.3 Further evidence from corpus data

Since then, the S-curve has been attested in numerous corpus-based studies, which we shall not mention exhaustively. However, these studies seldom rely on an analytical fit of the data, and rather are contented with the qualitative shape of the curve to identify it as an S one. A recent, non-exhaustive survey of these findings has been given by (Blythe and Croft, 2012). In the following, we will briefly look upon a few works not covered by this review, especially in relation with the Constant Rate Hypothesis.

Empirical support for the Constant Rate Hypothesis

The paper of Kroch (1989b) met a rather consensual approval from the community, though it was at odds with the opinions previously held. Far from being challenged, the Constant Rate Hypothesis quickly became the new dogma. Though one might regret that the innovative conceptions of Kroch did not lead to a more empirical discussion of hypotheses in language change and were reduced to an unchallenged certainty on the Constant Rate Hypothesis, the thrive to support it fortunately led to quantitative corpus-based accounts of the S-curve in language change.

One of the first papers to apply this method (Santorini, 1993) focuses on the syntactic change of the inflexion from medial to final in Yiddish. Data is provided from the fifteenth century to the nineteenth, with a time window of fifty years. According to the authors, the S-curve is to be found and the rate is the same in the two linguistic environments under scrutiny (simple verbs and complex verbs). However, we were not able to reproduce the same result on the basis of the data given in the paper (the given slopes are respectively 1.01 and 1.19 in the first and second environment). Furthermore, we show a strong dependence on the way the logit is applied. Indeed, the logit function diverges for 0 and 1; whenever the frequency reaches either one or the other value in empirical data, one has to decide how to treat the datapoint. There are two straightforward ways to do so: one can either remove this point, or assign to it a frequency close to 0 (resp. 1) such as 0.01 or 0.001 (resp. 0.99 or 0.999). The variation of the slope between these three procedures (remove, assign 0.01/0.99, assign 0.001/0.999) is actually greater than the statistical error of the linear fits (Table 3.1).

One can note that in our three procedures, the slope is greater in the first environment than in the second, while it is the opposite in (Santorini, 1993). In any case, we have found a rather poor linear fit (the r^2 parameter is at best equal to 0.93 for the first environment and 0.88 for the second).

One might also, as I suggested in the previous section, take into account the fact

X	Slope 1	Slope 2
Remove	1.52 ± 0.49	1.40 ± 0.36
Assign (10^{-2})	2.08 ± 0.59	2.13 ± 0.42
Assign (10^{-3})	2.65 ± 0.96	2.93 ± 0.65

Table 3.2: Different values of the slopes of the two changes studied in (Santorini, 1993), according to the way the extreme values 0.0 and 1.0, producing divergence in the logit transformation, are dealt with. The logit transform has been modified so as to account for a possible non-zero initial frequency of the new variants.

that the curves do not start at zero frequency. In this case, the Constant Rate Hypothesis seems more closely corroborated by the data and the slope can be greater in the second environment (Table 3.2). However, we do not find the same numbers as (Santorini, 1993) (slope of respectively 0.519 and 0.525 for the first and second environment). No matter what, it shows that these numbers are to be taken with caution and that the Constant Rate Hypothesis should not be considered valid unless tested on more reliable data.

One extreme example of this blind trust in S-curves and logit transform to assess the relevance of the Constant Rate Hypothesis is to be found in (Pintzuk, 1995). In this paper, data are obviously not compatible with a sigmoid function (Fig. 3.24). Yet, the authors apply a logit transform on these two ‘sigmoids’ and find the corresponding slope of its linear fit. Once again, I have been unable to reproduce the slope values on the basis of the data they give; what is more, the slopes obtained from the logit transform lead to a significantly higher value in the first environment than in the second one, no matter how the extreme values were dealt with. Pintzuk (1995) find that both slopes are equal, a conclusion whose relevance can be questioned on the ground that 1 - the data obviously does not conform to a sigmoid, and 2 - it is not easily reproducible and heavily depends on the specific treatment of extreme values.

We can also note that the unusual interpretation of the intercept is reassessed here, where the initial time is set at the fourth century, and the intercept interpreted as reflecting a non-zero ‘input’ frequency of the new form in each environment, at that time $t = 0$.

Anyway, these two studies show how well accepted the Constant Rate Hypothesis was by that time. Actually, it quickly became, rather than an empirically-grounded fact, something akin to a theoretical principle, whose violation was indicative not that the hypothesis was not supported by the data, but that some other linguistic phenomenon was interfering with these data:

This divergence is contrary to the Constant Rate Hypothesis, which we have proposed and which is supported by several other studies, according to which different linguistic contexts in which a single grammatical change is manifested should change at the same rate. This violation of the Constant Rate Effect leads us to wonder whether the grammars in play after the mid-16th century are the same as those in play earlier. (Kroch, 2006, pp. 12-13)

Oddly enough, the case discussed in (Kroch, 2006) is the same as in (Kroch, 1989b), with the same data taken from Ellegård.

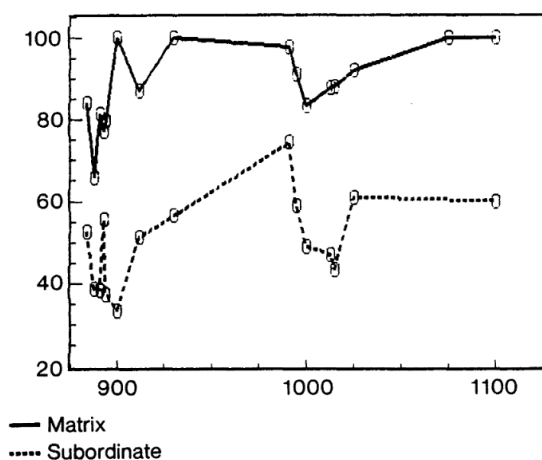


FIGURE 2. Frequency of *INFL*-medial versus *INFL*-final structure in Old English clauses with auxiliary verbs (884-1100).

Figure 3.24: Figure 2 of (Pintzuk, 1995, p.247).

Resurgence of lexical diffusion

In a completely different line of research, we also find an interesting and worth-mentioning resurgence of the lexical diffusionists' ideas in the field of corpus-based studies, starting with the work of Ogura and Wang (1996). Interestingly, one of the two authors, William S.-Y. Wang, has the paternity of the notion of lexical diffusion (Wang, 1969). As for the title, 'Snowball Effect in Lexical Diffusion', it is a direct reference to a metaphor originally used by Weinreich et al. (1968) to describe the S-curve. Ogura and Wang (1996), studying the emergence of the new morphological marking of third person singular in the verbal system of English, have tried to disentangle the diffusion from speaker to speaker ('S-diffusion', not to be confused with the S of the S-curve common to all kinds of diffusion, according to the authors) and the lexical diffusion from word to word ('W-diffusion'), to which they also add a diffusion from dialect to dialect. However, this work, while theoretically innovative, is a huge step backward in terms of empirical analysis: the 'S-curve' nature of the change is mainly reflected by the fact that, at each time period considered, and for each linguistic environment, the proportion of the novelty is either low or high, and never to be found in-between. The data provided, though collected with great care, does not allow to draw any firm conclusion regarding the sigmoidal nature of the curve or the Constant Rate Hypothesis.

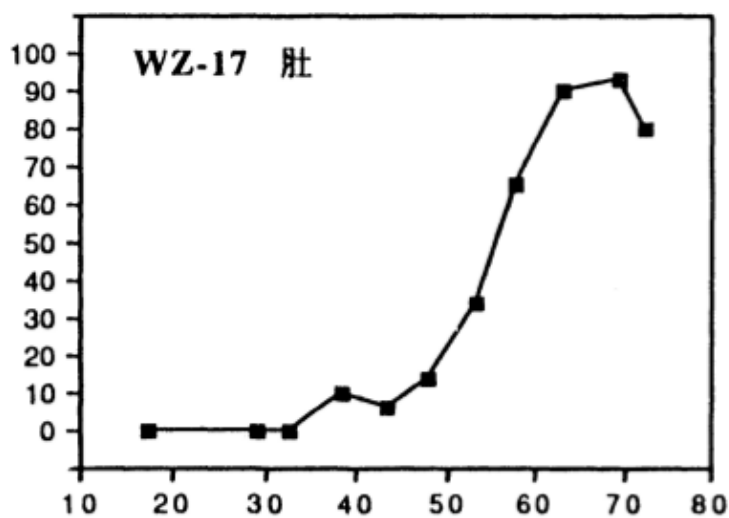
A perhaps more interesting approach is that of Shen (1997). In an extensive monograph, he explores the notion of Lexical Diffusion, and, with reference to the S-curve, claims that there are two ways for the lexical diffusion to occur. The first one would be Kroch's Constant Rate Hypothesis: a phonetic change affects different words at the same rate, but at different times. Therefore, if all changes are described by the linear fit of the logit transform associated with the S-curve, they would all have the same slope, but different intercepts. The alternative hypothesis is that of Bailey (1973),

which Kroch (1989b) explicitly opposed; it states that contexts affected later will also be affected faster. In short, the rate of change accelerates as a greater number of words comes to be converted. The corresponding quantitative signature would be a positive correlation with the slope, and the intercept, since the latter, as we have seen, can be interpreted as a time at which the change occurs (more precisely, the time at which the change reaches its maximal intensity, i.e. the time at which the S-curve reaches its inflexion point).

To decide which of these two hypotheses account best for linguistic change, Shen investigated a phonological change in Shanghai and Wenzhou. In line with the Sociolinguistics method of gathering empirical data, he proposed lists of words to be read by speakers. From this data, he sorted the speakers by age, and for each different word, for each age class, considered the number of speakers having adopted the change. From these percentages, he was able to draw S-curves, an example of which is given on Fig. 3.25a. Admittedly, this has nothing to do with corpus data and this work rightly belongs to the Sociolinguistics tradition, yet it was more meaningful to present it along the other works on the Constant Rate Hypothesis.

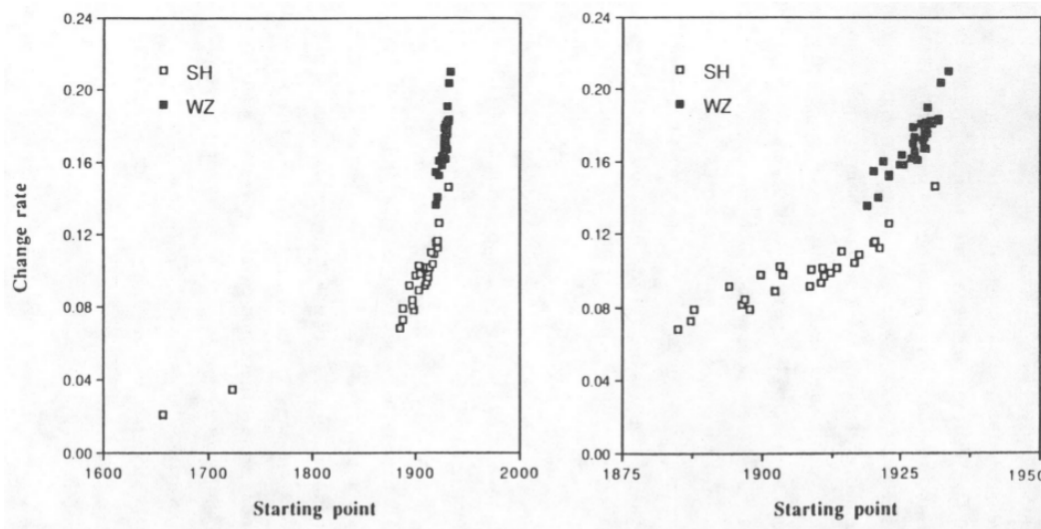
Indeed, Shen applies to these S-curves the same procedure as Kroch did, and compares, for each context of diffusion (i.e., for each word), the slope (rate) and the intercept (time) of the change. It appears that they are strongly positively correlated (Fig. 3.25b), as would be predicted by Bailey's hypothesis. This would be a clear empirical invalidation of Kroch's Constant Rate Hypothesis. However, it should be noted that both works are addressing very different kinds of change, and in very different ways. Kroch (1989b) investigates syntactic change, based on corpus data. Shen (1997) focuses on phonetic change, and favors a Sociolinguistics approach of the phenomenon. It may well be then that the two S-curves of change are associated with different kinds of phenomena. Indeed, corpus data allows for the possibility that speakers alternate between variants, and that they all do so in a more or less similar way at a given moment in time. In Sociolinguistics, informants make binary choices between variants. The first S-curve would therefore describe a diffusion within language use, the second, within a community of speakers. There is no need for these two phenomena to be equated. Thus, one can follow the Constant Rate Hypothesis, and the other, Bailey's rate acceleration law.

As it turns out, the evaluation of the Constant Rate Hypothesis was at that time hindered by the scarcity of available corpus data. It is only with the rise of digitalized corpora, in the 2000s, that convincing analytical undertakings of this question would have become easily tractable. However, it does not seem that these new technical opportunities led to a reassessment of Kroch's initial insights, nor to a significant large-scale analysis of various cases of language change. Most of the next-generation corpus studies — e.g. (Nevalainen and Raumolin-Brunberg, 2003; Fagard and Combettes, 2013) —, while broadly supporting the S-curve picture, remained qualitative in that matter: whereas the data provided was reliable and could have led to such an inquiry, no further mathematical analysis was performed and the confrontation to the Constant Rate Hypothesis left pending.



(a)

Figure 5-5. A comparison of the two changes. Fig. 5-5b is a detailed plot of the major clusters in 5-5a. The x-axis is the change rate and the y-axis is the starting point.



(b)

Figure 3.25: (a) Figure 4.4e of (Shen, 1997, p.101). The x-axis corresponds to the 'birth year' of the different age classes (i.e., the higher this number, the younger the speakers of the age class). The y-axis shows the percentage of speakers having adopted the phonetic change for the particular word coded WZ-17. (b) Figure 5.5 of (Shen, 1997, p.129) showing the correlation between the rate of change and the time at which it occurs.

3.4 The S-curve challenged

The story of the S-curve led from the expected pattern of the social diffusion of a new variant, to the empirically based template of frequency changes in cases of linguistic replacements. However, as its dominion expanded, discontent also stirred and grew, and soon its absolute sovereignty as the sole and unique pattern of language change would be questioned.

3.4.1 Theoretical challenging

Interestingly, these doubts did not first come from corpus-based studies, exhibiting counter-examples. As we saw throughout this history, the S-curve was never explicitly and firmly confirmed by the data (with the possible exception of the work by Altmann et al. (1983) which displays a remarkably fit S-curve): It was a theoretical belief. It is thus not too surprising if the discrepancy did not arise from empirical evidence, but from mathematical modeling of language change, starting with (Niyogi and Berwick, 1995):

[I]n contrast to Kroch (1989) and others, who mimic population biology models by imposing an S-shaped logistic change by assumption, we explain the time course of language change, and show that it need not be S-shaped. Rather, language-change envelopes are derivable from more fundamental properties of dynamical systems sometimes they are S-shaped, but they can also be non-monotonic.

(Niyogi and Berwick, 1995)

Their model lies explicitly on the Generative Grammar framework: Language is a set of Boolean parameters whose generations successively learn by inferring the parameters from a set of sentences. As each language draws sentences from the same ensemble, there is some overlap, in terms of possible sentences, between the different languages. The particular learning algorithm, as well as these overlaps, determine transition probabilities from one language to another. Hence, syntactic change is seen as a parameter change, guided by those biased transition probabilities. This mathematical framework leads to the depiction of the dynamics by an S-curve, *if* the proper conditions are met (here what distinguishes the different curves is the strength of an initial perturbation initiating the transition from one language to another). The S-curve is then but one possibility among a family of curves (Fig. 3.26).

The authors claim that their mathematical findings concur to the loss of the verb-second syntactic trait in Old French, yet they do not provide any empirical data to support this claim. We do not know then which curve would appropriately fit the data (though the assumed complete disappearance of this trait would be more likely compatible with $p = 0.75$). The approach here is then very different from Kroch (1989b), who tried to propose a parametric model of the frequency pattern of change, so as to get information about the change from the fitting of these parameters — a difference which the authors themselves do not point out:

Diachronic envelopes are often logistic, but not always. Note that in some alternative models of language change, the logistic shape has sometimes been assumed as a starting point, see, e.g., Kroch (1982, 1989). (Niyogi and Berwick, 1995)

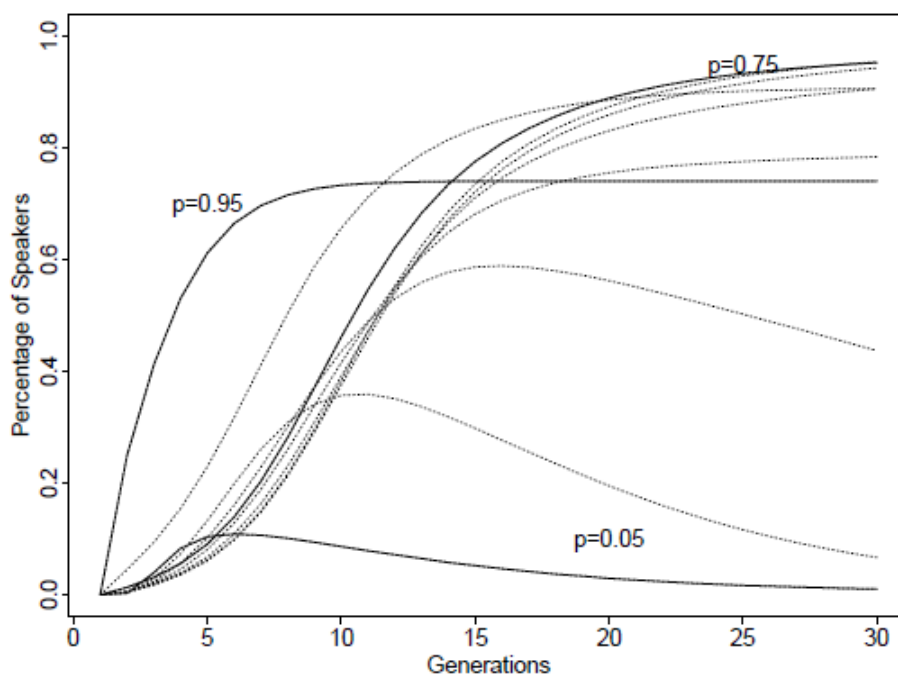


Figure 5: The evolution of V(erb) O(bject) S(ubject) +Verb second speakers in a community given different sentence distributions, P_i 's. The P_i 's were perturbed (with parameter p denoting the extent of the perturbation) around a uniform distribution. The algorithm used was single-step, gradient ascent. The initial population was homogeneous, with all members speaking a V(erb) O(bject) S(ubject) – Verb second type language. Curves for $p = 0.05, 0.75,$ and 0.95 have been plotted as solid lines. If we wanted the population to completely *lose* the Verb second parameter, the optimal choice of p is 0.75 (not 1 as expected).

Yet, Kroch did not take the S-curve as a ‘starting point’, but as a convenient enough tool to investigate quantitatively language change. He did not, furthermore, try to provide a model of language change, so that these two works should not be compared as they are in (Niyogi and Berwick, 1995).

The next line of critics targeting the S-curve took a completely different approach, and is due to Denison (2003), a researcher in Historical Linguistics. In his paper, he humorously questions the relevance of the S-curve and discusses it with great theoretical clarity; it may well be, incidentally, one of the best introductory paper on the question. He starts by explaining what an S-curve is, using a specific example (the transition from passival to passive in 18th century English); then he tries to provide a plausible theoretical yet very qualitative account of the S-curve, by mentioning some kind of a ‘natural selection’ between the competing variants, assessing that “[t]here would be no change at all unless there were some small advantage in the new form, whether structural or social.” (Denison, 2003, p.58). This idea is interesting and we shall see, in section 6, that recent modeling works on the S-curve goes along that way too.

The most interesting part, however, comes when Denison expresses his concerns on the S-curve:

However, not everything about S-curves is clear to me. I want now to turn to the vertical scale in those hypothetical graphs. What are the percentages percentages of?
(Denison, 2003, p.59)

Indeed, we have seen that the S-curve can instantiate many kinds of change: Lexical diffusion, Social diffusion, or merely the empirical frequency-patterns derived from a corpus study. Even in corpus studies, it can be the proportion of uses compared to another variant, or a raw frequency of use (which then does not add up to 100% as we have no clue whether the new variant has replaced all its competitors in all contexts or not ; in this case the notion of percentage has no longer any relevance). This delicate question of the competitor is as well addressed by Denison:

I do have other problems with S-curves. The original justification for them was the ecological competition between two variants competing to perform the same linguistic function. Immediately we have problems with sound changes in the vowel space, where at the phonetic level at least there is the possibility of continuous variation and hence no two competitors. But even with the syntactic phenomena I am more at home with, it is a rarity to find competition between one old form and a single replacing form.
(Denison, 2003, p.65)

We shall briefly return to this particular point later on in Chapter 9. Let us note for now that Denison, thanks to his linguistic insights, brings forth an idea which may be quite enlightening on this topic, that is, that the number of competing variants may depend on the context.

Due to this diversity of S-curves (which, apparently, was never stressed with such awareness before), Denison concludes in a somehow provocative way:

The catch-phrase ‘slow, slow, quick, quick, slow’ refers, I gather, to the fox-trot. Here’s what one of the OED’s citations has to say about that dance: “The Fox-Trot is a dance of many steps, and to the casual observer everybody seems to have

different ones.” (1919 [OED]) Much the same goes for S-curves and the scholars who draw them. The S-curve is neither as simple nor as uniform a phenomenon as is sometimes assumed. Given too the simplistic picture of variation it sometimes reflects (and requires), the S-curve should not be seized on too readily as the general shape of language change. (Denison, 2003, p.68)

This statement recognizes at last that the theoretical domination of the S-curve is not grounded on any true theoretical consensus (though the apparent one is strong, once we dismiss all the crucial differences between all possible interpretations of the S-curve), nor on any firm empirical basis. We may wonder at this point whether the S-curve is interesting at all and does constitute a phenomenon as such.

3.4.2 Other templates for change

The idea, put forward by Niyogi and Berwick (1995), that the S-curve might not be the one and only pattern of language change, has resurfaced in the late 2000s from a different, unrelated approach, this time based on quantitative analysis of empirical corpus data. Kallel (2007) explores the loss of negative concord in Early Modern English, and shows the decrease of this form in six different successive ‘stages’ (successive time windows of 25 years). In doing so, he performs the same logit transform as Kroch (1989b), yet his way of assessing the relevance of the Constant Rate Hypothesis is slightly different. Instead of comparing the value of the slopes, he determines, through a statistical procedure, that the rate is independent from the different contexts. Furthermore, in his case, the change characterized is a true competition, with data for both competitors.

However, though Kallel finds that the Constant Rate Hypothesis is corroborated—a result which we confirm using Kroch’s method (Fig. 3.27)—he is not satisfied with the quality of the linear fit of the logit. He proposes then to fit his data with a quadratic function, which he finds to be much more adequate. However, fitting six data points with a 3-parameters mathematical function is certainly an instance of overfitting, and henceforth does not seem to be justified. Plus, whenever one increases the degree in a polynomial fit, then the fit is bound to become better; the fact that it does so is not a sufficient argument in favor of this second model.

Anyway, the author finds a most striking conclusion (Fig. 3.28a): the curvatures of the fits is the same for the different contexts through which they track the change. They thus propose to replace the Constant Rate Hypothesis (which they treat as a ‘null hypothesis’ since the rate is constant through time) by a ‘Context Constancy Principle’:

On modelling this curvature, more significant chi-square values were obtained and the model clearly indicates a better fit for our data. Although the figures from the second model now illustrate curves, rather than straight lines, the findings uphold Kroch’s constancy effect across contexts. The new model and the logistic curves indicate the same pattern of decline of NC in all contexts; this is demonstrated by the parallel curves obtained, something we shall refer to as parallel curvature. [...] Accordingly, we would like to argue that what constitutes a principle of linguistic change is not constancy across time but constancy across contexts. Based on this distinction, we shift focus to contexts’ similar behavior as the key issue in linguistic changes. (Kallel, 2007, p.45)

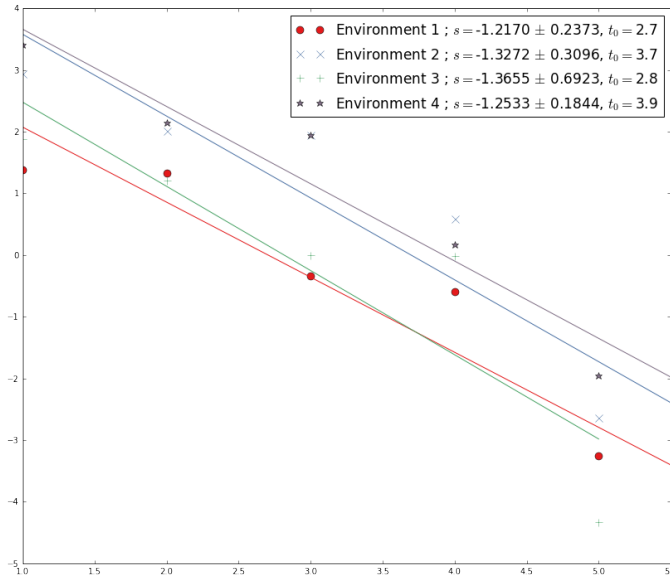


Figure 3.27: Linear fits of the logit transform of the data from Tables 3-6 of (Kallel, 2007).

We were not able to reproduce this finding (Fig. 3.28b).

In a similar spirit, Gabriel Altmann proposed much earlier, in 1983, to extend the initial view of the sigmoid to account for more diverse empirical evidence (Strauss and Altmann, 2007), starting with a general time evolution equation for the proportion p_t of the new variant, so as to derive a family of possible mathematical curves :

$$dp_t = k_t p_t (C - p_t). \quad (3.10)$$

In a recent and short survey of these findings, Strauss and Altmann (2007) discuss several cases; if $C = 1$ and k_t is a constant, then one finds a sigmoid whose rate is given by the constant. They then propose C to be different from 1 so as to account for cases in which there is no clear competitor so that the proportion of occurrences of the new variant comes to a saturation value C smaller than 1. Finally, they make a suggestion akin to the one of Kallel (2007), with $k_t = A - Bt$. They call this third case a ‘reversible change’ as the curve is non-monotonic, so that the proportion of the new variants decrease after having reached a peak.

For each of these changes, they provide one example taken from empirical data, usually of good quality (e.g. 26 data points to illustrate the reversible change). It appears that the typology of changes they propose is fully justified; the reversible change, for instance, strikingly deviates from a sigmoid (Fig. 3.29) and it would be pointless to force such a fit. On the other hand, one might always argue that this curve is made of two sigmoids whose rates are of opposite sign, so that the proportion of the novelty starts to decrease once it comes to be itself replaced. The ‘reversible’ change could thus be interpreted as a confluence of two subsequent, overlapping changes.

However, the authors do not justify these three different forms of the k_t term in the evolution equation, and do not explain why it would change from one case to another.

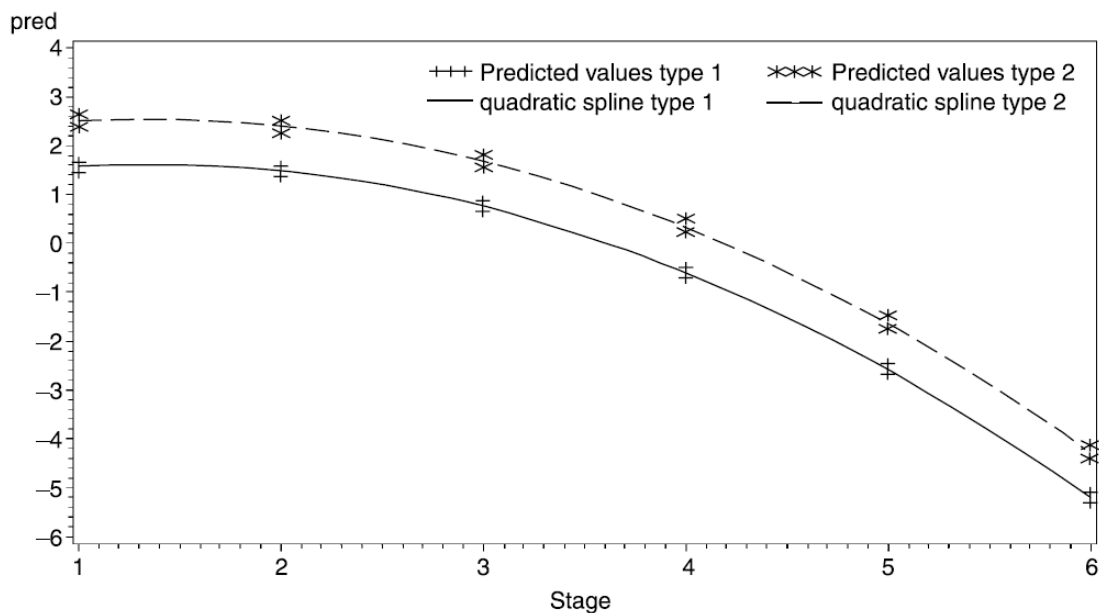
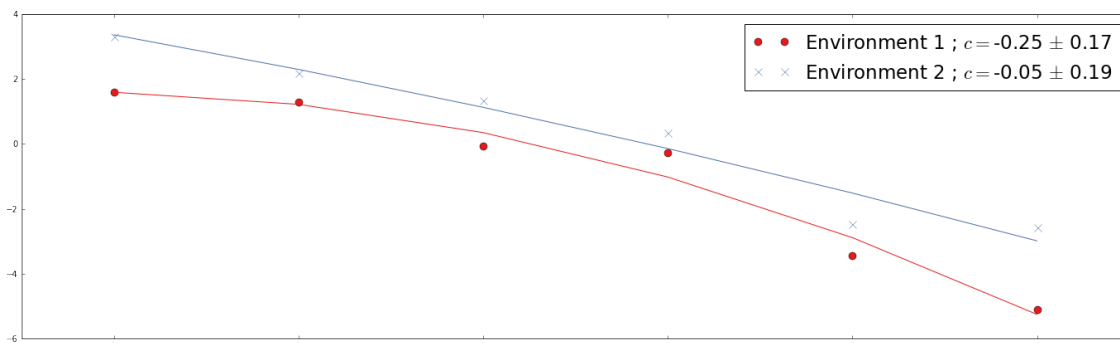


FIGURE 9. The plot of data for noncoordinate and coordinate contexts.

(a)



(b)

Figure 3.28: (a) Fit of the logit transform by a second-order polynomial; Figure 9 of (Kallel, 2007, p.42). (b) Our attempt to reproduce this result with the data provided in Tables 1-2 of (Kallel, 2007). Parameter c is the coefficient of the second-order term, and therefore quantifies the curvature.

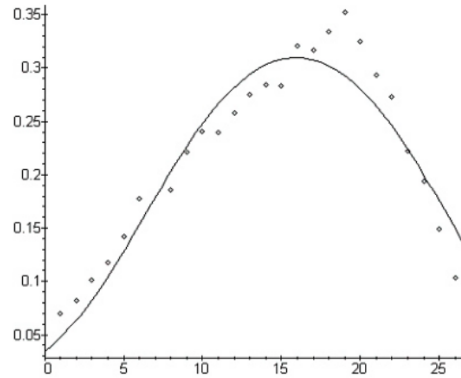


Fig. 3. Fitting formula (4) to data in Table 3

Figure 3.29: Figure 3 of (Strauss and Altmann, 2007), adapted from (Best et al., 1990).

Also, the coefficients A and B in the reversal change do not receive any interpretation.

What these two works commonly show is that a language change, as Niyogi and Berwick (1995) said, does not need to be sigmoidal. The S-curve is only one kind among a whole family of curves. This immediately raises a handful of crucial questions: What would be this family? Which are the mechanisms underlying it? What governs the onset of one type of change rather than another? What does it tell us about the change? We shall see very soon one successful example of such an approach.

3.4.3 A Lotka-Volterra model of competition

We should, before this, turn towards another attempt to go beyond the simple picture of the S-curve, due to Grieve-Smith (2009). In this work (a Ph.D thesis), Grieve-Smith notes that the S-curve is not neutral in its depiction of the change as it focuses on the rise of one novelty, without taking into account its competitor. It is true that, in the case of the dynamical system often used to support the sigmoid, there is an explicit asymmetry between the two variants : if the novelty has a non-zero frequency, it is bound to rise while its predecessor must eventually fail and fade away. In some sense it is an inescapable eviction rather than a competition, and the old variant has clearly no active role in the process.

This is why Grieve-Smith proposes to use a Lotka-Volterra equation to model this competition. He makes use of the following set of equations to describe the evolution of the proportion of variants a and b simultaneously (respectively x_a and x_b):

$$\begin{cases} \dot{x}_a = x_a(1 - x_a - \alpha_{ab}x_b) \\ \dot{x}_b = x_b(1 - x_b - \alpha_{ba}x_a) \end{cases} \quad (\text{Grieve-Smith, 2009, p.99}). \quad (3.11)$$

However, if we take into account that one must always have $x_a + x_b = 1$, then the

first equation turns to :

$$\begin{aligned} \dot{x}_a &= x_a(1 - x_a - \alpha_{ab}(1 - x_a)) \\ &= x_a(1 - \alpha_{ab} - (1 - \alpha_{ab})x_a), \\ &= (1 - \alpha_{ab})x_a(1 - x_a) \end{aligned} \tag{3.12}$$

which is exactly the usual equation:

$$\dot{p} = rp(1 - p), \tag{3.13}$$

where the rate r is simply equal to $1 - \alpha_{ab}$. This provides, actually, a nice interpretation for the rate (and explains why the evolution would favor one variant rather than the other in terms of these competing coefficients), but it is not an alternative to the usual logistic framework.

3.4.4 Information encoded in the S-curve

We will now turn to one work which goes slightly beyond the S-curve and explicitly presents it as one outcome among a family of possible ones. In this work (Ghanbarnejad et al., 2014), the authors propose to add a constant term in the evolution equation of the proportion ρ of speakers having adopted the new variant at time t :

$$\dot{\rho} = (a + b\rho)(1 - \rho). \tag{3.14}$$

There does not seem to be much difference with the previous situation. However, a and b receive clear interpretations in this sociolinguistic setting: a and b respectively weight the exogenous factors of change (if positive, it biases towards the novelty no matter how much speakers have adopted it), and the endogenous ones. If a is 0, then we go back to the traditional sigmoid. If b is 0, then we have a linear increase at first followed by an exponential saturation. Interestingly, a being infinitesimal is a sufficient condition for the new variant to rise even with zero initial frequency: There is no stable fixed point for $\rho = 0$ unless a itself is 0.

One of the most interesting features of this proposal is that this simple model can be directly used to fit empirical corpus data (their own data is obtained from the Google Ngram database) and to assess, for each of the studied changes, whether endogenous or exogenous factors explain most of the change (Fig. 3.30). They thus show that, in the case of orthographical reforms in German, the change is completely exogenous; while in the regularization of the preterit for English verbs, it is mainly endogenous. Though these results are not surprising, they are remarkably consistent and corroborate the relevance of the model. This method could then be efficiently used to probe older changes, for which we do not know with such certainty whether there was an external force (e.g. political) behind a language change or if it was spontaneous.

This paper unites two remarkable features that we already encountered: like the work of Kroch (1989b), they use the S-curve as a tool to extract information on a given language change; and they do not regard the S-curve as the only possible pattern of the change, but instead, show that it is one member of a larger family. Furthermore, they provide a detailed and quantitative answer to all the questions we asked previously: the mechanism underlying the extended S-curve is social mimicry ($b\rho$ term) with an

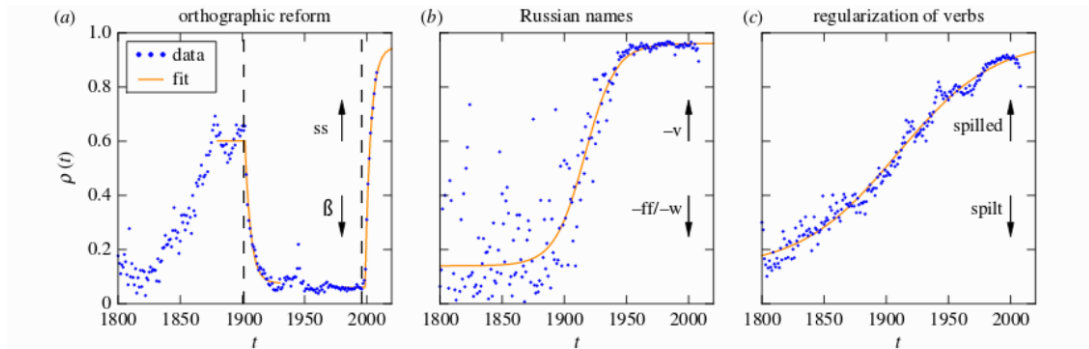


Figure 1. Examples of linguistic changes showing different adoption curves. We estimate the fraction of adopters $\rho(t)$ by the relative frequency as $\rho(t) = \sum_w n_1^w / \sum_w \sum_q n_q^w$, where n_q^w is the total number of occurrences (tokens) of variant q for the word w at year t . (a) The orthography of German words that changed to 'ss' ($q = 1$) from 'ß' ($q = 2$) in the orthographic reform of 1996 (many words changed from 'ss' to 'ß' in the 1901 reform). (b) The transliteration of Russian names ending with the letter 'b' when written in English (Latin alphabet), changed to an ending in 'v' ($q = 1$) from endings in 'ff' ($q = 2$) or 'w' ($q = 3$) (e.g. $w = \text{'Caparoe'}$ is nowadays almost unanimously written as 'Saratov', but it used to be written also as 'Saratoff' or 'Saratow'). (c) The past form of the verb spill changed to its regular form 'spilled' ($q = 1$) from the irregular form 'spilt' ($q = 2$). The light curve shows the fit of equation (1.2). The estimated parameters a and b are (a) $\hat{a} = 0.218$, $\hat{b} = 0.000$ in 1901, and $\hat{a} = 0.229$, $\hat{b} = 0.000$ in 1996; (b) $\hat{a} = 0.000$, $\hat{b} = 0.099$; and (c) $\hat{a} = 0.001$, $\hat{b} = 0.030$. The corpus is the Google-ngram [3,4] plotted in the minimum (yearly) resolution, see the electronic supplementary material, section I, for details on the data and section IIIB for details on the fit. (Online version in colour.)

Figure 3.30: Figure 1 of (Ghanbarnejad et al., 2014).

external bias (a term), the specific member of the family would be chosen according to the relative weights of these two terms, and it tells us whether the change is externally imposed or not.

Of course, we might not agree with the interpretation provided (and we certainly will not), notably because, as we saw throughout this extended review, the ρ quantity is not necessarily interpreted as a number of converted speakers: it can also reflect lexical diffusion, or some cognitive switching towards the new variant. At the very least, it reflects the evolution of a number of occurrences in a corpus: it is an empirical quantity, whose theoretical meaning is not granted. In that sense, it becomes much less clear whether the a and b parameters can still be interpreted as 'exogenous' and 'endogenous' factors or not. Yet it matters not. This analytical proposal can be used as an efficient criterion to empirically classify different instances of language change (the a -based and the b -based); it yields a new typology of phenomena, and as such, it is clearly of uttermost interest and can serve as a basis to devise and test further hypotheses on one particular change.

3.5 Summary

The story of the S-curve, as we have seen it, is a twisting one. It was first introduced as the common pattern of the social diffusion of a language change, as an analogy to many other cultural or technological changes. It then became a 'principle' of language change, theoretically assumed, but never empirically confirmed. In the meantime, the diffusion space of a language change was discovered to be more complex than was first thought, as a lexical diffusion was happening alongside the social one, and it was not clear anymore which of these diffusions the S-curve was accounting for. The wave-based nature of the social change also allowed for new interpretations of the S-curve,

where it was seen as a synchronic pattern in space. The idea of an apparent time also emerged, reflecting the idea that speakers were frozen the state of language as it were by the time of their childhood -speakers of different ages being then seen pretty much as living fossils of a language past. Theoretically, S-curves were ubiquitous; yet, no empirical work supported them in a clear, definitive way.

By the 1980's, S-curves were used to mathematically fit empirical data extracted from corpus, and the logistic function was proposed, mainly for its simplicity of use, but also for its relationship with simple dynamical equations of competition. The S-curve was then regarded as the sign of an ecological competition between two linguistic variants, in a view which was at odds with its sociological debuts. One cannot dismiss this discrepancy: as the S-curve was first theoretically assumed through a legitimate analogy, it was emerging in a way which was unexpected and no longer in line with this analogy. It became a phenomenon within the language itself, rather than the social realization of an arbitrary language change.

This empirical, analytical perspective on the S-curve also allowed to interpret its parameters and use these results to discuss hypotheses. Unfortunately, the hypothesis came to be seen as theoretically granted, and was often forced onto the empirical evidence. Kroch's Constant Rate Hypothesis, though innovative in its time, inhibited the theoretical understanding of the quantitative features of the S-curve. By this time, some empirical works started to hint that the S-curve was perhaps not the unique pattern of language change, which nonetheless did not lead to new theoretical insights on change. As Denison (2003) has shown, the S-curve was wrapped in different layers of confusion, especially since the data obtained from corpora was neither reflecting a lexical diffusion, nor a social one, but more probably a puzzling entanglement between the two—hence the ecological view of the language which was getting rid of the question of the diffusion space. Recently, the S-curve has also been questioned through modeling works, which we shall review more specifically in an oncoming chapter 6. Interestingly, those three approaches -modeling, empirical analysis, theoretical understanding of the parameters- have been tied in in a recent attempt to weight, for a given instance of a language change, the importance of the endogenous and exogenous factors behind it.

This survey has successively shown that the S-curve evolved from a theoretical, social phenomenon, to an empirical, linguistic one, and finally to a modeling tool among many used to fit data, but also to extract information from it. We can now wonder whether the S-curve is a phenomenon as such, whether it is worth mentioning, whether it belongs to the study of language change specifically. My own position towards these different concerns must now be made explicit.

The S-curve is not interesting because it is an S, but because it is a curve. It shows that there is some persistent consistency in a language change, a consistency which spans the whole duration of a change, which can last decades, even centuries. This consistency cannot be explained by sociolinguistic effects alone; we would expect more variation, and a shorter consistency length. Nor can it be explained by structural effects alone; though changes are often interrelated, many of them also occur independently and seemingly indifferently from what happens next in the rest of the language. In that sense, the specific shape assumed by a frequency change -which is the most salient of changes in corpus studies, and the one who most conveniently allows for quantitative analysis- is not of uttermost importance; the very existence of such a shape is.

On the other hand, we should expect a huge diversity of patterns. We can think of many different situations for a change, of many set-ups which would lead to very different mathematical curves. Changes may be interfering with each other, they may be influenced and biased by an external influence, they may or may not meet appropriate sociological conditions to unfold. And yet, as we shall see quite soon, the S-curve is pervasive. In the light of how messy language is, of how diverse the changes are, this regularity is striking. It is indicative of something which has escaped understanding so far. To understand the S-curve, I believe that one should first and foremost understand the mechanisms of the change. In trying so, the S-curve is, in some way, a challenge, a standard to meet. We can devise hypotheses and theories; but if they don't lead to an S-curve, they cannot pretend to be relevant.

Yet, we shall also see that a good diversity of models manage to capture the S-curve. This also is to be expected, as the S-curve, though non trivial, still remains quite generic. A great number of processes, of mathematical functions, can fit a roughly S-shaped curve. In some sense, the S-curve is the result of the most basic non-linear process of competition, presenting a bias towards novelty, and a saturation mechanism. This is exactly what is captured by the simple dynamical equation of which the logistic function is a solution. However, neither the bias nor the saturation have received definitive interpretations in a cohesive linguistic framework of language change. Any model with those three ingredients — non-linearity, edge for the new variant, saturation — is expected to present an S-curve, be it logistic or not, as the regular pattern of language change. However, not all models are equally fit to give a convincing representation of language use.

The S-curve as an empirical observation is thus a crude one and has to be completed. It will be our goal, then, to take the measure of the S-curve, so as to see which S-curve really appears in language change, and if this is the same as the one which appears in cultural, technical diffusions among the different members of a society. Only a broad, statistical and quantitative analysis of the S-curves of language change can answer this challenge. And yet, chasing the S-curve, it is easy to dismiss other important empirical evidence. This is why we will try to identify, as well, other empirical regularities, whose shape might be less striking, but their importance much equal.

Chapter 4

Extracting patterns from frequency profiles

Language change is characterized by an S-curve; or so it is claimed. As we reviewed in the previous chapter, the S-curve is more a shared and consensual theoretical belief than a carefully grounded empirical fact which could be safely considered as universal and symptomatic of the change. True enough, so far most of the cases of language change investigated from a quantitative perspective have shown that the S-curve was a good fit of the data. Yet, the S-curve suffers from a serious lack of agreement concerning which phenomenon it is supposed to represent.

For instance, the S-curve represents an increasingly affirmed presence of a novel linguistic variant, presumably with time, but even this last point is not clear, as some authors are making use of the concept of represented time, while others still speak of a spatial diffusion. We saw that the quantity which is increasing during the S-shaped growth is far from being a settled matter, either. Does it represent a proportion of converted speakers to the novel variant? An opaque evolution within the language itself, quite independently of the users? Or a lexical diffusion of some change?

In the context of semantic change — a family of changes which happen to be also characterized by an S-curve for the most part, see for instance (Kroch, 1989b) or (Fagard and Combettes, 2013) — the idea of a lexical diffusion is not so straightforward anymore (we can imagine a diffusion across different contexts of use, though, which would be far from absurd). Inter-speaker variation is also a bit odd in the case of semantic change, for it would mean that some speakers use a given word with a particular meaning, and some others use it with another meaning. It would imply a high level of misunderstanding within the community during the time of the change, a misunderstanding which can of course play the role of an explaining mechanism for the fast growth in the mid-part of the S-curve, yet this mid-part can last for close to a century in some cases and it seems difficult to assume that the community would be split in two during such a long period of time.

In this chapter, we will adopt an empirical perspective on the question. First of all, we will try to assess the robustness of the S-curve pattern in the accounting for diachronic data, obtained from textual corpora. However, we do not know a priori, except from a linguistic analysis of all utterances, at which time exactly a change

starts, and at which time it can be said to end. Hence, the search for patterns need to rely not on a specific time period, in line with the concerns raised by Gries and Hilpert (2008), and must be automatized. This extracting pattern procedure will be presented in details, and then applied to a wide array of cases of semantic change, based on data provided by the Frantex database. As we shall see, the S-curve is to be complemented by another phenomenon, which is a latency during which the frequency of the novelty remains low and stable, even if it is already attested in corpus. Finally, we will focus on two case studies, the same as those we discussed in chapter 2, to assess the relevance and the limits of this approach.

4.1 Extracting the S-curve from frequency data

Let us consider a time period T which covers the whole corpus, and divide it into L time windows t_i of equal duration Δt , such that $L = T/\Delta t$. In the COHA, this time window Δt is set to a decade by default, but one can achieve a resolution of one year. In Frantext, as well as in most corpora (Google Ngram, CORDE, COCA, etc.) the minimal resolution Δt is also of size one year. The total number of words in the corpus associated with the time window will be noted N_i . It should be noted at this point that Δt must be chosen in order to avoid the situation where $N_i = 0$.

A first concern may arise concerning the ‘size’ of each time window. Indeed, it might not be relevant to quantify it in terms of words. One might argue, for instance, that to study an n -gram, the size has to be evaluated in terms of how many n -grams are to be found. Yet, in a sentence of K words, should we count K/n n -grams, or $K - (n - 1)$, if we consider, as is the case in Google Ngram for instance, that the n -grams can overlap? For instance, the first sentence of this paragraph, composed of $K = 12$ words, displays 4 non-overlapping 3-grams and 10 different 3-grams in total. What is more, a form can be a n -gram in some constructs, and a n -gram of a different size in some others, (e.g. ‘out of his head’ and ‘out of a network of red veins’, both attested in the COHA, are two constructs of one and the same [out of {N}] construction, yet one is a 4-gram and the other 7-gram). What matters is only to provide a consistent evaluation of the corpus slices size over the different time windows. Hence, there is no valuable reason to argue against the choice of words count as a suitable measure of the size.

Frequency of a studied form is then given, for the time window t_i , by the ratio $x_i = n_i/N_i$ where n_i is the total number of occurrences of the form during the time window t_i . A closely related quantity is often defined, which is the number of occurrences of the form among an arbitrary number H of occurrences (e.g. number of occurrences per one million occurrences). This alternative measure of the frequency is simply given by $x_i * H$. In our own corpus studies we have set $H = 10^5$, yet as it is of no incidence on what follows, a new notation to distinguish the two different definitions of frequencies seems superfluous.

The simplest possible diachronic data provided by a corpus when studying a given form is thus the set $\{x_i\}_{i=1:L}$ of the raw frequency of this form in each time windows t_i .

4.1.1 Smoothing data

Very often, this frequency is too fluctuant and needs smoothing. A possible way to do so is to replace the frequencies x_i by their average on an window of size W , defining the smoothed frequencies y_i by:

$$y_i = \frac{1}{W} \sum_{k=\max(1, i-W+1)}^i x_k . \quad (4.1)$$

It follows that the difference between two consecutive data points, $y_{i+1} - y_i$, is given by the difference:

$$y_i - y_{i-1} = \frac{1}{W} (x_i - x_{i-W}) , \quad (4.2)$$

It thus does not modify too much the step-wise progression of the curve, only the time depth on which this progression is considered.

This method was already used by Best et al. (1990), using a window of size $W = 7$ over the data points. However, as the interval Δt covered by each data point is not given, we cannot know precisely over which total duration the average is performed.

One must decide how to treat the extreme values (for i between 1 and $W - 1$ in this particular proposal). A simple way would be to average over whatever is available, as we personally did. Hence, the first point would not be averaged, the second will be an average of the first and the second, and so on. We could also discard those points entirely.

Other possibilities of smoothing the curve by averaging are numerous. One can average over points on both sides, for instance two points backward and two points forward; or weight the different data points on which the average is performed given their distance to i , for instance:

$$y_i = \frac{1}{Z(W)} \sum_{k=i-W+1}^i \frac{1}{i-k+1} x_k , \quad (4.3)$$

with Z a normalizing constant depending on W .

There are many ways to perform such a weighting. The particular choice adopted in this work is an averaging window of size $W = 5$ and a smoothing described by (4.1), without any weighting. This particular asymmetry towards the past is supposed to reflect the fact that the community of speakers is still pretty much aware of the language produced in the recent decades. Of course, this averaging procedure introduces a bias, and gives more inertia to the process.

At this point, we define two quantities, S and M :

$$S = \sum_{i=1}^L y_i \quad (4.4)$$

and :

$$M = \arg \max_i (y_i) , \quad (4.5)$$

respectively the total weight of the form in the corpus and its peak of frequency.

4.1.2 Identifying growth periods

From this set $\{y_i\}$, it is possible to compute what we call ‘relative’ derivatives, d_i , in order to estimate the growth rate of the frequency:

$$d_i = \frac{y_{i+1} - y_{i-1}}{2y_i}, \text{ for } i \in [2 : L - 1] \quad (4.6)$$

and:

$$\begin{aligned} d_1 &= \frac{y_2 - y_1}{2y_1} \\ d_L &= \frac{y_L - y_{L-1}}{2y_L} \end{aligned} \quad (4.7)$$

This expression is chosen to be symmetric with respect to both sides of each data point. It differs from the usual discrete derivative by the $1/y_i$ factor. This latter trick is proposed to deal with the fact that the discrete derivative is proportional to the level of frequency, so that highly frequent words will be associated with high derivatives. Indeed, a growth from 1 to 101 between two datapoints has not the same significance as a growth from 10000 to 10100, yet the two are associated with the same discrete derivative. One could also have rescaled to divide the usual derivative by either S or M , so that all different cases of change would be set on similar grounds, but then different time periods of the same change will fail to be comparable.

We now have an estimate of the relative growth of the frequency of the form at each time period t_i . Next, we can set an arbitrary threshold above which the relative growth will be deemed as significant. In this work, this threshold is fixed at 0.15 (relative growth of 15%). We then look for period of repeated growth, which are sets of adjacent time windows such that the growth is not inferior to the threshold during more than two consecutive decades (two being chosen arbitrarily). This procedure allows then to identify the parts of the frequency pattern associated with a time-extended growth.

In order to capture the tails of the S-curve pattern, which are associated with a very low relative growth, we add, to these growth periods, all time windows comprised at a distance of 11 from the boundaries of the growth period (we could have chosen a smaller range than 11, but we wanted to be sure to catch the tails of the S-curve whenever applicable). For instance, if the growth period is $[51 : 57]$, the time period on which the pattern will be looked for will be $[40 : 68]$. Those ‘tail windows’ have deliberately chosen to be quite large: as the growth threshold is set up quite high, growth period will usually be limited to the growth peak of the S-curve, which covers only a limited width of the whole pattern. Moreover, we do not want to restrict a priori how extended the pattern can be.

4.1.3 Extracting the S-curve

Once those investigation periods T_g have been identified, we try to see if the growth can be described by an S-curve of equation:

$$\tilde{y}(t) = y_{min} + \frac{y_{max} - y_{min}}{1 + e^{-\alpha(t-t_0)}}, \quad (4.8)$$

where y_{min} and y_{max} are two data points corresponding to two decades t_{min} and t_{max} , both belonging to T_g and by which the pattern is bounded. We explain more below what are these values and how they can be obtained.

Note that the S-curve we defined is not classical (i.e. it does not go from 0 to 1), but a proper rescaling of \tilde{y} gives back the more familiar curve. Our proposal accounts for the possibility, first, that we do not know to which frequency would correspond ‘100 %’ in actual data, and second, that the ‘0 %’ of the process may not be associated with a null frequency of the form. Hence, this variant is formally equivalent to the usual S-curve, only slightly more general.

To extract such an S-curve, we must fulfill three requirements:

- to determine the most optimal bounds t_{min} and t_{max} (so that the pattern covers all data points corresponding to decades in the $[t_{min} : t_{max}]$ interval);
- to extract the two S-curve parameters from the data, α and t_0 , given the bounds t_{min} and t_{max} ;
- to devise a criterion so as to reject the pattern if it is not good enough.

In the following, we discuss one particular method which we used to extract the S-curve pattern from corpus data. We shall also discuss alternative methods which may prove to be more efficient, though we did not test them on a large-scale dataset.

Standard logit procedure

We now detail a method based on the logit method of extracting S-curve patterns as devised by Altmann et al. (1983). Note that this only covers step 2 of the whole procedure. Steps 1 and 3 seem not to have received significant discussion in the literature. Most often, the time period covered by the pattern was set a priori through linguistic expertise. As we already stressed, one of the main contributions of the present empirical study is to extract the S-curve patterns automatically without positing that the pattern is to be found on the selected data range. As for step 3, which is necessary in this data-driven survey, it was not discussed extensively either in the literature we surveyed in the previous chapter, and the fits were mostly deemed satisfactory from a visual perspective.

For each investigation period T_g , we create all subsets $\sigma_{ij} = [y_i; y_j]$ such that $i \in T_g$, $j \in T_g$, $j - i \geq 5$, and $\forall k \in]i; j[, y_i < y_k < y_j$. Then, for each subset σ_{ij} , we perform its logit transform $y_k \rightarrow \lambda_k(\sigma_{ij})$, for all k such that $i < k < j$:

$$\lambda_k(\sigma_{ij}) = \log\left(\frac{y_j - y_k}{y_k - y_i}\right). \quad (4.9)$$

An explanation is in order concerning the two conditions (i) $j - i \geq 5$ and (ii) $\forall k \in]i; j[, y_i < y_k < y_j$. The latter is a mere check that the argument of the logarithm in equation (4.9) always falls into its definition domain. The former is more arbitrary, but of crucial importance, and shall be explained later on.

The logit transform of equation (4.9) has the property that, if a curve follows an S-curve of expression:

$$\tilde{y}_k = y_i + \frac{y_j - y_i}{1 + e^{-(\alpha k + \beta)}}, \quad (4.10)$$

then it turns this S-curve into a linear curve:

$$\tilde{\lambda}_k(\sigma_{ij}) = \alpha k + \beta. \quad (4.11)$$

Then, a curve can be said to be sigmoidal, if and only if its logit transform is linear.

Now that we have built the logit transform of all subsets σ_{ij} for a given growth period T_g , we can evaluate which couples (i, j) , or equivalently which time boundaries of the phenomenon, delimits best the S-curve pattern, provided such a pattern really is to be found. To do so, we perform a simple linear fit of all logit transforms $\lambda(\sigma_{ij})$ and keep only those of sufficient quality. We define in this regard an arbitrary threshold that the r^2 estimate of the fit quality must pass over for the fit to be considered good enough. In this work, this threshold has been set up at 0.98, which is pretty high and restrictive.

The good fits having been filtered out, it remains to choose the best one. Instead of picking up the fit with the best r^2 value, we first restrict the candidate couples to the ones associated with the longest pattern, i.e. such that its width $j - i$ is maximal. Among them, we finally elect the one with the greatest r^2 value. This gives us our S-curve. Of course, if the threshold is reached for none of the (i, j) pairs, then there is no S-curve to be found.

One might wonder at this point why we bother with the search for growth parts. It could indeed have been possible to try out all pairs (i, j) verifying conditions (i) and (ii), and then to extract the best S-curve through the procedure detailed above. However, this would amount to assume that there is only, at best, one single S-curve in the whole frequency evolution of the form. On the contrary, by first locating the growth regions, we ensured that one S-curve at best could be found for each of these growth periods T_g . For instance, we have found up to three S-curves described over time by the frequency of one single form. In some cases, two consecutive S-curves were also found, a phenomenon which had never, to my knowledge, been presented before. Note that, because we extend the period of relative growth by ten decades on each side, it is also possible that the same S-curve is found for two different T_g , as those may significantly overlap.

We now return to the condition $j - i \geq 5$ to finally explain it.

4.1.4 Quality of the fit

The logit transform relies crucially on the determination of the two boundaries, y_i and y_j , since all points of the logit transform depend on these two values (hence, the systematic search for the ‘best’ boundaries). It also entails that, if the length of the subset σ_{ij} is l_{ij} , then the logit transform contains $l_{ij} - 2$ points. If we then set the condition $j - i \geq 5$, it means that the logit transform must be sampled by at least four points. A lower value, for instance $j - i \geq 3$, would have been ridiculous, since it would have led to the possibility to perform a linear regression of two points — a guaranteed perfect fit, no matter the curve. A less strict condition $j - i \geq 4$ would have been possible, but a linear fit of three points does not make much sense either, for a quality fit can too easily be obtained.

To quantify this bias introduced by a low number of points for the linear fit, we

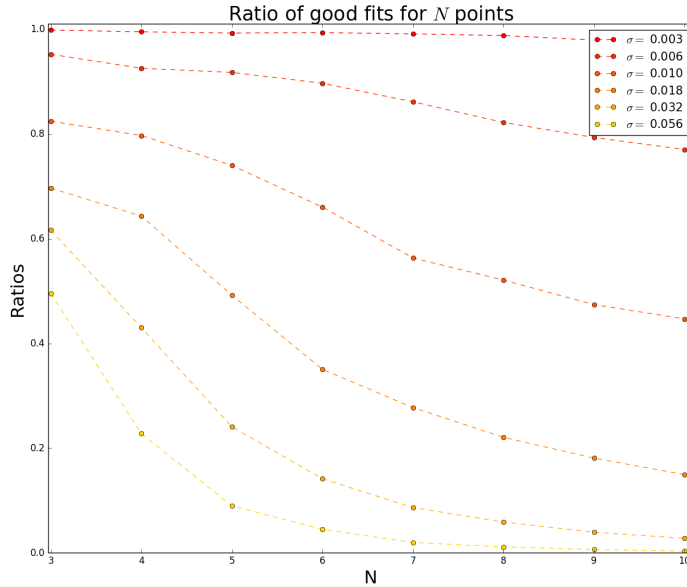


Figure 4.1: Ratio of samples of N points generated by an S-curve and then perturbed by a white gaussian noise of intensity σ , for different values of noise intensity and for different numbers N of points.

built sigmoids of N points (x_k, y_k) with:

$$\begin{aligned} x_k &= -7 + k \frac{14}{N} \\ y_k &= \frac{1}{1 + e^{-x_k}}. \end{aligned} \quad (4.12)$$

We then generated 10^5 samples of curves perturbed by a Gaussian noise, with different amplitudes of noise (from 10^{-5} to $10^{-0.5}$). For all these curves, we applied the logit transform, performed a linear fit, and kept only the curves for which $r^2 > 0.98$. The ratio of ‘good enough’ curves is then computed for each different number N of points, with N varying from 4 to 10, for each value of the noise amplitude. The higher the number of points, the more sensitive the pattern is to the level of noise (Fig. 4.1).

There is thus a higher probability of false negatives at high number of points: the more decades the S-curves spans, the more easily noise can make it fail to pass the criterion. Furthermore, low number of points are associated with a higher probability of being false positives. If we sample N points between 0 and 1, sort them, and try to fit those points by a linear curve, then once again, we see on Fig. 4.2 that low number of points (mostly, lower than five) are associated with a higher chance of passing the criterion. If, furthermore, we weight those chances by the probability to be sorted by chance, then the probability of false negatives is overwhelmingly higher for low number of points. Not only then regular growth can be mistakenly taken for a sigmoidal growth, but a random generation of points can be assumed to describe a growth, while it is a mere side effect of chance.

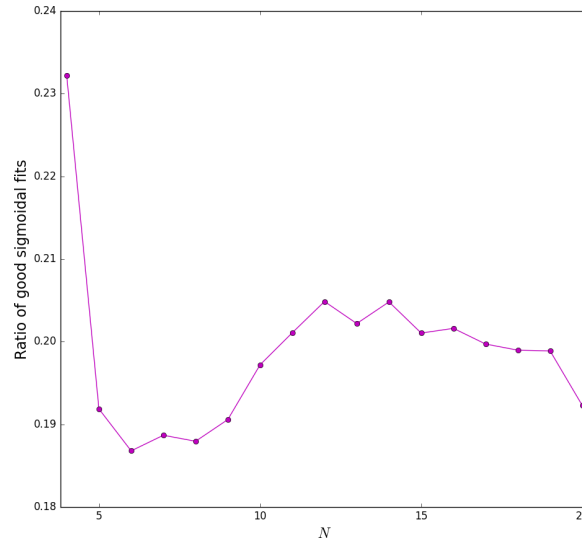


Figure 4.2: Ratio of randomized samples of N points considered as described by an S-curve according to our criterion, for different numbers N of points.

For this reason, our procedure is not so efficient at distinguishing S-curves for low number of points. We yet fix the threshold $j - i \geq 5$, leading to a minimal number of four points (the two extremities are not accounted for in the linear fit), because it already represents a long period of time (more than half a century), and a higher threshold will discard all ‘fast’ changes entirely. One could grasp more easily fast changes by setting a lower Δt (of five years, for instance, so that we can catch S-curves as short and fast as thirty years), yet this is impossible given the sparseness of the data in earlier centuries. A specific investigation of changes in the late nineteenth and twentieth century, with a Δt of five years, would be interesting to give an idea of how many fast changes we miss in proportion by setting this $j - i \geq 5$ threshold.

Another possibility to assess the quality of the fit is to compute directly how far away the frequency pattern is from an S-curve. First, the parameters of the S-curve are computed thanks to the linear fit of the logit transform. Then, we can compute the total squared error Y between the data and the fit:

$$Y = \frac{1}{N} \sum_{k=1}^N (s_k - \tilde{s}_k)^2, \quad (4.13)$$

with $N = j - i - 1$ and where the s_k 's are the data points after scaling:

$$s_k = \frac{y_{i+k} - y_i}{y_j - y_i}, \quad (4.14)$$

while the \tilde{s}_k are given by :

$$\tilde{s}_k = \frac{1}{1 + e^{-(a(k-1)+b)}}, \quad (4.15)$$

a and b being the two parameters of the linear fit of the logit transform.

Similarly, we can define a threshold (say, $Y_{th} = 0.05$), keep all pairs (i, j) associated with a total error Y inferior to the threshold, then picking out, among all biggest subsets σ_{ij} , the one with the lowest Y . This gives results which seem much more encouraging than before: the sensitivity to noise depends fairly weakly on the number of points (interestingly enough the dependence is non-monotonic). However, the fact that the goodness of fit is not quite perturbed by the noise in this case may actually be concerning; it entails that an S-curve would be recognized as such, even if noise has blurred it exceedingly. It means that the procedure would be prompted to recognize S-curves, even if there are not.

Indeed, this variant gives very poor results when applied to actual data, and the procedure relying on the r^2 criterion is much to be favored, despite its own flaws. Furthermore, the low error which is to be found shows that the midpart of the S-curve is well fit by the parameters found through the logit transform; otherwise, the error would have been much greater, for the high-slope region is associated with the widest variations. This is encouraging of the fact that the parameters found thanks to the logit transform are indeed robust, a result which applies as well for the r^2 method.

4.1.5 Limitations and alternatives

One of the main limitations of relying upon the logit transform of the data to perform the fit is that the logit transform is most sensitive to the extreme values, i.e. to the regions when the S-curve is closest to 0 or 1 and varies the least. This sensitivity is due to the divergence of the logit transform derivative at those two values (Fig.4.3). This is clearly problematic, for the most salient part of the S-curve is the mid-part, the part of fast growth, not the staling ones at the beginning and the end of the process. Thus, a logit-based fit of the data is bound to be more sensitive to the least interesting parts of the change, which is somehow problematic. As we saw in the previous chapter, depending on the specific treatment of those extreme values in Kroch's data, the obtained results can widely vary.

In our procedure, we strove to be as less exposed to these flaws as possible. First, we only consider sets of consecutive points whose frequency values are all between a lower and a higher bound (which respectively translates to 0 and 1 once the rescaling is done), ruling out all problematically divergent values (such as the ones appearing in Kroch (1989b), for instance). As all points take their values between those boundaries, it restrains the focus around the mid-part and we may expect the fit to work well. Furthermore, trying out different pairs of points for these extreme values and picking out the most satisfying fit (given it encompasses the highest possible number of points) allows to reduce the exposition to these extreme values problematic behavior.

The fit-error-based method, on the contrary, will tend to avoid the most interesting midpart of the S-curve, as any deviation from the sigmoid in this region will produce a large contribution to the total error. Hence, it will tend to avoid the midpart, and will retrieve a pattern which encompasses either the lower or the upper half of the logit. As the curve does not grow much in those regions, the contribution to error will indeed be very low. Hence, the patterns found through this method are skewed towards flat, less interesting regions than the proper bulk of the growth.

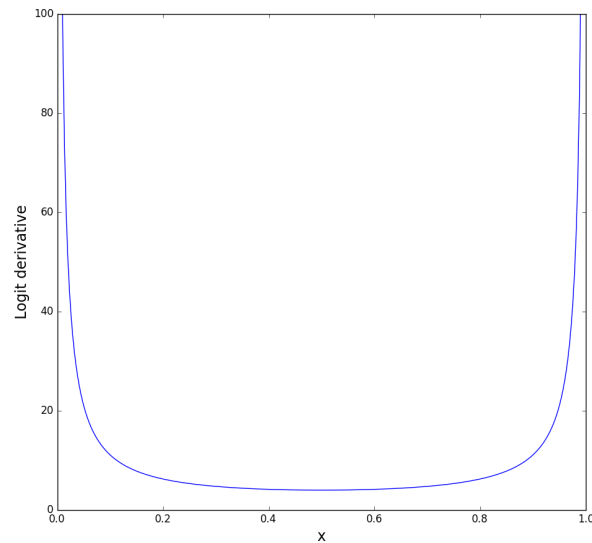


Figure 4.3: Derivative of the logit transform.

An interesting alternative to our approach is to abandon the hyperbolic sigmoid entirely, and to use instead the cumulative Gaussian distribution. According to this perspective, the derivative of the data must be Gaussian, and one can find the optimal parameters of the curve through a simple Gaussian fit of this derivative. The a parameter (slope of the logit transform) is then roughly equivalent to the variance σ^2 of the obtained Gaussian (the higher the variance, the smaller the slope), while the b parameter (intercept of the logit transform) is closest to the mean μ of the Gaussian, as they both state the actual position of the S-curve within the time line of the process.

This method is for instance favored by Michard and Bouchaud (2005), who compute the parameters of different S-shapes curves for various sociological transitions (spread of a technology, wearing off of clappings in concerts, etc.) by performing a Gaussian fit of the discrete derivative of their data. However, in our case, the derivative of the data happened to be extremely noisy, and the Gaussian fit was therefore quite poor if not infeasible, while the logit transform usually gave satisfying results and nicely linear outputs. As a conclusion, among all variants that we tried, the logit transform method, as developed long ago by Altmann et al. (1983), indeed happens to be the most appropriate one for the investigation of S-curves in language change.

4.1.6 Improving the extraction procedure

Now that we have presented the method that we applied for our data survey, we would like to present ways to improve this method. Indeed, the more successfully we extract quantities out of raw data, the more reliable they will become. Though these improvements were mostly not used in the analyses conducted in this thesis, we believe it remains important to present what can be done in future work to obtain better and more trustable results.

In the following, we will then propose ways to improve each of the requirements that the extracting procedure should fulfill: determining the optimal boundaries of the pattern, fitting the parameters of the S-curve as described by equation 4.8, and finally, devising a quality criterion so as to rule out the S-curve as a correct depiction of the frequency rise.

Comparing the S-curves

The boundary determination step is the most crucial one, since the quantities we can extract from the S-curve, especially the width (total duration of the pattern) and the rate (slope of the logit transform) are extremely sensitive to it. Actually, the error interval on the slope which comes with the linear fit is usually much smaller than the difference between the slopes obtained from two different pairs of pattern boundaries.

Note that, from equation 4.8, these boundaries could be obtained from a four-parameters fit of the curve (α , t_0 , y_{min} and y_{max}). The problem would be that fitting four parameters with a limited number of datapoints (say around ten) would be a clear instance of overfitting, and would raise convergence issues as well. Worse, even if we define the times t_{min} and t_{max} by:

$$t_{min/max} = \arg \max_{t \in T_g} |y_t - y_{min/max}|, \quad (4.16)$$

then we would face the problem that the decades obtained that way may not be part of the range on which the parameters have been computed. For instance, if we performed the fit for decades from 13 to 30, and obtain values for t_{min} and t_{max} which are equal to 9 and 27, then should we take for the width the range over which the data is fit, or the decades encompassed by the $[t_{min} : t_{max}]$ interval?

As it stands, we therefore did not find out a better method than treating y_{min} and y_{max} as the extremal values of the S-curve reached at the decades t_{min} and t_{max} which bound the data we are trying to fit. What we proposed earlier, namely trying out all candidate pairs (i, j) of a given interval T_g , and then picking up the pair corresponding to the ‘best’ S-curve, is certainly inelegant, but remarkably simple and efficient.

The actual question becomes then: what would be the ‘best’ S-curve? We already discussed several ways to assess the quality of a fit and from this basis compare one boundary choice to another: the r^2 coefficient of the linear fit, or the average square error, in both cases combined with a width maximization. There could actually be a better one, which is the Cramér’s V error. It is defined by the square root of the average χ^2 error of each data point:

$$V_C(t_{min}, t_{max} | \alpha, t_0) = \sqrt{\frac{1}{w} \sum_{k=t_{min}}^{t_{max}} \frac{(y_k - \tilde{y}_k(\alpha, t_0))^2}{\tilde{y}_k(\alpha, t_0)}} \quad (4.17)$$

where $y_k(\alpha, t_0)$ is given by equation 4.8 and w is the width of the pattern ($w = t_{max} - t_{min} + 1$).

Actually, since we try to find the couple (t_{min}, t_{max}) which minimizes this quantity, we can drop the square root. Furthermore, to be sensitive to the fact that a wider pattern has more chances to be noisier, and to give a compensating edge to these

wide patterns, we can divide the average χ^2 by an additional factor n . There is no need, then, to first maximize the width over a selected set of ‘good’ pairs, and next to minimize the error among the widest patterns: minimizing the quantity $\langle \chi^2 \rangle / w$ right away is now sufficient.

This quantity has on the other hand the disadvantage that the start of the process will be a high source of error (since \tilde{y}_k is very low then). The method seems nonetheless quite promising, and the selected patterns closely match the visual intuition: whenever the different procedures disagree, the pattern obtained through this latter variant looks usually nicer (it must not be forgotten that the human eye has so far outmatched any automatized algorithm to detect patterns, so that the importance of this latter check is not to be dismissed). We therefore believe that the data analysis would benefit from this new choice, and that it should at least be favored in future work on the topic.

Parameters extraction with the maximum likelihood method

To extract the parameters α and t_0 from the data in the $[t_{min}; t_{max}]$ range, we used the method put forward by Altmann et al. (1983), which consists in performing a linear fit of the logit transform of the data. Then, parameter α is given by the slope of the linear fit, and t_0 by the ratio $-\beta/\alpha$, where β is the intercept of the linear fit. This method is simple and efficient, but it is not the only possible one. One drawback of the method is that it weights all points equally, while some of them can be more representative than others (if they are associated with data of better quality, for instance).

Though the quality of the data is a complex mixture of quantity, genre representation, author diversity, and many other textual and sociological factors, it is almost impossible to give a quantitative estimate of it. However, if we restrict ourselves to the sheer quantity of data, we can weight the datapoints according to their reliability, in this bare quantitative sense. To do so, it suffices to maximize the likelihood \mathcal{L} of the S-curve fit of the data:

$$\mathcal{L}(\alpha, t_0 | y_{min}, y_{max}, \{n_t\}, \{N_t^{tot}\}) = \prod_{t=1}^{w-1} \binom{N_t^{tot}}{n_t} \tilde{y}_t(\alpha, t_0)^{n_t} (1 - \tilde{y}_t(\alpha, t_0))^{N_t^{tot} - n_t}, \quad (4.18)$$

where n_t would be the number of occurrences of the form in decade t and N_t^{tot} the total number of occurrences in this decade, and $\tilde{y}_t(\alpha, t_0)$ is the sigmoid as defined by equation (4.8). The problem is, we work with average frequencies, and here the frequencies are not averaged, unless we define the average frequency y_t as:

$$y_t = \frac{\sum_{k=t-4}^t n_k}{\sum_{k=t-4}^t N_k^{tot}}. \quad (4.19)$$

Then, we can define:

$$n_t \leftarrow \sum_{k=t-4}^t n_k \quad (4.20)$$

and:

$$N_t^{tot} \leftarrow \sum_{k=t-4}^t N_k^{tot} \quad (4.21)$$

so that equation (4.18) remains unchanged, and accounts for average frequencies.

Now, instead of maximizing the likelihood, it is usual to maximize its logarithm, which gives:

$$\log \mathcal{L} = C + \sum t = 1^{w-1} N_t^{tot} [y_t \log \tilde{y}_t(\alpha, t_0) + (1 - y_t) \log(1 - \tilde{y}_t(\alpha, t_0))] . \quad (4.22)$$

Maximizing the log-likelihood amounts to solve the equations:

$$\begin{cases} \frac{\partial \log \mathcal{L}}{\partial \alpha} = 0 \\ \frac{\partial \log \mathcal{L}}{\partial t_0} = 0 \end{cases} \quad (4.23)$$

or more explicitly (dropping the variable dependencies of $\tilde{y}_t(\alpha, t_0)$ to lighten the notations):

$$\begin{cases} \sum_{t=1}^{w-1} N_t^{tot} \left[\frac{y_{max} - \tilde{y}_t}{1 - \tilde{y}_t} \frac{\tilde{y}_t - y_{min}}{\tilde{y}_t} \right] (y_t - \tilde{y}_t)(t - t_0) = 0 \\ \sum_{t=1}^{w-1} N_t^{tot} \left[\frac{y_{max} - \tilde{y}_t}{1 - \tilde{y}_t} \frac{\tilde{y}_t - y_{min}}{\tilde{y}_t} \right] (y_t - \tilde{y}_t) = 0 \end{cases} \quad (4.24)$$

This can be achieved through a gradient descent method.

We present on Table 4.1 the results obtained for both methods (linear fit of the logit transform and maximum likelihood performed directly on the frequency), for ten randomly selected examples. Note that the linear fit of the logit transform has been performed on the data averaged according to equation (4.19), so that the parameters obtained here can differ from those presented on table A.1.

Form	Logit method	Maximum likelihood
<i>à condition de</i>	(1.87, 1.77)	(1.81, 2.12)
<i>à la longue</i>	(0.37, 9.11)	(0.39, 9.04)
<i>à l'encontre de</i>	(0.45, 8.51)	(0.43, 8.18)
<i>à l'instant</i>	(0.83, 1.86)	(0.84, 1.84)
<i>à l'instar de (i)</i>	(0.46, 3.51)	(0.46, 3.53)
<i>à l'inverse</i>	(1.37, 3.36)	(1.28, 3.42)
<i>dans une large mesure</i>	(0.90, 2.52)	(0.94, 2.46)
<i>de toutes parts</i>	(0.84, 4.13)	(0.89, 4.17)
<i>de temps en temps</i>	(0.39, 7.04)	(0.40, 7.08)
<i>tout au long de</i>	(0.70, 5.81)	(0.63, 5.93)

Table 4.1: Parameters (α, t_0) describing the S-curve as obtained from the linear fit of the logit transform method, and the maximum likelihood method, for ten randomly selected forms. Time t_0 is defined with respect to the beginning of the pattern, not the beginning of the whole 1321-2020 time period.

The parameters obtained from both methods are remarkably close. Therefore, weighting the data points according to the quantity of available data has little impact on the extraction procedure. That the same parameters can be obtained in two different ways comfort at least the idea that the results we obtain are robust and reliable.

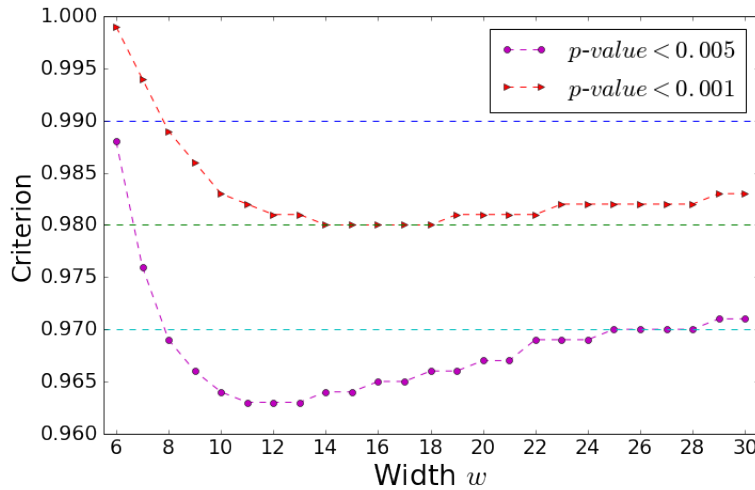


Figure 4.4: Width-dependent criteria for two different aimed p -values: 0.05 (magenta dots); 0.01 (red triangles).

A null model of frequency rise

So far, we have set an arbitrary threshold ($r^2 > 0.98$) above which the fit was considered as good enough to validate the S-curve. However, we saw the shortcomings of such a choice; in particular, the criterion will filter differently the patterns according to their width. For short widths (e.g. of six or seven decades), the criterion was not as selective as for longer widths.

We can address this issue more rigorously by positing a null model of frequency rise to compare the S-curve with. A null model would be such that, for a given width w , the frequency rises from 0 to 1 in $w - 1$ steps, not necessarily monotonously. A wide range of models would however be consistent with this rather bare requirement, so that the choice of a null model is not neutral. One could, for instance, think of a linear increase, but this would already be quite a strong assumption on the specific nature of the growth.

We tried two different null models. The first one assumes a random gaussian process such that the frequency y_k is obtained iteratively according to:

$$y_{k+1} = y_k + \frac{1}{w-1} (1 + \eta_k) \quad (4.25)$$

with the initial condition $y_0 = 0$ and η_k is drawn from a normal distribution of mean 0 and variance 1. The average process is therefore a linear increase from 0 to 1.

We can now evaluate the p -value associated with our criterion, which is the probability for surrogate data generated from the null model to pass the said criterion. Alternatively, we can compute which criterion should be set, for any value of the width w , to obtain a given p -value. Doing so, we can obtain a criterion for the r^2 which is no longer arbitrary, but corresponds to a particular wished for p -value of the sigmoidal fit. Different p -values lead, expectedly, to different criteria for each width (Fig. 4.4).

It appears thus, as we already discussed, that the standard to which evaluate the quality of a pattern depends on the width of this pattern. The arbitrary criterion that we set, $r^2 > 0.98$, corresponds to a very good p -value (0.01) for high number of points, but would be associated with poorly significant sigmoidal fits for low number of points, for which a much higher constraint must be set (0.987 for a width of six decades). It would therefore be commendable to favor a different criterion for each value of the width, as given by the two curves of figure 4.4, instead of choosing a unique criterion for all patterns indistinctively.

Two remarks should be added. First, the choice of the null model is certainly arbitrary and cannot be held as fully reliable. We tried another null model, with a monotonous growth made of steps drawn from an exponential probability, so as to obtain an average process different from the linear increase. It turned out that the results were extremely similar between the two null models, so that the results are weakly dependant on the choice of a particular null model. Second, this method (devising the criterion so as to pass below a p -value significance threshold) is not specific to the use of the r^2 . We could use any measure of error characterizing the S-curve pattern as well: what counts only is to evaluate the probability that a null rise process would pass the chosen criterion. Therefore, this null model procedure of setting a particular width-dependent criterion could be compatible with other error evaluation methods, such as the Cramér's V error previously discussed.

We shall consider, later in this chapter, how different the results can be if we consider such width-dependent criterions instead of the uniform $r^2 > 0.98$ criterion.

4.1.7 Summary

The goal of this section was to devise a procedure to extract an S-curve pattern from frequency data, with the requirement that no a priori period of change was to be fed into the procedure. The one we presented here is simple and straightforward. Though it depends on multiple arbitrary parameters, they have little incidence on the result. In this study, they have been calibrated using a handful of actual instances of change, with a Δt of a decade.

The main important steps of the procedure are the following:

1. Smoothing data by averaging over a gliding window.
2. Identifying time periods of frequency growth, characterized by repeated relative growth.
3. For each period of growth, try out all suitable pairs of times and compute the logit transform of all points in-between. Keep pairs only if the linear fit of this logit transform is good enough.
4. Finally, for each period of growth, pick out the pair associated with the best S-curve, if there is one.

We can now use this procedure to track down the S-curve pattern over a great many instances of semantic change in French, thanks to data provided by the Frantext textual database. Before turning to the results, we shall present this database and explain how the data was actually obtained.

4.2 The Frantext corpus

To perform our statistical and quantitative investigation of the S-curve in semantic change, we relied exclusively on the Frantext textual database. As such, we shall present it in details.

4.2.1 Presentation of the database

Frantext is a French language database developed and managed by the ATILF laboratory, and is accessed upon registration. The base contains over 5000 texts totalizing 300 millions occurrences by the time of June 2017, and new texts keep being added every year (a drawback of this is that results are not exactly reproducible from one year to another). It covers a wide period of time, from the xth century to the XXIst, and various literary genres, from drama to correspondence, from novels to science treatises, including, in a small extent, translated works. There are, however, no newspaper article, nor any material which has not been properly published, except in the earlier decades (e.g. manuscript registers in the XIVth century).

Contents of the database

As we already mentioned, the database covers French language starting with the xth century, up to nowadays. However, there is only one text in the whole xth century, dated of year 950, titled *Passion de Jésus-Christ ou Passion de Clermont*, and containing only 3500 words. There are two texts for the whole XIth century, both approximately dated of year 1100. One can thus consider that the database really starts with the second half of the XIIth century.

Expectedly, there are much many more texts in the most recent centuries (especially the last couple of ones), and the distribution of texts over decades roughly follows an exponential increase (Fig. 4.5). To give an idea of this unbalanced repartition, the period 1801-2013 accounts for about two thirds of the database (66%). One might regard this skewed distribution of texts as a major flaw, yet pragmatically, it seems better to include more texts whenever possible, rather than to level down the number of texts to the poorest decade. 100,000 words per decade (the smallest number of words per decade of the whole 1321-2020 period) would have felt rather limited. It is less than the total number of entries in the last published version of the reference French usual dictionary (*Le Petit Robert 2017*), which is 300 000.

Various literary genres are represented in the corpus, including novels, drama, epistolary correspondence, essays, treatises, poetry, memoirs, travel literature, etc. However, no effort was made to provide a time-robust repartition of the documents among those different genres. In 2017, the total number of different genres is 70, with no overlap (all texts are given one and only one genre). It includes some rather odd categories, such as ‘journal fictif’ (‘fictive diary’), ‘récit personnel’ (alongside with ‘autobiographie’, ‘récit autobiographique’, ‘récits personnels’, ‘notes et fragments autobiographiques’, ‘écrits personnels’), ‘dit poétique’ and ‘fiches et fragments non fictionnels’, all of those four genres being represented by only one document. The six most represented genres (over 300 texts) are ‘roman’ (novel) with 1252 texts, ‘théâtre’ (drama, 680 texts), ‘traité ou essai’ (treaty or essay, 730 texts), ‘poésie’ (poetry, 485

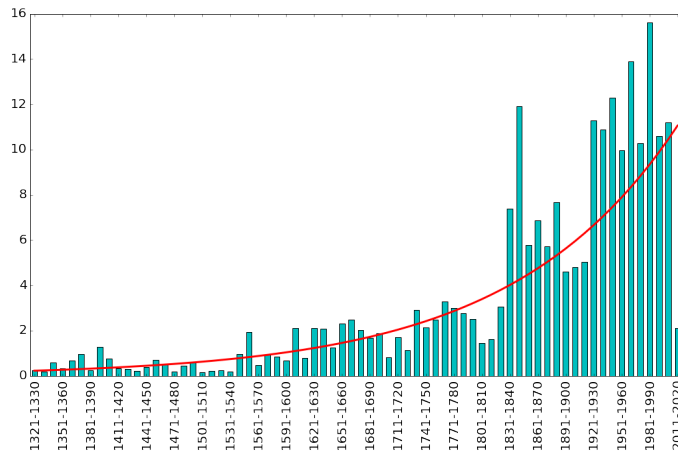


Figure 4.5: Total number of words (in millions) within each decade of the Frantext database for the whole 1321-2013 period. An exponential fit, in full red line, shows the overall increase of contents as we approach the present time.

texts), ‘essai’ (essay, 355) and ‘genre non renseigné’ (unknown, 339 texts). The genre is thus to be chiefly considered as informative of the nature of the text, rather than a criterion accounted for in the compilation of the database.

Some works appear several times in the database, with no justification. *Le Cid*, by Pierre Corneille, appears three times, respectively dated 1637, 1637 and 1682, with three slightly different titles, *Le Cid : tragi-comédie (1637)*, *Le Cid (1637)* and *Le Cid : tragédie, 1636*. This, however, is a minor issue, as it remains highly exceptional.

Concerning, but statistically insignificant, is the presence of anachronistic editorial notes. For instance, in *L’Olive*, by Joachim Du Bellay, we find occurrences such as:

Du Bellay imite un sonnet de Francesco Sansovino
 Du Bellay is mimicking a sonnet by Francesco Sansovino

which obviously does not belong to the original text. Yet, this remains peripheral, and even for this document, the XXIst editorial material makes a small proportion of the total content. This little caveat is finally here to remind that individual occurrences behaving unexpectedly, especially in older times, must be carefully checked for they can be irrelevant.

Our corpus

To perform our statistical investigation over semantic changes, we choose to limit ourselves to the period 1321-2020 (2013), covering thus 70 decades. It is indeed the longest period for which there are at least 5 texts per decade in the database.

We kept most of the texts, excluding a few of them because the dating was too inaccurate: *Le Canarien, pièces justificatives* (Frantext ID 6205), whose associated date is ‘1327-1470’, *Chartes et documents de l’abbaye de Saint-Magloire* (Frantext ID 8203), dated ‘1330’, but in reality encompassing a far too long period of time, and at last *Registre criminel du Châtelet* (Frantext ID 8201 and 8202), dated ‘1389’, for the same reason as the previous one. It was particularly crucial to remove those

documents, for they represent a large part of the total number of occurrences of their respective decades, and it would have biased the data accordingly. We have been less strict regarding inaccurate dating in the XXth century, though we encountered some, because the error represented by those texts would be statistically insignificant. Finally, as the study was performed in 2016, it does not include any of the texts added in the 2017 update.

We did not distinguish between different literary genres. The first reason of this choice is that the repartition among different genres is not homogeneous across time. For instance, the XVIIth presents 196 drama pieces, totalizing 3.7 millions occurrences, which represents 15% of the total, while the XXth contains 150 of them, summing up to 4.2 millions of occurrences, a mere 3.4 % of the total. A second reason is that not so many authors contribute to a given literary genre. Drama, to stick with our example, is quite often due to a few authors. In the two decades 1531-1540 and 1541-1550, the drama production acknowledged by Frantext is, for seven out of nine drama pieces, the work of the sole Marguerite de Navarre. This would heavily bias the data towards the idiolectic peculiarities of a given author. This would be also true for travel literature: over the 55 texts of this genre that Frantext contains, 18 are due to the same individual, Pierre Loti.

The third and final reason is that the content of the different literary genres is not itself homogeneous. Drama in the XVIth century plays a different role in society than drama in the late XXth century, as it is not aimed to the same audience in terms of cultural class. Thus, we expect represented orality close to spoken language in the former, and sophisticated, literary language in the latter. The motivation to consider a ‘homogeneous’ corpus by restricting it to a single genre is thus weakened by such an observation.

Query structure

Frantext is not accessed as raw data, but through queries to the database. A great advantage of those queries is that it allows for Booleans (OR and NO, respectively represented by the signs | and ^, the AND operator being always implied) as well as empty slots of various lengths (&q(*n,m*) represents an empty slot to be filled with any number of words between *n* and *m* included). For instance, the query:

```
par ^ouvrir &q(0,1) (fenêtre|fenestre)
```

will retrieve all occurrences of the form *par* DET (ADJ₁) *fenêtre*, also accounting for the obsolete spelling *fenestre* (note that ^ouvrir counts as any word which is not *ouvrir*, so it is the equivalent of a wildcard with a particular restriction; and indeed, apart from this verb, only determiners are to be found in this position). We can thus find all occurrences of *par la fenêtre* (‘through the window’), but also less expected ones as *par quelque secrète fenêtre*. There are no limits to the number of booleans operators that can be added in a single query; it will only increase drastically the processing time at some point. Some words can also be marked as optional by using the sign &? in front of them.

The database is lemmatized: one can encompass all different forms of a given verb using the coding &c right before the infinitive form (e.g. &cvendre covers *vendons*, *vendaient*, *vendue*, etc.), and those of nouns and adjectives using &m (e.g. &mcheval

`&mblanc` will retrieve both *cheval blanc* and *chevaux blancs*).

However, the data is not tagged (contrary to other databases such as all Mark Davies' corpora or Google Ngram). This is, to be sure, a clear limitation, but it should not affect us that much in the present study, for we are dealing with semantic changes and semantic expansions, which entail that the category of words under investigation is subject to shift, and the scope of its collocates can also either shrink or expand across different categories. As an example, *venir de* ('to come from PLACE' > 'to just have V_{PARTICIPATE}') was followed first by nouns (locations), then also by verbs (actions). Tagging the corpus would allow to track down more easily the transition time at which the second use has come to be developed, but it would also be unable to catch blurry transitions often characteristic of grammaticalization. There may also have been scientific reasons behind this absence of tagging, as it avoids to project on past centuries linguistic production the intuitions and categories specific of the present state of the language. The syntactic surface of the occurrences is thus our sure and only guide, which is quite in line with Construction Grammar claims, especially those of Radical Construction Grammar, which we already encountered.

There exists, however, the possibility to create lists of words (e.g. the list of all determiners) which can be called as an entry. Once the list of all determiners (which is written into the query thanks to the command `&ldeterm`, where `determ` is the name given to our example list) has been created, one can ask the following query in order to only get the meaning 'to come from PLACE':

```
&cvenir &?juste (de|d') &ldeterm
```

and find occurrences such as:

Je peux dresser une liste de ce que je sais, et qui **vient de cette** lettre
 Le faible jour **venu de la** cuisine éclaire de biais un invraisemblable amas de vieux bidons puants
 Et ceux des déportés qui **venaient de la** campagne connaissaient le sort réservé au bétail...

Even this query is not without flaws, as definite articles can serve as reflexive pronouns, e.g.:

elle fit repasser la robe d'une camarade qui **venait de la** recevoir de Paris
 she ironed the robe of a classmate, who **had just received it** from Paris

Furthermore, the lemmatizer includes forms which can serve both as participles and nouns, such as *venue* (either 'she has come'/'she came' or 'the coming of'), an expected source of spurious occurrences, for instance:

Mais aucune confirmation n'*est venue de* ce canard.
 Yet no confirmation *came from this* rag.

vs. the spurious:

En évoquant la **venue de mon** père à Paris,
 Speaking about the **coming of my** father in Paris,

This kind of interferences can be, in this particular case, easily gotten rid of by adding `^1a` at the beginning of the query.

A still more refined option would be the creation of a 'grammar', which is the equivalent of a list, but can include more complex items, and even have some parameters.

The grammar remains manually created and can be a useful tool to design elaborated queries.

Also, it must be added that there exists a smaller version of Frantext, restricted to the two last centuries (1821-2013), but entirely tagged by syntactic category. It contains slightly less than 2000 texts, which covers about 130 millions of words. The time period covered and the tagging makes it more similar to the COHA, though the latter is significantly bigger (400 millions of words). Unfortunately, the tagging code is not without imperfections; for instance, the command `&e(g=V)` should represent any verb, yet it shows some oddities. In the query:

```
&e(g=V) voulu
```

the command conforms to the expected behavior, but for some reason in the query:

```
&cvenir (de|d') &e(g=V)
```

it finds only proper nouns (which is also true for the simpler query `venir de &e(g=V)`). The position of the command in the query is not the explanation either (`bien &e(g=V)` lists verbs in second position).

Another type of query is called ‘expression régulière’, which relies on a different query syntax. As the first type of query allows for syntactically complex queries, this one aims at catching the morphological diversity of a single linguistic form. It allows to leave unspecified parts of a given word, so as to grasp all possible different spellings, or all forms with a given morphological ending. Thus, the query:

```
souventes?f*
```

retrieves all possible spellings of *souventesfois* (‘oftentimes’), including for instance *souvente fois*, *souventes foiz*, *souventeffoiz*, *souventez fois*, *souventefoys*, etc. and the other one:

```
*erie$
```

which will find all words finishing by *-erie* (the sign `$` at the end of the query is used to mark the end of the word) such as *serjanterie*, *poterie*, *messagerie*, *chevalerie*, *ribauderie*, *chapitainerie*, *prothoficerie*, *chancelerie*, *talemellerie*, *pelleterie*, *truanderie*, *pitancerie*, *faërie*, to name of few from the two decades 1321-1340.

However, this second type of query does not combine easily with the first one. Excluding particular combinations of words or investigating specific collocates is made extremely difficult within this particular syntax. We thus used it much less often as the regular query type, but used it anyway in a handful of cases.

Query output

Once the query is submitted to Frantext, the obtained results is a list of all occurrences compatible with the query, including for each:

- the reference of the text document the occurrence has been found in, including a classification mark specific to Frantext;
- the name of the author of the occurrence, whenever available;
- the date of publication;

- the occurrence and its context of use, which can be expanded up to 700 words for texts under copyright, up to 1500 otherwise.

Note that the date of publication is sometimes problematic. Usually, Frantext gives the date of first publication, instead of the publishing date of the used edition (much more recent in most cases). Yet, it is problematic in some cases. For instance, the work entitled *Mémoires de l'oubli (1927-1933)* has been first published in 1997, so that Frantext attributes to this text a date which is seventy years too late.

Another limitation of Frantext is that the output of a query can encompass a strict maximum of 50,000 occurrences. To avoid this overflow, one can either ask a more specific query, or restrict the corpus to a shorter period (the exported data can then be concatenated without difficulty). However, this procedure, when it comes to deal with hugely frequent words such as *dans* ('in', 'into'), becomes extremely cumbersome, for the corpus must be sometimes split into years instead of decades.

Once obtained, this online output can be exported through a `.txt` file, including the desired pieces of information. In our study, we kept only the date and the classification mark. Thus, for each occurrence, we get in which text it appeared, and in which year. Thus, a query associated with N occurrences will be recovered through the data of the N couples $\theta_i = (\text{year}_i, \text{document}_i)$ for $i = 1 : N$. The frequency of the form to which the query correspond is thus given for the decade k ($\text{year}_k - \text{year}_k + 9$) by the ratio:

$$x_k = \frac{\mathcal{C}(\{\theta_i | \text{year}_i \in [\text{year}_k : \text{year}_k + 9]\})}{N_k}, \quad (4.26)$$

where N_k is the total number of words in the corpus for decade k .

4.2.2 Alternatives to Frantext

There is no satisfying alternative to Frantext. It is the only database which covers extensively and consistently a time period as long as seven centuries. Apart from very few exceptions, the corpus has been digitalized with care, and spurious errors are deleted every year. Texts usually appear in only one version and are given a date consistent with their first publishing. Yet, it is good to know that there are alternatives to Frantext, each of them with their own advantages and limitations.

One of these alternatives is the giant database Google Ngram. The French 2012 database contains as many as 100 billion of words for the whole period 1547-2009, which is huge, almost three hundred times bigger than Frantext. This impressive statistical weight would invite to rule out Frantext entirely in favor of Google Ngram. However, several issues of this gargantuan database make Frantext a better option for the study we have performed.

Why not using Google Ngram?

Some limits of the Google Books database (to which the Google Ngram is based in an unclear way, as detailed below) have already been stressed in the recent past (Pechenick et al., 2015). They pointed out that the corpus is unbalanced in the sense that the most productive authors of their time are more represented than others, biasing the representation of the overall language towards their own idiolectic specificities. Yet,

this argument can also be addressed to Frantext, and many other corpora as well (CORDE, for instance). Another limitation evidenced by Pechenick et al. (2015) is that Google Ngram contains many academic texts, accounting for a language which is specific to academic customs, especially in the terminology, so that the frequency of uses are biased towards a specific sub-use of language, not representative of its most common uses by the majority of people. A third concern is the fact that the frequency of a word found in literary texts does not reflect its cultural saliency. The authors give the example of Frodo, the main protagonist of the *Lord of the Rings* series, published in 1954. In the database, this name is thus only frequent for this particular year, while it reached cultural popularity thereafter (especially since the late sixties when the book knew a renewal of success, partly due to the American publishing of the book, partly because it was resonant with the New Age movement), so that it exhibits a gap between the contents of the book and the living culture, hence between the texts and the living language. Yet, once more, this matter is not specific to Google Ngram and applies to almost any text-based dataset.

The criticisms that we want to level at the Google Ngram database do not concern the overall design and orientation of the database, but regard the fact that this tool is not quite appropriate to investigate language change diachronically and quantitatively.

The first of these issues is that there is no way to probe the actual contents of the database. Indeed, Google Ngram is a data repertory which gives, for each n-gram (with n going from 1 to 5), for each year, how many occurrences of the n-gram are to be found for this year, in how many different texts. For instance, in the French 2012 Google Ngram database, the word *współczesnej* is to be found 49 times in 1997. Actually, almost all words in the file containing all 1-grams starting with letter 'w' are German, English, Polish, Flamish, or something else, a fact which already hints at the overall low quality of the corpus. Knowing that Google Ngram includes n-grams as long as they appear more than 40 times in the whole corpus, it means that all those spurious words are not isolated occurrences.

Yet the actual problem is, this data does not give us any clue about the context of the occurrence, or the kind of text it is to be found in. The only way to probe the contents of the database is by using the Google Books engine. However, to which extent the Google Ngram database and the Google Books one overlap is entirely unknown. At least, we can be sure that the overlap is not exact, since there are many books in Google Books dated after 2009 while the 2012 Google Ngram data stops at this year. This is a very concerning point, because it means that we have basically no idea regarding what is the actual linguistic contents of the database. We do not know on which books the n-gram counts are based, nor can we have access to them in details for the most part (only a small minority of books can be fully visualized, for copyrights reasons independent of Google Ngram).

A second issue is that, for what we actually see of the database, we have much to be concerned with regarding the scientific value that it can actually have. First of all, in Google Books at least, many texts appear much more than one time, with very different datations. To take back the example of *Le Cid*, Google Books displays about thirty different versions of it, with publication date ranging from 1775 to 2013, some of them being ascribed to Jean Racine (as they are found in several editions of a book called *Oeuvres de J. Racine et de P. et T. Corneille*). The case of *Le Cid* is,

in Frantext, quite an exception, while in Google Books, most famous classical novels from past centuries are found in a dozen versions at least. Once again, there is no actual possibility to know if this flaw affects Google Ngram to the same extent.

Also, as highlighted by Pechenick et al. (2015), Google Books over-represents academic literature, which adds another bias to the database. For instance, among the fourteen results of the request ‘par ma barbe’ on the French Google Books subdatabase, for the years 1950-2000, only three of them are relevant, two being modern translations of older texts (Don Quixote and a nineteenth century German play by Töpffer). The third one comes from an anthology of French folktales. All other occurrences are academic quotes and glosses of past works, or reprints of such works. On that account, it means that only one fifth of the occurrences would be reliable as a reflect of language use in this time period (two of them being borderline cases). Frantext, on the other hand, has two occurrences of ‘par ma barbe’, one of them from the song lyrics of singer Georges Brassens, the other from a 1988 translation of a Shakespeare play (and so more debatable). There are thus almost as many relevant occurrences in Frantext and Google Ngram (two versus three), while none in Frantext are completely irrelevant.

We also stressed that, in the provided data for the ‘w’ head-letter 1-gram of the French Google Ngram data, many words were not French words. The letter *w* being infrequent in French, it is understandable that the noise (say, from bilingual works) may be more important in this case, but as things are, it completely overshadows the relevant data. In the letter *n*, we find hundred of occurrences of ‘nordamerikanischen’ (not a French word), hundred of occurrences of ‘nourril’ (probably a scan error of some other word), both of them on the whole time period 1800-2000. The English ‘newsworthy’ is to be found almost every year between 1961 and 2009, in 117 different texts. It clearly indicates that a lot of texts in the French database are not actually written in French.

A last point on this matter is the very poor quality of the scanning. Texts older than 1800 are completely unreliable, as the algorithm of character recognition has been clearly optimized for contemporary fonts only. For instance, the following sentence from *The royal dictionary abridged, in two parts*, by Abel Boyer, 1715: ‘Parler avantageusement de quelqu’un, to speak well of one, to speak much to his advantage, to give a good character of him, to speak honourably of him.’ has been transcribed as: ‘Parler avantageusement e quelqu’un, 1° speak well of one, te steak much to his advant lge, to ive a gead characier of him, to steak h2nourably of him.’ Some words, such as ‘steak’ and ‘rince’, consequently appear much more frequently than they should, as they are mistaken for ‘speak’ and ‘Prince’. Another example of this poor scanning quality can be seen in the comparison between: ‘I found that the New-modelling of this Story, would force me sometimes on the difficult Task of making the chiefest Persons speak something like their Characters, on Matter whereof I had no Ground in my Author.’ and ‘I faura that the Ne: -we kling of this Story, troi’i fr e ve { ctives on the di ili 7 k of making ti e li fist Perffns steak { like their Carefiers, en -i/attro sviereof. I had no Gréard in » , Author.’, to be found in *The History of King Lear, A Tragedy. Acted as the King’s-Theatre.* by Nahum Tate, 1736. The original text is admittedly hard to decipher, yet any posterior check on the scan would immediately detect such nonsensical concatenations of characters. By comparison, every text in the Frantext

database has been manually checked and such blatant errors are not to be found.

In conclusion of this second issue, if we were to restrict Google Ngram to the relevant occurrences only, it might not be that huge. First, it would only cover the last two centuries, second, it would not be significantly bigger than other available corpora. The third and last issue we want to briefly address is the data structure. Contrary to Frantext, which allows for empty slots, elaborated queries, boolean research over long and complex syntactic strings, in Google Ngram we are only given the n-grams and the number of their occurrences. Concatenation of grams, up to a length of 5, and a manual boolean research, can help find the equivalent results. For instance, the French construction [*a {X} reprises*], with *{X}* being a paradigm of quantity, is associated with very different constructs: *'à deux reprises'*, *'à deux ou trois reprises'*, *'à plusieurs reprises'*, *'à de nombreuses reprises'*, etc. This search is made all the more difficult by the fact that *'à'* did not always take an accent in older texts. In Google Ngram, we would thus have to search all 2-grams, 3-grams, 4-grams and 5-grams compatible with this construction. Another example: if we want to investigate the paradigm of *dans la mesure de*, as this construction is already a 4-gram, we could find only the first following word, which will always be a determiner, and thus we would miss the content name necessary to give an idea of the preferential semantic collocates of the construction (e.g. *'dans la mesure de ses moyens'*, *'dans la mesure de tes talents'*, *'dans la mesure de mes forces'*, all found in Frantext).

To summarize, it is patent that Google Ngram is not an adapted tool for a quantitative retrieval of word frequencies, for there are far too many biases: data is opaque and its relevance (context, origin) cannot be checked; data is noisy and of poor quality overall; its specific structure limit the scope of possible investigations. Actually, the driving aim of Google Ngram and Google Books seems to be the encompassing of anything that anyone might look for. To achieve this valuable goal, it is thus of no consequence if a lot of spurious and undesired content is added along the way. It should be stressed, however, that the idea that the statistical weight of the database is powerful enough to balance the possible biases is fundamentally misleading: as we saw, there are at least as good a chance to find in raw data a relevant occurrence as to find an inappropriate one, a lesser chance still that this occurrence is correctly dated, so that the obtained frequency of forms would have certainly little value. One can still hope that the wide noise is anarchic enough to pull in all possible directions, so that an actual trend will show up anyway. It might certainly be the case, but as long as there is no possible posterior check on the data, it seems safer to rely to other corpora for any quantitative investigation.

Mark Davies' corpora

Mark Davies of the Birgham Youth University has gathered several corpora of interest. We will only mention the most relevant for our diachronic study.

The COHA corpus (Corpus of Historical American English) is remarkable on many points. It covers a substantial period of time (1810-2009), contains a total of 400 millions of words (one hundred million more than Frantext), more or less evenly allotted over the covered decades. An effort has been made to balance the representation of different genres for each decade. The corpus is furthermore tagged and lemmatized.

As Frantext, it is probed through queries, which admit Booleans, but do not reach the elaborated levels of Frantext lists and grammars. Once a query is submitted, the UI (User Interface) provides either the number of occurrences in each decade, or the corresponding frequency. All occurrences can be checked, so as to see their context of use and the document in which they are to be found. Finally, one can also directly search for collocates; the query returns the list of collocates, the total number of their occurrences, and the number of their occurrences per decade.

The COHA is thus a reliable tool, extremely powerful, supported by an impressive database, both quantitatively and qualitatively valuable. Its main limitation for our study comes from the period covered, the two last centuries, which is much shorter than Frantext. However, this limitation allows for a consistent tagging, and the possibility to balance the different literary genres — a luxury that cannot be achieved in earlier centuries where the available documents are much more sparse. We made occasional use of the COHA and all examples from English are directly based on the data it provides.

Another interesting database is the Hansard corpus. As it belongs to the series of Mark Davies' corpora, it presents the same interesting features as the COHA (tagged, lemmatized, useful research tools, access to the occurrences' context). It covers a similar period of time (1803-2005), and is a recording of the speeches held in the British parliament. The contents of the database is thus obviously biased towards the specific way of speaking in this very singular context, yet it has also the advantage to provide a much homogeneous dataset. Furthermore, it is bigger still than the COHA, with 1.6 billions of words, and it covers another variety of English, which is also interesting.

In the same series of corpora, we find, in other languages, two databases suited for diachronic investigation: the Corpus del Español (100 million words from the XIIIth century), of a size comparable to that of Frantext, and the Corpus do Português (45 million words from the XIVth century). However, there is no database provided for French.

Among those corpora, the COHA covers too limited a time period for our purposes, but remains occasionally quite useful (we already provided examples from the COHA data). The best alternative would be the Corpus del Español, to perform a similar study based on Spanish examples of semantic expansions. However, the study of Spanish is made all the more difficult because of the different varieties of this language. There exists, for instance, two sets of translation of the *Asterix* of comicbooks, one for Spain, the other for South America, with two different Obelix's motto (*¡Están locos estos romanos!* for Spain and *¡Estos romanos estan chiflados!* for South America), and this is true for most contemporary cultural productions translated in both geographical areas (the Simpsons cartoon series, for instance, is dubbed twice). In South America, though the translated works (especially TV series) are dominated by Mexico, huge differences in speech are also to be found: Peruvian Spanish, for instance has the grammaticalization Perfect > Past (Heine and Kuteva, 2002, p.231) almost accomplished, while Colombian Spanish still distinguishes the perfect use and keeps using the preterit in the daily language. Peruvian Spanish is therefore further into this particular path of change than Colombian Spanish, which is likely to get down this cline as well, although there is no necessity for it to occur.

Though the Corpus del Español does carefully distinguish all Spanish varieties, they

must be studied separately. Hence, the actual size of the database is much reduced, for it must be divided among all those varieties. The same issue is also encountered in Portuguese. Frantext, on the contrary, focuses on metropolitan French, and does not include any other variety of the French language (as would be Quebec French).

The set of corpora provided by Mark Davies is thus most valuable. Though they are not as fitted for us as Frantext, they remain reliable tools, so they still offer the opportunity to test our findings on other databases, and other languages. It should be also stressed that Mark Davies corpora are free of use and can be accessed on an open basis, without registration. One can pay a fee to download the database, but the query system already allows to probe the entirety of the corpus and to extract data of very satisfying quality.

Other corpora

There are other available corpora online which are worth mentioning. The Helsinki corpus allows to investigate language change in English over a long period of time (730-1710) and provides a database of reasonable size (1.5 million words), yet two hundred times smaller than Frantext (though the three last centuries, left aside by the Helsinki Corpus, accounts for the most part of Frantext contexts, the French database still gathers about 50 million words for the corresponding time period, 950-1710). It must be noted that the Helsinki corpus is one of the oldest of its kind, as it has been compiled in the eighties.

In French, the BFM database covers the period between the IXth and the XVth for a total of four million words (to be compared to the 13 million of Frantext for the same time period). Its contents, for the most part, are also part of the Frantext database. Its main interest lies not in the opportunities offered for a wide statistical and quantitative use, but as it provides an impressive array of research tools, it is chiefly used for detailed and precise linguistic investigation of a given change.

Finally, in Spanish, the CORDE corpus, provided by the Real Academia Española, covers the whole history of Spanish language up to 1974 for all its different varieties, and gathers 260 million of words — half the size of Frantext. It is thus a valuable alternative tool to the Corpus del Español: albeit bigger, it is neither tagged nor lemmatized.

From this non-exhaustive review, it thus appears that Frantext is one of the best database available for a diachronic linguistic investigation, almost on par with the COHA in terms of size and functionalities, yet covering a much wider time period. It seems to be the best compromise between the size of the database, and the size of the time period. Finally, I can also personally benefit from the fact that, French being my natural language, working on it it sharpens the intuition often required to grasp the subtle semantic nuances characteristic of an on-going grammaticalization process.

4.3 Four hundred instances of semantic change

Now that we have described our method of data analysis and the Frantext database, it is time to finally turn towards the actual questions motivating this large-scale study:

- How established is the S-curve pattern in language change?

- Can the S-curve be quantitatively characterized?
- What does the quantitative characterization tell us about a given change?

The third question is certainly interesting, but shall be mostly left aside in this thesis. Indeed, individual changes can certainly be quantified and better understood through this effort; yet, these quantitative features of a change would only make sense by comparison with statistical established tendencies. For instance, we can easily obtain the rate of a particular change, but cannot say whether this change is slow or fast as long as we do not have a set of data to compare the numbers with. This is the reason why, in this work, we shall more specifically focus on the statistical quantitative properties of language change. In the next chapter though, we shall provide a few empirical tools to analyze in more detail individual instances of change.

In this section we shall limit ourselves to the first two questions. To tackle with them, we decided to investigate many instances of language change, and more specifically, of semantic change.

4.3.1 Why focus on semantic change?

Indeed, this focus could seem unjustified. The S-curve is supposed to describe any language change; even if we evidence the robustness of this pattern for the sub-category of semantic changes, it will not establish the S-curve as the guideline of language change in general. We shall therefore outline the reasons underlying this choice.

A non-restrictive choice

First of all, semantic change is not that restrictive. In some extreme perspective, we could even consider that all linguistic changes are either phonetic, sociological (including orthographic conventions), or semantic, especially if we include constructionalization as a possible semantic change (which we do). Thus, if a change is ongoing, it must either be phonetic, sociological, or semantic.

Second, studying semantic change is easier, in a way. Indeed, the only historical data which is straightforwardly available is frequency of use of linguistic forms. Now, this frequency of use displays an interesting behavior, for it exhibits bursts through which the frequency can increase tenfold, a hundred fold, a thousand fold. This is clearly not insignificant and can hardly be ascribed to noise. There must be a trigger of this phenomenon.

A lot of triggers can be proposed, actually. The repeated use of a word can be the result of a new trend. It can be the (possibly unintentional) doings of a particularly influential individual, such as a King, a Pope, a President, or a pop singer, or it can just stem from a Fashion Game balancing a tendency to mimic others, and to particularize oneself. This would be the case with first names (Coulmont et al., 2016), which obey a logistic dynamics with an increase first and a decrease second — a decrease which is seldom immediately subsequent to a rise in language change. Other sociological triggers are possible, leading to different dynamics — see e.g. (Ke et al., 2008; Castellano et al., 2009).

External triggers are also possible. Yet, as shown by Ghanbarnejad et al. (2014), when a change is driven by an external influence, such as an orthographic reform from

a language Academy, the change follows an exponential instead of an S-curve. This kind of case is actually quite rare, and the S-curve change is at least qualitatively in better agreement with the strong increases of frequency.

Another possible trigger is semantic expansion. When the meaning of a linguistic form expands, its uses become more diverse, hence more numerous, and the frequency increases. Conversely, studying the empirical signature of a semantic change amounts to find a frequency burst. Indeed, in semantic change (but see below), the form remains unchanged, only the meaning does, while corpus data is form-based, not meaning-based. If it were not for these strong frequency variations, semantic change could only be accounted for through an additional, external expertise, working on the detail of the individual occurrences. To refine this statement, it must be recalled that semantic change can also be detected through a change in collocates (Hamilton et al., 2016), but still, it implies a frequency rise of the involved co-occurrences.

One might argue that we mistake, at this point, an implication (semantic change entails frequency increase) for an equivalence. However, if the frequency is increasing, then, with the very serious exception of the previously discussed case where it can be ascribed to sociological factors, it is usually because something has changed in the language, which allows for such a rise. Even in the sociolinguistics literature, it is often assumed that there is a motivation behind the change, that the form has some intrinsic edge which strengthens its spread and its social reproduction. Broadly speaking, either the form has been always better and people have just been oddly slow to notice it, or something in the structure of the language has changed which gave it this particular edge. More generally, there is no sheer novelties: everything new has a history, it emerges out of the possibilities given at some point in time by language, instead of popping up out of nowhere (Tria et al., 2014). It is thus not so daring to consider that frequency rise must be associated with a language change. Then, since we are tracking the frequency of a form in a textual database, it concerns only written testimonies of language use, and as long as the form is one and the same, a phonetic change is not so likely. Hence, the change has a good chance to be semantic. To summarize: when frequency rises, it is probably due to a semantic change. As the saying goes: *cuando el río suena, piedras lleva*.

We can defend further the idea that frequency change implies almost certainly a semantic change. Later on in chapter 9, we shall define the overall meaning of a linguistic form by a weighted sum over all its different attested uses, or contexts, where the weights are given by the frequency associated with each of these uses. For a frequency change of a form not to involve a semantic change, it would require that the frequency increase (or decrease) homogeneously in all contexts of use of the said form. This seems highly improbable. It can happen, however, if some linguistic form is highly specific and so associated with only one very specific use (a given technological device, for instance). Then, if the frequency of the form increases or decreases, it will do so in relation to this specific use only, and so its meaning will be left unchanged (e.g. *typewriter*, which suffered a significant drop in frequency since the nineteen eighties).

Under this assumption, semantic change has not to be assumed a priori to be found. Phonetic change would imply to first identify a change, and then to track down the two phonetic forms. In semantic change, even if we don't know that a change has occurred to begin with, we can identify shifts anyway thanks to frequency changes.

This a major advantage of semantic change, as it does not rely on an input from the investigator. Of course, the actual occurrence of a semantic change can be checked a posteriori through linguistic expertise, but it is nonetheless interesting to be able to detect a possible semantic change from the mere datum of the frequency of use.

Focus on functional semantic change

It should be immediately stressed, however, that we did not investigate just any semantic change. We rather focused further on semantic changes towards a more functional use — in other terms, mostly, but not exclusively, on grammaticalizations. Apart from being at the core of our work, this comes with interesting advantages. For instance, contrary to the use of a word in a new context, which can be initiated through mere trend, grammaticalization affects items which are not quite noticeable and thus not very prone to be used in such a new, shiny way. Also, it would be very odd for a linguistic institution to grammaticalize a linguistic item by reforming the language. Orthography, as it is but a convention, can be reformed by such means, but Semantics, not quite so. An academic institution can sanction an on-going grammaticalization, but it will take act of the change, not initiate it.

Grammaticalization, or more generally, semantic change towards functionality, is thus a ‘cleaner’ language change, as it chiefly involves mechanisms which are specific of language change. But there are very many semantic changes which are, similarly, not ascribable to a temporary trend, or an external influence. The remarkable semantic expansion that *weird* followed, for instance, does involve, as any language change, a strong social diffusion component, yet it can still perfectly be considered as prototypic of semantic change. Albeit its contemporary use comes from a successful drama piece by Shakespeare, namely *Macbeth* and its infamous ‘Weird Sisters’ triad of characters, the semantic expansion that followed was in no way enforced by that work, and indeed happened about two centuries later. The mechanism of this expansion is actually typical of the ‘green sun’ effect evidenced by Tolkien (2014).

There are similarly many other semantic expansions which, while they involve specifically linguistic motivations, fall most clearly outside of the scope of grammaticalization. This would be the case of the word *courage*, which shows a tenfold increase of frequency through the xvth and the xvith centuries (Fig. 4.6). The semantic history of this word is actually quite complex, and it gained and lost uses all across its history. In the late xivth century, *courage* has become a near-synonym of *vaillant* (‘brave’), the two words being frequently used as a pair; but in the xiiith century, the meaning of *courage* was more complex and multifaceted, with a meaning still often related to the etymological *cœur* (‘heart’), e.g.:

Promis t’avoie en mon courage Que je feroie le voiage
I had sworn to you in my heart that I would do the trip.

DE DIGULLEVILLE, Guillaume, *Le Pèlerinage de vie humaine*, p.94b, 1330-1331
(Frantext)

Gradually *courage* probably ceased to be a personal attribute to become an independent, sometimes measurable quality, which can be given and taken (contrary to *courage* as related to the heart of a person, unless we live in the Temple of Doom). It came to be associated with expressions such as *prendre courage*, *donner courage*,

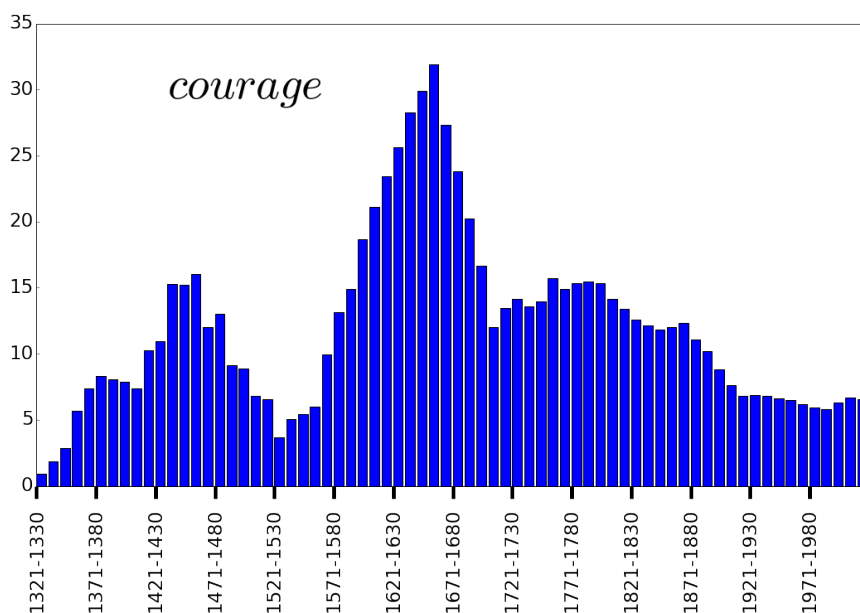


Figure 4.6: Frequency profile of *courage*, after average. Frequency is measured in number of occurrences of the form per a hundred thousand words.

reprendre courage, *redonner courage*, and also *encourager* ('to infuse with courage'). Indicative of this shift, we start to find occurrences such as *faute de courage* ('for lack of courage'), since 1578. The idiomatic *homme de courage*, which is incompatible with the personal meaning of *courage* shown in expressions as *en mon courage*, also appears in 1456 in Frantext but becomes frequent only after 1601. As it becomes a quality whose intensity can vary, it makes also sense to say that someone have *beaucoup de courage* ('a lot of courage', since 1593; one can suspect that no occurrences of this kind is found in earlier centuries because *beaucoup* itself is not completely established at these times, but we also searched *moult*, *mout*, *mut*, *mult* courage and found nothing).

All those signs are truly indicative of a semantic shift which coincides with the frequency rise. Therefore, focusing on functional changes, which would leave aside a case such as the previous one, certainly amounts to discard many interesting and relevant instances of language change. There is, however, a good reason to do so: to track down the frequency of a rich content word is tricky and subject to biases, especially in corpora of reduced size based on literary production. Indeed, a lexical item might be the topic of a given book, and reach then hundreds of occurrences, making for an unreliable data point representing the decade in which it appears. True, our averaging procedure tends to weaken this kind of peak, but sometimes the gap between the ordinary use of a word and its use in one single specific work is so huge that it cannot be fixed, unless we remove it entirely. Yet it raises the question: beyond which threshold must a text be removed from the corpus? How can we distinguish an individual, isolated choice, from an actual trend? And at which point does this removal procedure lead to results which are, in a way, specifically designed to conform as the expected pattern?

We are to take an example of this dilemma with the frequency evolution of the

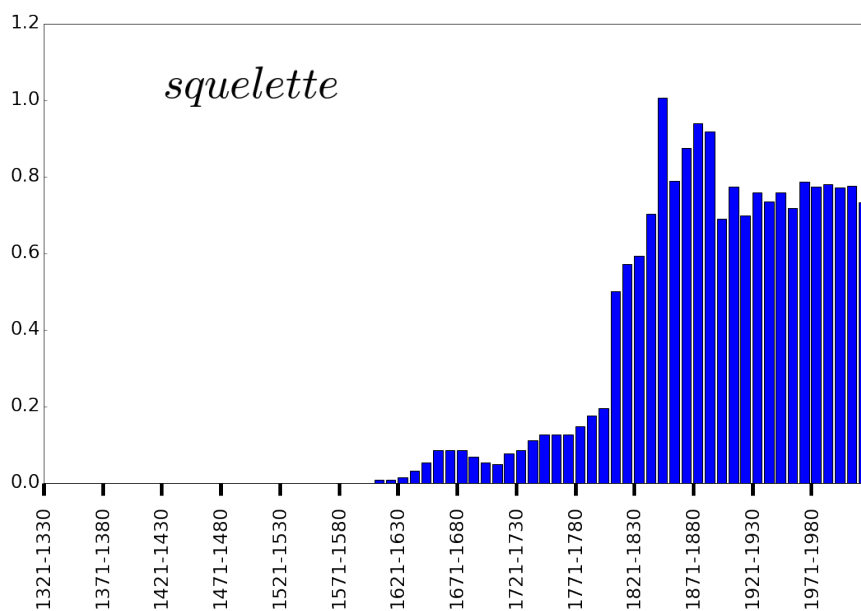


Figure 4.7: Frequency profile of *squelette*, after average. Frequency is measured in number of occurrences of the form per a hundred thousand words.

word *squelette* (Fig. 4.7). There are two peaks of frequency, which both affect the five next data points because of the averaging procedure. The first one, in 1801-1810, is due to the *Leçons d'anatomie comparée*, by Georges Cuvier (a naturalist) and to the *Philosophie zoologique*, by Jean-Baptiste Lamarck (another naturalist), totalizing respectively 33 and 34 occurrences of *squelette*. It must be said that the decade 1801-1810 is one of the worst in Frantext, for it is loaded with these two enormous texts of a very peculiar nature, which account for about one fourth of all occurrences of that decade. What is more, they are written in an academic style and tend to use over and over the same idiomatic formulae, so that their contents is much less diverse than would be a novel of equivalent size. The second peak is due to *Les Mystères de Paris*, by Eugène Sue, which stages a ruthless assassin named 'Le Squelette'. As *Les Mystères de Paris* is a particularly big novel (700 000 words), we get for this sole document 155 occurrences of *squelette* (to be compared with the 1842 occurrences over the seven centuries covered by our corpus), which explains the peak, and why the four next data points are messed up in the wake.

Focusing on semantic change towards a functional meaning shields us from such a problem, for a functional construction is not likely to be the main character of a book. Of course, this possibility is not entirely ruled out (especially with the repetitive idiomatic sentences of Cuvier in the 1801-1810 decade), but it should be less a worry for this family of semantic changes.

Another advantage of functionally-oriented semantic changes is that, as we saw in chapter 2, it is very often (if not mandatorily in the case of primary grammaticalizations) accompanied with a constructionalization, meaning that the syntactic behavior of the form itself is changing, or that a new functional form is emerging. For instance, *par ailleurs*, once grammaticalized, comes to be used in the beginning of sentences;

the functional *à la base* is not followed by *de* anymore; *à cause de*, *en vertu de*, *partant que*, emerge out as grammatical forms. In all those cases, the formal change and the semantic change are consubstantial, so that a rise in frequency of the newly forged form is necessarily indicative of its semantic expansion. Contrary to cases such as *courage* where evidencing the semantic change required to dive into the occurrences and to provide some detailed investigation of the contexts of use, the reality of semantic expansion is, in those cases, granted. The price to pay is that we lose the ‘a priori’ character of semantic change (not entirely though, for a functional form is also subject to further semantic expansions which will be detected in some cases through subsequent frequency rises; but then there is no certainty that the frequency rise is indeed a semantic expansion, no matter how likely such a hypothesis can be).

Functional vs grammatical: a short terminological note

We owe, at this point, to briefly stress the contrast that we want to introduce between ‘grammatical’ and ‘functional’. In the previous chapter, we saw that scholars quite often freely alternate between the two, in order to distinguish their use of the term ‘grammatical’ from other possible acceptations (e.g. from the idea that ‘grammatical’ is whatever is sanctioned by acceptability judgments from competent speakers). Here, as we proposed quite a specific notion of grammatical (a construction is grammatical if its meaning cannot be explained without making reference to its linguistics status), and as we want to consider a broader array of changes than mere grammaticalizations (which, anyway, cannot be strictly identified as such in mots cases), it seems important to state carefully the difference that we make between the two notions.

First of all, any linguistic form can be functional, while only constructions can be said to be grammatical. For instance, *le lendemain* (‘the following day’) is a simple noun phrase, which is, by all accounts, a fairly standard example of what a lexical item can be. Its semantics presents no specific reflexivity towards its own structure or the utterance. Yet, it can serve functional uses, as in the sentence:

Et l’endemain, des qu’il sera cler jour, il doit retourner a ses brisiees et
requerir son cerf, quar j’ay veü prendre trop de cerfs **le lendemain** qu’ilz
avoient esté failliz le jour devant.

And the following day, when it will be daylight, he must return to the
tracks and search for the deer again, for I have seen too many times deers
caught the next day, whereas they had been missed the day before.

PHÉBUS Gaston, *Livre de chasse*, 1387, p. 202 (Frantext)

This functional use is made only possible through a specific argument structure construction, which we shall roughly name the ‘adverbial’ construction for the sake of the discussion. This argument structure is used to insert almost any circumstantial specification either at various locations within the utterance structure (as for the question if it each different location counts as a separate construction, we leave it aside). Within this construction, to which *le lendemain* shows a very strong colligation preference, this linguistic form assumes an operative function in the utterance, providing a component of its orientation in the sense of Chafe (2002).

As such, *le lendemain*, which has emerged as a noun in the XIVth century, can be said to have known a semantic expansion (through the lexicalization of *l'en demain*). It is, certainly, associated with a functional use, yet it is not a grammatical construction, and its diachronic evolution is not an instance of grammaticalization.

Another example which highlights the broader scope of the notion of a ‘functional’ item is given by recurrent constructs in grammatical constructions. For instance, many nouns in French, such as *foule*, *tas*, *soupçon*, *flopée*, are all lexical items, yet they are quite conventionally enrolled in the quantifier construction [un/une/des {N}₁ de {N}₂] (e.g. ‘une tripotée de fêtards’), which can itself be safely considered as grammatical. The linguistic forms appearing in the first noun paradigm {N}₁ serve, through the grammatical quantifier construction, clear functional purposes, yet they are, by any account, lexical.

As the colligation of these lexical nouns to the quantifier construction increases, they are all subject to a semantic expansion towards new functional uses; yet none of them becomes grammatical independently of the construction they figure in, nor these evolutions can be identified as grammaticalizations. This is true for many recurrent constructs which have not been subject to a further speciation process. In general, the apparition and possible entrenchment of a new construct of a given grammatical construction will lead to a semantic expansion towards functionality of its components, even though these components are not themselves grammatical.

Semantic change and semantic expansion

In the preceding part, we used alternatively the expressions ‘semantic change’ and ‘semantic expansion’, as if they were equivalent. For what we are interested in, they might just be. Semantic change occurs only in two cases: because a form gains new meanings, or because it loses some. Since in the latter case is not associated with any frequency rise, but with frequency drop (which we do not study in this work, though they have an interesting and non-trivial phenomenology), the changes that we grasp through our extracting pattern procedure will be instances of semantic expansions.

There are also less trivial reasons. In a semantic change, the meaning of the form is seldom, if never, translating over some semantic territory (in the geometrical sense of the term). Usually, the form retains its source meaning. It might then lose it in the process, but this is never instantaneous, and new and old meanings can coexist for a long time period. In the case of *courage*, constructions such as *en mon courage*, related to the original meaning of *courage*, remain fairly frequent up to the second half of the XVIth century. Hence, it really is the case that the meaning has expanded, so that the form became more polysemic than it was before the change.

Also, if a detailed balance was kept between the gain of new meanings and the loss of old ones, then the frequency would have no particular reason to rise. The rise can only be explained because the dominion of the form over the semantic territory kept growing, making it compatible with an increasing number of contexts of use.

Finally, semantic reduction is probably a different phenomenon than semantic expansion, and it might not be entirely relevant to call it a semantic change. Indeed, semantic expansion requires links between meanings, rat-holes in the conceptual structure. Semantic reduction, on the contrary, is just a wearing off of what once has been.

There is no innovation involved, no associated unpredictedness. To predict a semantic expansion would be to predict a path of change in the semantic territory, which is not an easy task, in addition with all the dynamical details of the process; to predict a reduction would only amount to analytically describe the decline. A relevant question that can be asked (Aitchison, 2013, p.192), and that will only be superficially addressed in this work, would be to know whether a semantic reduction is due to a spontaneous, intrinsic tendency of the form, or would happen because of a semantic expansion occurring elsewhere in the language (so that the older form would be pushed off by the new one). In the first case, a specific mechanism would need to be posited, and semantic reduction would be a kind of semantic change in its own right, rather than a mere corollary of another semantic expansion.

4.3.2 A note on French

We acknowledge that we restricted ourselves to instances of semantic expansions in French, a choice which may appear to restrict the scope of our findings. Yet we believe this is not the case. We already argued that few corpora are as efficient as Frantext to achieve such a goal. Spanish or English could have been studied as well, but those two languages are closely related to French, both historically and phylogenetically. Though Spanish, being a Romance language, would be much closer in that regard, it is also well-known that English has been hugely influenced by French, especially in terms of vocabulary (Minkova and Stockwell, 2009), to the point that, except for the most frequent words, the English vocabulary can be said to be inherited more from Middle French than from Old English (Williams, 1975, p.67), but also in the more functional realm of prepositions (Corisco, 1997), and morphology (Dalton-Puffer, 1996). Therefore, if we suspect the focus on French to be too much language specific, there is little point in considering languages as close to French as English or Spanish. Studying semantic change in Chinese, for instance, would have been much more interesting, but the only substantial diachronic corpus of Chinese, the Sheffield corpus, contains less than 500,000 words, making it a thousand times smaller than Frantext, COHA, or CORDE.

Also, it seemed more crucial to consider a long time period, rather than different languages. Indeed, a frequently asked question is whether or not recent technological advances (radio, TV, the Internet) had an influence on the way language changes. Sociologically, this influence is obvious: languages tend to homogenize over greater geographical areas and dialects have constantly declined throughout the twentieth century. Yet, the pattern of change of an established language is something entirely different. Our statistical survey shows that the pattern of change is the same, no matter in which century it may happen. It is furthermore consistent with recent findings establishing that the rate of change did not increase in the most recent decades (Dubossarsky et al., 2016). It also goes along our claim (in chapter 7) that the pattern we exhibit is cognitively driven by memory retrieval and conceptual organization, two cognitive mechanisms that the most recent technological evolutions could not have significantly altered.

The time period we studied covers seven centuries, and at least three diachronic varieties of French, Middle French, Classical French and Modern French. Middle French,

especially from the XIVth, is only partially understandable for Present-Day speakers; in sentences like the following one:

Si a tantost un message tremis a Rome a son pere, et d'ilecquez a Ardee a son mari, qui estoit en l'oust, et leur a mandé que chascun de eulz avecquez un seul et loial compaignon venissent a lé hastivement, car besoing estoit, et que une chose espoentable li estoit avenue.

BERSUIRE Pierre, *Les Décades de Titus Livius*, 1354 (Frantext)

the general meaning can easily be retrieved (a woman has sent a message to her husband and her father to summon them on the ground that something terrible happened to her), but some syntactic constructions such as *si a tantost un message tremis* do not exist in Present Day French, and words such as *ilecquez* have completely disappeared from use. Versified texts, of which we have plenty for those times, are still harder to process. We can thus rightfully consider that the language covered is not homogeneous; yet we found no significant differences between the different grammaticalization cases, even though there are strong dissimilarities among those time periods from a data quality perspective.

A last remark is in order: We deliberately do not provide any English translation of the studied forms (listed in Appendix A). Indeed, these forms have all undergone a semantic expansion, so that a translation would be most mistaking as it would concern only one among several meanings adopted by the form. The only satisfying way of glossing the items we studied would have been to find forms which not only have the same meaning, but have also undergone (at least roughly) the same meaning shifts, as in the case of *anyway* and *de toute façon* for the later stages of their respective semantic evolutions (albeit they have different lexical origins, they developed the same pragmatic uses out of their previous functional ones). Obviously, this would have been possible only for a handful of cases, so we left the matter unattended.

4.3.3 How were the forms chosen?

To provide statistical results about the S-curve, we investigated data frequency from about four hundred of different linguistic forms. They were chosen so as to be functional, and most of them are constructions in the pre-theoretical sense of the term (i.e. they are made up of different words), complex constructions in the Construction Grammar terminology. This choice was motivated by the fact that complex constructions necessarily arise out of a constructionalization process, which, by its very definition, involves a semantic change. The frequency of these constructions was usually different from zero even before the start of the process, for there are constructs which have exactly the same surface form (and which, in fact, are required for the constructionalization to happen).

We chose mostly functional forms still in use nowadays (with a few exceptions: *fors*, *aucunefois*, or *donc*), otherwise we would have been likely to witness only a disappearance of the form. Functional forms usually having a long life period, it would have meant that they arise quite early in the history of French. The rise of frequency was thus likely to have happened before the XIVth century. Too recent forms have been discarded, for they must be at least 60 years old for their evolution to fit into the pattern we were investigating.

The chosen forms cover all parts of the functional load of the language:

- pragmatic markers: *décidément, au pire, j'imagine, m'est avis, pour tout dire*
- discourse markers: *par contre, par conséquent, par ailleurs, sur ce thème, au bout du compte*
- temporal markers: *de jour en jour, dans un instant, à un moment donné, au jour d'aujourd'hui, à l'heure actuelle*
- spatial markers: *de ce côté, aux alentours, de proche en proche, d'outre en outre, en long et en large*
- other complex adverbials: *dans son ensemble, d'emblée, d'une voix {adj}, par à-coups, sans ambages*
- connectives: *vu que, tandis que, à mesure que, en sorte que, c'est alors que*
- temporal prepositions: *au bout de, en passe de, sur le point de, le temps de, au terme de*
- spatial prepositions: *à l'orée de, de point en point, en face de, à bord de, en bordure de*
- other prepositions: *à la lumière de, à base de, en guise de, de l'avis de, en dépit de*
- quantifiers: *un tas de, un surcroît de, faute de, tout plein de*
- intensifiers: *en particulier, notamment, très très, carrément, outre mesure*
- interjections: *ouille, que dalle, zut, voilà, bah*

and many others.

It must be stressed that not all investigated changes are instances of grammaticalization. Some of them would be best regarded as lexicalizations: *à tout prendre, chemin faisant, de mèche, lendemain, si ça se trouve*, etc. Some of them are repeated colligations of an item in a schematic grammatical construction or pattern, and do not necessarily correspond to new, grammatical constructions: *à la lumière de, faute de quoi sur ce thème, parmi d'autres, pour l'essentiel*, etc. We also kept one lexical word with no linguistic function whatsoever, *liberté*, to show that it could be treated in a similar fashion than functional semantic expansions. This would hint that our results are not specific to the particular class of semantic expansions we focused on, but could apply to the semantic expansion phenomenon in general.

Also, we did not want to restrict the choice of observed forms by a specific definition of grammaticalization. There are two reasons behind this: first, we expected some differences in frequency behavior to emerge, allowing to devise an empirical criterion to distinguish a subclass of functional semantic expansions (which one could have coined as 'grammaticalizations'). Unfortunately, we observed nothing of that sort and our results form a clear and consistent continuum. Second, if we had selected only what could be safely consider as grammaticalizations, then all empirical features emerging from our statistical survey would have been 'specific' of grammaticalization (only because we would not have looked at anything else). On the contrary, our conclusions, which we now present, hint to the fact that the empirical features of frequency do not allow to distinguish specifically a grammaticalization process from any other semantic process. This is a negative result, but a crucial one nonetheless: the specificity of grammaticalization must therefore lie in the specific semantic ties which channel the change, not in its frequency profile.

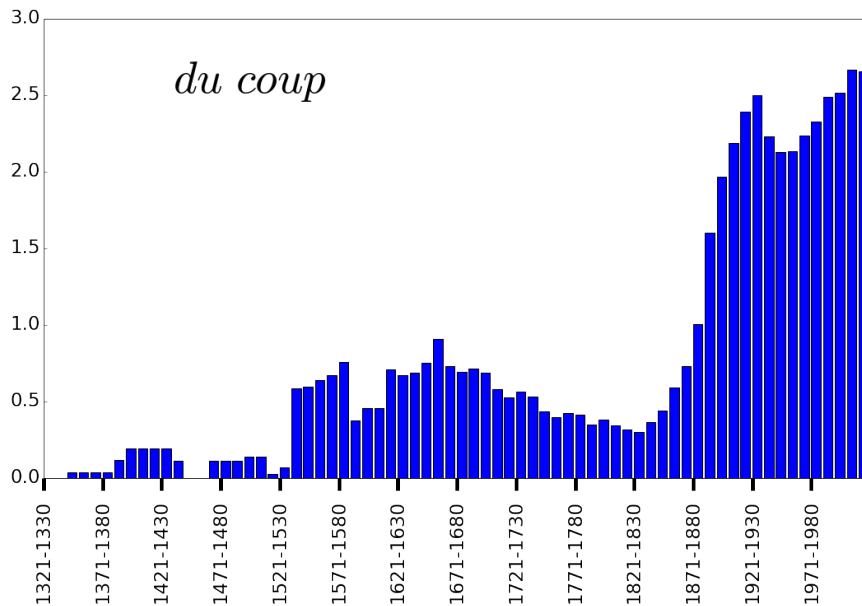


Figure 4.8: Frequency profile of *du coup*, after average. Frequency is measured in number of occurrences per a hundred thousand.

4.4 Qualitative results

Before turning to the statistical study of the S-curve based on 400 hundred cases of semantic expansions, we shall first consider a few chosen individual cases, most of them not included in the statistical survey, in order to take a closer look at the S-curve.

4.4.1 Case studies: *du coup*, *way too*

Since we already investigated the constructionalization of *du coup* as a connective marker of consequence, we know exactly the time at which the novelty started to appear in Frantext. We can thus compare the quantitative pattern automatically extracted from the complete set of the frequencies of the form over all decades (Fig. 4.8) —which is almost never zero, since ‘du coup’ appears occasionally as a construct, as we have seen at length in chapter 2— to the qualitative occurrence-based study we proposed, and see if they coincide. Remember that we identified the first occurrence of the grammaticalized *du coup* as dating from 1842.

Our algorithm detects several phases of significant growth: 1331-1400, 1451-1470, 1521-1540, 1601-1620 and finally 1831-1890, most of them being clearly spurious and imputable to the low number of occurrences. The form *du coup* being rare in the first studied centuries, it suffices to have a single occurrence in a decade to show a sudden increase of frequency. However, those growth parts are only helpful to restrain the search for interesting patterns. Actually, the second filtering (keeping only the pairs separated by at least four data points, such that all frequencies in-between are comprised between the two boundaries) selects all pairs between 1821 and 1930 (note that the search for the pattern goes beyond the boundaries of the growth part). The third filter selects the curves which are sigmoid shaped ($r^2 > 0.98$) and among them,

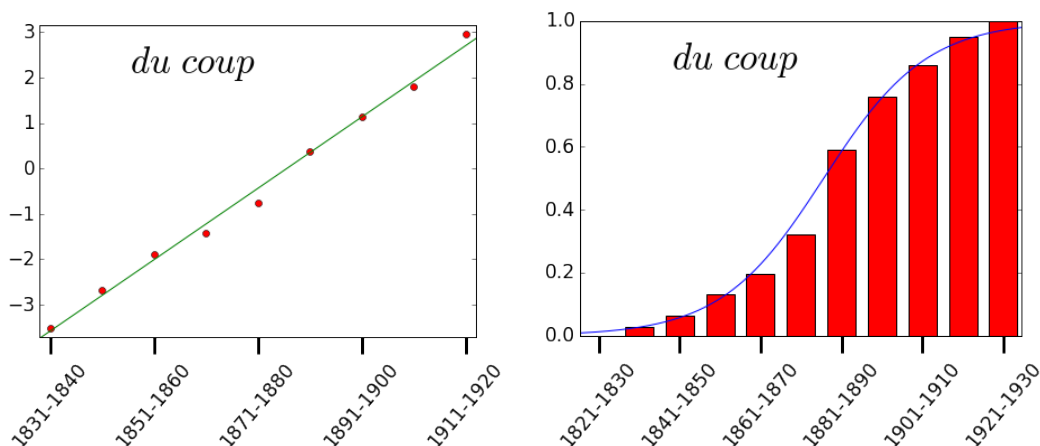


Figure 4.9: Left: logit transform of the extracted growth part of the frequency profile if *du coup*. Right: Corresponding sigmoidal fit of the extracted growth part.

the best one according to the total duration of the growth criterion and then to the r^2 criterion. We thus end up with a sigmoid covering the period 1821-1930 (Fig. 4.9).

The sigmoid starts with the decade 1821-1830, two decades sooner compared to what we expected from our past investigation. It seems thus that the semantic expansion was triggered by a prior increase of frequency of the form. However, the Frantext query *du coup* encompasses constructs such as:

Remerciez bien votre mère du coup de poing...

Thanks your mother for her punching me...

SAND, George, *Correspondance : 1831*, 1831 (Frantext)

where *du coup* appears as a construct of the construction [remercier de + {N}] ('to thank for'). Hence the search for *du coup* interferes with other processes going on at the same time: the semantic expansion of *coup de*, which also follows a sigmoid starting from the decade 1811-1820, and the semantic expansion of *coup d'œil* (a glance, a glimpse, a connoisseur look), both of those constructions possibly appearing preceded by *de*.

It should be thus best advised to refine the query, so as to only keep wanted occurrences of the grammatical *du coup*. The query:

$\hat{(\text{auteur|force}) \text{ du coup } \hat{(\text{de|d' |du|que|qu' |qui|dont|duquel})}$

solves this problem; yet it does not lead to a nice sigmoid, the best r^2 obtained lying around 0.965 (Fig.4.10). However disappointing this might be, this second picture is actually closer to what we discovered, that the occurrences of the grammatical *du coup* remained sparse even after the construcionalization was first evidenced.

We can now start to do the same with *way too*. However, the semantic expansion of *way too*, starting around the 1930s, does not lead to a frequency rise until the 1980s (Fig. 4.11). Hence, there are not enough data points to build the S-curve. This phenomenon of latency between the semantic expansion and the frequency rise is actually not uncommon and shall be discussed much later on.

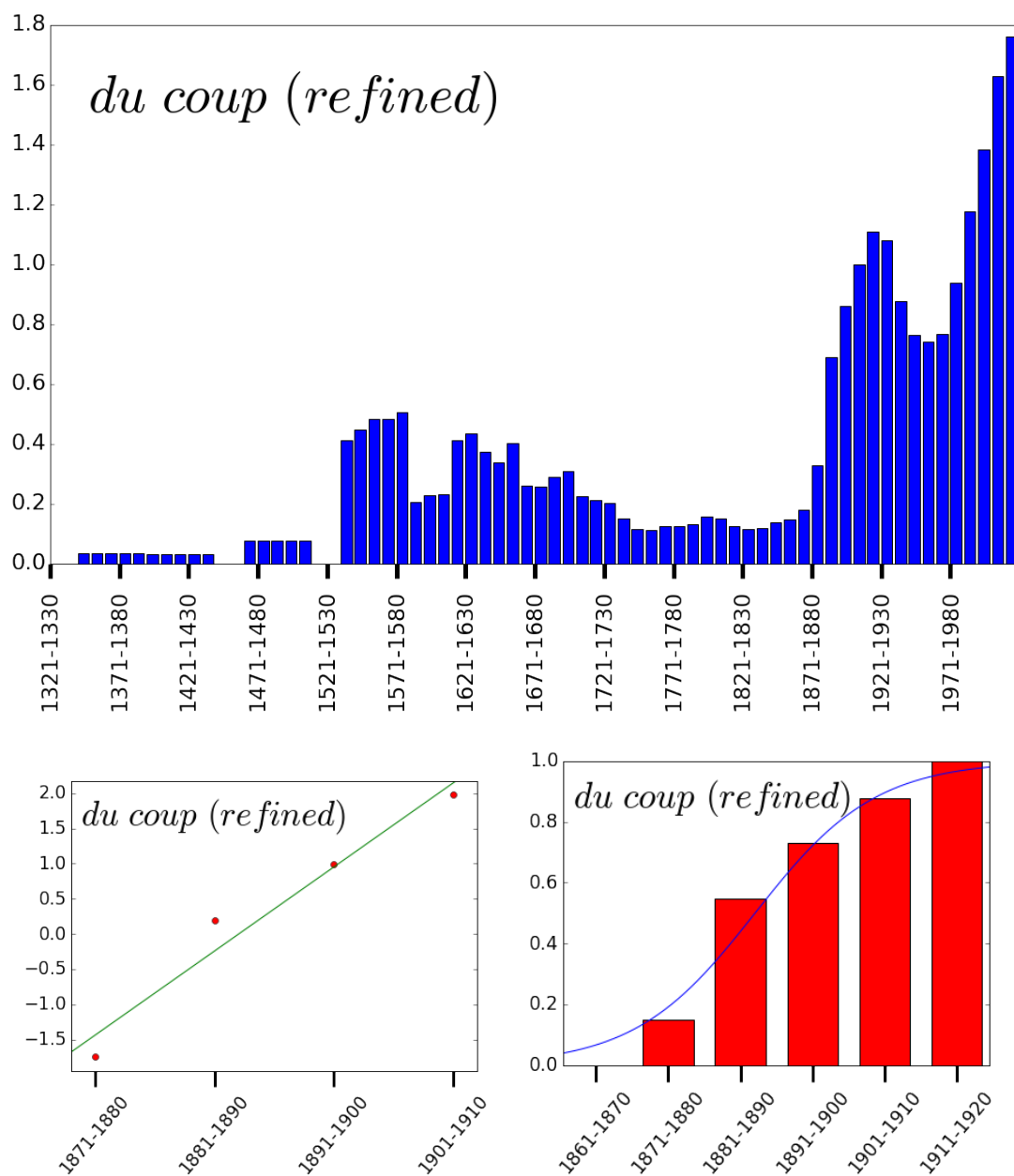


Figure 4.10: Top: frequency profile of extraclausal *du coup*. Bottom left: logit transform of the extracted growth part of the curve. Bottom right: Sigmoidal fit of the extracted growth part.

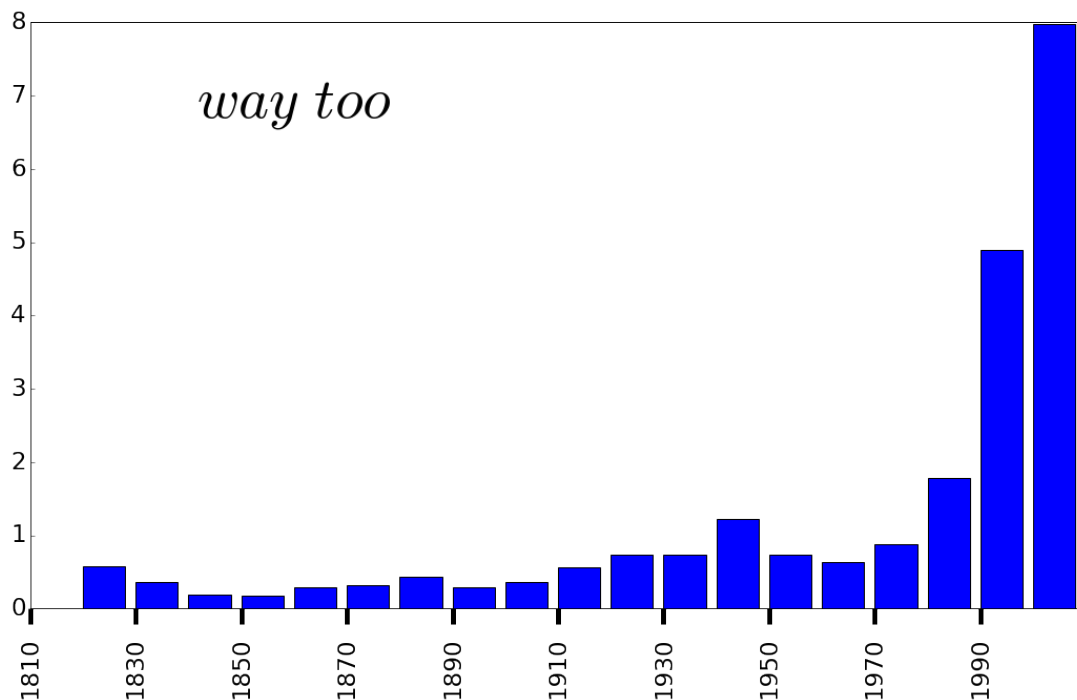


Figure 4.11: Raw frequency profile of *way too*, without any average performed over the data.

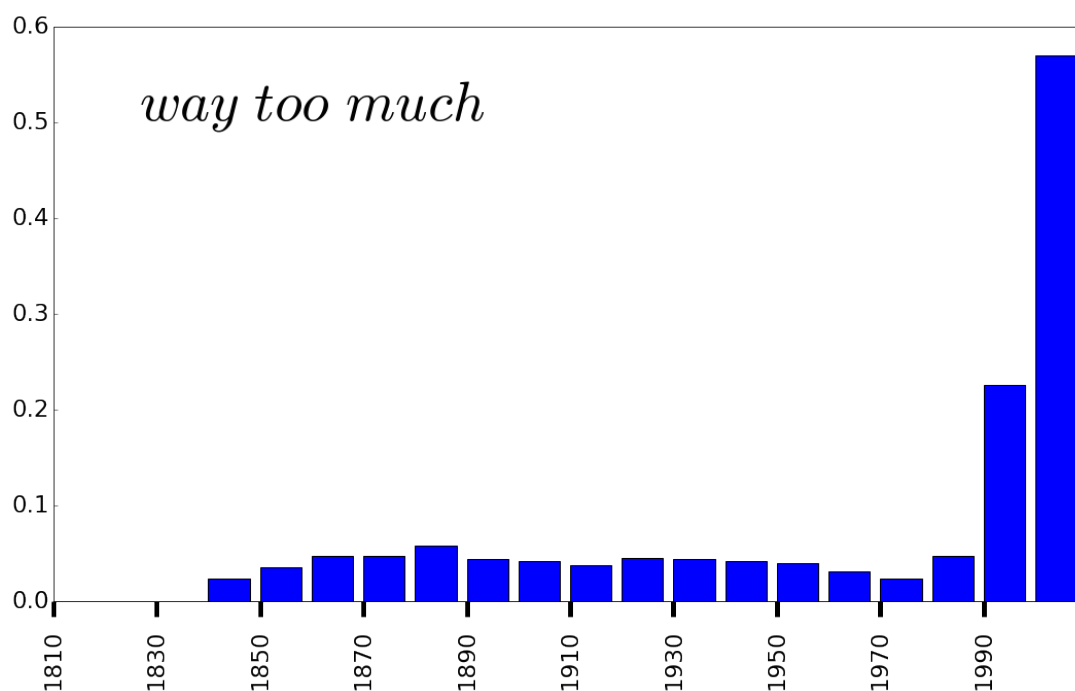
This behavior is confirmed by the study of *way too much*, cleaner than *way too* in that it should be associated with an exceedingly low number of spurious occurrences, and accounts for a good part of the uses of *way too* (97 out of 467 according to the COHA). The frequency remains indeed extremely stagnant up to the 1970s (Fig. 4.12).

4.4.2 Cases of interest

Those two examples might suggest that the S-curve is an approximation and is not to be found in actual data. Yet, we find striking examples of the S-curve in the corpora. For instance, *weird*, which we already presented, followed a semantic expansion leading to its use as an adjective meaning ‘odd, queer’. This semantic expansion is accompanied by a remarkably S-shaped rise of frequency (Fig. 4.13). That language change can follow so closely, for 11 decades, a mathematical curve, is astonishing. The frequency of *anyway* also follows a remarkable S-curve over the twenty decades of the COHA (Fig. 4.14).

As evidenced by the case of *weird*, a semantic expansion can be followed by another one, also sigmoidal (the second expansion not being completed by 2009). However, the restricted time window of the COHA prevents from observing this phenomenon at ease.

Very often there are interferences between different processes involving a given individual form, as we saw with *coup*. Hence, a detailed request and a more general one, albeit the occurrences of the first will be included in the second, may lead the algorithm to focus on different time periods. For instance, if we are looking for *une es-*

Figure 4.12: Frequency profile of *way too much*, averaged.

pèce de ('a kind of', (une|ungne|un|ung) (espece|espèce) (de|d')), then we will find a sigmoid describing the constructionalization of the form in the late XVIth century (Fig. 4.15), while the query (espece|espèce), while showing a peak of frequency correspond to *une espèce de*, will extract a later pattern (Fig. 4.16), corresponding to the semantic expansion of the term in the XVIIIth century, which indeed correspond to a renewal of the scientific and philosophic conception of the notion of species (Morange, 2016, pp.94-100). It is also remarkable that *une espèce de* is not sensitive to this frequency evolution of *espèce*, showing that it indeed acts as an independent construction, a separate linguistic form with a life of its own, as would consider Construction Grammar. This also comforts, at least in this case, the idea that functional forms are not quite likely to follow the cultural shifts of the society in which they are used.

However if we look for:

$$(\text{un|ung|une|ungne})(\text{espece|espèce})|(\text{un|ung|une|ungne})(\text{espece|espèce}) \text{ } ^{\wedge}(\text{de|d'})$$

which excludes all occurrences of *une espèce de*, it appears that the frequency peak corresponding to the rise of *une espèce de* still remains, indicative that a semantic expansion was concomitant to the entrenchment of this construction. However, the semantic expansion of the XVIIIth still remains the more outstanding pattern of this complex and multifaceted semantic evolution (Fig. 4.17).

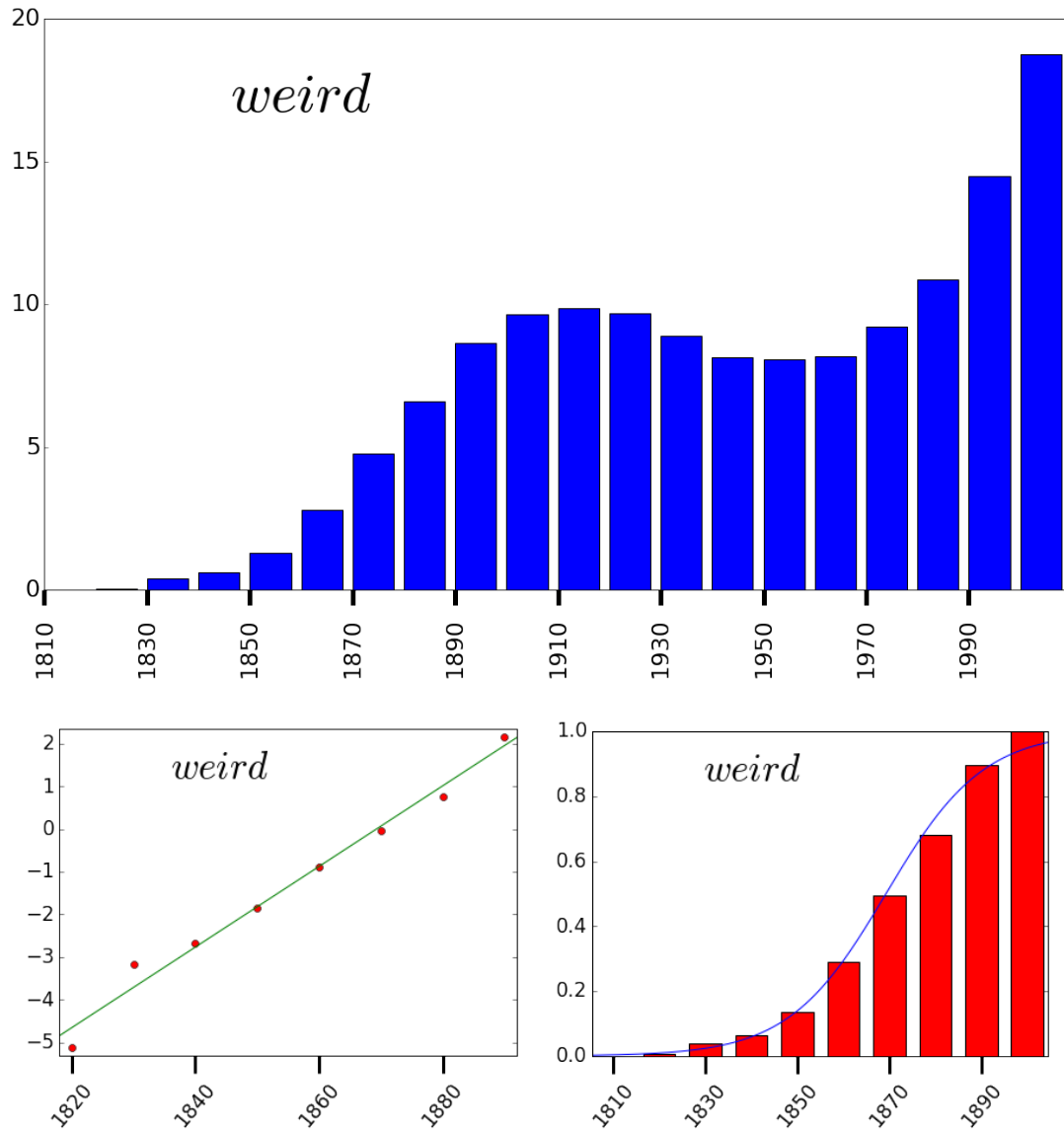


Figure 4.13: Top: frequency profile of *weird*. Bottom left: Logit transform of the extracted growth part of the curve. Bottom right: Sigmoidal fit of the extracted growth part.

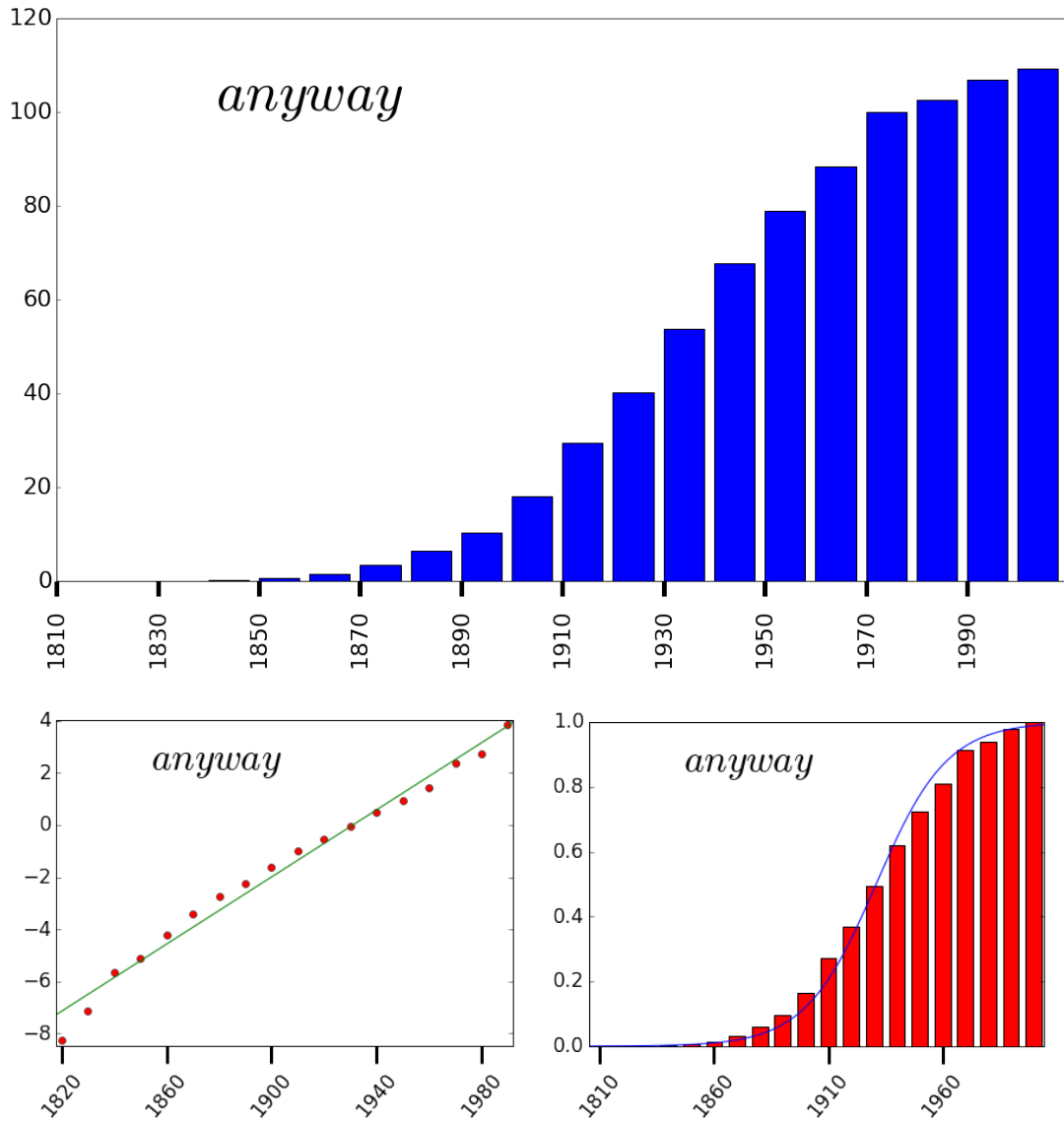


Figure 4.14: Top: frequency profile of *anyway*. Bottom left: Logit transform of the extracted growth part of the curve. Bottom right: Sigmoidal fit of the extracted growth part.

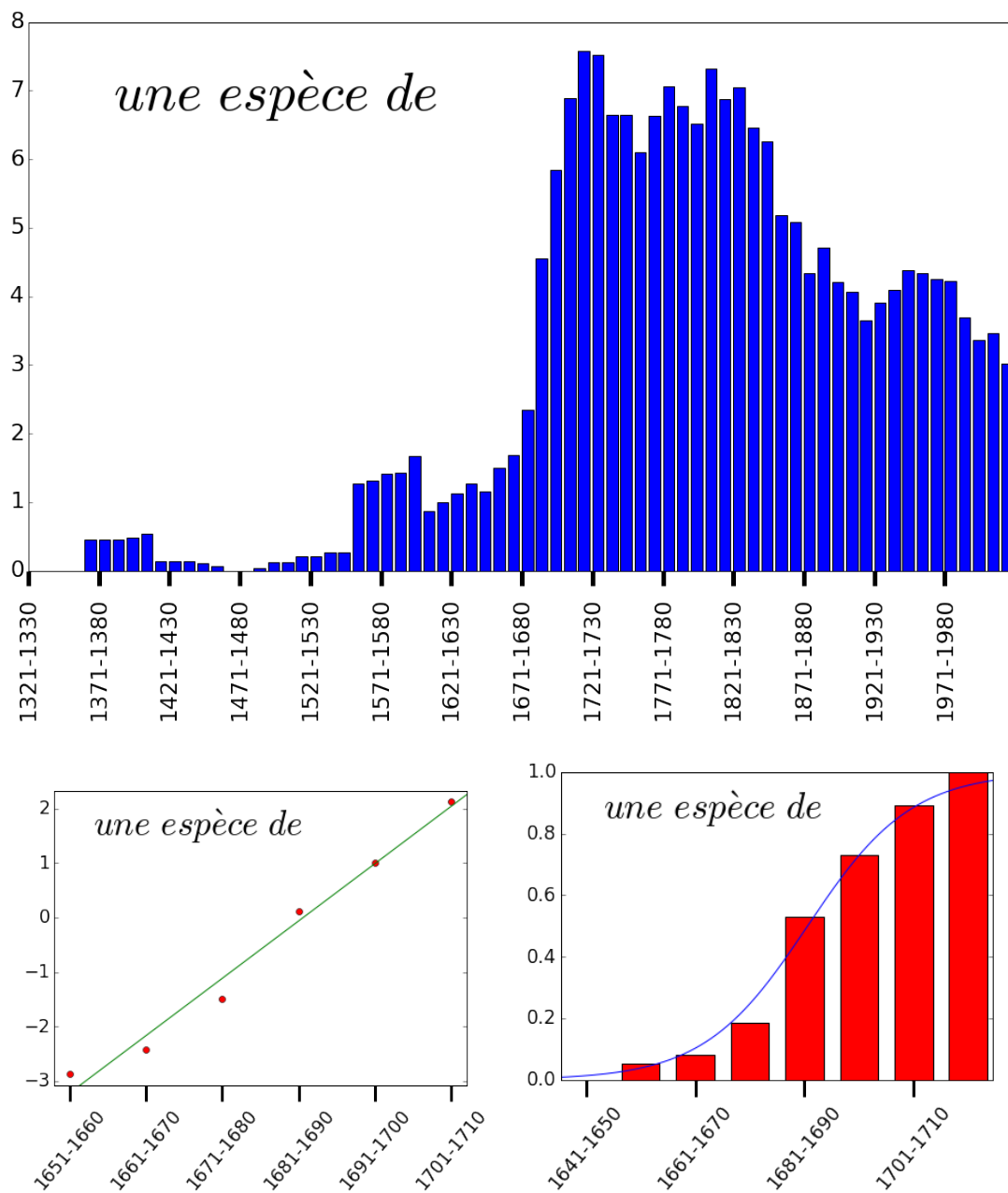


Figure 4.15: Top: frequency profile of *une espèce de*. Bottom left: Logit transform of the extracted growth part of the curve. Bottom right: Sigmoidal fit of the extracted growth part.

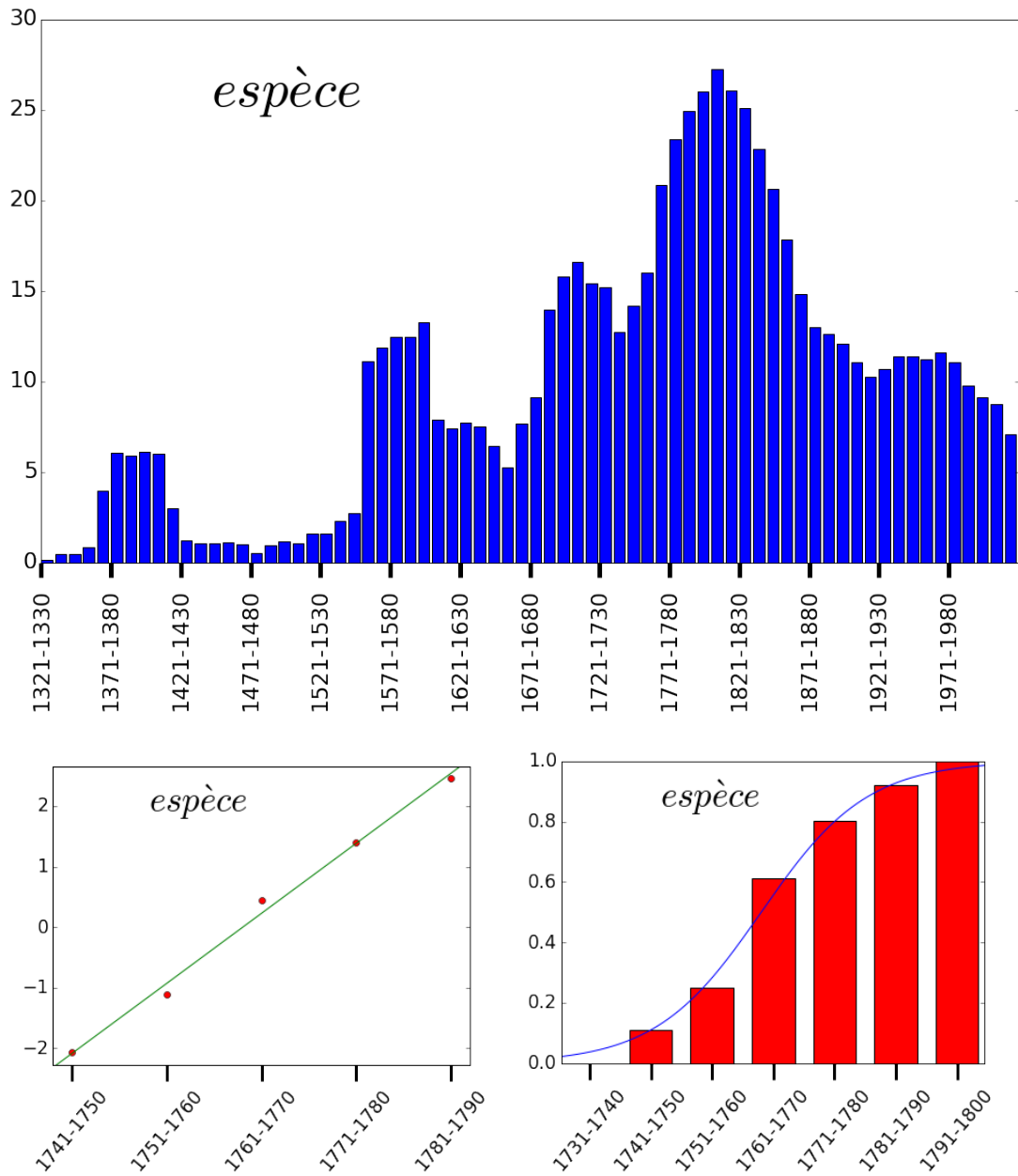


Figure 4.16: Top: frequency profile of *espèce*. Bottom left: Logit transform of the extracted growth part of the curve. Bottom right: Sigmoidal fit of the extracted growth part.

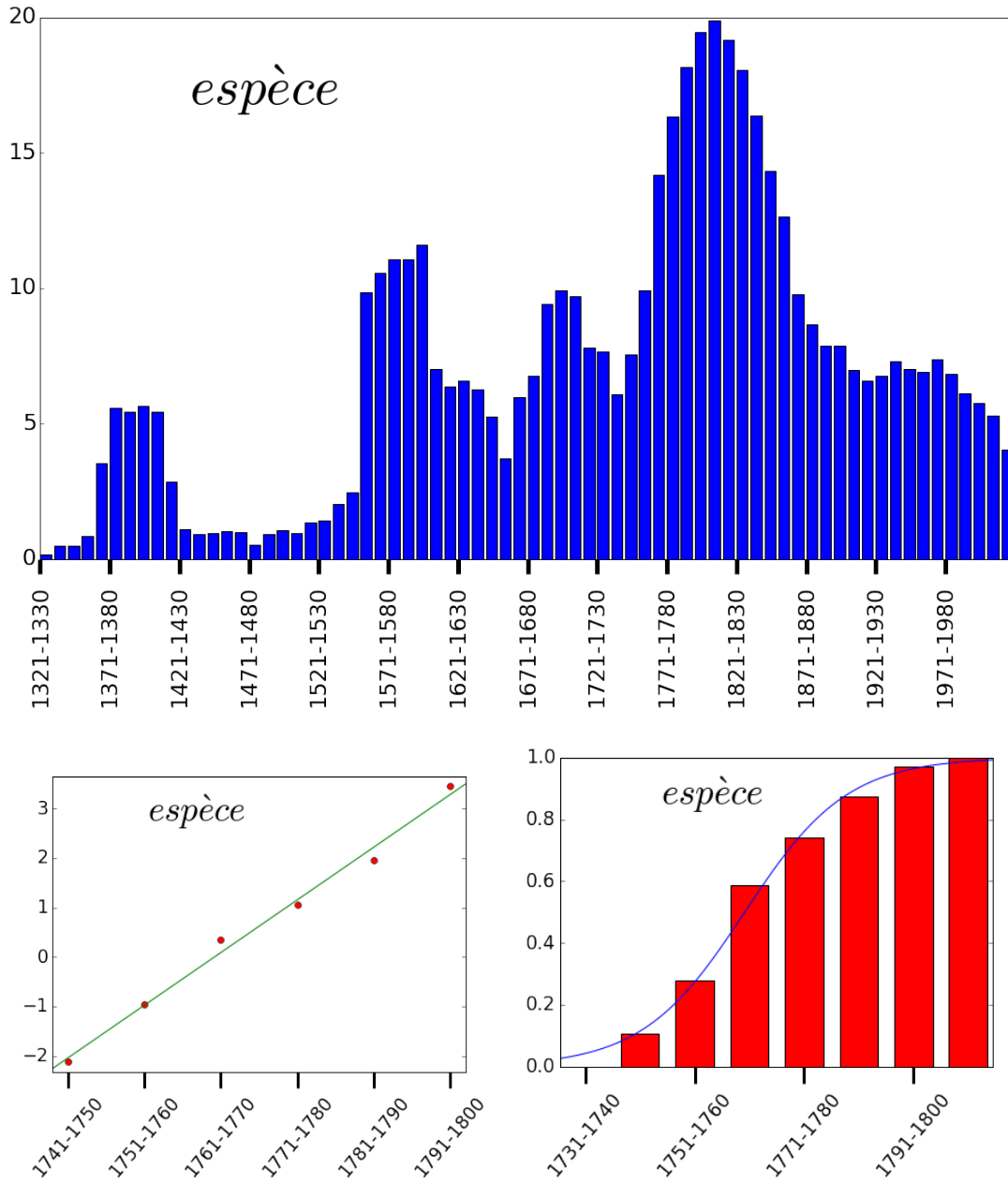


Figure 4.17: Top: frequency profile of *espèce* to which have been excluded all instances of *une espèce de*. Bottom left: Logit transform of the extracted growth part of the curve. Bottom right: Sigmoidal fit of the extracted growth part.

4.4.3 S-curves in other linguistic domains

Actually, the S-curve pattern is extremely pervasive and can be found in the frequency evolution of very different linguistic forms. We already encountered it in the lexical realm, but there are still other domains in which the S-curve is still a good picture of change. One are proverbs. The problem with proverbs is that they are, most often, very infrequent compared to words or constructions, so their study is not possible. Yet, our method managed to find an S-curve conforming to all our criteria for at least one: *On ne fait pas d'omelette sans casser d'œufs* ((omelette|omelettes) sans &q(1,3) &moeuf), meaning that nothing is achieved without some amount of collateral damage (also, there's no free lunch), which also exists in English, as a translation from French. The pattern extracted (Fig. 4.18) is not as clean as the examples we provided previously, and, considering the few decades on which it runs, and the extremely low number of associated occurrences (29), it might be due to sheer luck. Yet, we found near sigmoids for many other proverbs. Thus, proverbial expressions, as any linguistic constructions, are expected to ride S-curves of frequency rise.

We also looked for proper names. The frequency rise of a proper name can be due to a cultural evolution, or to the societal popularity of the individual. Compared to functional words, very few lead to clean and clear S-curves according to our criteria, yet it was expected, as we explained while justifying our preference for functional forms. Proper names are too much subject to idiolectal and cultural variations, and so the frequency is more spurious than it is for the more innocuous functional words. Yet, we found at least two cases in which our procedure managed to extract an S-curve. The first one is for the French polemicist *Voltaire* (Fig. 4.19). Interestingly, the rise started with the decade 1721-1730, while he was still pretty much alive, for he died in 1778. Actually, the peak of frequency (non-averaged) comes with the decade 1771-1780. It is thus safe to assume that this S-curve is due to a purely social phenomenon and reflects the high popularity of Voltaire during his life.

Our second example is more interesting. It is given by the famous Gaulish war-chief, Vercingétorix, who fought against César, as the latter himself related in the seventh book of his *De Bello Gallico*. The name shows up first in the corpus with the decade 1801-1810, and rises from 1881 to 1940, roughly following the time of the French Third Republic (1870-1940), and culminating with the Second World War (Fig. 4.19). This, of course, did not happen by chance, and reflects deep cultural tropisms, reeling with patriotism and carried by a phantasmal vision of History. In this case, the S-curve is thus pretty revealing.

Interestingly, the S-curve has often been regarded as the depiction of a competition between two variants. In the two previous cases, proverbs and proper names, it seems nearly impossible to think of proper competitors. Though one might be curious to know which other noble figure Vercingétorix might have overthrown to achieve its popularity, it seems dubious that such a one-to-one replacement actually happened. This observation at least incidentally suggests that the S-curve may occur in contexts which have little to do with competition, and thus might require alternative explanations.

Interjections are also worth noticing, for they follow S-curves too when they rise into usage. The interjection *oh* (Fig. 4.21), for instance, describes at least two S-curves

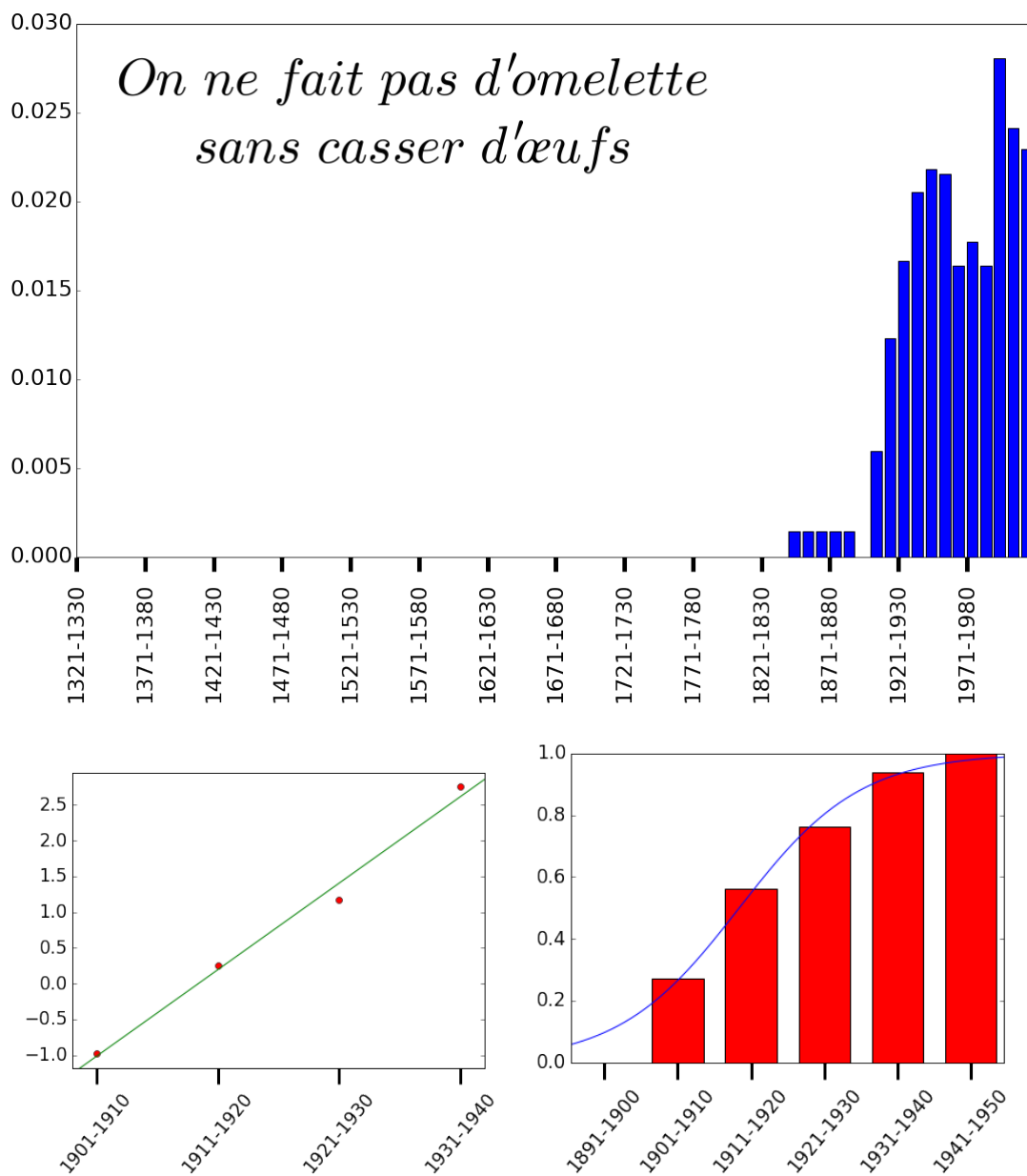


Figure 4.18: Top: frequency profile of *on ne fait pas d'omelettes sans casser d'œufs* and its possible variants. Bottom left: Logit transform of the extracted growth part of the curve. Bottom right: Sigmoidal fit of the extracted growth part.

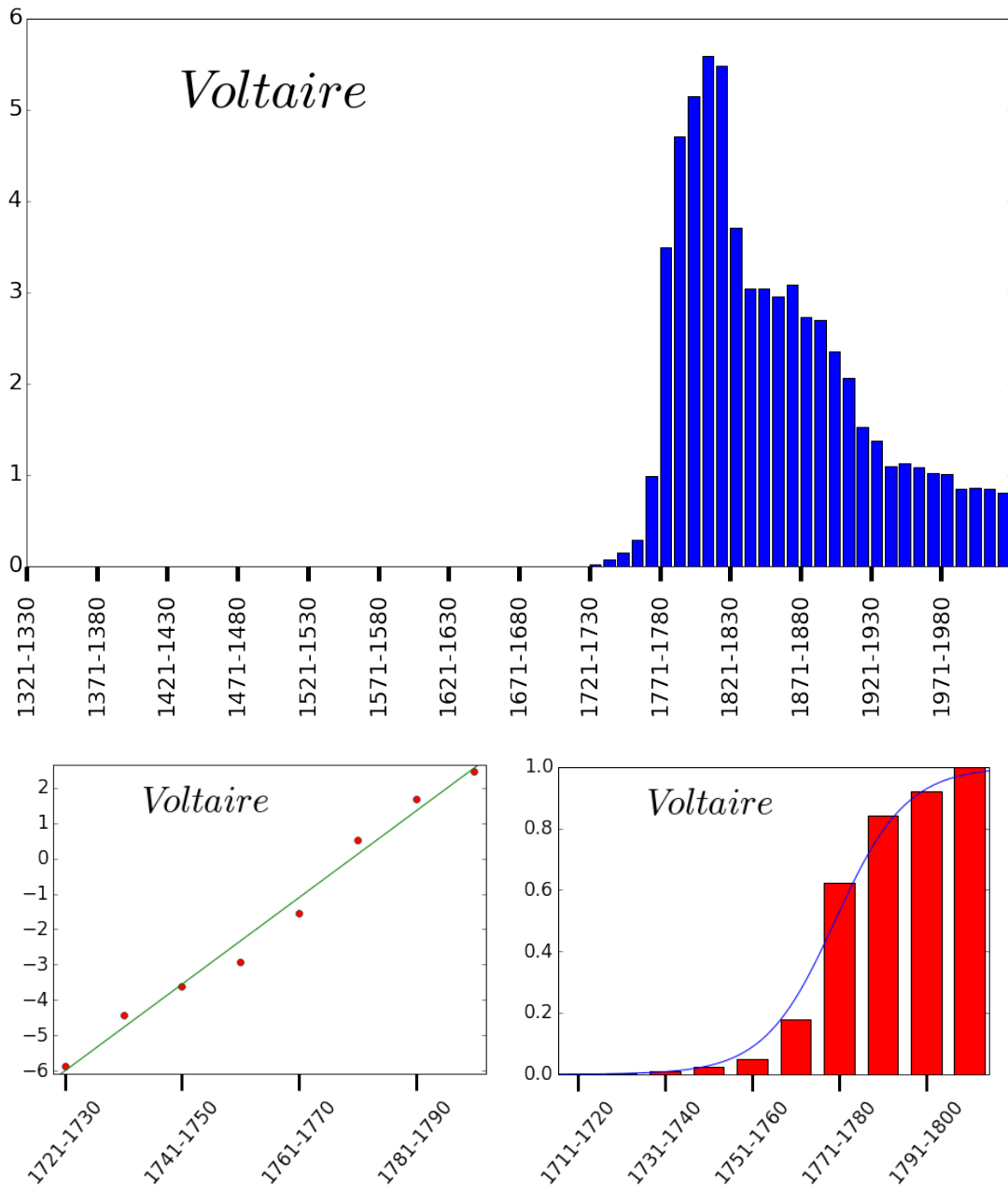


Figure 4.19: Top: frequency profile of *Voltaire*. Bottom left: Logit transform of the extracted growth part of the curve. Bottom right: Sigmoidal fit of the extracted growth part.

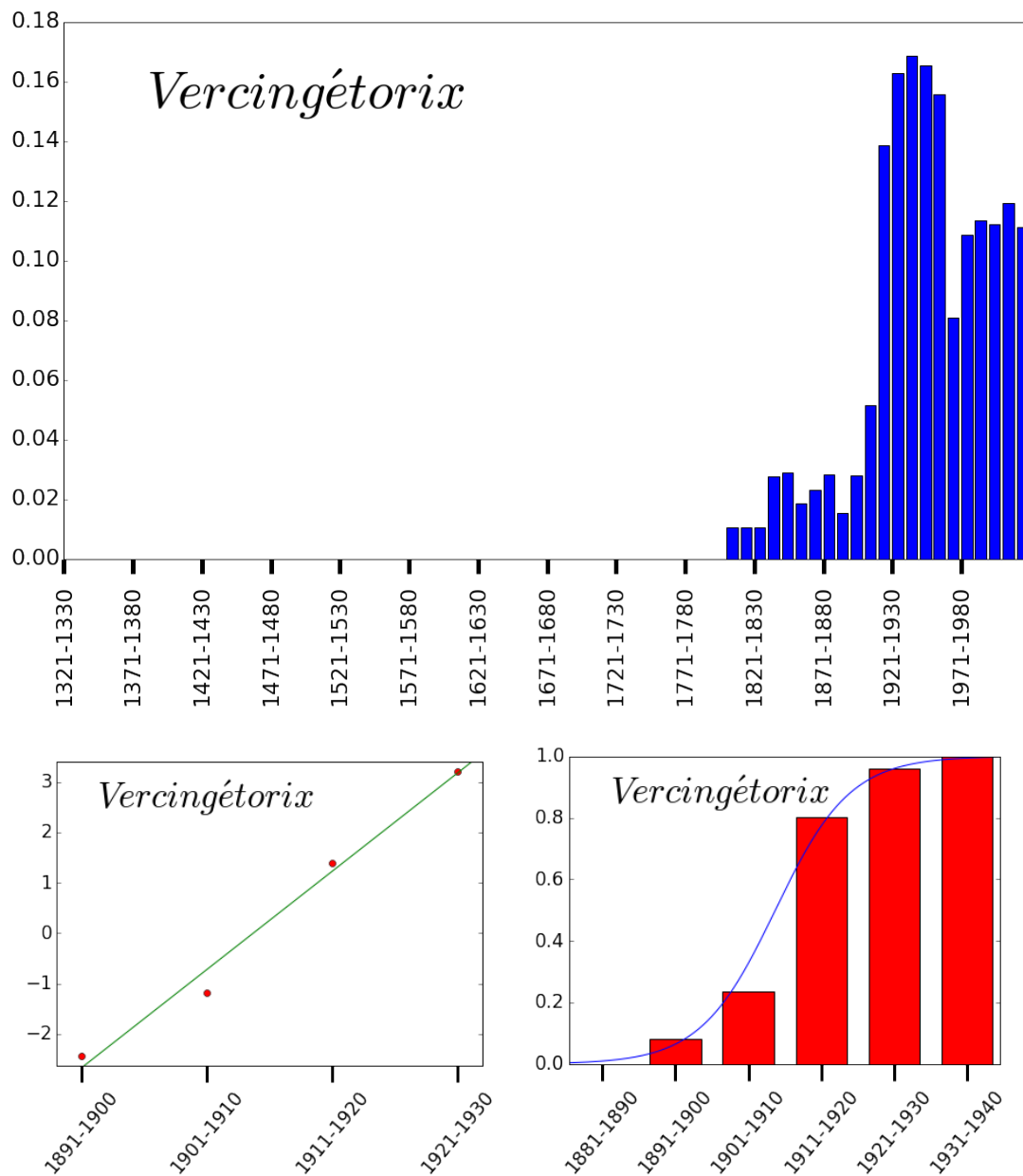


Figure 4.20: Top: frequency profile of *Vercingétorix*. Bottom left: Logit transform of the extracted growth part of the curve. Bottom right: Sigmoidal fit of the extracted growth part.

according to our extracting pattern procedure (we only displayed the second, most important one). In some cases, the rise can be relatively sudden and then immediately followed by a disgrace, as is the case with *fichtre* (Fig. 4.22). In this case, the behavior of *fichtre* reminds of the dynamics of first names, as if it were a trend. Interjections, and particularly insults, could indeed be described as a Fashion Game (Coulmont et al., 2016), for they tend to lose their power as their frequency increases and their use becomes more common.

As a conclusion, though S-curves are to be found almost everywhere in language change, they do not necessarily correspond to the same mechanism every time. It is possible that there are several alternative and legitimate explanations for the S-curve, all of them holding for a subset of cases. Social diffusion and imitation can lead to fast S-curves, but does not seem to be a suitable mechanism to explain grammaticalization instances. It is thus a desiderata — which unfortunately will not be much more than formulated in this work — to find empirical criteria so as to distinguish the different families of S-curves, precisely as Ghanbarnejad et al. (2014) did. In this study, we shall also propose one.

4.5 Statistical survey

We shall now present the results of our statistical survey over the 400 hundred instances of semantic expansion. For each studied form, we ran the procedure, and for each S-curve (provided they were consistent with the overall evolution), we stored the following information:

- the total duration of the latency (see below);
- the total duration of the growth, including the two boundary data points, w ;
- the slope of the logit, h ;
- the r^2 value, indicative of the quality of the fit;
- the total squared error:

$$err = \sqrt{\frac{1}{w} \sum_{k=1}^w (s_k - \tilde{s}_k)^2}; \quad (4.27)$$

- the total number of occurrences.

Results have been reported on Table A.1 in the appendix.

4.5.1 How generic is the S-curve pattern?

As we saw in the previous section, the frequency of use of a form is subject to many interferences, some being due to cultural influences, idiolectic influences, or to the unbalance of the corpus. Linguistic interferences are prone to affect the frequency, especially if the query does not carve out the occurrences with the required precision. Thus, many frequency patterns are too spurious for anything to be extracted at all.

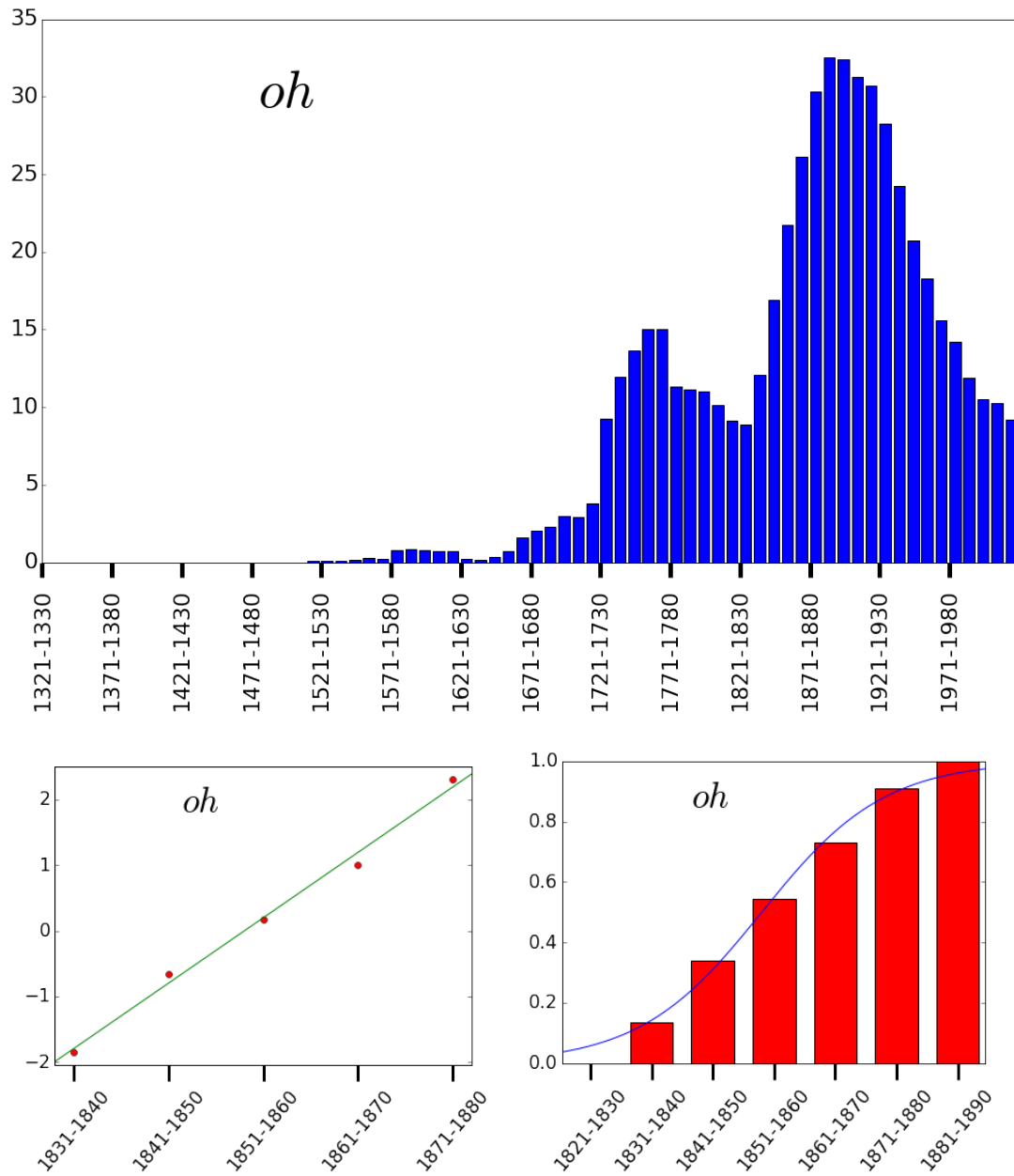


Figure 4.21: Top: frequency profile of *oh*. Bottom left: Logit transform of the extracted growth part of the curve. Bottom right: Sigmoidal fit of the extracted growth part.

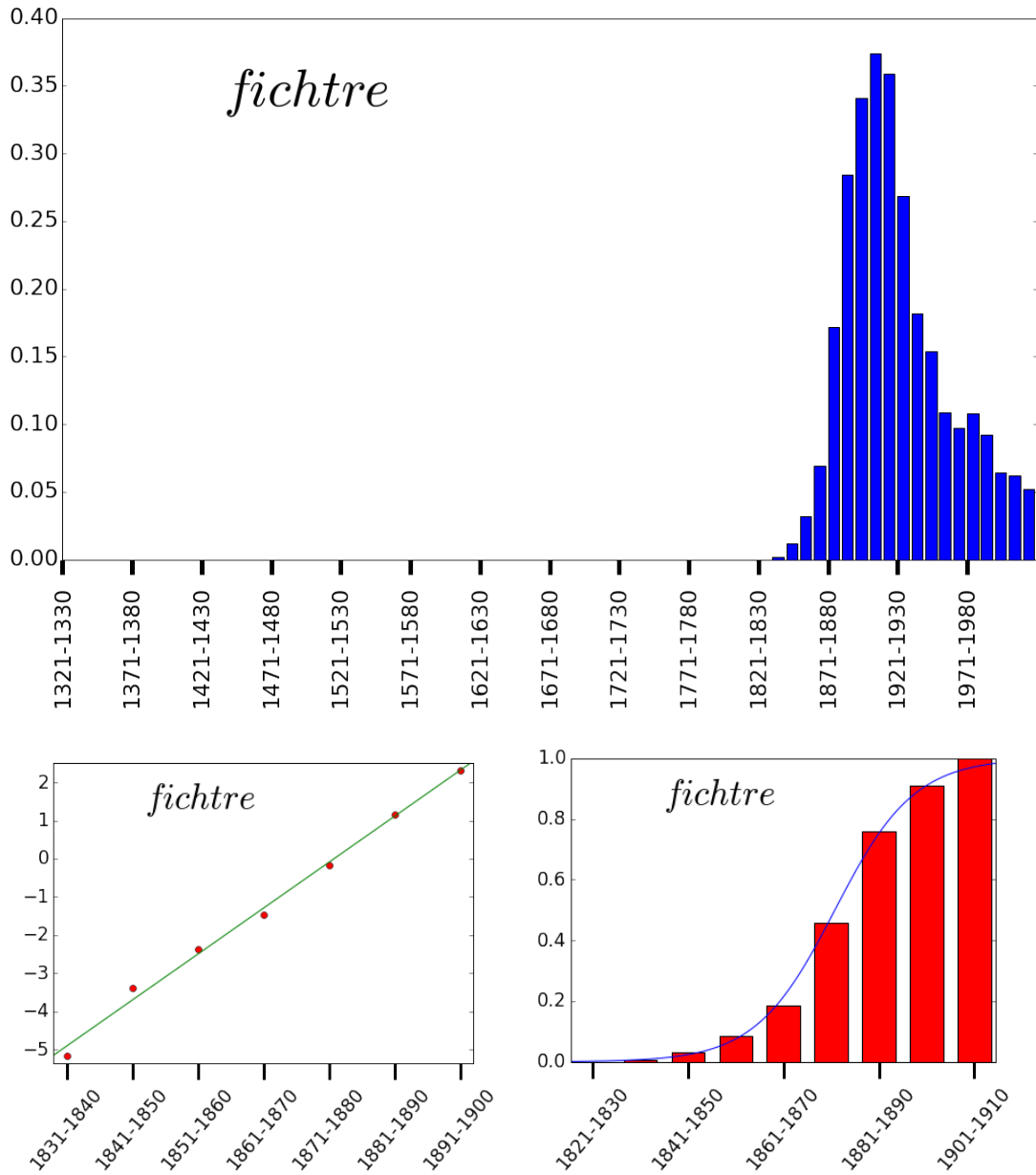


Figure 4.22: Top: frequency profile of *fichtre*. Bottom left: Logit transform of the extracted growth part of the curve. Bottom right: Sigmoidal fit of the extracted growth part.

Yet, our method was able to find at least one S-curve (and up to three in some cases) in the frequency evolution for 69 % of the studied forms. Had our criterion been lower, we could have found many more, but we wanted to select only S-curve of satisfying quality, so as to get a reliable set of values for the slope h and the width w .

We believe that a greater quantity of data could have allowed to extract nicer S-curves (or to extract S-curves at all) in some cases; especially when some individual texts tend to bias the frequency profile, as they weigh too much in their associated decade. Yet, some S-curves were found while relying on a very low number of occurrences. The form *à plus d'un titre*, for instance, follows an S-curve (Fig. 4.23), even if it reaches a small total of 62 occurrences over the corpus (in 2017; 59 in 2016). Conversely, some forms with very many occurrences (e.g. *à cause de*, 24840 occurrences in 2016) did not lead to any S-curve (Fig. 4.24). Actually, improving the curves is not so much a matter of data quantity, as of data quality. More balanced decades would be certainly of help, but the fact that we find such beautiful curves already inclines us to think that a better way might lie elsewhere. Indeed, very often, the frequency profile is blurry, for the evolution of the form is tangled up with many other processes of language change. The best results are obtained for isolated evolutions.

Thus, to improve data, it would be certainly more efficient to exert further linguistic expertise, so as to grasp precisely which constructional changes are actually going on, and focus on each of them separately. If we focus for instance on a subset of constructs of the [à cause de {Pronoun}] construction, the subset $N = \{\text{moi, toi, lui, etc.}\}$, associated with the query:

$\hat{c}(\text{n}'|y) \text{ a}|\hat{\text{a}} \text{ cause (de|d')} (\text{moi|toi|lui|elle|eux|nous|vous|elles})$

then we find at least one sigmoid in the process (Fig. 4.25), though less than a tenth of the initial occurrences remain (we can observe, actually, that the curve of *à cause de* displays a small S-shaped increase corresponding to this S-curve, yet it is too small to be detected as this little change is drawn into the mass of occurrences). This is why we believe that Construction Grammar could be an efficient framework to investigate language change in general, as it would allow to identify which changes are actually going on, instead of blending them all together by tracking too broad a construction.

The conclusion of this is that the S-curve indeed seems to be a generic pattern of language change, or at least of functionally-oriented semantic expansion. Whenever the S-curve is not to be found, a refined query, a more careful and accurate delineation of the ongoing processes, should allow to bring into light the S-shape eventually. There is probably no pattern completely contradicting the S-curve in our dataset, only a merging of several linguistic changes which are intimately mixed up. Of course, once independent and individual S-curves are identified, a good question would be to understand the possible interactions and influences between the S-curves of different yet dependent and related changes. This is only a broad perspective of research, which we shall not explore in the present work.

4.5.2 Distributions

Besides allowing for a confirmation of the generic character of the S-curve pattern, the quantitative scale of this survey presents the notorious advantage to build statistical distributions of the different observables (slope of the logit and growth time). Whereas

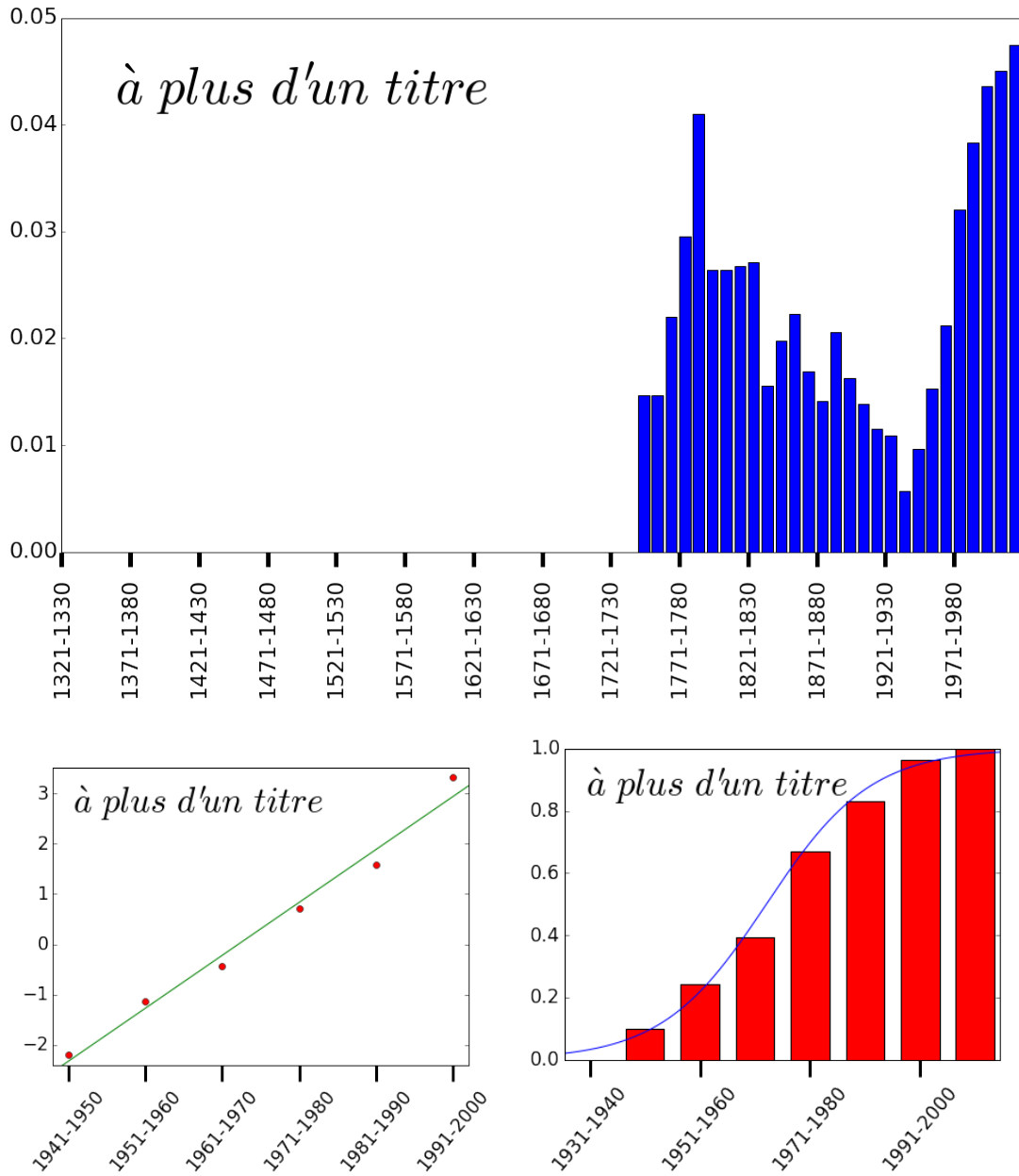


Figure 4.23: Top: frequency profile of *à plus d'un titre*. Bottom left: Logit transform of the extracted growth part of the curve. Bottom right: Sigmoidal fit of the extracted growth part.

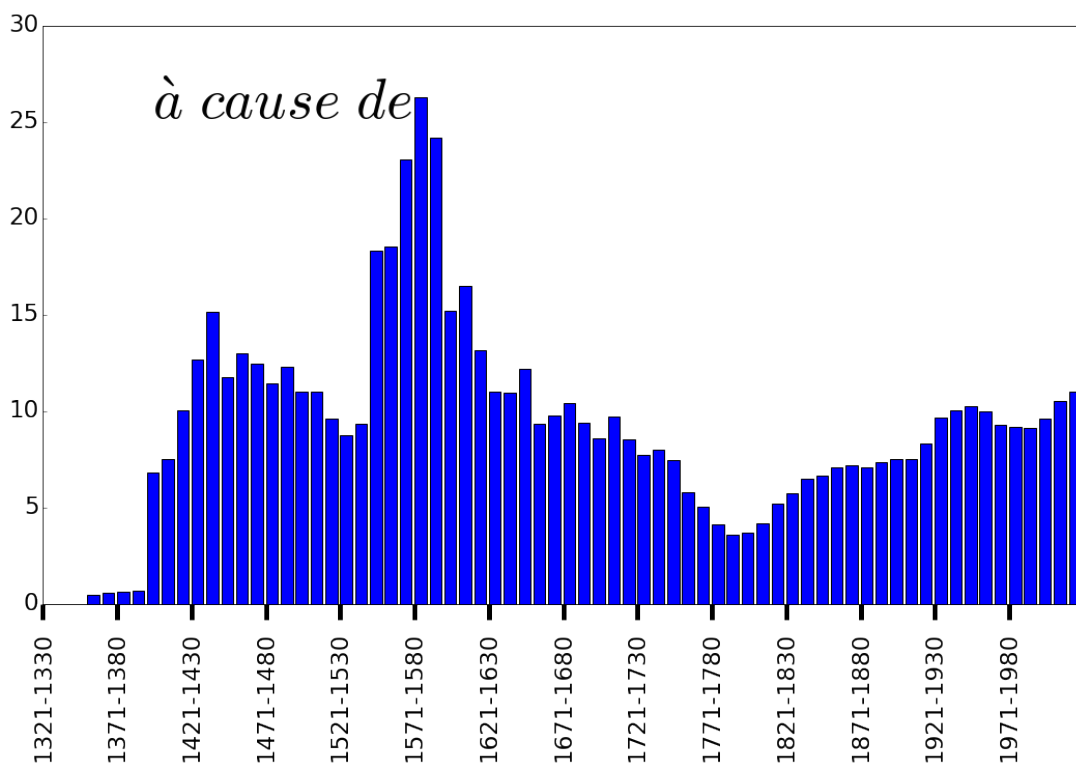


Figure 4.24: Frequency profile of *à cause de*. No sigmoidal pattern has been detected by our procedure.

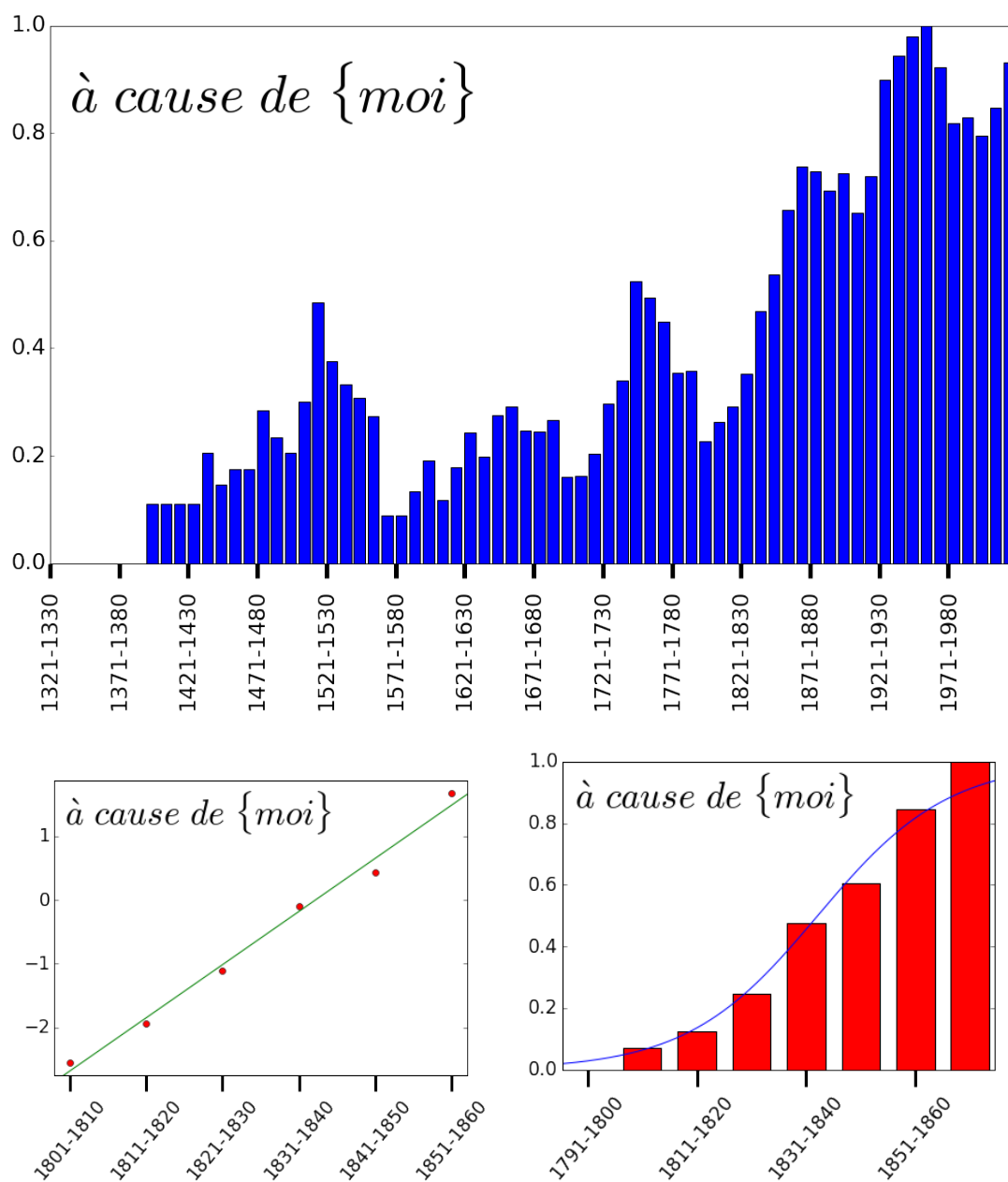


Figure 4.25: Top: frequency profile of *à cause de moi* + *personal pronoun*. Bottom left: Logit transform of the extracted growth part of the curve. Bottom right: Sigmoidal fit of the extracted growth part.

a large array of mechanisms and processes can lead to an S-curve, those distributions are much more specific to the phenomenon considered. Hence, they provide us with precious insights on the quantitative signature of semantic expansion.

Growth times

How long is a semantic change? The answer depends on the perspective adopted. Usually, a semantic change is instantaneous; more or less suddenly, new occurrences of the form appear with a meaning incompatible with the ones it used to be associated with. Yet, it takes time for the semantic change to be entrenched in language use, and this is why there is an S-curve pattern of frequency rise, starting with a semantic expansion. The length of the S-curve can cover several disparate phenomena: social diffusion, entrenchment in use, and even lexical diffusion. All three take time and are expected to produce an increase of frequency in the database. However, we can expect that the two first processes are more or less of constant duration: social diffusion and entrenchment in use of a given change have no particular reason to yield much different timescales from one change to another (they are quasi-deterministic processes and, for a given time period at least, rely on the same setting). Hence, we can hypothesize that lexical diffusion (or the acquisition of new semantic or syntactic features) is chiefly responsible for the variations in duration between two processes of semantic expansion: the longer the process, the more broadly the use and meaning of the form expand. Anyway, it is possible to plot the distribution of growth times (Fig. 4.26).

One might immediately notice two things: 1 - there are no growth times of less than six decades (because we prevented it in our pattern extraction procedure, for purely technical reasons); 2 - the data point associated with six decades is considerably higher than all the others. There are good reasons to believe that this data point is not reliable, and there also good reasons to keep it nonetheless. It should indeed be reminded that a growth of six decades would lead to four data points of the logit to be modeled with a straight line.

We already stressed that there were problems with such a low number of points as four. Indeed, they are associated with both a much lower number of false positives, and a higher number of false negatives, leading to a possible over-representation of four-points logits (or six-points sigmoids) in a statistical distribution. We previously saw that a four-points logit was not very robust, since the removal of about a hundred of occurrences (about 1% of the total of occurrences), in the case of *après tout*, was enough for the S-curve to fail passing the criteria. Hence, this high value associated with a six-decades growth in the distribution, value which is about twice bigger than the next highest value, is most likely an artifact of our pattern extraction procedure.

Why keep it then? The short answer is: it gives better results. A more reasonable justification is that our procedure leaves aside many instances of growth of either three, four, five decades or lower still. Thus, this high-value associated with the six-decades growth compensate with this forced removal and probably improves the estimation of the mean of the distribution. Indeed, if we miss, say, the lower half of the distribution (as the Inverse Gaussian suggests, see below), then the average will be overestimated. Hence, the six-decades peak artifact is a good way to balance this other issue.

One might still be concerned and wonder if the peak might be real, or if the

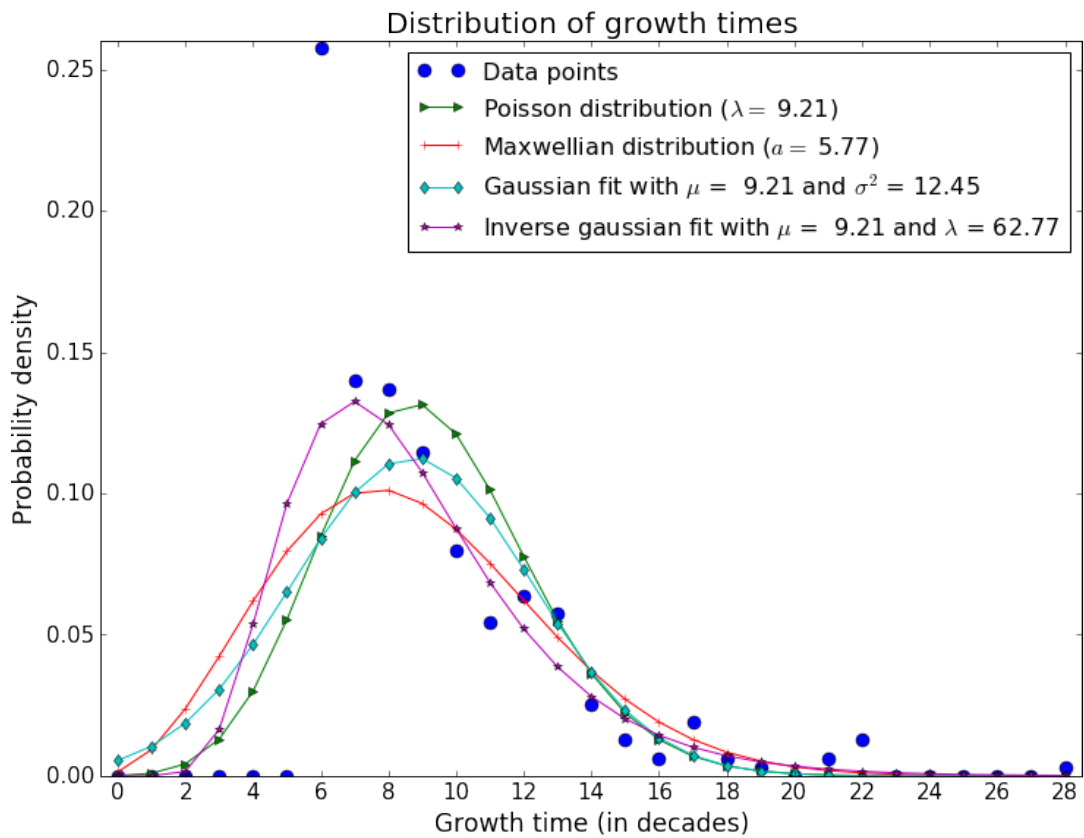


Figure 4.26: Statistical distribution of the growth times of over four hundred instances of semantic change in French. The distribution has been fit by various mathematical functions.

three, four, five decades-length growth could be bigger still. First, if we only consider the distribution from a seven-decades growth, it shows a clear inflexion of the curve. We can therefore expect this decrease of the derivative not to be counter-acted for lower values. It is reasonable to assume that the derivative will at least continue to decrease as we reach lower values for the growth time. Second, if there were not such a decrease, the number of predicted changes would explode, which would be quite concerning. Indeed, language is stable enough at the timescale of the daily life: if forms were constantly changing their meaning and new forms and idiosyncratic constructions kept appearing, communicating would soon become an intractable task. Hence, we do not expect too many changes associated with a very short growth time. Third, of all the forms we covered, we already found many S-curves and most frequency rises are already covered by our pattern extracting procedure. As a result, we should not have missed that much changes, so that the changes of the missing part of the distribution should not be more numerous than the ones we already caught. It would certainly be preferable to include it, but we can legitimately believe that, even if we do not, we are not missing the essential part of the growth times distribution.

We can also try to assess which statistical distribution is most fitted to account for this distribution. We first tried out several common statistical distributions:

- Poisson distribution: the probability $P(n)$ to observe a growth time of n decades would be given by (the notations $P(n)$ and n will remain the same for all distributions):

$$P(n) = \frac{\lambda^n}{n!} e^{-\lambda}, \quad (4.28)$$

where λ is the mean of the distribution.

- Maxwellian distribution:

$$P(n) = \sqrt{\frac{2}{\pi}} \frac{n^2 e^{-n^2/2a^2}}{a^3}, \quad (4.29)$$

with a , the mean of the distribution multiplied by a factor $\sqrt{\pi/8}$.

- Gaussian distribution:

$$P(n) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(n-\mu)^2/2\sigma^2}, \quad (4.30)$$

the parameters μ and σ^2 being, respectively, the mean and the variance of the distribution.

To fit the data by one of these distributions, we thus need to estimate the value of their defining parameters. However, as they are all related to the mean and the variance of the distribution, we always used the actual mean and variance of the empirical distribution to compute those parameters. It should be furthermore noted that the last two of those distributions are actually continuous. Hence, we can define the discrete estimates of each of these distributions by summing on all the correspond values:

$$P(n) = \int_n^{n+1} P(x) dx. \quad (4.31)$$

Growth times (width)				
Test	Poisson	Maxwellian	Gaussian	Inverse Gaussian
D_{KL}	0.34	0.52	1.05	0.27
AIC	336	396	730	311
BIC	340	400	738	318
Slopes (rate)				
Test	Maxwellian	Gaussian	Inverse Gaussian	Scaling law fit
D_{KL}	0.16	0.23	0.11	0.12
AIC	252	297	220	232
BIC	256	305	228	239

Table 4.2: Results of three statistical tests (D_{KL} : Kullback-Leibler divergence; AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion) for different fits of the growth times distribution and the slopes distribution

It makes, however, little difference.

We also consider another distribution, the Inverse Gaussian, for theoretical reasons which shall be put forward in chapter 8:

$$P(x) = \sqrt{\frac{\lambda}{2\pi x^3}} \exp\left\{-\frac{\lambda}{2x\mu^2}(x-\mu)^2\right\}, \quad (4.32)$$

where λ is given by, with the same notations as for the Gaussian:

$$\lambda = \frac{\mu^3}{\sigma^2}. \quad (4.33)$$

The best fit according to the Kullback-Leibler divergence (abridged D_{KL}) is given by the Inverse Gaussian (Table 4.2). It should be noted, however, that the Inverse Gaussian (and the Gaussian as well) has two parameters, which theoretically would allow for a better fit (yet the parameters are not fit in this situation but imposed by the empirical distribution). Therefore, we should make use of the Akaike Information Criterion or the Bayesian Information Criterion to evaluate the different fits. The results, however, do not change, and the distribution seems indeed well fitted by an Inverse Gaussian distribution.

Slopes

We can perform the same study for the slope of the logit (equivalently, the highest value of the S-curve derivative). As the slope takes value on a continuum, contrary to the growth time, we divided the range on which it takes its values (between 0.27 and 2.5) into 50 slots and counted how many values fall within each of these slots. We were thus able to obtain a distribution for the slopes (Fig. 4.27).

Once more, we can look for the most fitted distribution (Poisson is now ruled out as it can only be used for integer values). It turns out that Inverse Gaussian is the best choice, followed by the Maxwellian and the Gaussian at last (Table 4.2).

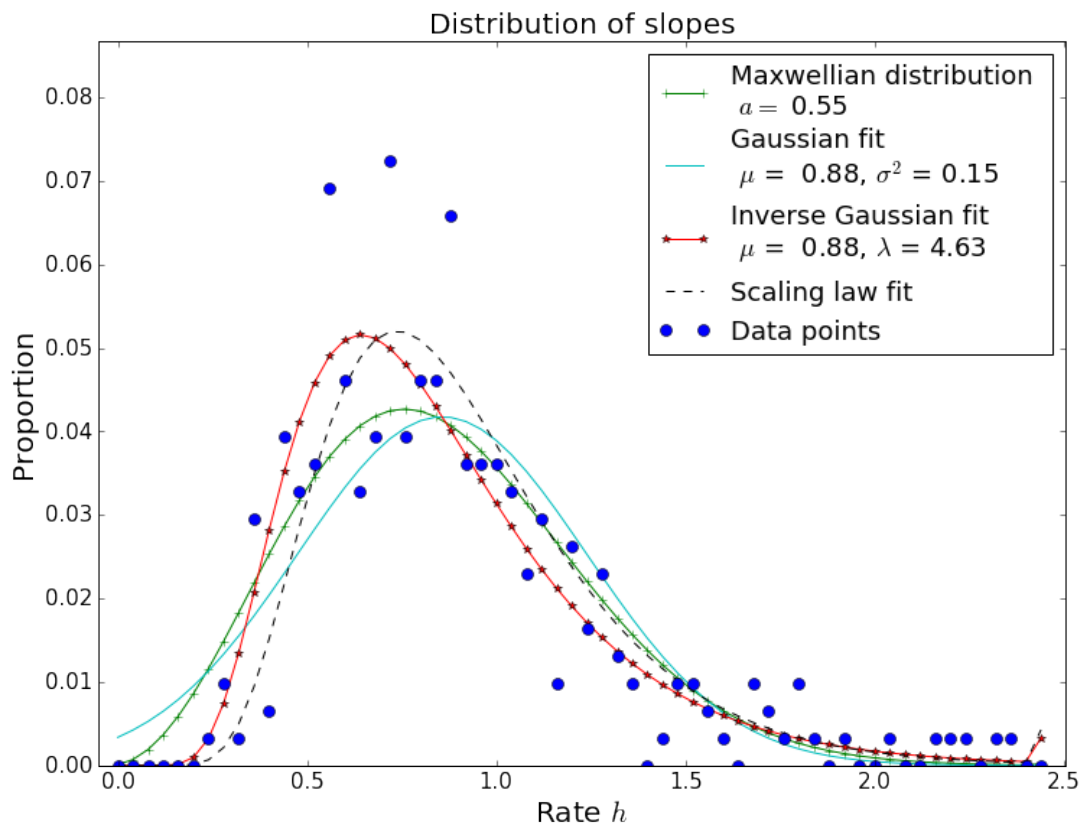


Figure 4.27: Statistical distribution of the slopes of over four hundred instances of semantic change in French. The distribution has been fit by various mathematical functions.

We can also compute what the distribution would be assuming the width and the rate were related by a scaling law (see below):

$$h = \frac{A}{w}, \quad (4.34)$$

where A would be equal to $e^{2.01}$. Then, if w is distributed according to ρ_w , the distribution $\rho(h)$ of the slopes should obey:

$$\rho(h) = \frac{A}{h^2} \rho_w \left(\frac{A}{h} \right). \quad (4.35)$$

Then, assuming that the width follows an Inverse Gaussian, we can fit the empirical distribution of slopes by the distribution given by equation (4.35). This is the fit we refer to as the ‘Scaling law fit’ on Figure 4.27. The agreement is all the more remarkable that only the parameter A has been computed with the slopes data (A is the exponential of the intercept obtained from the linear fit of the scattering plot $\langle h \rangle$ vs w).

Scaling law

As expected, the slope of the logit and the width of the sigmoid (growth time) are negatively correlated (we obtain a Pearson coefficient of -0.57), meaning that the longer the growth, the less intense it gets. This is an expected feature of the S-curve. Indeed, the reciprocal function of:

$$f(x) = \frac{1}{1 + e^{-(ax+b)}} \quad (4.36)$$

is:

$$f^{-1}(x) = \frac{1}{a} \left[\log \left(\frac{1-x}{x} \right) - b \right]. \quad (4.37)$$

If we now want to compute the width of the S-curve, we set an arbitrary criterion, Δ (say, 0.05), so that the width w corresponds to the values for which the S-curve reaches Δ and $1 - \Delta$. The bulk of the S-curve can now be considered to be comprised between those two values, so that:

$$\begin{aligned} w &= f^{-1}(1 - \Delta) - f^{-1}(\Delta) \\ &= \frac{1}{a} \left[\log \left(\frac{\Delta}{1 - \Delta} \right) - \log \left(\frac{1 - \Delta}{\Delta} \right) \right] \\ &= \frac{2}{a} \log \left(\frac{\Delta}{1 - \Delta} \right) \end{aligned} \quad (4.38)$$

Hence, the width scales with the slope a (or h) with:

$$w \propto h^{-1}. \quad (4.39)$$

Conversely, the slope h scales with the width according to $h \propto w^{-1}$.

Actually, since the changes are described by a sigmoidal S-curve but do not necessarily follow it exactly, it may be the case that the scaling behavior between those two variables would be different. In this case, the scaling exponent may be different

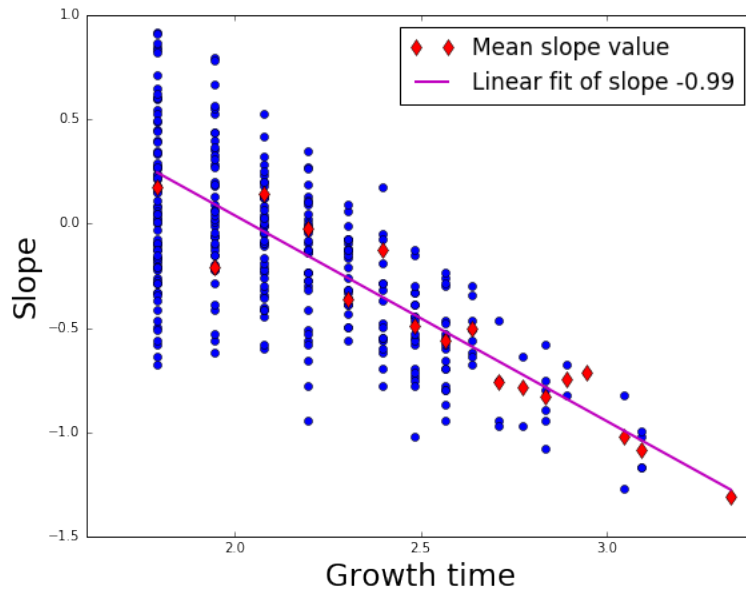


Figure 4.28: Log-log scatter plot of the slopes vs the growth time. For each value of the growth, the mean slope is represented by a red dot. The red line represents a fit of these mean slope data points.

from the trivial value -1 . Michard and Bouchaud (2005) have for instance shown, in the case of the social diffusion of technological goods, natality behavior, and concert clapping interruption patterns, that the width of the S-curve scales with its ‘height’ h according to $h \propto w^{-2/3}$. Hence, it is worth investigating whether or not we find a non-trivial scaling behavior.

We thus plotted a log-log scattering graph of the height h versus the width w (Fig. 4.28), and then computed, for each different value of w , the mean height h . Those values were then fitted with a linear curve, whose slope gives an empirical estimate of the scaling exponent. We happen to find an exponent equal to 0.99, hence extremely close to the trivial expected value. Though this result may not be worth noticing by itself, it is at least a confirmation of the consistency of the S-curve pattern throughout our dataset. More crucially, it clearly shows that the S-curve of language change is probably not the same as the S-curve of social diffusion.

There may also be another interesting fact. Indeed, the scattering plot shows some sort of an upper rim, as if, for a given growth time, the slope couldn’t exceed a given threshold. If we thus take the maximum slope value for each value of the width, and fit them by a linear curve, it seems that they align pretty neatly (Fig. 4.29): the r^2 value is equal to 0.99, to be compared with the r^2 of 0.97 found in the previous case. This alignment, and the confinement of nearly all data points below this line, are extremely uncommon. Furthermore, the associated slope is not trivial, roughly equal to -1.5 .

This would suggest the existence of a ‘saturating’ scaling law, whose consequence would be that frequency cannot rise too fast for too long a time; the longer the time of growth, the more restricted the speed. The interpretation of this saturating law is rendered difficult by the fact that the unit of the slope is unclear: it should be

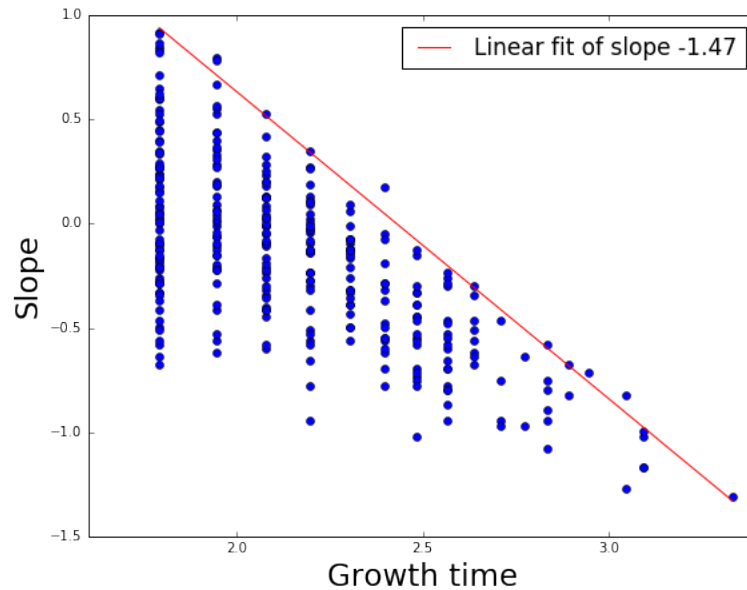


Figure 4.29: Log-log scatter plot of the slopes vs the growth time. For each value of the growth, we selected the maximal slope and performed a linear fit of those points, represented by a red line.

in frequency per decade (more precisely: in number of occurrences per a hundred thousand words per decade), but the rescaling of the frequency to make the sigmoid runs from 0 to 1 blurs a bit the picture. We can consider that the unit of the slope reflects, in a competition framework, the proportion of use of the form per decade, i.e., the replacement rate. It would follow that the total replacement would be limited by $w * h_{max}(w)$, i.e. $w^{-1/2}$. This is a bit surprising, for it means that the longer the change, the less easily the form can prevail: either you take over quickly, or you struggle sluggishly. This might mean that coexistence —what has been called ‘residue’ in the lexical diffusion literature (Wang, 1969)— of a dominating form and a lingering one would be all the more favored if the replacement is slow. It is as if the speakers had grown accustomed to this coexistence and thus had no reason to delete one entirely from the language. However, as there might not be any actual one to one competition, the exact interpretation might be much different. At least, this result crucially shows that there must be some kind of change-resilience built in language, for a change cannot be both long and intense; sharp changes are tolerated as long as they are short enough.

However, this result could be imputed to a mere coincidence. To test the robustness of this finding, we considered slightly different methods of extraction for the slope and the growth time. They do not give the same value for the saturating slope, and this structure actually does not even appear for other extracting procedures. Fig. 4.30 reports the scaling law obtained when choosing a width-dependent criterion for the r^2 so as to pass a significance threshold characterized by a p -value of respectively 0.05 and 0.01, according to the random growth null model. The specific values of the criterion, for each width, can be read on Figure 4.4.

Though the ‘saturating’ law is not robust, the trivial scaling relation $h \propto w^{-1}$

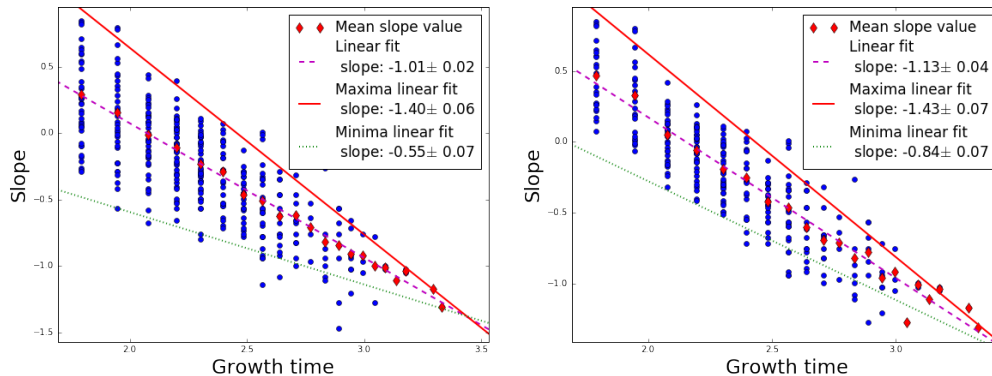


Figure 4.30: Log-log scatter plot of the slopes vs. the growth time, with linear fits of the maximal, minimal and mean values of the slope for each growth time. These plots have been obtained by changing the criterion $r^2 > 0.98$ used in the extraction procedure, making it now width-dependent so as to ensure a homogeneous significance p -value of 0.05 (left) or 0.01 (right). The latter presents less data points than the former since the criterion is more restrictive.

holds in any case (less so for the more restricting case of 0.01 p -value, but the data is sparser in this case since the criterion is more restrictive). An interesting feature emerges nonetheless from this comparison: in all cases, the three lines intersect in one point. This may be indicative of an upper bound for the growth point, given by the corresponding abscissa of this intersecting point. We find a conjectural maximum growth time of 25, 31 and 33 decades for each of the three extracting pattern procedures. This is indeed close to the longest patterns we found in our survey (the record belongs to *à part*, ranging 28 decades from 1361 to 1640).

Meta-evolution of language change

Finally, we can superficially address the question: does the way language change itself evolve? Is there a meta-evolution of language change? This is actually assumed most often, for the world has grown more and more international, connecting more closely different language-speaking communities; also, the emergence of mass-media, such as radio, television, and more recently the World Wide Web, could be expected to encourage a fast diffusion of novelties. We can therefore investigate a possible meta-evolution of language change. A simple way of doing it would be to compare the average slopes and the average growth times for different centuries (Fig. 4.31), an evolution being said to belong to a century if the beginning of the growth falls within this century. The speed of change can be said to increase if either the average slope becomes bigger, or the average growth is reduced. Here the centuries are defined with respect to the corpus, starting with 1321-1420 and ending with 1921-2020.

What we observe is that there is no significant change in these averages. One might yet perceive some augmentation of the average slope running for the three last centuries, concomitant with a small decrease of the average growth time. Yet, as the corpus stops with the decade 2011-2020, changes in the twentieth century cannot be

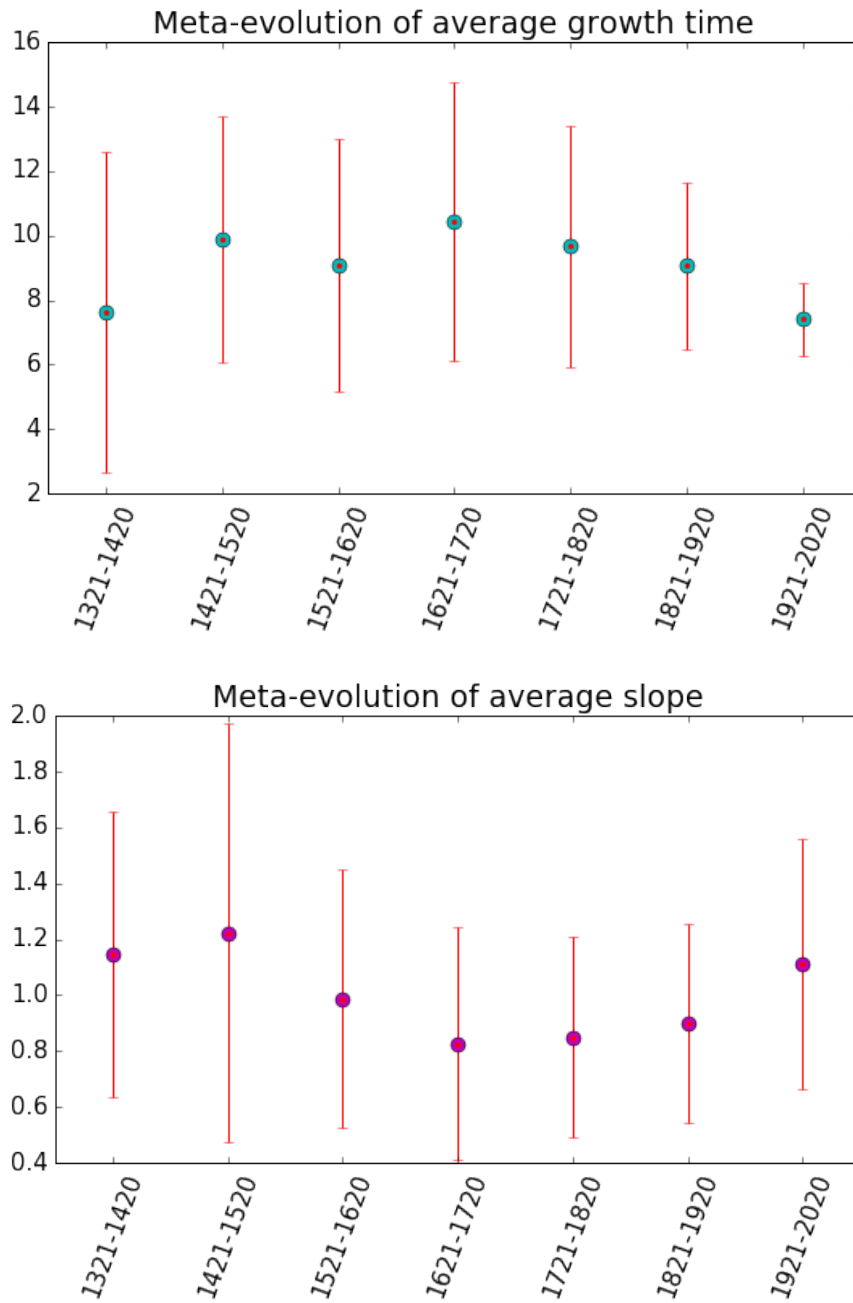


Figure 4.31: Top: evolution decade by decade of the average growth of all sigmoidal patterns starting in this decade. Bottom: evolution decade by decade of the average slope of all sigmoidal patterns starting in this decade. Error bars correspond to the standard deviation.

longer than ten decades. This limitation is still perceived in the eighteenth century (e.g. a change starting in decade 1871-1880 cannot be longer than fifteen decades). Hence, this bias alone could explain the weak change in average slope and growth time for these two last centuries.

It seems thus impossible to conclude, given the corpus, the procedure, its limitations, and the special kind of changes under investigation, that any meta-evolution has taken place in the evolution of language in the last century.

4.6 The full pattern of semantic change

We have previously established the generic character of the S-curve as a pattern for language change. Yet, we want now argue that, for semantic change, this pattern is only part of a picture, and should be completed by another phenomenon, which is latency. Indeed, most changes can be described by the following three steps:

1. The frequency slightly increases to a low value. The first occurrences of the form (or then construction) clearly linked to the new meaning start to appear in the corpus. This rise does not typically last more than a decade.
2. The frequency remains constant for an arbitrary time, either short or long. This step can last several decades. Occurrences characteristic of the new meaning continue to pop up in the corpus.
3. The frequency rises to its final value, following an S-curve.

The first step is actually nearly impossible to extract empirically, for it is very short. Also, if the form is already attested as a construct before becoming an actual construction (as was the case with *du coup*), then this frequency increase is typically too small to be noticed and the latency phase is a bit mixed up with the preceding parts of the curve.

It should be noted that, to my knowledge, latency has never been noticed as such before, even though it has occasionally been observed that the beginning of a given process might be further way in time that it appears first. Denison (2003) notes for instance, in the case of the progressive passive, that ‘earlier examples keep turning up’ and that ‘early use of the progressive passive *is* sporadic’ (emphasis original). Before trying to deal automatically with the latency part of the pattern, we shall thus provide a few selected examples so as to illustrate this phenomenon.

4.6.1 A few examples of latency

Latency can be of varying duration and can actually be relatively short. However, as long as it is short, one can always argue that it is part of the S-curve, and so that there is nothing worth noticing about it. Yet, some instances show spectacularly long latencies, making it hard to argue that this corresponds to approximately nothing. One must remind indeed that, at least in some cases, the new meaning is already present, so that we observe a new variant, which survives and holds on before an eventual rise which can happen up to several centuries later. This lingering persistence is indeed remarkable, and deserves an explanation. We shall provide one in chapter 8.

One of the most beautiful latencies observed in our statistical survey is given by *de toute façon* (Fig. 4.32) and lasts 35 decades (which is half of the complete time period covered by our corpus). It is then followed by an S-curve which is, most probably, unfinished, so that the process is still going nowadays. The meaning found during the latency period is closer to ‘in any/everyway’ or sometimes ‘in all manners’, and later to ‘in either way’, than to its pragmatic use, most comparable to ‘anyway’, which appears as soon as 1845:

je sais ses intentions et son ardeur d’acheter vos terres, il n’en démordra pas, et pour en avoir trois cent mille francs, il faut lui en demander trois cent cinquante mille. Vous les auriez si vous vous obstinez ; mais, de toute façon, il ne faut pas qu’il paie le bien au-dessous de sa valeur.

I know his intentions and his eagerness to buy your lands, he will not renounce, and to get three hundred thousand francs of it, you have to ask him three hundred and fifty thousand. You would get them, were you to persist; but, anyway, he must not pay the good much less than it’s worth.

SAND George, *Le Meunier d’Angibault*, 1845 (Frantext)

The latency lasts then at least for six decades (admittedly much less than the 35 observed). Still, six decades is equivalent to the time of subsequent growth.

Furthermore we can look for a more refined query, focusing on a syntagmatic context which is chiefly compatible with the pragmatic meaning. Doing so, we discard a lot of acceptable occurrences, but we also get rid off almost all the spurious ones:

(. | , | ; | : | ! | ! ?) de toute façon (. | , | ; | : | ! | ! ?)

We then find a latency of seven decades (Fig. 4.33), coinciding closely with the ‘first’ occurrence in corpus, which is to be compared with the actual time of growth (11 decades). Hence, the latency part of the pattern is, in this case, of a size comparable to the S-shaped growth part.

Another beautiful example (Fig. 4.34) is given by the word *bureau* (which was not included in the statistical study), showing that the latency phenomenon is also to be found in non-functional semantic expansions (the word *bureau* has no functional uses). The meaning of *bureau* is actually quite broad, covering meanings such as ‘office’, ‘desk’, ‘board’ and ‘study’. Most of these meanings are present throughout the whole period of latency, indicative that the main semantic expansions of the word has taken place before this latency. The period of frequency rise which follows does not seem to be associated with any significant further expansion. Note that, in that case, the automatized procedure to extract the latency overestimates the latency, which is closer to 18 decades than to 28 decades. Anyway, a latency of 18 decades is quite remarkable.

As this phenomenon can be controversial, it would be a good thing to provide examples from another corpus, another language. We have been able to automatically extract the pattern from data provided by the COHA database, and our findings supports the actual existence of the latency phenomenon. We already see with *weird* that a period of latency follows the first S-shape growth and precedes the second (which, being unfinished, has not been extracted by our method). The same goes for *actually* (Fig. 4.35), though in both cases, a more detailed analysis of the occurrences would be needed to check if there really has been a semantic expansion at the beginning of this latency period. To make a clearer case, we have to find a form whose appearance is

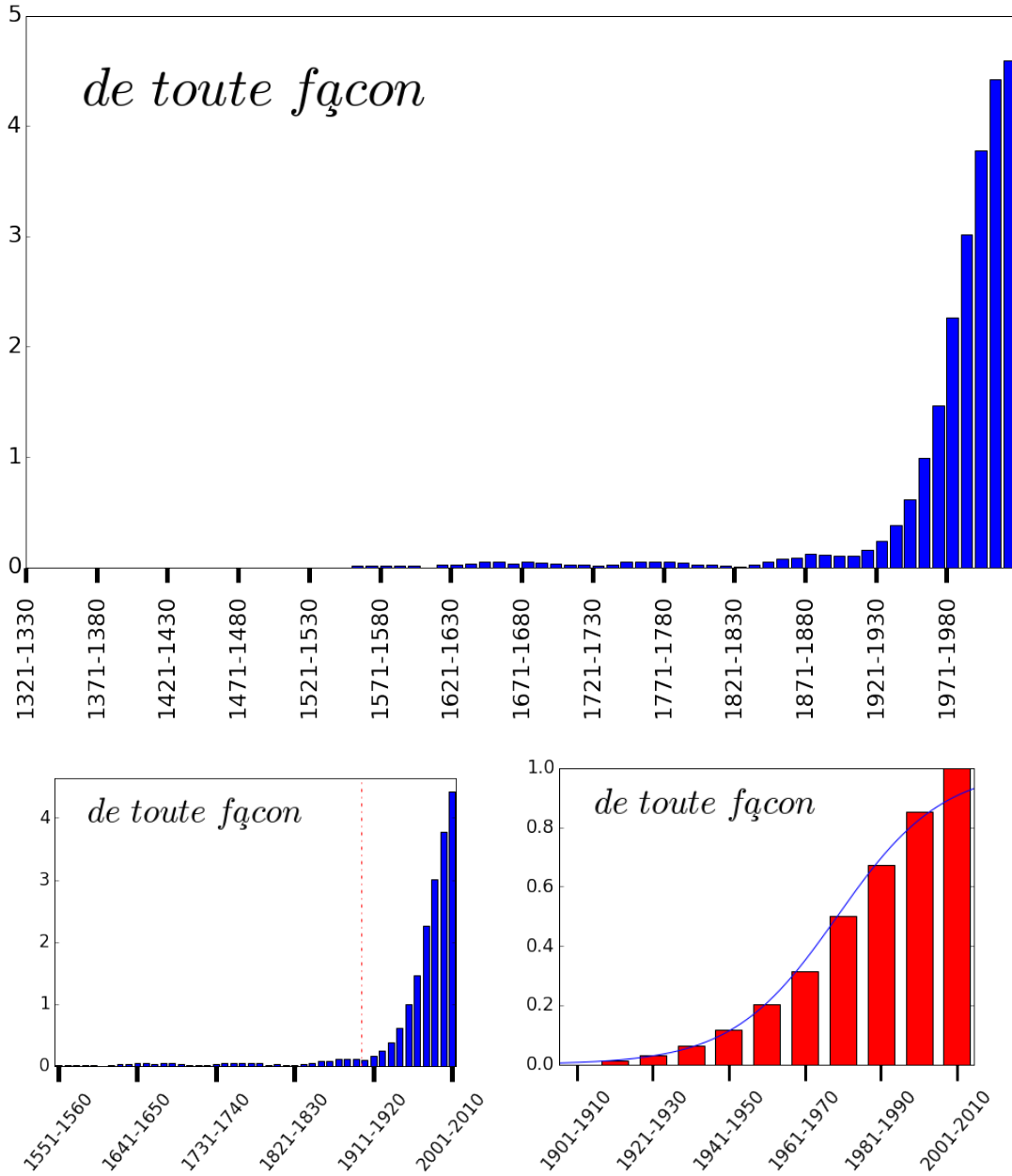


Figure 4.32: Top: frequency profile of *de toute façon*. Bottom left: full extracted pattern, with a dashed red line separating latency part from growth. Bottom right: Sigmoidal fit of the extracted growth part.

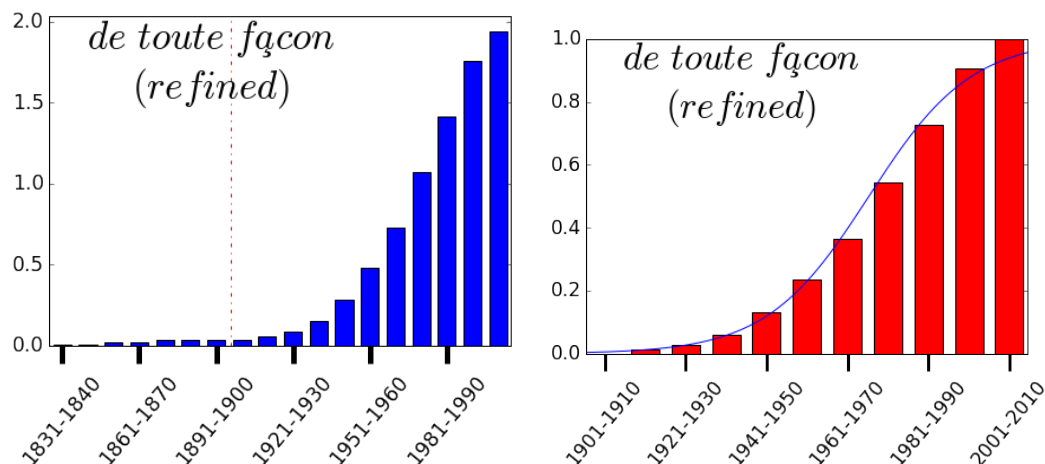


Figure 4.33: Left: Full extracted pattern from the refined query of *de toute façon*, with a dashed red line separating latency part from growth. Right: Sigmoidal fit of the extracted growth part.

already indicative of a new meaning. A fitted candidate is the construction *the thing is*, (comma included), whose new syntactic behavior is the result of a constructionalization. All occurrences of *the thing is*, are thus theoretically related to the new meaning of the new construction. As we observe a nice 9-decades latency in this case (Fig. 4.36), the pattern is confirmed.

Actually it was particularly complicated to find a working example since the two-centuries window of the COHA is very short. For a long enough latency (say, of five decades) to be observed, at least eleven decades of the pattern must be covered by the observation window of two centuries. So, the semantic expansion giving rise to the latency must occur around the mid-nineteenth century, which severely limits the chances to find a suited change. This confirms the need to favor a corpus covering a long time period, such as Frantext.

Finally we provide a last example, *ouille* ('ouch') which also complies to a new form/new meaning correspondence, so that all occurrences that can be found result from the semantic expansion. As it is an interjection, it is unexpected that it shows a latency period too (Fig. 4.37). Yet, it actually results from an unfortunate combination of a spurious occurrence:

Qu' est-ce qui paiera la goutte à la pa..., à la patrou... ou... ouille !

Who will pay the pa... the patro... o... ol its drink !

ZOLA Émile, *L'Assommoir* 1877 (Frantext)

and of our averaging procedure which will detect this occurrence for the fifty following years. Hence, the first occurrence of *ouille* in the corpus is found in 1927 —in the exact decade triggering the beginning of the S-curve.

This survey of examples shows thus that the latency is trickier to track than the S-curve, for it is associated with ambiguous occurrences and low frequencies. Yet, we tried to propose an empirical procedure to automatically give an estimate of the latency time for each S-curves. Its results may not be entirely reliable (e.g. *ouille*),

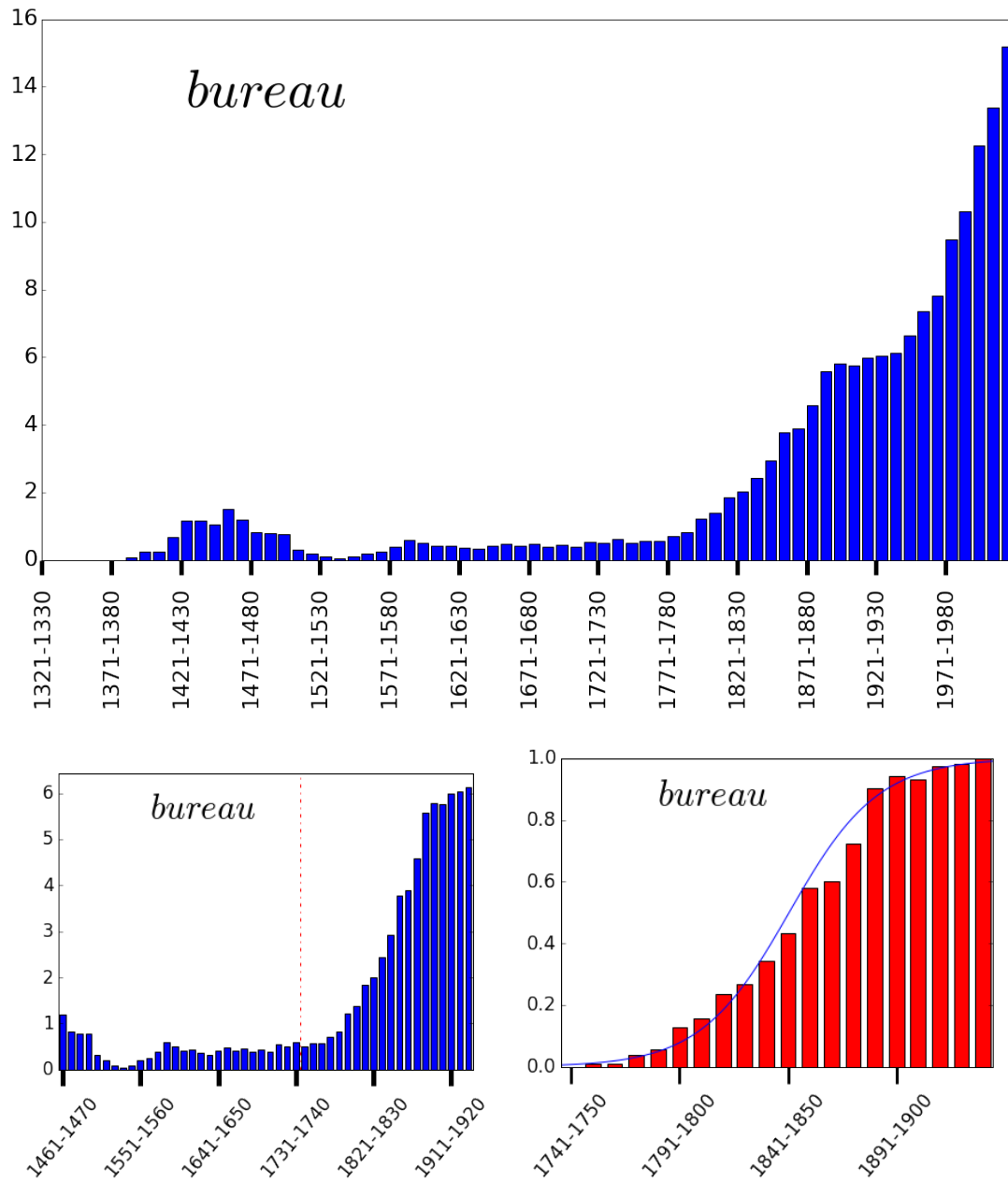


Figure 4.34: Top: frequency profile of *bureau*. Bottom left: Full extracted pattern, with a dashed red line separating latency part from growth. Bottom right: Sigmoidal fit of the extracted growth part.

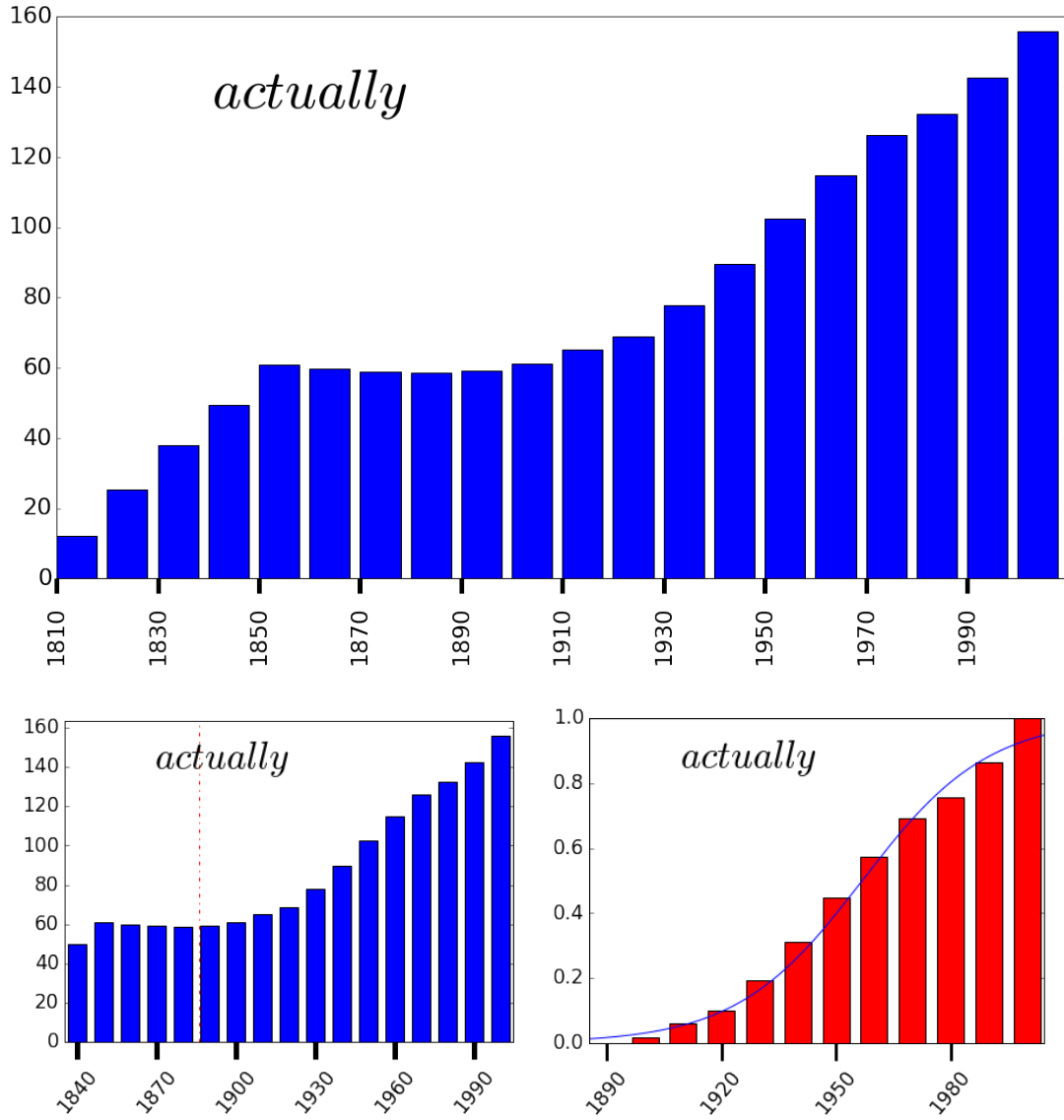


Figure 4.35: Top: frequency profile of *actually*. Bottom left: Full extracted pattern, with a dashed red line separating latency part from growth. Bottom right: Sigmoidal fit of the extracted growth part.

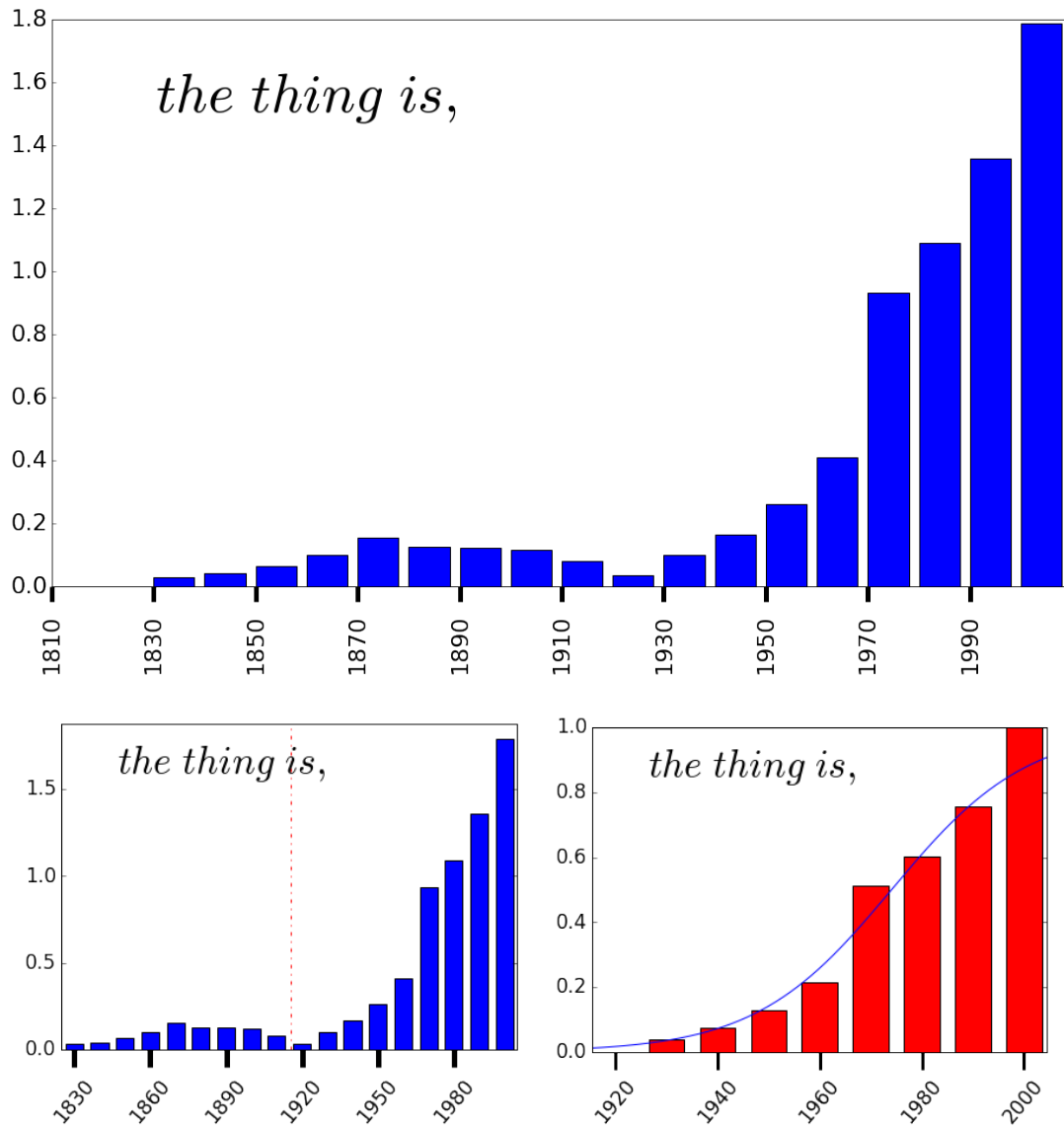


Figure 4.36: Top: frequency profile of *the thing is*. Bottom left: Full extracted pattern, with a dashed red line separating latency part from growth. Bottom right: Sigmoidal fit of the extracted growth part.

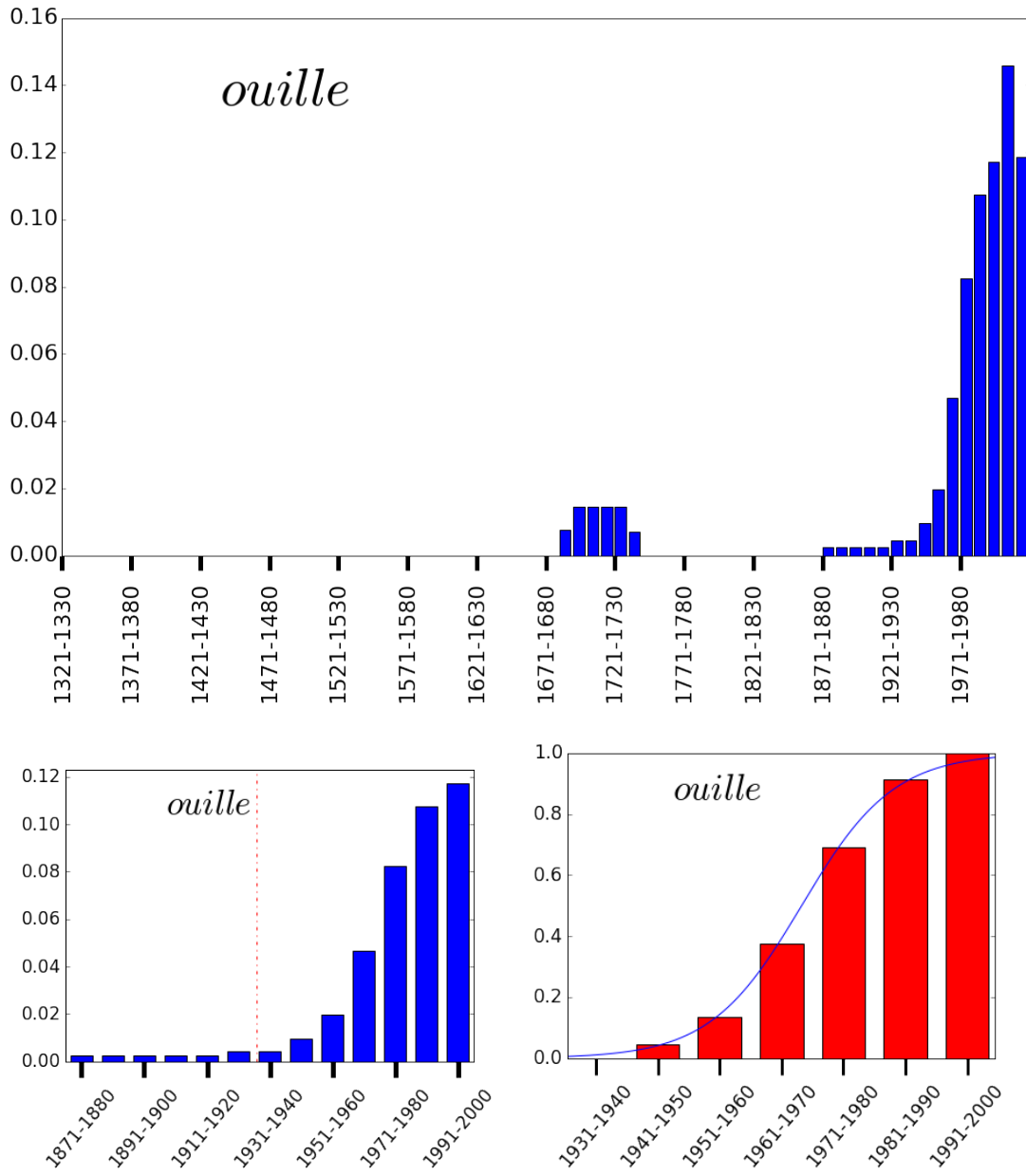


Figure 4.37: Top: frequency profile of *ouille*. Bottom left: Full extracted pattern, with a dashed red line separating latency part from growth. Bottom right: Sigmoidal fit of the extracted growth part.

but at least it is entirely data-driven.

4.6.2 Extracting latency

Extracting the latency part of the pattern is a delicate matter: it is easily destroyed by interferences with other processes; as it is associated with a low number of occurrences, it can fluctuate a lot and be heavily idiolect-dependent... A first intuition would be that latency is a period where the frequency approximately does not grow, so that any possible growth is below a given threshold. However, we cannot define such a threshold for all instances of change. Indeed, a latency can happen for a form already established (e.g. *weird*), so that any variation will be associated with a significant growth compared to the variation of a newly appeared form. The trick of relative frequency is of no use here, since the frequency in latency time can be close to zero; hence a minimal variation in frequency will be associated with a gigantic relative growth, if not to a divergence.

The solution we proposed therefore do not rely on growth. Once we have the boundaries y_{max} and y_{min} of the S-curve, we can draw two horizontal lines of equations:

$$\begin{cases} y_{up} = & y_{min} + a * (y_{max} - y_{min}) \\ y_{down} = & y_{min} - a * (y_{max} - y_{min}) \end{cases}, \quad (4.40)$$

with a a parameter (set to 0.15 for Frantext and to 0.10 for the COHA in our examples and statistical survey). Then, for all data points y_k before the S-curve, we consider that they belong to the latency pattern as long as $y_{down} < y_k < y_{up}$. As soon as one point drops out of these boundaries, the latency pattern ends. Thus, the pattern is built up backwards, starting with the beginning of the S-curve.

A problem raised by this procedure is that, if the form is appearing and if y_{down} is smaller than 0, then all data points with zero frequency will be enrolled to the latency pattern. However, we can have a latency pattern associated with sporadic occurrences so that there are indeed gaps of zero frequency. We thus tolerated gaps of zero frequency no longer than five decades, so that if, going backwards, we find a zero frequency for the fifth decade, we come back to the last non-zero frequency point and end the latency there.

Crucial to this method is the choice of the a pattern. 0.15 is quite high and will frequently overestimate the latency by considering as latent a previous process of a lesser magnitude than the S-curve for which the latency pattern is constructed (as happened with the few first decades of the latency pattern of *bureau*). 0.10 is quite restrictive and the latency must be very flat to count as such. In our statistical survey based on Frantext examples, as we wanted many latency values to be able to study their statistical distribution, we favored a non-restrictive criterion. At least, the bias is uniform throughout all our examples, so that it should not affect too much the distribution.

4.6.3 Distributions and correlations

As for the growth time and the slope we tried to fit the distribution of the latency time with several usual statistical functions, and the Inverse Gaussian as well. We tried out

Growth times (width)			
Test	Exponential	Gaussian	Inverse Gaussian
D_{KL}	0.21	2.08	0.10
AIC	873	1391	756
BIC	877	1398	764

Table 4.3: Results of three statistical tests (D_{KL} : Kullback-Leibler divergence; AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion) for different fits of the latency times distribution

the traditional Gaussian and the Exponential distribution as well, for the shape of the latency times distribution is reminiscent of the latter:

$$P(n) = \int_n^{n+1} \frac{1}{\beta} e^{-x/\beta} dx, \quad (4.41)$$

the β parameter being equal to the mean of the distribution.

We also removed from the distribution the data point corresponding to no latency time. Though it is expected that some changes have a latency time of less than a decade, there are so many empirical reasons for the latency to be missed that this point has been assumed to be not representative. Including it could however speak in favor of the exponential, so we considered both distributions: one with all data points (Fig. 4.38), the other with the exclusion of the zero-latency data point (Fig. 4.39). It appears that even the exponential distribution is unable to catch both the zero-latency datapoint and the bulk of the distribution in the latter case, so we shall focus on the former. Once again, the parameters have not been computed so as to optimize the fit, but are fed in by the empirical values of the mean and the variance of the distribution. It results that the Inverse Gaussian is once more the most fitted distribution, followed by the exponential (Table 4.3). The Gaussian is obviously inadequate.

That the distribution of the latency time is not trivial, and actually follows pretty well a known statistical distribution with a very good score (better indeed than for the growth time and the slope), comforts the idea that the latency is an actual phenomenon, separated from the S-shaped rise, and that it should complete the pattern alongside the subsequent S-curve. It may be less remarkable, but it nonetheless obeys an interesting law which can provide us with new insights, both quantitative and qualitative, about language change in general.

It is also interesting to check if there are any correlations between the latency time and the growth time. One could indeed expect that the longer time the change wait to unfold, the more intense its impetus, so that growth time and latency time could be positively correlated. Actually, they are, with a Pearson coefficient of 0.24, but this correlation is very weak. We shall argue later that this correlation can result from the procedure, which builds the latency part with reference to the S-curve part of the pattern, introducing ipso facto a correlation between the two. Similarly, the correlation between the latency time and the slope is weak (-0.19) and concurs with the growth-slope negative correlation already observed.

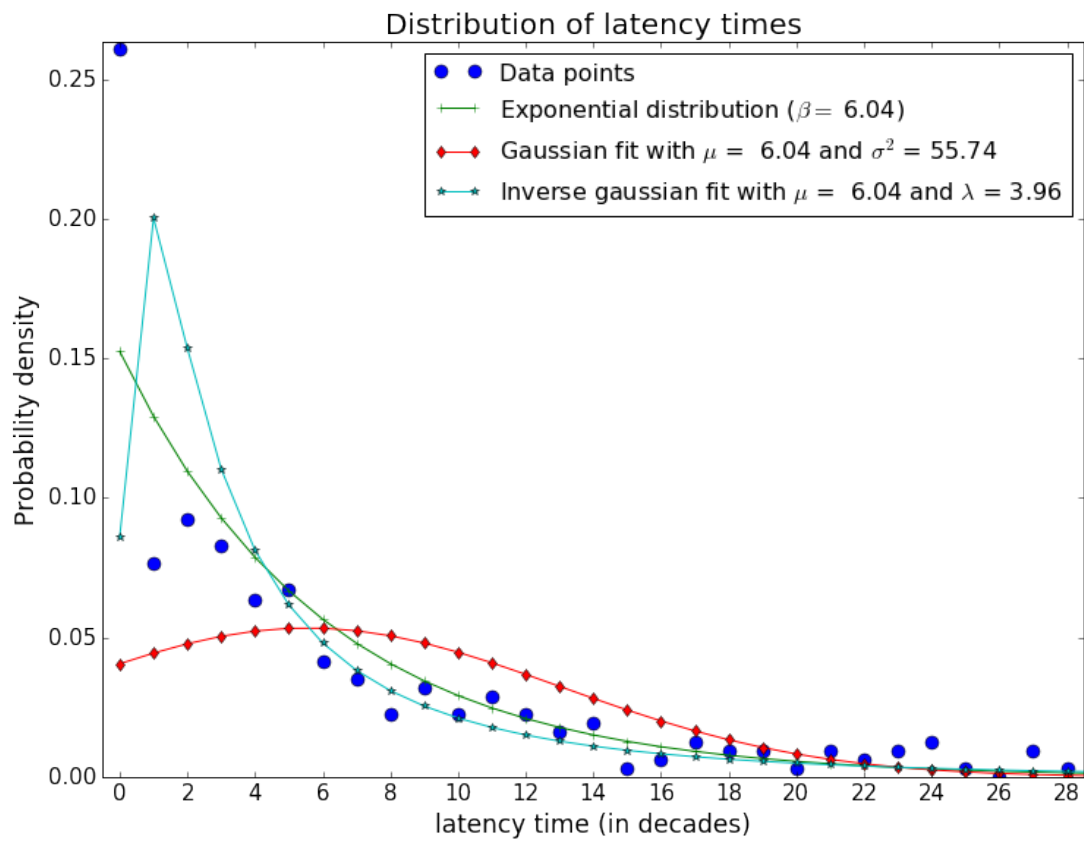


Figure 4.38: Statistical distribution of the latency times of over four hundred instances of semantic change in French. The distribution has been fit by various mathematical functions.

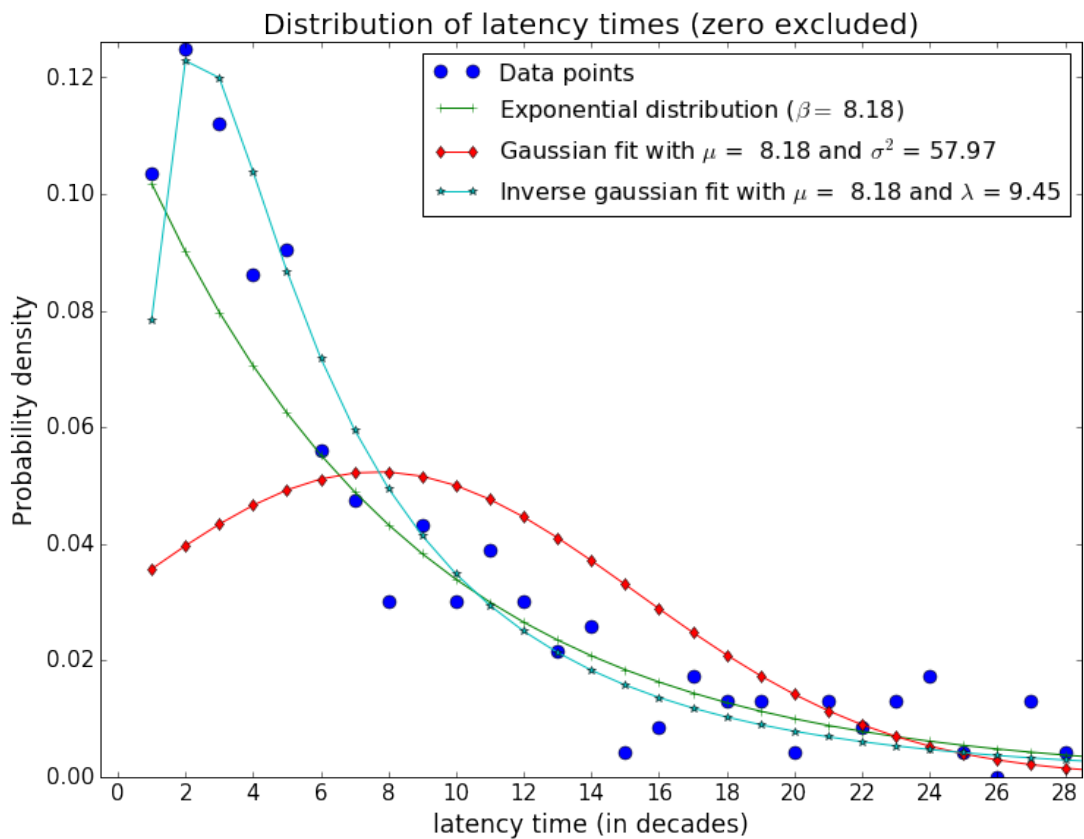


Figure 4.39: Statistical distribution of the latency times of over four hundred instances of semantic change in French, where the data point corresponding to an absence of latency has been excluded. The distribution has been fit by various mathematical functions.

4.6.4 A specificity of semantic change?

We shall conclude on the latency phenomenon by pointing out that it might not be as generic as the S-curve. Indeed, the S-curve is found for all sorts of change, and can be triggered by different mechanisms. The latency, on the contrary, seems to be more specific of semantic changes, changes involving a cognitive junction between two concepts. Interjections, as they are likely to be social-based phenomena, follow nice S-curves, but almost never show latency (remember *ouille*). However, not all interjections are alike. Interjections such as *ouille* has no prior history and is of onomatopoeic origin; *zut* has a more elaborated (and somehow unclear) origin, but chiefly results from its humorous phonetics. On the other hand, *bah* has a rich history prior to its reappearance in the XVIIIth century (Fig. 4.40), and indeed derives from an earlier *ba* (of onomatopoeic origin) dating back to the twelfth century. Hence, its frequency rise pattern results, as for regular semantic expansions, from a cognitive process of meaning reinterpretation. We can thus distinguish new forms carried on by a past linguistic history, and those carried only by a social trend. In the former case, we expect a latency, while we do not in the latter.

The same asymmetry is to be found for proper names. While *Voltaire* solely refers to the person of that name, so that the frequency rise in the use of its name can be understood as resulting from social imitation (hence with no latency), people had no reason to speak about Vercingétorix in the XIXth century (*De Bello Gallico* was read much before that time), if not for complex cultural reasons. As his figure came to embody the French patriotic identity through a long and complex associative process, the frequency rise of its name is thus not only due to social diffusion and imitation, but also to cognitive reconfiguration of knowledge. Thus, we find no latency for *Voltaire* while we find eight decades of latency for *Vercingétorix*.

It would thus seem that, if the S-curve is the empirical signature of the diffusion of a change, the latency is a sign that a process of cognitive reorganization has taken place, and can thus be considered, in the case of language, as characteristic of semantic change. Those ideas will be developed further in chapter 7, where we shall argue that semantic change indeed results from the passing through of a linguistic form from one cognitive domain to another. This mechanism being associated with a latency, as we shall see in chapter 8, we can propose that latency is an effective criterion to distinguish S-curves chiefly due to social imitation (without latency) from S-curves motivated by a cultural, semantic, or cognitive shift (with latency).

4.6.5 Robustness of the results

We tried to assess the robustness of these results by considering a slightly different procedure, applying the width-dependent criteria of Figure 4.4 to select the patterns. The difference is mainly that, for a maximum p -value of 0.05, it becomes harder for processes associated with a low number of points (i.e. covering a small number of decades) to be deemed as significant, while for higher numbers of points, it should be easier. As a consequence, most processes of six or seven decades will be rejected, whereas longer processes, previously set aside, will now be included.

We first consider the distribution of growth times 4.41. The Inverse Gaussian fit remains, among a selected set of statistical distributions (Poisson, Maxwellian,

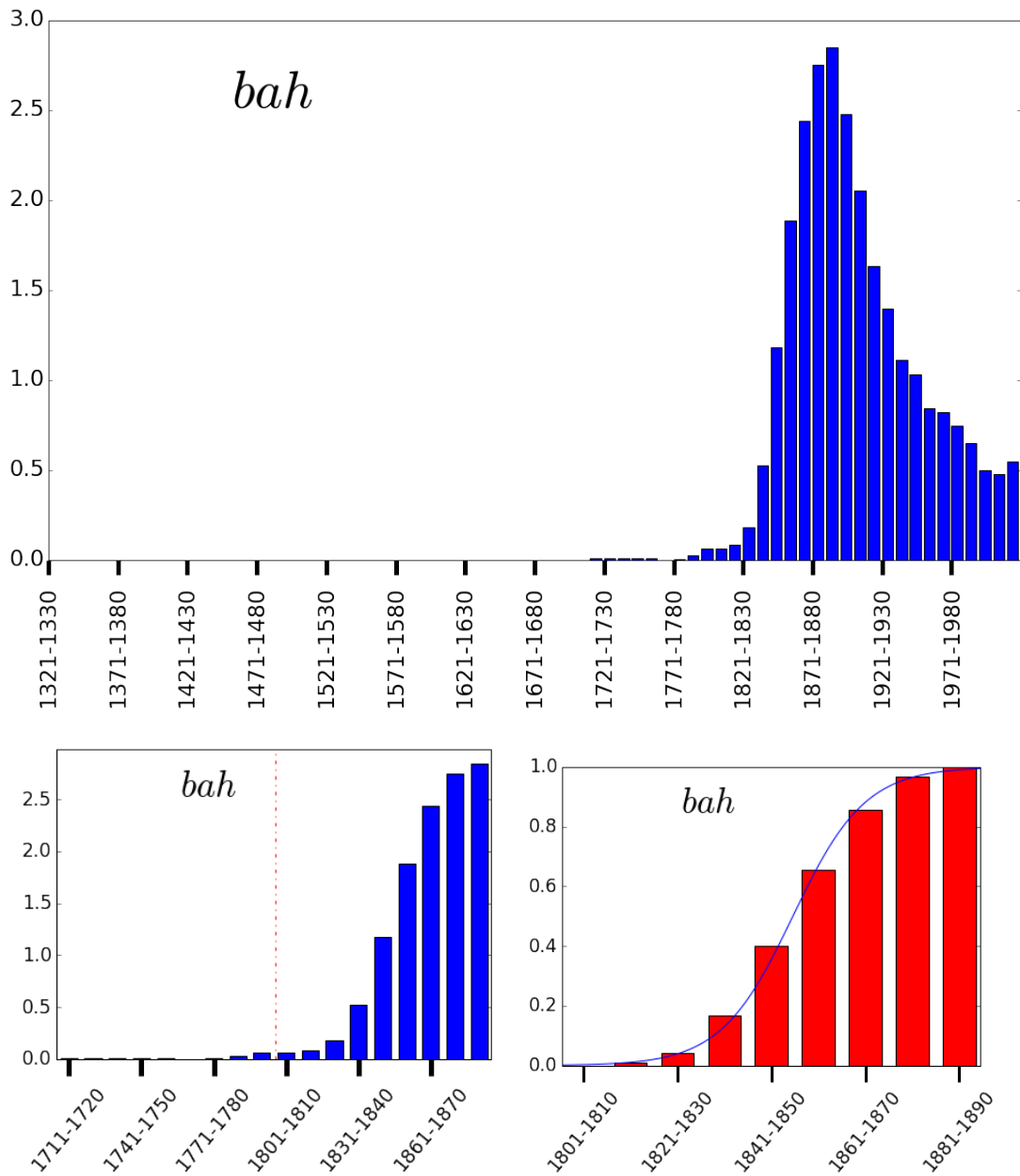


Figure 4.40: Top: frequency profile of *bah*. Bottom left: Full extracted pattern, with a dashed red line separating latency part from growth. Bottom right: Sigmoidal fit of the extracted growth part.

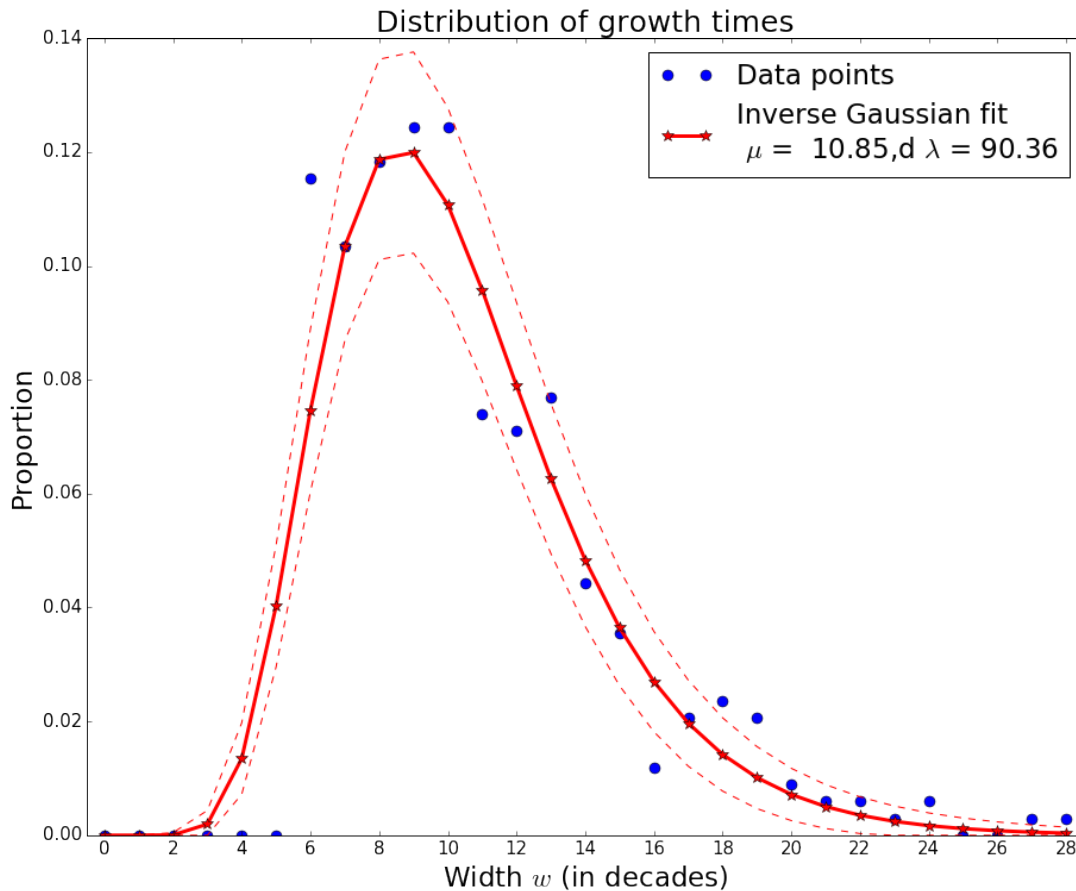


Figure 4.41: Statistical distribution of the growth times of over four hundred instances of semantic change in French, selected with a width-dependent criterion associated with a maximum p -value of the sigmoidal fit equal to 0.05. The distribution is best fit by an Inverse Gaussian. Error bars, in red dashed lines, correspond to the standard deviation.

Gaussian), the most appropriate one, according to the Kullback-Leibler divergence, which is now much better. All quantitative results are summarized on Table 4.4. The mean growth time is now longer, which is expected since short patterns have now a higher propensity to fail passing the criterion.

The two other distributions, that of the latency 4.42 and that of the slopes 4.43, are also still best fit by the Inverse Gaussian, and present smaller Kullback-Leibler divergences than before. The quantitative characteristics of these distributions, such as the mean or the Péclet number, are comparable as well to what was obtained for the $r^2 > 0.98$ criterion.

Thus, we see that the results can be improved: the data points, obtained from the sigmoidal fits with a p -value higher than 0.05, are more closely fit by the theoretical statistical distributions than what was the case with the previous procedure. However, our main findings, regarding the shape of the distributions (given by their Péclet number), their mean values, and the correlations between them, do not differ from one

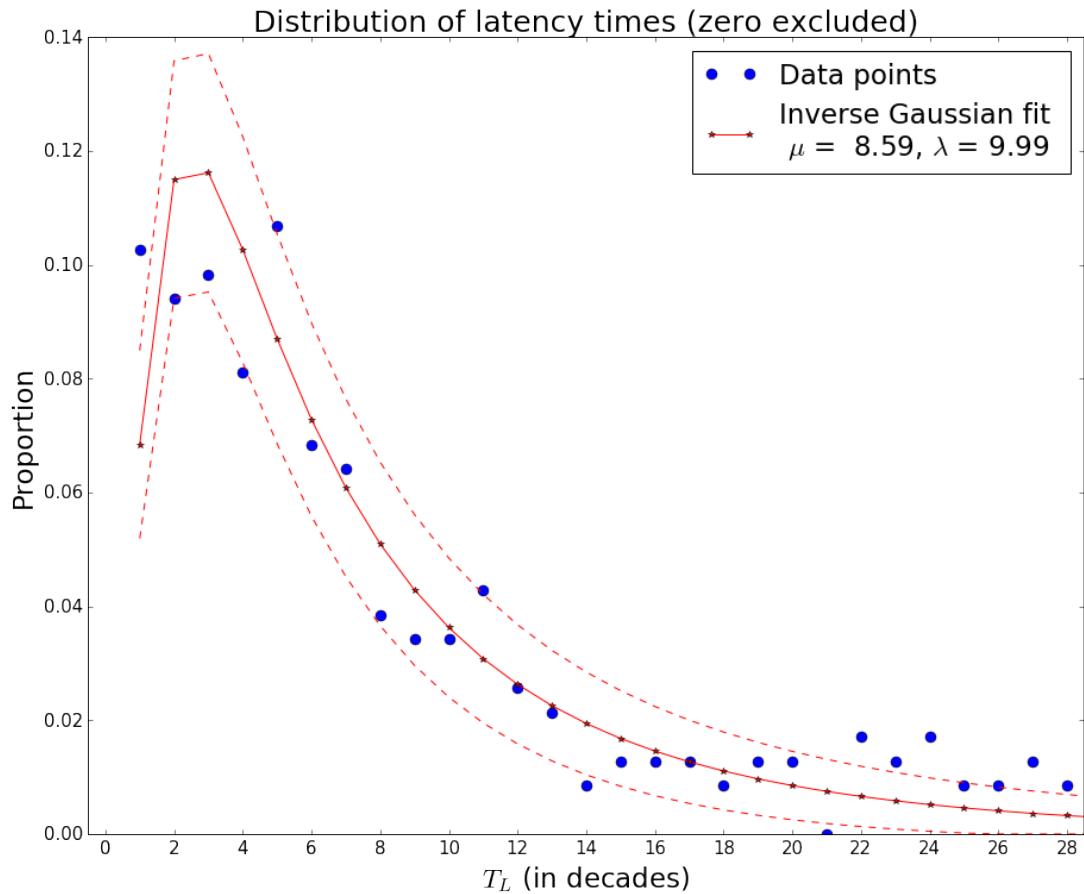


Figure 4.42: Statistical distribution of the latency times of over four hundred instances of semantic change in French, selected with a width-dependent criterion associated with a maximum p -value of the sigmoidal fit equal to 0.05. The distribution is best fit by an Inverse Gaussian. Error bars, in red dashed lines, correspond to the standard deviation.

Criterion	$r^2 > 0.98$	$p\text{-value} > 0.05$
D_{KL} (growth)	0.27	0.10
Growth Péclet number	13.63	14.27
Mean growth time	9.21	10.85
D_{KL} (latency)	0.24	0.10
Latency Péclet number	2.71	2.33
Mean latency time	9.74	8.59
D_{KL} (slope)	0.20	0.10
Slope Péclet number	12.46	10.38
Mean slope	0.71	0.80
Latency-slope correlation	-0.08	-0.15
Latency-growth correlation	0.16	0.16
Slope-growth correlation	-0.63	-0.68

Table 4.4: Comparison .

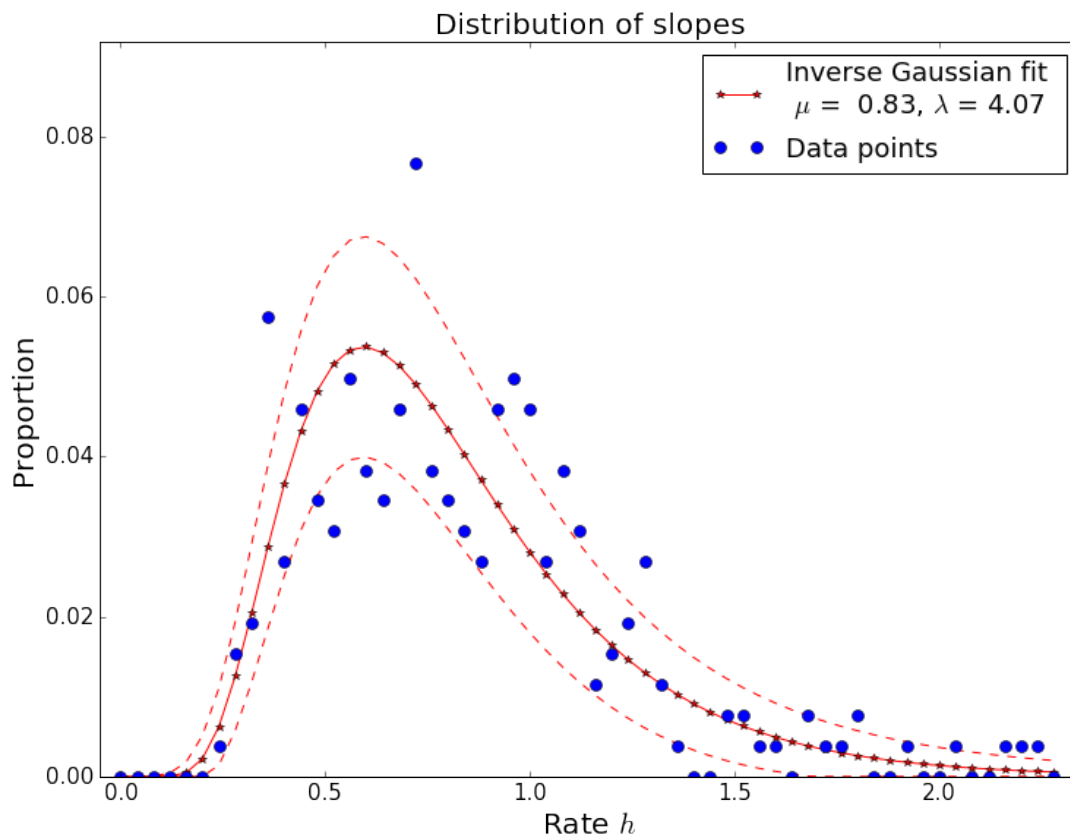


Figure 4.43: Statistical distribution of the slopes of over four hundred instances of semantic change in French, selected with a width-dependent criterion associated with a maximum p -value of the sigmoidal fit equal to 0.05. The distribution is best fit by an Inverse Gaussian. Error bars, in red dashed lines, correspond to the standard deviation.

method to another. Therefore, this is a good hint that these results are robust and reliable. A further check of their trustability would be to change the time window Δ_t on which the occurrences are counted, here set to a decade ($\Delta_t = 10$). Extracting the patterns with a time window of a quinquennium ($\Delta_t = 5$), or better still, of a year ($\Delta_t = 1$), would indeed be a more crucial test. The difficulty is that these choices hinge upon the requisite that the corresponding corpus divisions still encompass a sufficient number of texts (which, for Frantext, is clearly not the case when $\Delta_t = 1$).

4.7 Summary

In this chapter, we proposed an automated procedure to empirically extract S-curve patterns from frequency profiles. Performing a statistical survey over 400 cases of semantic change thanks to the Frantext database, we established the generic character of the S-curve, which was lacking a broad, large-scale empirical confirmation, though it was theoretically an accepted feature of language change.

Yet, the problem with the S-curve is that it may be too generic. Indeed, the S-curve can be the empirical counterpart of many different kinds of change phenomena, relying on different mechanisms, and therefore is in no way specific to language. This issue justifies our survey, since the statistical distributions of the quantitative features of the S-curves obtained from the 400 semantic expansions is highly more specific than the S-shape alone. Furthermore, we brought to light another aspect of the frequency pattern of semantic change, which is a latency preceding the S-curve. This latency can be of various length (from one decade to one century or even more) and is characterized by a temporal mismatch between the actual occurrence of the semantic change in the corpus, and the frequency rise which is expected to result from this expansion of meaning. As we argued, this latency might be specific of semantic change, and could be the sign that a cognitive reanalysis has taken place.

Also, this statistical survey, and especially its limitations (most patterns are blurry, even though an S-curve can still be extracted of them), points out to a theoretical point of great importance. An S-curve is a sign that a change has taken place; but what exactly is that change is not always clear. It would seem that Construction Grammar is a powerful and efficient framework to think of the locus of the change: what is changing in language change are constructions, or constructional features. Most semantic expansions are either examples of a constructionalization (the addition of a new construction in the constructicon, such as *par le biais de*), or a constructional change (a change in the features, e.g. semantic or syntactic, of a given construction).

Thus, a desideratum for the study of language would be to isolate these individual changes, to describe them as such by extracting the quantitative features of their characterizing curves, and then to try to understand the relations and the influences that intertwines different changes in a single and complex frame. The constructionalization of *du coup*, for instance, is most probably related to the semantic changes undergone by *coup* at the same time, just as that of *way too* is probably guided by the rich paradigm of *way* + ADV constructions. Only then should it be possible to explore the specific interactions in change that allow far-reaching semantic expansions, such as grammaticalizations.

Chapter 5

Paradigms and constructions

To study the relationship between different constructions, empirical corpus linguistics has especially focused on co-occurrence frequencies. Different measures have been proposed to assess the strength of a collocate between two constructions (Gries, 2013; Desagulier, 2015). For instance, one can contrast the preferential collocations of two closely related constructions by looking at their collocational preferences (Desagulier, 2012). This can combine with a diachronic approach by dividing the whole corpus into several time periods, and then performing this kind of collocational analyses on each different time period. Therefore, one can investigate how the semantics of a construction has evolved, for instance in cases of attested grammaticalization processes (Hilpert, 2006).

In this short, exploratory chapter, I would like to propose a different family of tools to explore the features of a given set of constructions. The analyses evoked before are essentially synchronic, and they can be transposed to a diachronic perspective only by juxtaposing successive synchronic analyses. I shall suggest in this chapter to proceed the other way around, offering a way to shed light on the synchronic organization of language by using its diachronic genesis. Especially, I want to address the question whether we can know if two distinct forms are two different constructs of one and the same construction, or two separate constructions? The method is also useful to extract information on the diachronic process. For instance, it should enable us to detect speciation events (i.e. when a construct of a construction, or a paradigm member, takes its independence and becomes a construction on its own), as well as to identify competitors in a semantic expansion process.

A short disclaimer is in order. The following method relies on nothing which is strikingly new, and is certainly not very original. Yet, I did not encounter any mention of something close in the literature, and as it proved a useful way to get new insights on specific semantic changes, I have deemed relevant to include it in the present study.

5.1 Co-evolution of forms within a paradigm

The method is quite simple and relies on one, crucial hypothesis, that is: a construction evolves as a whole. It means that all members of a given construction should display similar frequency profiles. A weaker version of this hypothesis would be that, if two

linguistic forms show the same evolution, then they both are instances of the same construction. This would make room for the possibility that different members of a same construction occasionally show different frequency profiles, but the stronger hypothesis is required to pretend detecting speciation events.

5.1.1 Detecting correlations

The question is: how do we know if two frequency profiles are the same? We can simply use a standard measure of correlation between the two datasets, for instance the Pearson coefficient. Hence, as soon as we have two frequency profiles, preferentially averaged, we can obtain the coefficient. Let us remind the definition of this coefficient P_e , for two quantities X and Y :

$$P_e(X, Y) = \frac{\langle (X - \langle X \rangle)(Y - \langle Y \rangle) \rangle}{\sqrt{\langle (X - \langle X \rangle)^2 \rangle} \sqrt{\langle (Y - \langle Y \rangle)^2 \rangle}}, \quad (5.1)$$

where $\langle \cdot \rangle$ denotes the average. It can also be rewritten, introducing σ_X and σ_Y , the standard deviation of each quantities:

$$P_e(X, Y) = \left\langle \frac{X - \langle X \rangle}{\sigma_X} \frac{Y - \langle Y \rangle}{\sigma_Y} \right\rangle, \quad (5.2)$$

which highlights the fact that this correlation measure is independent of the actual scale of the considered quantities.

This correlation measure is symmetric with respect to X and Y , and treats all data points equally and independently (it does not account for the fact that they are a time succession). This coefficient varies between -1 and 1 . A value of 1 corresponds to an identity between the two datasets; the closer to 1 the coefficient gets, the more correlated they are. Conversely, negative values indicate an anti-correlation between the two datasets. A value of -1 would correspond to a case where Y is always as far as its mean than X is in the opposite direction.

A perhaps more relevant measure could account for the fact that if X and Y are close for say $N/2$ successive data points, where N would be the number of points of the whole process, and far away for the remaining of the process, then it is more relevant than if they are close one decade over two, and far away the other interspersed half. The Pearson coefficient has however the advantage of being straightforward, properly scaled, and widely used. Furthermore, it is scale-independent, which is an enormous advantage. Indeed, not all members of a paradigm have the same overall frequency: some are high-frequency members, some are only peripheral. These differences of frequencies do not matter though; the importance is that the relative variation of frequency is the same in both cases. Therefore, there is no need to rescale properly the frequency data in order to march the two profiles for them to be compared according to the Pearson criterion.

It should be noted that this Pearson coefficient depends on the period on which the correlations are computed. Two forms may be correlated at some period in time and live a life of their own later on.

5.1.2 Preliminary test

We tested this criterion with the following paradigm, on the period 1721-2020:

- par *souci* (de|du|des|d');
- par *crainte* (de|du|des|d');
- par *peur* (de|du|des|d');
- par (gout|goust|goût) (de|du|des|d');
- par *amour* (de|du|des|d');
- par *besoin* (de|du|des|d').

Correlations between paradigm members

From the Pearson coefficients (Table 5.1), it appears that *par crainte de* and *par peur de*, which share a very high Pearson coefficient of 0.94, are in fact one and the same construction, as well as *par besoin de*. Interestingly, *par souci de* follows its own evolution and is only weakly correlated with the three others (0.58, 0.73 and 0.75). It would also seem that *par souci de* and *par goût de* are correlated, while *par amour de* is marginal.

Form	<i>souci</i>	<i>crainte</i>	<i>peur</i>	<i>goût</i>	<i>amour</i>	<i>besoin</i>
<i>souci</i>	1	0.58	0.73	0.95	0.29	0.55
<i>crainte</i>	0.58	1	0.94	0.68	0.78	0.96
<i>peur</i>	0.73	0.94	1	0.81	0.68	0.86
<i>goût</i>	0.95	0.68	0.81	1	0.33	0.65
<i>amour</i>	0.29	0.78	0.68	0.33	1	0.83
<i>besoin</i>	0.55	0.96	0.86	0.65	0.83	1

Table 5.1: Matrix of the Pearson coefficients between frequency profiles of the test paradigm.

Is the hypothesis reasonable?

Independently of our hypothesis, how expected was this result? First of all, we might want to attribute this correlation to a correlation between the words *crainte* and *peur* themselves, which are near synonyms in French. However, those two words, on the period under scrutiny, are not only not correlated, they are also negatively correlated ($P_e = -0.82$) — note that we limited our research to *crainte de* and *peur de* because of the 50,000 occurrences limit of Frantext. The correlation between *par peur de* and *par crainte de* is therefore plainly imputable to a construction, or a pattern, [par {P_N} de]. This same pattern could actually correspond to several, different constructions, and even if it would correspond to only one single construction, different members of the {P_N} could compete between each other for the control of the construction.

This correlation was therefore not trivially predictable, unless we make use of our hypothesis. It is clear that *par peur de* and *par crainte de* are not entrenched as two separate constructions. Indeed, they both show the same, idiosyncratic meaning, only

modulated by the subtle nuances between *crainte* and *peur*. Then, as they are two versions of the same linguistic object, which is the construction, they should follow the same frequency profile. This is not a trivial claim, and it only makes sense in the framework of Construction Grammar which we recalled in chapter 1.

Even if we can consider that *par peur de* and *par crainte de* are members of the same overarching construction, the two show some labor division, and their own paradigms P_X , defined by $[\text{par } \{P_N = \text{peur, crainte}\} \text{ de } \{P_X(P_N)\}]$, show some differences, though the very low co-occurrence frequencies associated with the different collocates make the results neither robust nor reliable:

par peur de: *ne pas* (7), *se compromettre* (6), *la guerre* (3), *ne plus* (3), *se tromper* (2), *la mort* (2), *manquer à* (2), *la solitude* (2), *se rendre* (2), *la contagion* (2), *mourir sous* (2), ...

par crainte de: *ne pas* (7), *la mort* (6), *avoir à* (5), *ne pouvoir* (5), *se voir* (4), *ne plus* (3), *passer pour* (3), *n'être* (2), *faire trop* (2), *un mouvement* (2), *la noce* (2), *son maistre* (2), ...

It should be stressed that *par crainte de* is about twice as frequent as *par peur de* (534 vs 268 occurrences, even though *peur de* and *crainte de* show the inverse unbalance, with 15232 vs 7023 occurrences), therefore the preferences exhibited by *par peur de* are twice as significant. The most striking feature of this labor division would be perhaps that *par crainte de* shows a clear preference for indefinite articles compared to *par peur de* (conditional probabilities of 0.12 vs 0.06).

How exactly this labor division would come to take place, and how it would articulate with the diachronic evolution of the over-arching construction, is nevertheless an open question.

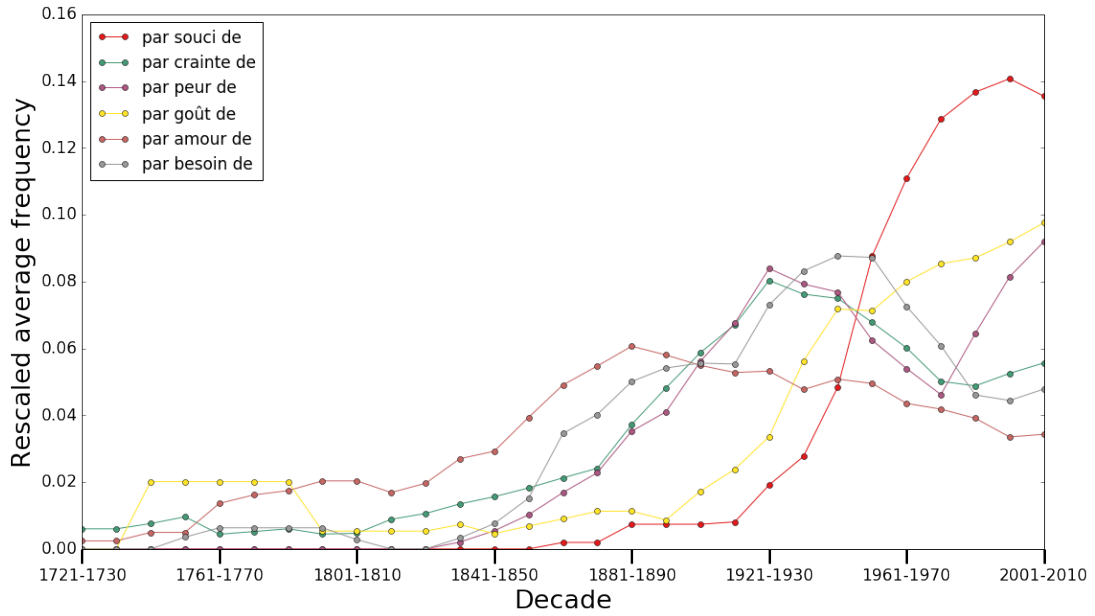
Are the results reliable?

We can wonder whether these correlation coefficients are reliable. They are at least consistent with the plot of the involved profiles (Fig. 5.1a). However, the agreement of the curves for *par goût de* and *par souci de* is not that good, despite their previous correlation score. A look at their derivatives shows for instance that *par souci de* displays a peak of growth (which corresponds nicely to the inflexion point of the S-curve) which does not coincide with the peak displayed by *par goût de* (Fig. 5.1b).

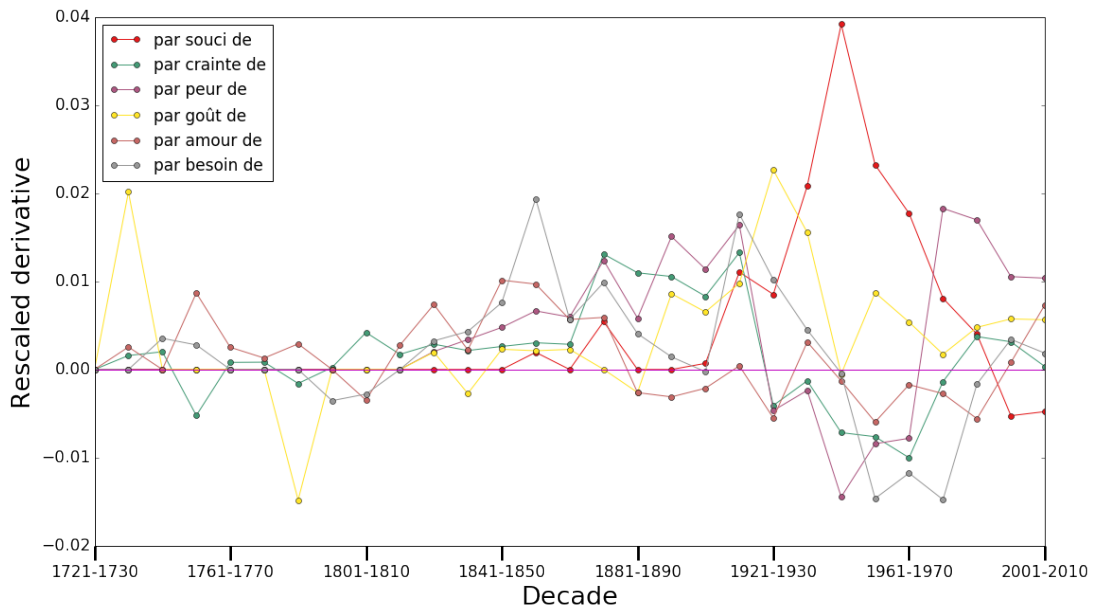
It seems therefore recommended to look also at the correlation between the derivatives of the frequency curves (Table 5.2). This is in line with a suggestion made by Koplenig (2015) to improve the reliability of time series correlations. However, because the derivatives profiles are more noisy than the frequency profiles, we expect lower Pearson coefficients to be still significant. In the present case, *par crainte de* and *par peur de* are still correlated on this level, as well as they are with *par besoin de* to a lesser extent, while *par souci de* and *par goût de* do not appear much correlated anymore.

Summary

This first test shows that looking at the correlation coefficients between the frequency profiles is a reliable way to find paradigmatic relationships between different linguistic forms. However, as two forms growing at the same time for independent reasons could



(a)



(b)

Figure 5.1: (a) Rescaled frequency profiles of all members of Paradigm 1. (b) Rescaled derivative profiles of all members of Paradigm 1. The scaling is obtained in both cases by dividing, for each decade, the frequency of each form of the paradigm in this decade by its total frequency over the studied period.

Form	<i>souci</i>	<i>crainte</i>	<i>peur</i>	<i>goût</i>	<i>amour</i>	<i>besoin</i>
<i>souci</i>	1	-0.46	-0.5	0.22	-0.33	-0.23
<i>crainte</i>	-0.46	1	0.73	-0.04	0.08	0.53
<i>peur</i>	-0.5	0.73	1	0	0.06	0.25
<i>goût</i>	0.22	-0.04	0	1	-0.24	0.11
<i>amour</i>	-0.33	0.08	0.06	-0.24	1	0.50
<i>besoin</i>	-0.23	0.53	0.25	-0.11	0.50	1

Table 5.2: Matrix of the Pearson coefficients between derivative profiles of the test paradigm.

tend to be correlated from the point of view of the Pearson coefficient, it is advised to consider an additional layer of empirical evidence. Only the correlations which are attested for both the frequencies and their derivatives can be deemed significant.

In the following, we shall consider as a thumb rule that a correlation is significant if the Pearson coefficient is both above 0.85 for the frequencies and above 0.45 for the derivatives (with the respective negative threshold for anti-correlations).

5.1.3 A case study

To give a flavor of what kind of linguistic phenomena this method allows to shed light on, we investigated the paradigm of quantifiers of the form [(un|une) {N} (de|d')]. To restrict the study, we only considered intensifiers associated with large quantities. Admittedly, only one single construction is involved here. Therefore, why this paradigm is evolving, and why some of its members would bundle up together and evolve similarly, is an open question. This shows at least that the ‘strong’ hypothesis (two forms are co-members of a construction if and only if they co-evolve) is untenable as such.

Definition of the paradigm

The members of the paradigm which we included are listed here. The number of occurrences of the form is given between parentheses:

- nombre (12563)
- tas (4700);
- foule (4629);
- infinité (3351);
- millier (3223);
- dizaine (3161);
- quantité (2837);
- centaine (2775);
- million (2353);
- multitude (2211);
- douzaine (1752);
- masse (1392);
- amas (1158);
- montagne (921);
- monceau (651);
- nuée (583);
- assemblée (477);
- milliard (344);
- légion (278);
- tonne (269);
- abondance (222);
- myriade (214);
- profusion (210);
- étalage (157);
- débauche (141);
- cohorte (83);
- kyrielle (51);
- foisonnement (50);
- festival (47);
- flopée (46);
- déballage (22);
- pléthore (12);
- max (10);
- foison (9);
- opulence (8);
- foulitude (4).

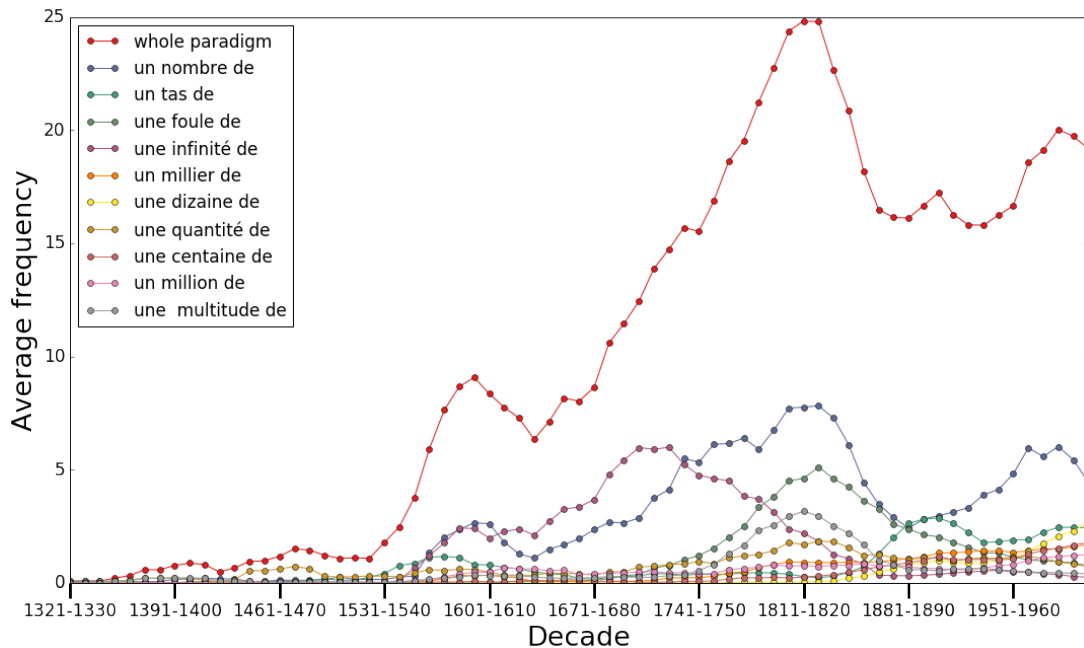


Figure 5.2: Frequency profiles of the ten most frequent members of the paradigm, and of the whole paradigm (frequency sum of all members), over the period 1321-2020.

We can first have a look on the frequency profile on the whole period 1321-2020 (Fig. 5.2). The frequency of the whole paradigm (i.e., the total frequency achieved by all members taken together) itself rises twice, then falls down and rise again. The initial rise can be explained by at least three reasons: 1 - a constructionalization occurred; 2 - we did not include older members of the paradigm, out of use nowadays, so it is just an effect of an incomplete query; 3 - the indefinite article was not used much before this date anyway. Underlying this overall picture, the individual members of the paradigm themselves have a life on their own, which is only remotely correlated with the paradigm as a whole. The profiles of the individual members of the paradigm, even the prominent ones which contribute for most of the total frequency, do not suffice to explain the more general behavior of the construction.

Remarkable correlations

We can now apply our method to detect co-evolutions of forms. We remind here that the method works best on a limited time period, so we have chosen the period 1771-1970, which covers the fall for its major part. We retained a few interesting facts from the different correlations revealed through our method:

- *kyrielle* and *myriade* are correlated, though they are both marginal;
- *assemblée* and *infinité*, which are both decaying, are also correlated;
- *tas*, *profusion*, *centaine* and *dizaine* are all mutually correlated, and *dizaine* and *centaine* are furthermore with *milliard*;

- *foule*, *quantité* and *multitude*; they furthermore correlate with the whole paradigm; also, *nombre* correlates with *quantité* and *multitude*
- last, *tas*, *dizaine* and *centaine* are anti-correlated with *multitude*; *tas* is furthermore anti-correlated with *nombre* and *dizaine* with *foule*.

The first correlation is remarkable in the sense that it is clean and convincing (Fig. 5.3), despite being supported by very small sets of occurrences (respectively 51 and 214 for *kyrielle* and *myriade*). Their Pearson correlation coefficient is equal to 0.83 for the frequencies, and 0.50 for the derivatives.

This co-evolution is interesting for several reasons, alongside with it being a pleasant empirical curiosity. First, one could think that the two words would be used by the same authors, but this is not the case; with the exception of the decade 1901-1910, where two authors, Pierre Loti and Romain Rolland, use both. This decade explains the peak of frequency at that time, and therefore the associated peak of derivative, as well as the decrease, five decades later, since the averaging procedure makes not feel anymore this particular decade. This certainly diminishes the reliability of the result, but still, if there is any particular association between *kyrielle* and *myriade* felt by the two authors, then other language users could have felt it too. This result is furthermore robust, as it holds even if we only include singular uses of *une kyrielle de* and *une myriade de* (the latter having three times more plural uses — *des myriades de* — than singular ones). Second, it makes sense that they are correlated. Though they are of extremely different semantic origins, both are ultimately derived from Ancient Greek, which the *y* letter recalls (most French words displaying a *y* come from Greek stems). They are therefore associated with scholarship and erudition and their use is exclusively literary. They are precious, exquisite words, therefore standing apart from other members of the paradigm.

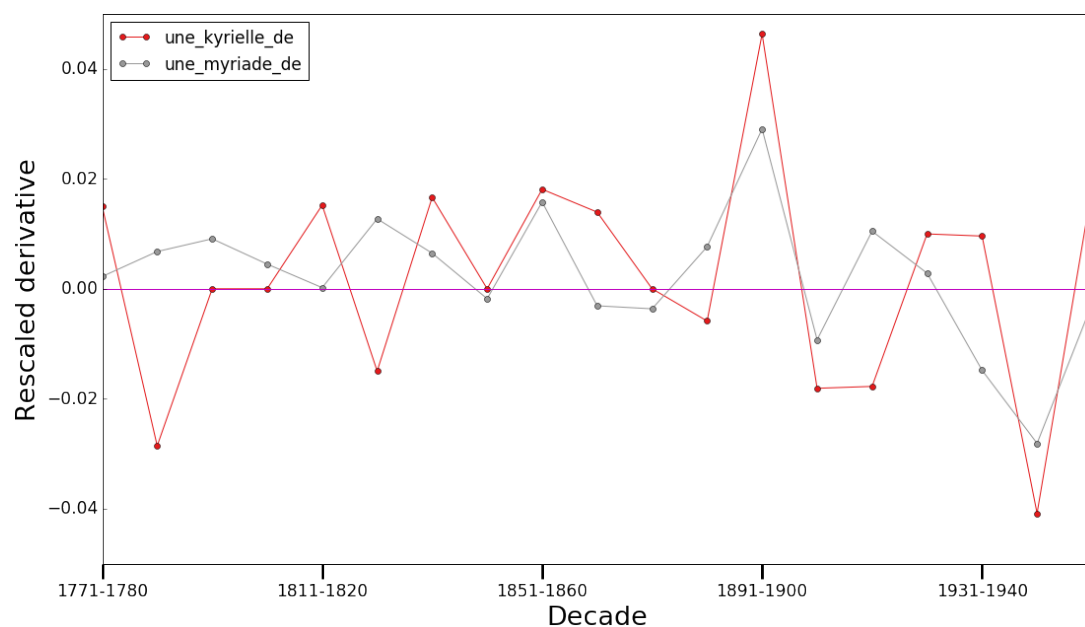
The second correlation we present, between *une infinité de* and *une assemblée de*, is less appealing, albeit a bit intriguing. We can actually consider that *une assemblée de* (‘a gathering of’) does not really mean ‘a great deal of’ as the others, though admittedly there is no gathering if there is only a few people. Therefore, even if *une assemblée de* was nonetheless partly belonging to the paradigm of quantifiers, its use was highly specific and conveyed a slightly different meaning than the other members of the paradigm, with also the rather bizarre possibility to be followed by a number (e.g. ‘une assemblée de cinquante gens de lettres’, ‘a gathering of fifty scholars’). This odd behavior can led us think that this correlation with *une infinité de* might be a coincidence, but the Pearson scores are very good (0.93 for the frequencies and 0.58 for the derivatives); furthermore, both are supported by a reasonable number of occurrences.

Clusters of correlations

The three last results are the most crucial to the understanding of the paradigm at this time. Yet, they immediately raise some concern. Indeed, we could have expected the correlation relations to be transitive: if the evolution of a form is correlated with a second, and the second with a third, then the first and the third must also be correlated. It happens this is not the case. One could therefore state as a rule that for two forms



(a)



(b)

Figure 5.3: (Co-evolution of *une kyrielle de* and *une myriade de*. (a) Rescaled frequency profiles. (b) Rescaled derivative profiles. The scaling is obtained in both cases by dividing the frequency of each form of the paradigm by its total frequency over the studied period.

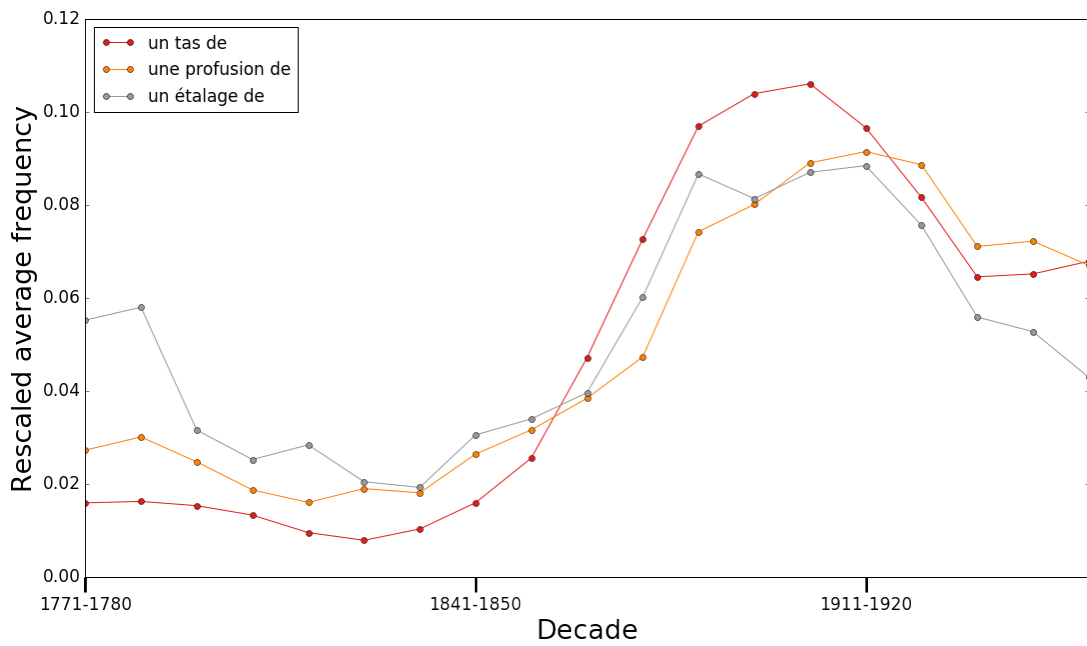
to co-evolve, they must be mutually correlated, and also share the same correlates. Only complete subgraphs would matter, and a form could belong only to one of these subgraphs. In this case, the correlation between the pair *une centaine de/une dizaine de* and *un milliard de* would be discarded. Or should we consider instead that the links between *un tas de/une profusion de* and *une dizaine de/une centaine de* should be severed, especially since *un tas de* and *une profusion de* belong to another cluster, semantically more uniform (Fig. 5.4)?

Actually, this would be missing the point. True, such a clique would be a clear indication that the correlations involved are more reliable than others. Yet, the complexity of the process suggests that there would be some semantic gradation between members of the paradigm, a gradation reflected in these intransitive correlation relationships. Indeed, the period we study here sees a shift of the paradigm, with the replacement of some previous prominent members (*quantité, foule, multitude*) by new ones, which may or may not have marginally existed before. All those members are not homogeneous though, and show semantic differences. As they are all more or less part of the competition at work, they co-evolve, yet they also form groups and individual attachments, and the details of their co-evolution are more diverse (Fig. 5.5). That those differences are captured by our approach seems worth noting by.

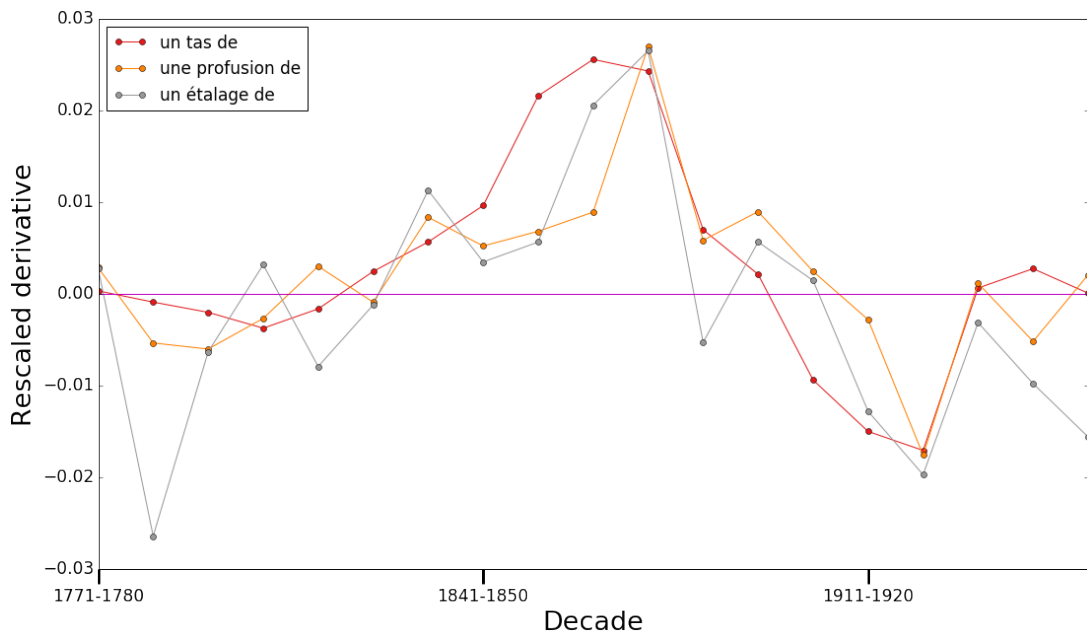
We can try to understand a bit more those differences. The ‘numbers’ members of the paradigm (*centaine, dizaine, milliard*) are more precise in that they give an order of magnitude of the quantity (even if it is used in a most bleached way, with little reference to the actual scales involved). Conversely, *un tas de* and *une profusion de* are extremely vague. Also, *centaine, dizaine, milliard* and *tas* are often used with *des* (the difference between *des tas de* and *un tas de* would be the same as between *lots of* and *a lot of*), but *profusion* and *étalage*, much less likely. Therefore, the actual logic behind all those correlations is not very clear.

The other cluster of interest (Fig. 5.6), made of *une foule de, une multitude de* and *une quantité de*, arises the same puzzlement. The correlation between *foule* and *multitude* could be easily understood: both are nearly synonymous, tend to be preferentially used for people, and, aside from this particular construction, are found to be used with a singular article: *la foule* and *la multitude* are generic forms to refer to ‘the people’ as an indifferentiated whole. Yet, if *la multitude* correlates avec *une multitude de, une foule de* and the whole quantifiers paradigm, *la foule* correlates with none. So, if the two forms share a lot of semantic features, they differ as well in unexpected ways and the relationships between *une foule de* and *une multitude de* must be ascribed to the constructions themselves, and not to their individual items.

As for why *une quantité de*, ‘a quantity of’ — which is, by the way, delightfully uninformative, but corresponds to a usual semantic change, alongside with *un nombre de*, see also the English *a number of* —, would be correlated with the two, it is quite a mystery. Yet, the additional correlation between *une quantité de* and *un nombre de* is expected, as they are semantically extremely close. Both have followed the same semantic channel, from the indication of a quantity to the, probably pragmatically inferred from their uninformativeness, indication of a rather large one. This common origin might nonetheless have led to believe that one would have replaced another through a repeated cycle of pragmatic strengthening, wearing off, and recycling (Haspelmath, 1999). On the contrary, they evolved together, seemingly striding on

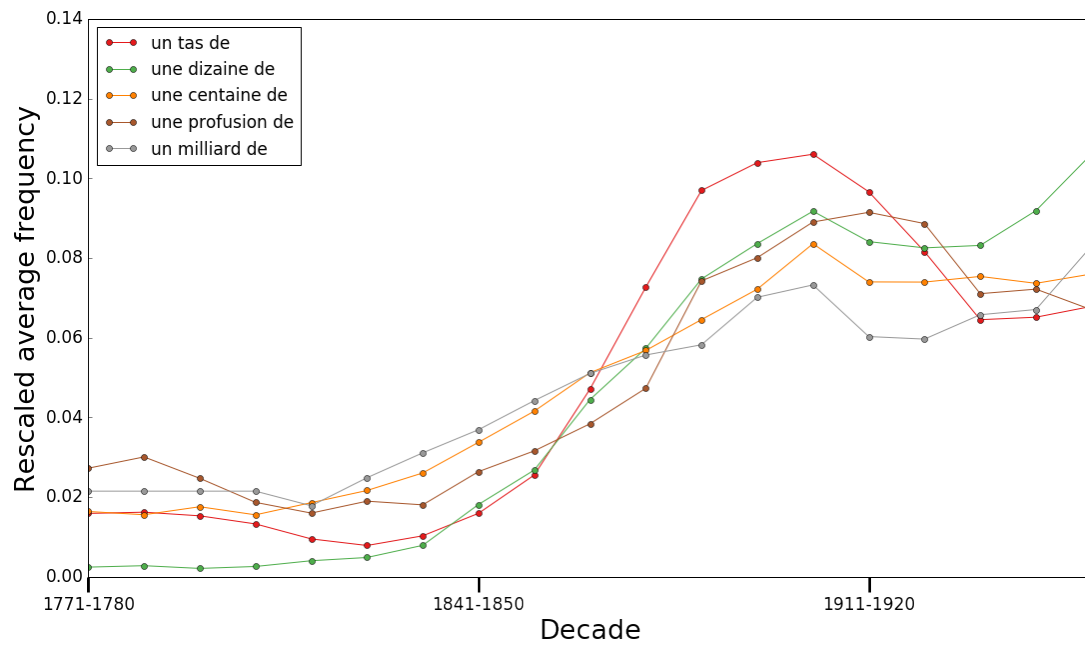


(a)

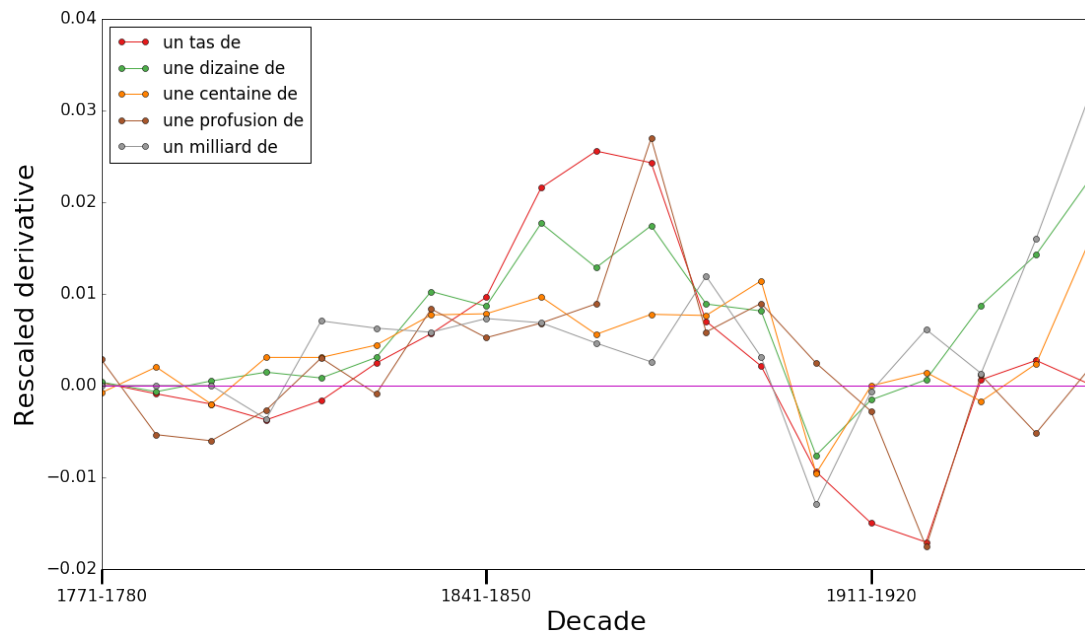


(b)

Figure 5.4: Co-evolution of *un tas de*, *une profusion de* and *un étalage de*. (a) Rescaled frequency profiles. (b) Rescaled derivative profiles. The scaling is obtained in both cases by dividing the frequency of each form of the paradigm by its total frequency over the studied period.

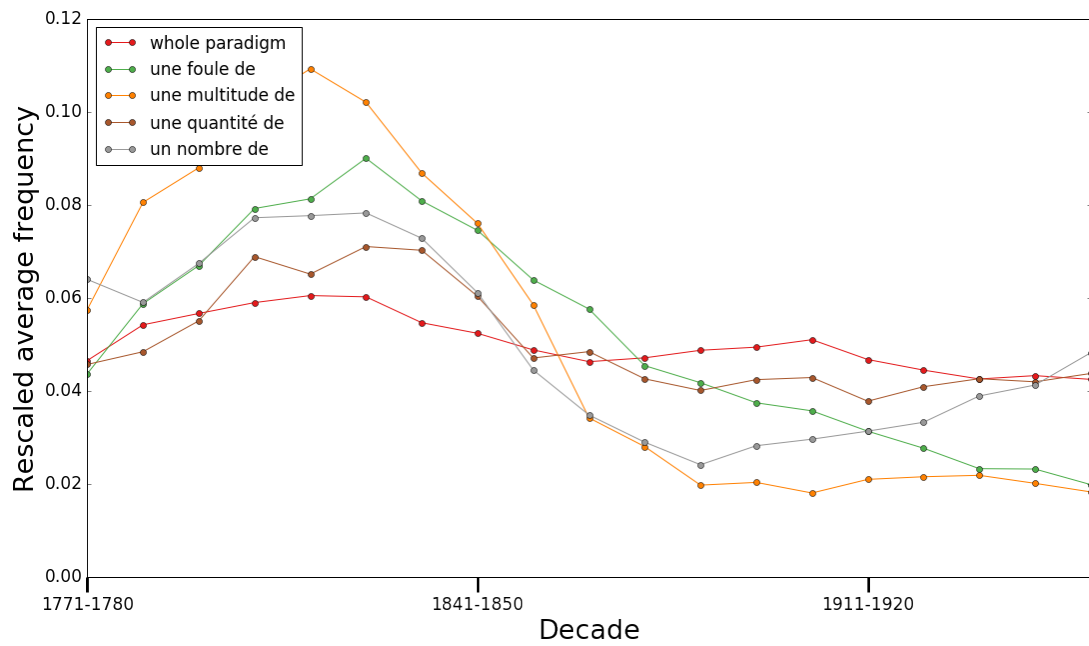


(a)

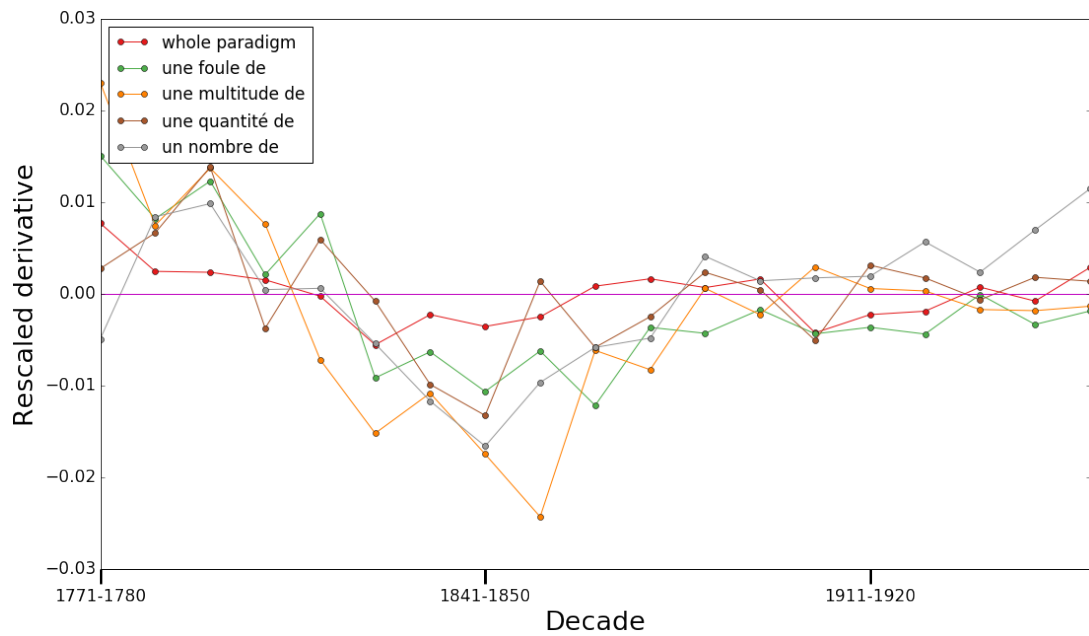


(b)

Figure 5.5: Co-evolution of *un tas de*, *une profusion de* and *une dizaine de*, *une centaine de* and *un milliard de*. (a) Rescaled frequency profiles. (b) Rescaled derivative profiles. The scaling is obtained in both cases by dividing the frequency of each form of the paradigm by its total frequency over the studied period.



(a)



(b)

Figure 5.6: Co-evolution of *une foule de*, *une multitude de* and *une quantité de*, *un nombre de* and the whole paradigm. (a) Rescaled frequency profiles. (b) Rescaled derivative profiles. The scaling is obtained in both cases by dividing the frequency of each form of the paradigm by its total frequency over the studied period.

the same semantic path. This is one more hint that language change occurs at several, entangled levels.

Anti-correlations

The same kind of transitivity issues are raised in the case of the anti-correlations, and one might wonder whether adding a principle of completeness to filter out the spurious correlations.

However, this issue is a little bit deeper than it seems. Indeed, the same transitivity principle should also be applied in the case of anti-correlations. That is, if a form anti-correlates with a second form, it must also anti-correlate with all cluster members of this other form. Moreover, all correlates of the first form must themselves anti-correlate with all correlates of the second. This seems a requisite impossible to meet; first, because we deal with limited empirical data; second, because it might not be relevant in all situations.

In the present case, we have an interesting situation with multiple anti-correlations between our two main clusters (Fig. 5.7). On the decaying side, we have *une foule de*, correlated with *une multitude de*, correlated with *un nombre de*; on the growing side, every one is mutually correlated (*un tas de*, *une centaine de* and *une dizaine de*), but the latter two are much more closely correlated. The three are anti-correlated with the central node of the opposite cluster. Then, *un tas de* is also correlated with *un nombre de* while *une dizaine de* is correlated with *une foule de*, and we could consider that *une centaine de* also, since the Pearson coefficients are close to the threshold (Fig. 5.8). Interestingly, the anti-correlation between *un tas de* and *une quantité de*, which would be expected given the high proximity between *une quantité de* and *un nombre de*, is very close to the significant threshold.

Therefore, though we can consider that the two clusters overall replace each other, they do so while maintaining some sort of a semantic gradation between their members. This gradation may not be motivated by the same semantic traits on the two sides of the replacement though, and the complete picture may be extremely complex. Yet, our method is able to highlight some interesting phenomena about the replacement process which would have been hard to detect otherwise. Furthermore, these findings are entirely data-driven and the only theoretical input we fed the investigation with was the list of paradigm members.

A proposal of explanation

Further investigation and a true linguistic expertise would be required to understand what is happening in this particular case. I will nevertheless propose some sketch of an explanation, by looking at the collocation of each of these two clusters members (Table 5.3). In this table, the tightness of the paradigm is the contribution to the total number of occurrences of a form from the twenty first members of its collocates paradigm. If the paradigm is very spread, then this number is expected to be very low, while if the paradigm is specific and the form only compatible with a restricted number of collocates, it will be close to 1.

From this table, one can infer what are the prototypical uses of each of the members: *un nombre de* was chiefly used for denumerable entities, *une quantité de* was used

Forms	First	Second	Third	Fourth	Fifth	Tightness
<i>un nombre de</i> (2684)	hommes (0.035)	années (0.028)	personnes (0.013)	idées (0.009)	individus (0.008)	0.21
<i>une foule de</i> (2119)	choses (0.048)	gens (0.024)	hommes (0.023)	détails (0.013)	individus (0.010)	0.28
<i>une quantité de</i> (921)	eau (0.039)	choses (0.018)	chaleur (0.018)	gens (0.010)	individus (0.009)	0.23
<i>une multitude de</i> (781)	hommes (0.017)	faits (0.015)	choses (0.014)	êtres (0.013)	causes (0.009)	0.25
<i>un tas de</i> (1397)	choses(0.115)	pierres (0.026)	gens(0.009)	ordures(0.007)	bêtises (0.006)	0.25
<i>une centaine de</i> (597)	mille (0.084)	francs (0.070)	pas (0.060)	années (0.049)	mètres (0.034)	0.54
<i>une dizaine de</i> (466)	années (0.191)	jours (0.146)	mille (0.073)	hommes (0.017)	pas (0.011)	0.43

Table 5.3: Five first members of the paradigm of collocates of each members of the two competing clusters (separated by a double line) for the XIXth century. The total number of occurrences of the form is provided between parentheses. The conditional frequency of the collocate is given between parentheses. Paradigm tightness is also provided.

1871-2020

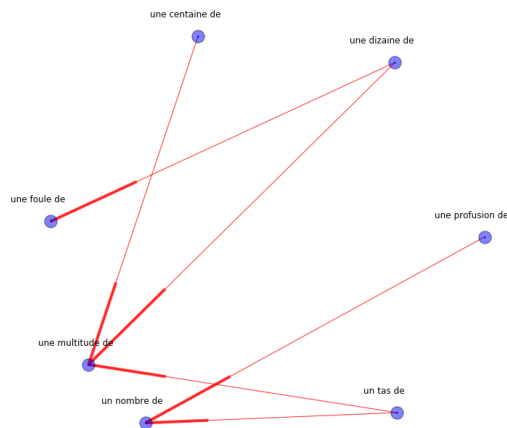


Figure 5.7: Competition network between the two clusters $\{ un\ tas\ de,\ une\ dizaine\ de,\ une\ centaine\ de \}$ and $\{ une\ foule\ de,\ une\ multitude\ de,\ un\ nombre\ de \}$. Links indicate that the two forms are negatively correlated, both in terms of frequencies and their derivatives. The bold extremity points towards the form which frequency is decreasing.

for uncountable entities (all kinds of fluids especially), *une multitude de* was associated with countable things, as well as *une foule de*, with maybe a more particular focus to people, due to a phenomenon of semantic retention from its original use. Both are associated with large quantities of ordered, homogeneous entities. The semantic difference of the two paradigms is subtle, to be sure, but *foule* tends to be associated with more abstract things: we find, in the remaining of the paradigm, words such as *questions*, *mots*, *idées*, *circonstances*, *exemples*, while *multitude* shows *oiseaux*, *insectes*, *plantes*, *filets*, *nerfs*. It would therefore seem that *multitude* conveys an additional idea of diversity and heterogeneity, but not of chaos or disorder.

The paradigm was thus clearly organized when the first cluster was dominant and each member had some specific semantic role within, though they were all quite bleached (the tightness of the parameter is pretty low). Contrary, the second cluster shows very tight collocate paradigms for *une dizaine de* and *une centaine de*. Both are associated with countable, denumerable things. Next to this, *un tas de* ('a heap of') is pretty much unspecific (see how much it collocates with *choses*, 'things') and refers to disordered, disorganized large amount of things. It seems therefore that the whole [un {N} de] paradigm has been reshuffled and its semantic scope has been polarized between the quantifiable, homogeneous entities, and the other stuff, which counterbalances somehow what could be felt as an excess of precision. Thus, we will find *un tas de souvenirs* ('a heap of memories'), *un tas d'idées* ('a heap of ideas'), but never *une centaine d'idées* ('hundreds of ideas'), and conversely, we find *une dizaine de voitures*

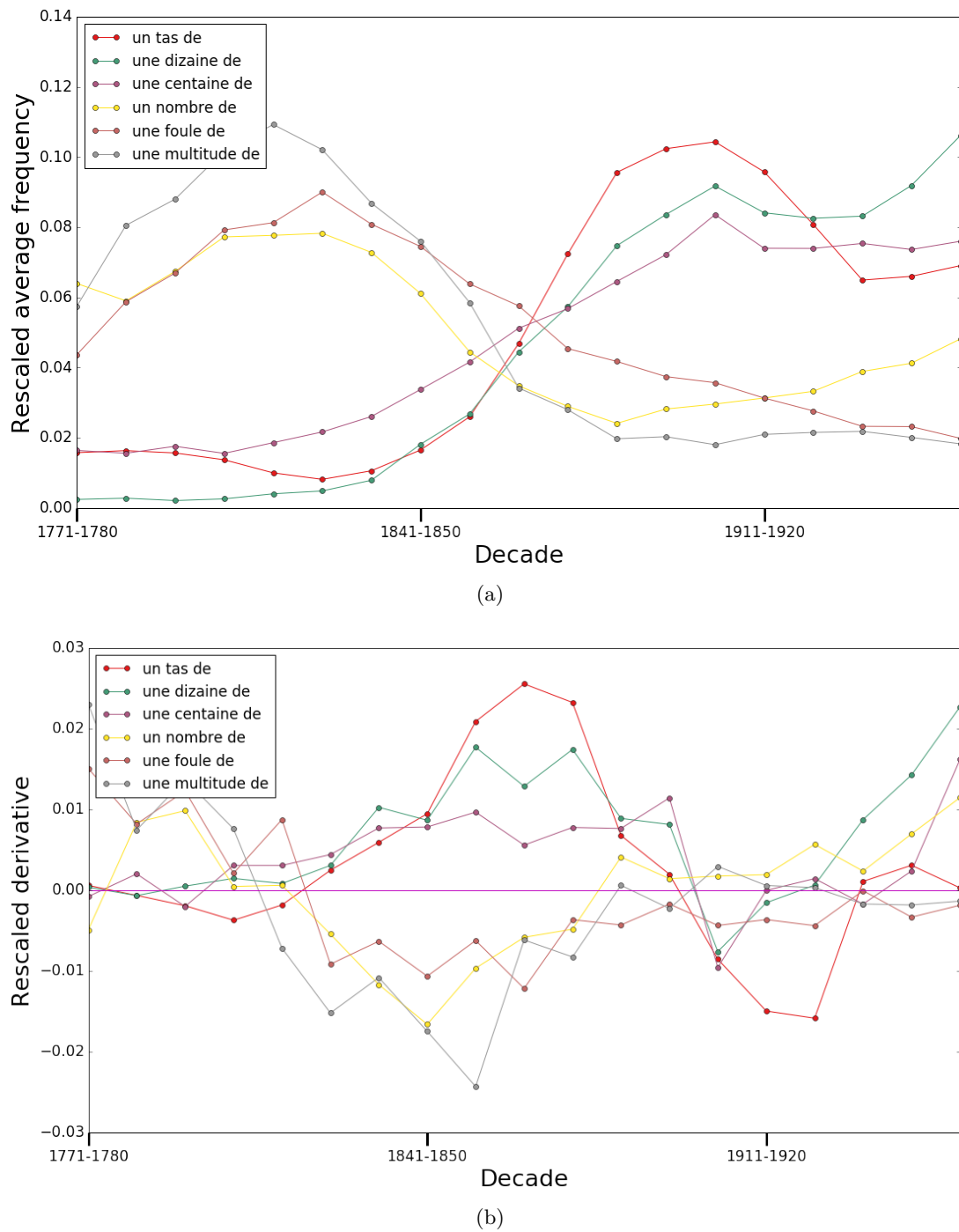


Figure 5.8: Competition between the two clusters $\{un\ tas\ de,\ une\ dizaine\ de,\ une\ centaine\ de\}$ and $\{une\ foule\ de,\ une\ multitude\ de,\ un\ nombre\ de\}$. (a) Rescaled frequency profiles. (b) Rescaled derivative profiles. The scaling is obtained in both cases by dividing the frequency of each form of the paradigm by its total frequency over the studied period.

but not *des tas de voitures* ('tens of cars'/'heaps of cars'). It is also not a coincidence if *des litres de*, which we did not include here because it comes only with a plural article, also appears at this time and follows *une dizaine de* and *une centaine de*. Liquids are not easily quantified, but they are still measurable.

Therefore, *un tas de* took from *un nombre de* the unorganized aspect, from *une quantité de* the indistinguishness, and from *une multitude de* the eclectic character, while the couple *une dizaine de/une centaine de* took from *une foule de* and *une multitude de* the neat, organized, vision of the things to be quantified. It is however surprising that *un nombre de* and *une dizaine/centaine de* are not correlated, since the latter clearly took a lot of uses from the former.

This sketch of a scenario opens the question of what caused the paradigm crisis in the first place. Was it an excess of bleaching? A failure in its internal organization? Did it crumple because of the pressure of the new cluster? This would be worth investigating.

Clusters, competition, and the S-curve

In regard to semantic change, the automatic detection of anti-correlation presents another, crucial interest. Indeed, it has been duly remarked that competitors, in a situation of semantic replacement, are not always obviously identified (Denison, 2003); true, there is a semantic replacement, but what actually is the new variant replacing?

In the previous chapter, we also raised the related question: what exactly is the new variant, what is really changing in an instance of semantic change? This may seem an artificial concern, but the complexity of the [un {N} de] quantifier construction should have shown that several, inter-related processes are at work, in such a way that different members of the paradigm partially bundle up, and yet remain individualized. Therefore, it makes sense to ask what exactly is changing, and more generally, in a case of the semantic change, which linguistic items are competing against each other. Of course, we already hinted at an answer: clusters of paradigm members can compete against each other.

To investigate it further, we propose to do the following. For each member of the *un tas de* cluster, we can obtain the slope of the associated logit, and the boundaries of the growth. Therefore, we can test these boundaries to build up a decreasing S-curve for the members of the opposite cluster (the frequency loss of the depleting variant is expected to mirror the frequency rise of novelty whenever a replacement occurs). True competitors are expected to have the same slope, with an opposite sign. As a technical detail, it may happen that the frequency of the members of the opposite cluster is not monotonic during the period of growth. In this case, we restrict to the largest interval for which all frequency values are comprised between the two boundary frequencies value .

From the results display on Table 5.4, it appears that the possible pairs of competitors are *un tas de* and both *un nombre de* and *une quantité de* (which confirms the correlation analysis), *une profusion de* and *un nombre de* (which are, indeed, negatively correlated according to our criteria, though we did not include it in our previous analysis), and finally *une dizaine de* and *une foule de*. It may seem surprising that *une centaine de* stands apart, but what it shows is that the slope may be extremely

sensitive to the boundaries, as we have already seen in chapter 3.

Yet, the most remarkable result is that the clusters themselves are suitable competitors. It is another hint that the competition may not lie between one variant or another, but between a sub-paradigm of variants and another. Therefore, the difficulty to find a competitor in a case of a semantic expansion may lie in the fact that the change happens not at the level of individual paradigmatic members, but at a larger level. Here, it may be that two paradigms are competing over a construction. That two clusters of co-evolving linguistic forms may be the true competitors of a semantic change is anyway a truly deep result, which could not have been obtained without our correlation analysis.

An ecological analogy

This situation is actually reminiscent of some networks found in ecological set-ups (Gracia-Lázaro et al., 2017), where the situation can be described by a two-layer networks. One layer represents plant species, the other one animal species. Links from one layer to another are mutualistic (the animal feeds from the plant and pollinizes it in return), while links within a layer indicate a competition (the different animal species compete from the different plant species, and the other way around). Here, we have a similar situation, where the members of the paradigm internal to the construction (*{tas, nombre, foule, dizaine, ...}*) could play a role similar to that of the animals, competing for a mutualistic relationship with the different members of the collocate paradigm (*{choses, hommes, détails, ordures, eau, ...}*) playing the role of the different plant species.

It would be interesting to investigate as well whether or not paradigmatic networks, as ecological ones, shows the nestedness property (Bascompte et al., 2003; Nielsen and Bascompte, 2007), which is a specific hierarchical organization ranging from generalist species to specialist ones, such that a specialist of one species type (here, internal members of the construction) interacts only with the generalists of the other species type (here, the different collocates). An alternative would be that the network is organized in small mutualistic specific clusters; i.e., there would be almost independent small sub-clusters corresponding from one layer to another; for instance, the cluster of numeral quantifiers would correspond to the cluster of measurable quantities. Here, there is at least one generalist collocate species, *choses*, but the overall feeling would be that the network does not show nestedness. For instance, the quite specific collocate species *chaleur* and *mouvement* seems only to co-occur with *quantité*, which is also specific; in a nested structure, we would have expected to be correlated with the seemingly more general *nombre*.

5.1.4 Summary

We hope to have shown that the method we propose – to track correlations between diachronic profiles of different linguistic forms — allows to unveil a rich phenomenology which can guide, complete and support a linguistic analysis. The paradigm we focused has been chosen because of the great richness, in any language, of quantifiers, which usually abound in all sort of co-existing variants. We tried to illustrate how the diachronic data could help us to understand and organize this extremely rich and

Forms	Slope	Boundaries	<i>nombre</i>	<i>foule</i>	<i>quantité</i>	<i>multitude</i>	Cluster 2
<i>un tas de</i>	1.22 ± 0.06	1821-1910	-1.12 ± 0.07	-0.80 ± 0.04	-1.37 ± 0.32	-0.97 ± 0.07	-1.02 ± 0.05
<i>une profusion de</i>	1.12 ± 0.08	1841-1920	-1.05 ± 0.03	-1.00 ± 0.07	-0.66 ± 0.55	-1.06 ± 0.20	-1.11 ± 0.01
<i>une dizaine de</i>	0.80 ± 0.02	1791-1910	-1.12 ± 0.07	-0.80 ± 0.04	-1.37 ± 0.32	-0.93 ± 0.05	-1.02 ± 0.05
<i>une centaine de</i>	0.56 ± 0.02	1801-1910	-1.12 ± 0.07	-0.80 ± 0.04	-1.37 ± 0.32	-0.93 ± 0.05	-1.02 ± 0.05
Cluster 1	0.96 ± 0.02	1821-1910	-1.12 ± 0.07	-0.80 ± 0.04	-1.37 ± 0.32	-0.97 ± 0.07	-1.02 ± 0.05

Table 5.4: Slope of the sigmoid of the forms of ‘Cluster 2’ = {*un nombre de, une foule de, une quantité de, une multitude de*} when we assume that they are the competitors of the forms of ‘Cluster 1’ = {*un tas de, une dizaine de, une centaine de*}, so that the frequency boundaries must coincide. The requisite that all intermediate frequencies must be comprised between those boundaries leads to a possible adjustment, which is why some values are identical for forms of Cluster 1 even though their growth boundaries are different. The growth boundaries and the logit slope of Cluster 1 members have been obtained as a result of our extracting pattern procedure. They are provided in the second and third columns of this table.

complex paradigm. It would have been interesting to add in the picture quantifiers of small quantities (*un peu de*, *un brin de*, *un zeste de*, etc.), and quantifiers relying on different patterns (*beaucoup de*, *plein de*, *masse de*), but the aim here was to illustrate the method, not to provide a consistent study on quantifiers.

Especially, we have evidenced that the competition, during a semantic change, might not act at the level of the individual linguistic items, but at a more global, paradigmatic levels, so that whole clusters of different forms can compete cohesively against each other. This is, to my knowledge, an observation which has never been put forward in the literature. It may also be the first time that a competitor is identified by empirical means only, and not posited a priori. It pleads in favor of a deeper investigation of the frequency profiles of linguistic forms, which can provide a remarkable amount of information on change. Although these frequency profiles are often blurry and seemingly devoid of any sense, we tried to argue that, by focusing on a limited period of time and by running different kinds of analyses (frequency correlation analysis, S-curve pattern analysis, collocational preferences analysis), it was possible to understand a bit more which processes were at work.

I must stress that, contrary to biological data, linguistic data can be extremely clean. If things appear noisy or a bit messy, it does not necessarily mean that data is too sparse, and that we need a bigger, more balanced corpus. Indeed, it can be that several processes are interfering, that a lot of things are simultaneously happening and influence each other in a complex picture, or that we simply do not look at the relevant level, where the game is really played. Here, we did not provide any concrete and general means to answer this question and to identify at which linguistic level the phenomena can be the most easily understood. We only suggested empirical ways to investigate those phenomena in more details, and to support the linguistic expertise.

Finally, there are certainly many interesting other empirical analyses that could reveal new and crucial insights on a given instance of change. Here, we superficially considered the collocates paradigms, but they probably encode a lot more information than a list of preferential collocations. For instance, one might advantageously look at the distribution profiles of those collocates paradigms. We saw for instance, in the case of *un tas de*, that the distribution was highly peak at *un tas de choses*, and then was pretty much a flat long tail, indicative of the function of *tas* — to serve as some sort of a joker when the stuff the speaker wants to point out as abundant is too messy to be described by a precise, scaled or measurable quantity. There are certainly universal features of language hidden in those distributions, as well as some idiosyncrasies which could support a functional typology of forms. It would seem, for instance, that they are pretty much exponential, except perhaps during times of change, when there are severe reconfigurations of the collocates paradigm. These claims are certainly speculative, but it would be worth investigating.

5.2 Diachronic networks of co-evolution

The correlation matrices we built out in the previous section can straightforwardly serve to draw networks between different linguistic forms from a single or several paradigms. Furthermore, as the coefficients of the correlation matrices depend on

the time period, we can consider a gliding time window on which to compute these coefficients, so as to make the networks evolve diachronically. This actually requires nothing more than what we did in the previous section, but proves to be a very convenient and adaptive tool to build diachronic networks between all sorts of linguistic forms, as long as one can get their frequency profile through the proper search query on a suitable corpus.

In the following, we should consider two kind of such time-evolving networks: the network built out of the correlations, which provides a picture of the internal organization of a paradigm, and the network built out of the negative correlations, so as to identify all possible instances of competition over time.

5.2.1 Internal organization

We consider the diachronic evolution of the [un {N} de] quantifier paradigm. To do so, we compute the correlation coefficients over a time window of ten decades, every thirty decades, starting from the period 1721-1820 and up to the twentieth century (1901-2000). To make the networks a bit less fuzzy, we raised the threshold to 0.85 for the frequency and 0.45 for the derivative. A link between two linguistic forms indicates that their respective time evolutions correlate, both in terms of frequency and derivative, with associated Pearson coefficients higher than their respective thresholds. The network is randomly built out of a spring layout, whose weights w are given by the mean of the Pearson coefficient: $w = (P_e(\text{frequency}) + P_e(\text{derivative}))/2$. In the following, we compute the Pearson coefficient over time windows of ten years, shifted of ten years from one to another (i.e. we consider the time windows 1721-1820, 1731-1830, etc.)

Quantifier paradigm

We distinguish three phases of the evolution of the paradigm. The first phase (Fig. 5.9) runs from 1721 to 1780, speaking in terms of the starting point of the time window on which the Pearson coefficients are computed (i.e., from the time window 1721-1820 to the time window 1771-1880). In this phase, the network is dominated by one giant component. The diachronic profile of the whole paradigm is part of a big, fully interconnected cluster. Only one little component made of the two members *une profusion de* and *un étalage de* are isolated, but they are extremely marginal in terms of frequency. There are more prototypical members of the cluster (*une masse de*, *une foule de*, *un monceau de*) and more marginal ones (*une débauche de*, *une montagne de*, *une centaine de*). Then, in 1751-1850, one of the marginal members, *une centaine de*, shifts away from the giant component, and forms a component of its own with additional members (*une dizaine de*, *une tonne de*).

In the second phase (Fig. 5.10), the isolating component grows bigger as some items are marginally attracted to the newly appeared pole. The semantic organization of the whole paradigm tears apart. Prototypical members of the paradigm are still densely connected, but their frequency is now decaying. Soon they join the already decayed members, *une infinité de* and *une assemblée de*, while the new, opposing cluster becomes larger and tighter. The paradigm itself decays with the fall of its prototypical cluster. This phase lasts from 1781 to 1870.

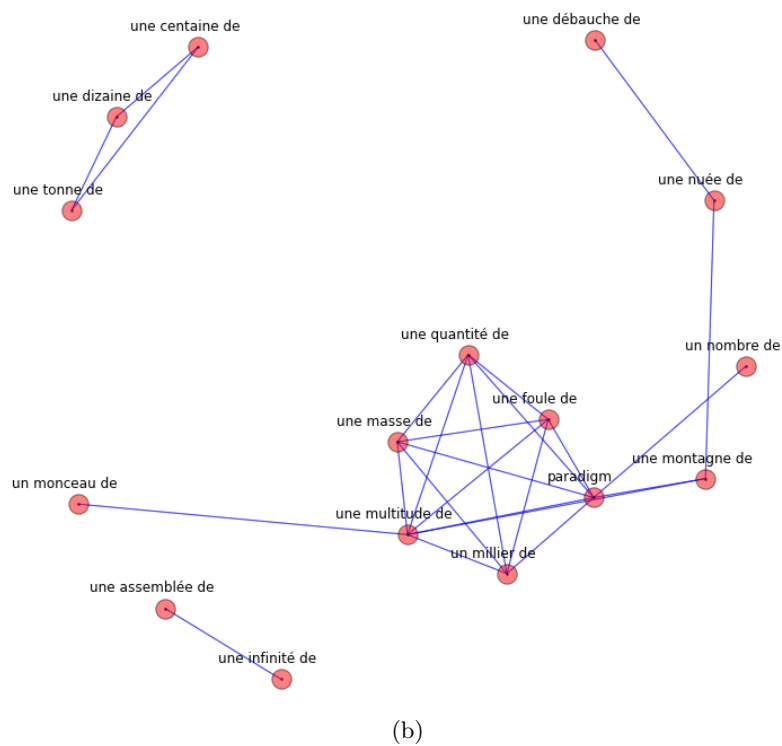
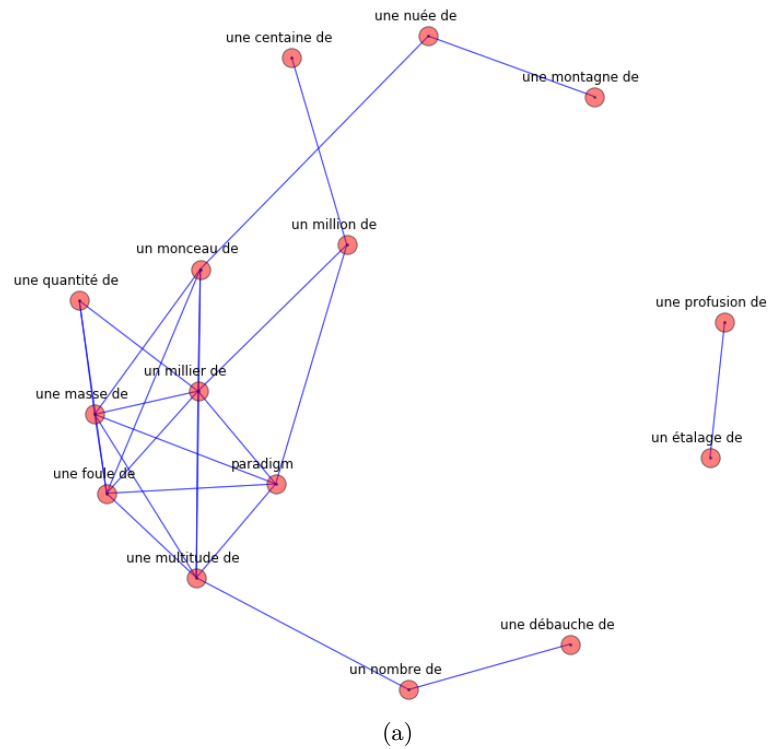
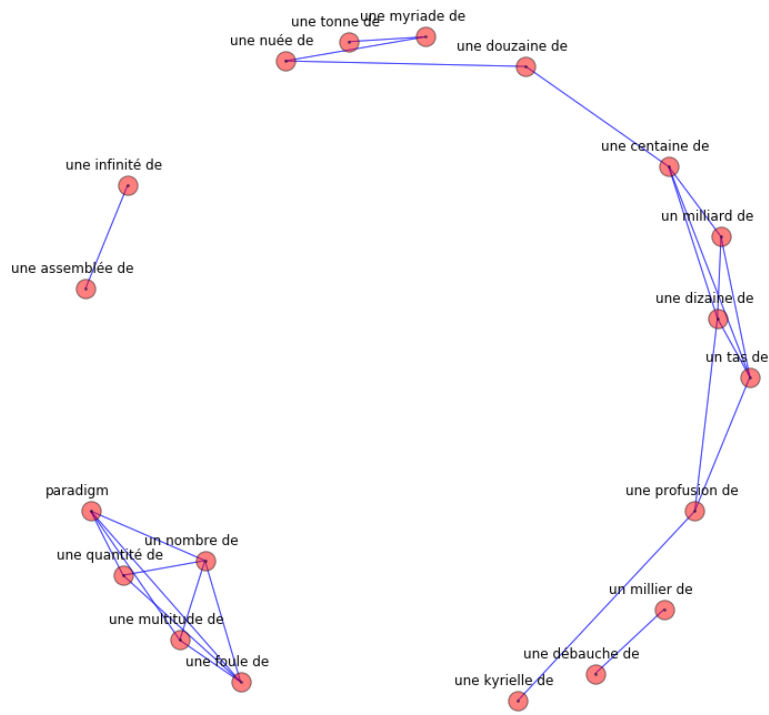
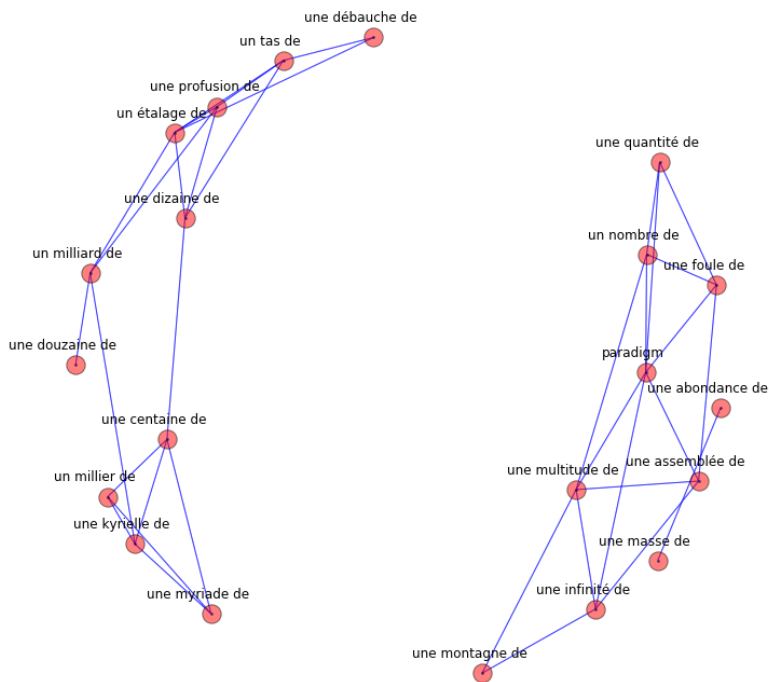


Figure 5.9: Correlation network between members of the $[\text{un } \{N\} \text{ de}]$ quantifier paradigm. The correlation is computed over two different time windows: (a) 1721-1820; (b) 1751-1850.



(a)



(b)

Figure 5.10: Correlation network between members of the [un {N} de] quantifier paradigm. (a) 1781-1880; (b) 1811-1910.

With the period 1871-1970, we witness how the newly risen cluster has become the foundation of the new paradigm (Fig. 5.11). It is now part of a the cluster of the ‘quantitative, measurable’ quantifiers, alongside with *une centaine de*, *une dizaine de*, *une tonne de*, etc. Interestingly, *un nombre de* has managed to change side and belongs now to the prototypical cluster. The cluster of ‘fuzzy’ quantifiers, which contribution to the arising cluster in Phase II may have been crucial to evict the former prototypical cluster, has now become independent. The member of the former dominating cluster, such as *une foule de*, *un morceau de*, are still forming a single component. We find a few other isolated small components as well. During the last period we investigated, this new polar opposition of the paradigm between the precise, quantitative quantifiers and the fuzzy, plentiful quantifiers, is more firmly marked. Interestingly, the former members of the protoypical paradigm (*une foule de*, *une montagne de*) are now part of the ‘fuzzy’ cluster. They have found a new place in the reorganizing configuration of the whole paradigm. However, the paradigm remains dominated by a still more densely connected cluster of quantitative quantifiers, as shows the correlation between the member of this cluster and the point representing the whole paradigm. The marginal attachment of *un foisonnement de* to this cluster seems a spurious effect, and may be due to the ambivalent status of *un nombre de*, which has proved adaptable to the new quantitative rule, but certainly retains some idiosyncrasies of the former dominant cluster of phase I.

Adverbial paradigm

We will now investigate a second example, given by the paradigm of the adverbial [par {N}] construction. This construction allows the means, the motive, the manner or the cause of an action, which is somehow ‘colored’ by the noun of the construction. English language also acknowledges some similar construction (e.g. ‘by chance’, ‘by surprise’), but there is not a [by {N}] construction as versatile and productive as the one we find in French. The exact semantics of the construction depends on its Noun argument, so we would expect different translations, depending on the noun, in other languages (e.g. ‘par surprise’ is translated to English with ‘by surprise’, ‘par amour’ with ‘out of love’, ‘par inadvertance’, with ‘inadvertently’, ‘par moments’ with ‘at times’, ‘par principe’ with ‘on principle’, etc.). This adverbial construction is widely attested in French as soon as the XIIth century, where we find occurrences such as ‘par amor’, ‘par proeiere’, ‘par franchise’, ‘par feid’, etc. The construction is likely to have been inherited from Latin, and is found as well in other Roman languages (e.g. Italian [per {N}], Spanish and Portuguese [por {N}]).

We provide here the list of investigated members of this paradigm. The associated Frantext queries were kept as simple as possible, all following the pattern *par* NOUN. We did not exclude a following *de*, and we did not include the possibility for the noun to be preceded by an article (e.g. *par une nécessité aveugle*).

- *exemple* (32826);
- *conséquent* (13062);
- *hasard* (6975);
- *suite* (4928);
- *moments* (2807);
- *excellence* (1850);
- *amour* (1726);
- *instants* (1365);
- *degrés* (1261);
- *malheur* (1253);
- *nature* (1243);
- *bonheur* (1045);

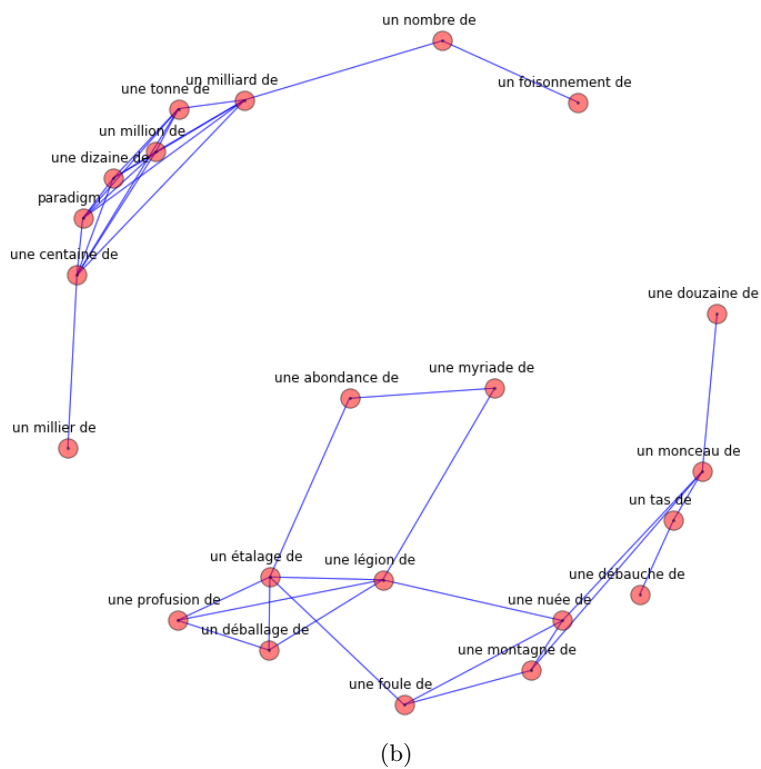
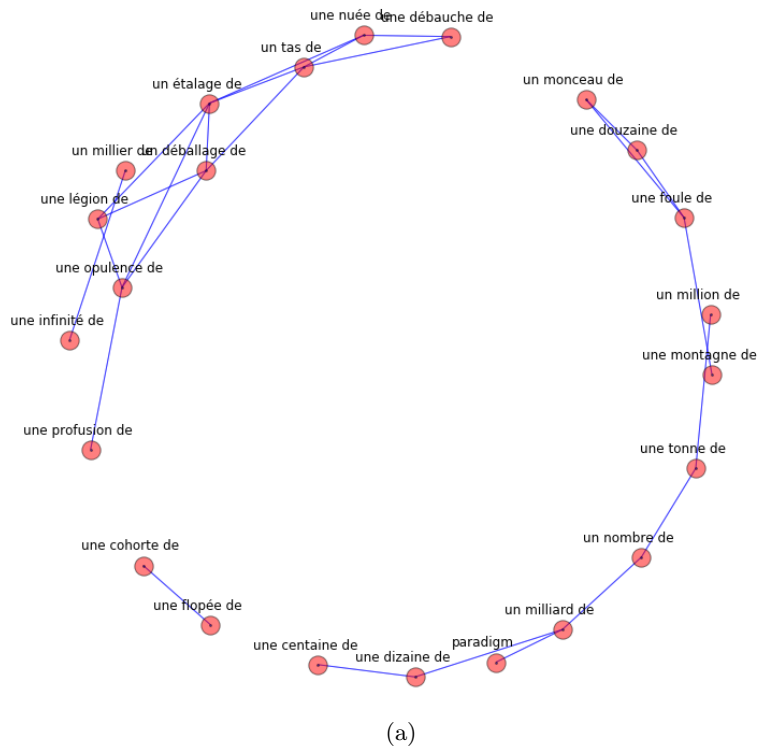


Figure 5.11: Correlation network between members of the $[\text{un } \{N\} \text{ de}]$ quantifier paradigm. (a) 1871-1970; (b) 1901-2000.

- crainte (1044);
- pitié (952);
- intervalle (884);
- habitude (768);
- nécessité (681);
- accident (642);
- surprise (604);
- surcroît (586);
- mégarde (579);
- peur (567);
- goût (559);
- prudence (516);
- chance (474);
- principe (468);
- orgueil (449);
- erreur (443);
- essence (413);
- amitié (402);
- intérêt (395);
- ignorance (383);
- vanité (329);
- dépit (314);
- manque (276);
- égard (272);
- besoin (269);
- faiblesse (260);
- pudeur (256);
- fortune (247);
- précaution (246);
- effet (242);
- anticipation (239);
- inadvertance (216);
- à-coups (213);
- envie (212);
- désespoir (210);
- souci (210);
- jalousie (207);
- bonté (204);
- désir (204);
- faute (196);
- honneur (193);
- hypothèse (191);
- magie (190);
- occasion (190);
- négligence (182);
- bonds (177);
- artifice (171);
- imitation (167);
- choix (161);
- sympathie (160);
- dérision (156);
- mépris (153);
- vengeance (144);
- fierté (129);
- fidélité (122);
- plaisanterie (108);
- défi (100);
- volonté (100);
- effraction (93);
- méchanceté (93);
- trahison (88);
- imprudence (83);
- coup (83);
- égoïsme (79);
- piété (79);
- honnêteté (72);
- bêtise (72);
- zèle (62);
- malchance (59);
- sauts (55);
- colère (53);
- folie (50);
- élimination (48);
- paliers (46);
- étourderie (46);
- gourmandise (37);
- dessein (37);
- cause (37).

Before turning to the diachronic networks, we can first have a look at the frequency distribution of this paradigm (Fig. 5.12). This distribution tells us three things: first, that the paradigm members obey locally a Zipf's law, with slope -1 (the intercept has been chosen so as to fit the curve best); second, that the two most frequent members of the paradigm significantly deviate from this law; third, that the least frequent members disobey the law, and are actually much less frequent than what would be predicted by Zipf's law.

According to us, the deviating behavior of the two most frequent members of the paradigm is illustrative of a speciation event. That is, they are no longer part of the adverbial [par {N}] construction, but are also independent constructions of their own. This is not only because, being more frequent, they should benefit from a special storage, but also because they have semantic and syntactic idiosyncrasies. Thus, *par conséquent* is no longer an adverbial, but a discourse marker, alongside with *du coup*, and could be translated as well by 'as a result'. It is most often used in extra-clausal positions and, contrary to other members of the paradigm, it does not make room for variation and the query *par &q conséquent* only send spurious occurrences, or a spelling alternative with an hyphen (which counts as a separated word in Frantext), indicative of the morphologic fixation of the form. On the other hand, *par exemple* is more semantically idiosyncratic, accounts for a small variation (e.g. the very marginal

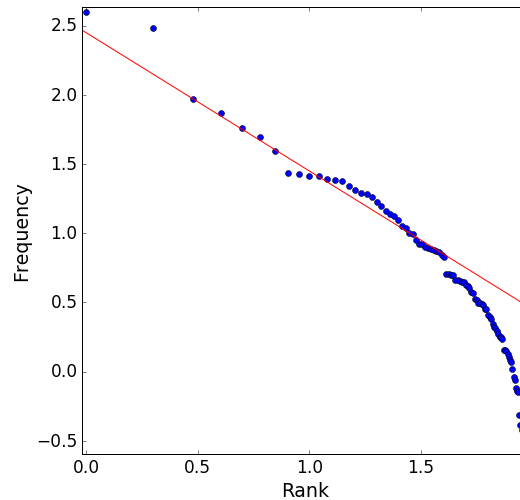


Figure 5.12: Log-log plot of the frequency distribution of each member of the adverbial [par {N}] paradigm. The full red line is a Zipf's law fit of the distribution, reduced to datapoints 3 to 39, with slope -1 and the intercept chosen to fit the first datapoint of the linear-looking part, associated with *par hasard*. $r^2 = 0.97$.

par cet exemple), has clear etymological roots within the construction (e.g. ‘et cecy je le preuve par exemple’, 1352), but still remains largely extra-clausal in its syntactic use. As it happens, it deviates less from the Zipf's law, but this may be a mere coincidence.

This mark of speciation must anyway be borne in mind when analyzing the diachronic networks. As we shall see, it might be possible to observe, in some extent, this speciation event. To visualize it more clearly, and because these two forms are extremely frequent, they have both been substracted from the frequency profile of the whole paradigm. Otherwise, they would have been correlated with the paradigm at all times without this correlation ever being meaningful.

Regarding the other, even more remarkable deviation from the Zipf's law, we would also like to venture an explanation. There might be a distinction between entrenched, frequent constructs of a construction, which would define its paradigm at a given moment in time, and occasional constructs, uttered by analogy with the existing paradigm members. This observation is akin to the claim of Hoffmann (2004), who suggested that some low-frequencies complex prepositions could have grammaticalized by analogy with more frequent prepositions obeying the same pattern. Here, we would have a way to distinguish between entrenched members, and the members who result from the analogical productivity of the construction. This dichotomy is however to be taken with serious caution; for instance, members of the ‘low-frequency’ part, such as *par effraction* or *par magie*, are quite idiosyncratic and their use as construction members may be stored apart from the use of the individual nouns entering the construction.

However, we can consider that the frequency distribution is not a power law, but an exponential one (Fig. 5.13), as we have already suggested earlier on. We would then get, as well, speciated members of the paradigm (the seven first), and a distinction between entrenched members of the paradigm, and those due to occasional aproductive analogy

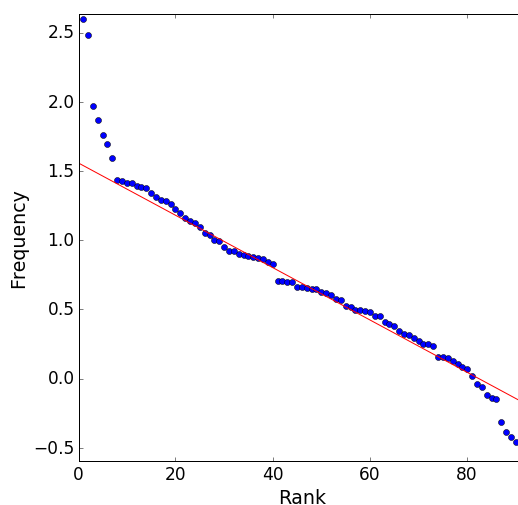


Figure 5.13: Semilog plot of the frequency distribution of each member of the adverbial [par {N}] paradigm. The full red line is an exponential fit of the distribution, obtained by leaving aside the seven first and fifteen last points. $r^2 = 0.99$.

(the ten last). This would still make sense, since *par suite* is a connective, *par hasard* has a strong, idiosyncratic, pragmatic meaning (e.g. in *si par hasard*), *par excellence* has adjectival uses (as its English calque) and *par moments* is a temporal marker. Only *par amour* would belong clearly to the adverbial paradigm, so it makes sense to consider this deviation from the exponential law as a sign of speciation. Furthermore, the infrequent, occasional members are indeed much less entrenched than many other members of the paradigm.

Yet, we could also fit the frequency f as a function of the rank r by the following proposal:

$$f = A \left(\frac{r - a}{r - b} \right)^\beta, \quad (5.3)$$

where A , a , b , β would be parameters to be fitted. This would much likely allow us to capture the whole curve, so that ‘speciated members’ and ‘non-entrenched members’ would be part of the flock as much as the other members. Therefore, all these analyses are, once more, to be taken with caution.

We can now consider the diachronic evolution of the whole paradigm through the correlations of its members. As the paradigm is wide and extensive, we set very high criteria to consider a correlation as meaningful (the Pearson coefficients must be greater than 0.9 and 0.7 for the frequencies and their derivatives, respectively). Despite of this, the networks are sometimes fuzzy, and extremely complex to read and to interpret.

As [par {N}] is an old construction, and covers our whole corpus, we could start the investigation as soon as 1321-1330. However, individual members of the paradigm are not as frequent as more general constructions, so that the data associated with the earliest centuries covered by our corpus is not always reliable, and extremely sensitive to the statistics of individual texts. Yet we find, for correlations computed over the period 1321-1420, a fully connected cluster which members are *par accident*, *par*

conséquent and *par nature*, all correlated with the paradigm as a whole. It shows that, despite the sparseness of the data, the method is still able to extract a central cluster of paradigm members. Additionally, it also appears that, at this time, *par conséquent* was fully connected to the adverbial paradigm. This cluster can furthermore be understood by the fact that all the concepts of this central cluster (*nature* et *accident*) were key philosophical concepts during this time, especially within the High Scholastics. Therefore, we find philosophical treatises (especially *Le Livre d'Aristote*) discussing these two concepts. As the Scholastics were versed in Latin, it likely that they would have calqued, in their French prose, the Latin *per consequentiam*. Hence, this cluster would be the reflect of an academic literature of this time. It may therefore be that the [par {N}] adverbial construction were revived, in Middle French, through the influence of scholars, which were reading and writing Latin as well as French.

The two next centuries (1421-1520 and 1521-1620) do not lead to interesting network structures. The fifteenth century, in particular, sees a severe decrease of the frequency of the whole paradigm. It might be that the list of forms we have investigated lack some of the key forms of this period. More likely, the explanation could be related to the fact that the academic treatises which most supported the use of the paradigm are not to be found anymore. A dominant cluster is still present, with the paradigm at its core, and related to *par désir*, *par négligence* and *par exemple*. This little fact is quite remarkable, since these three paradigm members all illustrate a different function: *par exemple* is mainly extra-clausal, *par négligence* is purely adverbial, and *par désir*, followed by *de*, can serve as a preposition, introducing an infinitive proposition (e.g. 'par desir de toutes choses cognoistre' 1404). Also, *par orgueil* and *par folie* correlate together, which may indicate that the adverbial *par* construction is no longer academically-oriented, but comes also to be more widely used in lay contexts. During the XVIth century, the network is reduced to a few independent components of minimal size (pair correlations). It is however worth noting that, in the XVIth century, one of this component shows that *par conséquent* is still correlated with the paradigm as a whole.

The network drew for the XVIIth century shows a large, densely connected cluster (Fig. 5.14a). The paradigm is, at this time, prototypically organized, with members such as *par nature*, *par désir*, *par inadvertance*, *par fortune* in the center, and more marginal members (*par dessein*, *par honneur*). What relate all these cluster members is unclear though. One might say that they are all fair instances of adverbial use, but that would not be even true: *par occasion*, for instance, is not much adverbial, and has been replaced, in Present Day French, by *à l'occasion*. At this time *par exemple* loses its adverbial uses, so it makes sense that it would be separated from the rest of the paradigm, and as it happens, *par honnêteté*, although extremely marginal, is chiefly extra-clausal as well. What could perhaps be the case is that, at this century, the paradigm is tightly organized, and changes happen especially on the margins of the paradigm.

The situation in the next century is no clearer (Fig. 5.14b). We still find a giant component, to which the paradigm as a whole is belonging, but it is now much loosely and sparsely connected. Aside from it, we find five little clusters of a few nodes each, one being more internally connected than the others (*par besoin*, *par habitude*, *par égard*, *par goût*, *par faiblesse*). Interestingly, *par conséquent* and *par exemple*, are

correlated as a pair. As we already argued that they were the two most speciated of the paradigm, this makes perfectly sense. The cluster (*par essence*, *par bonds*, *par excellence*) is mildly puzzling, as a correlation between *par essence* and *par excellence* could have been expected, since they have very close meanings. Yet, why would they correlate only through *par bonds* is a complete mystery.

In the XIXth, a new densely connected large cluster emerges (Fig. 5.15a). Its contents are not the same as the cluster found in the seventeenth century. It would appear thus that the eighteenth century was a period of transition, leading from one large dominant cluster to another. The most central members of this new cluster are *par mégarde*, *par hasard*, *par malheur*, *par amour*. The pair *par moments* and *par instants* is marginally connected with this giant cluster. Interestingly, the paradigm as a whole is not part of any cluster.

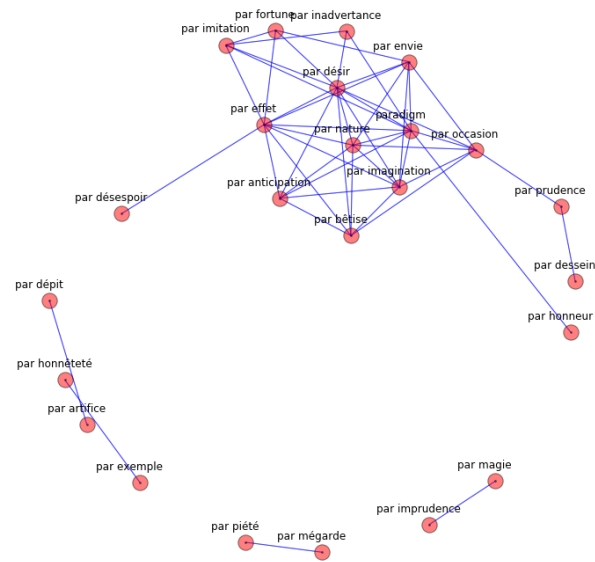
The last century under study displays, in terms of its network's shape, a picture closer to the one of the eighteenth century (Fig. 5.15b). There is a big cluster, but loosely connected, and several little ones. It has lost several of its members, notably the pair *par moments* and *par instants*, as well as *par hasard*, which may have speciated. This loosening may be an indication that the paradigm is expected to shift again. Another important cluster is visible as well, with *par inadvertance* playing a central role, and including the pair *par chance* and *par malchance*, which are both recent members of the paradigm.

A short summary is in order. As such, these networks lead to new puzzles of language change, rather than they help to understand the existing empirical facts. Admittedly, *par* is one of the most complex and versatile prepositions in French. It has many different uses, enters many different constructions, and conveys a wide array of different meanings. Worse, all these functions are interfering with one another. For instance, this [par { N}] paradigm is certainly analogically related with the [par {Prep}] paradigm, encompassing forms such as *par ailleurs*, *par dessus*, *par en bas*, etc. It is therefore too broad a topic, with too many interferences, to be investigated through diachronic networks, especially on a period of time as long as seven centuries.

To make the most of those networks, it would therefore be preferable, as we did with the previous [un {N} de] paradigm, to focus on a limited time period, so as to understand in detail a specific part of the evolution. Also, it would be more advised to select a limited number of linguistic forms, once they have been identified as belonging to the most relevant clusters of the network. Yet, large-scale networks can perhaps provide a new kind of data. For instance, it could be interesting to rank the forms as a function of their connectivity; either by counting the edges, as we hinted at previously, or by summing the weights. It would reveal which paradigm members are the most central at any given moment in time. A step by step evolution of this hierarchy might help to elucidate certain puzzling features of the change.

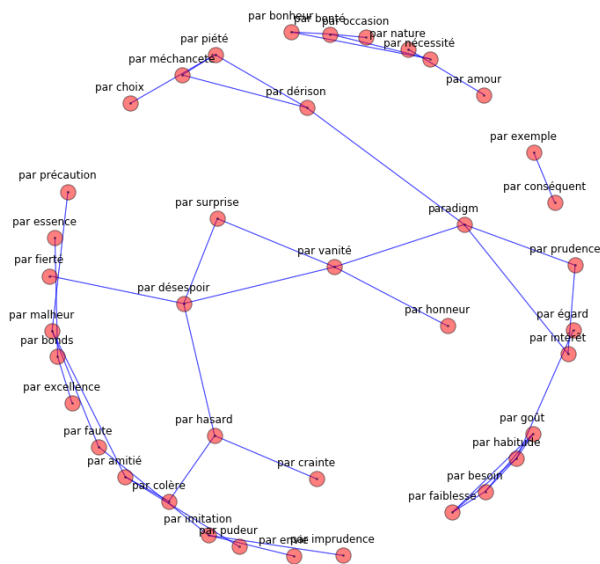
I must stress, as a word of conclusion on this matter, that these networks are not a result, but a tool. As such, they are more puzzling and head-scratching than anything else, just as would be any toolbox. Yet, appropriately used, and handled with enough linguistic expertise, I have good faith that they can be helpful and revealing. They allow to extract a new family of empirical facts on language change from diachronic corpus data which, combined with the methods already existing, might lead to interesting insights on specific instances of semantic change.

1621-1720



(a)

1721-1820



(b)

Figure 5.14: Correlation network between members of the adverbial [par {N}] construction. (a) 1621-1720; (b) 1721-1820.

5.2.2 Identifying competitors

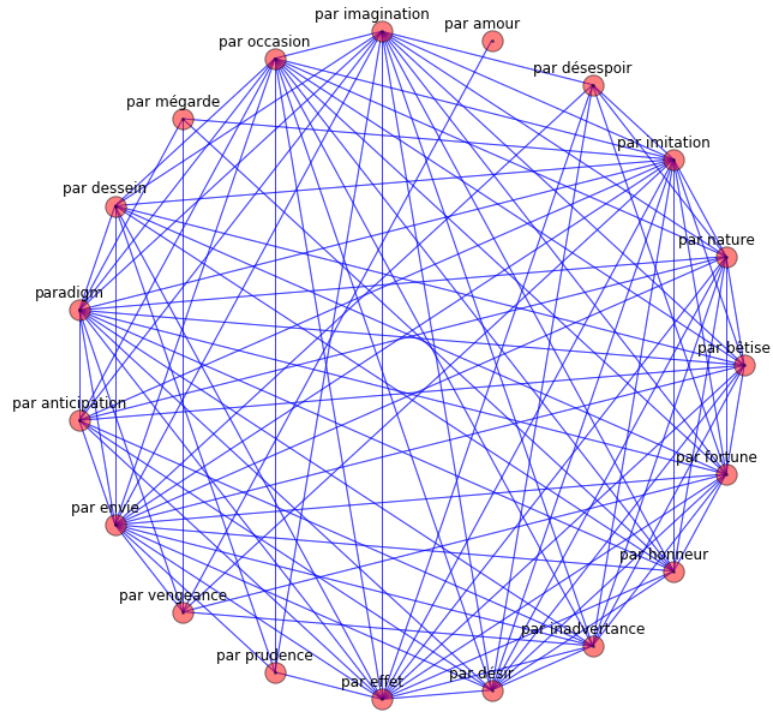
We have already seen, for the [un {N} de] paradigm, that the competition might not be one to one, so that identifying the competitor of a single linguistic form was less meaningful than identifying the competitor of a whole cluster of forms. We will now provide a second example, related to the [par {N}] paradigm. As we will investigate a more specific case, we set back the threshold values at 0.8 and 0.4, for the frequencies and the derivatives, respectively.

As we saw, one of the most remarkable features revealed by our diachronic network analysis is that the paradigm is organized around a densely connected cluster in the XVIIth century (1621-1720), as well as in the XIXth century (1801-1900). In between, the paradigm loses its tightness and the cluster is thinned down. For greater clarity, we kept only the members of these two clusters found in the previous analysis (now that the thresholds have been lowered, these clusters would probably encompass a greater number of members; by keeping only the members which were already belonging to the cluster, the latter appears all the more densely correlated). It must also be noted that most members of the two giant clusters are ‘pure’ adverbials, that is, they are members of the construction used as adverbs. Yet this is less true for the second cluster where we find members such as *par hasard*, *par malheur* and *par moments*, hinting at an overall slight reorientation of the construction.

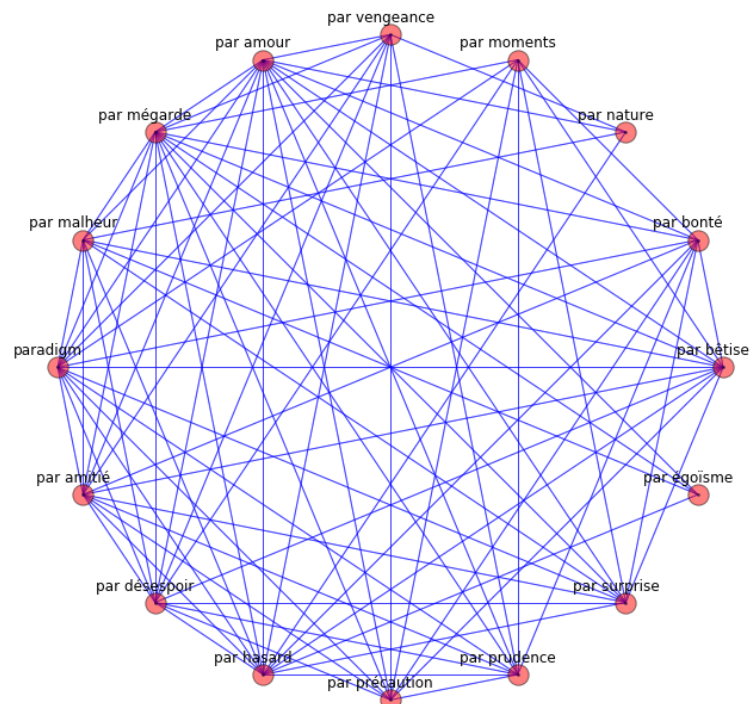
These two clusters have respectively 16 and 15 members (Fig. 5.16), and only five members in common (*mégarde*, *bêtise*, *amour*, *vengeance*, *nature*). It means that the two thirds of the contents of the giant component have been replaced. Furthermore, the giant component breaks apart as soon as the period 1651-1750, so that something has definitely happened. If we compare this process with the frequency evolution of the paradigm (Fig. 5.17), it appears that the first cluster is associated with a frequency breakdown of the paradigm as a whole (all forms are co-evolving since they all suffer from a frequency drop), while the second cluster corresponds to an overall S-shaped frequency rise.

Therefore, contrary to the former case where the quantifier paradigm was subject to an internal reconfiguration, it would seem here that the paradigm is competing against an external variant, so that the [par {N}] paradigm loses its main focus, and gains a new one two centuries later. Indeed, there is very few competition within the paradigm from 1621-1720 to 1741-1840. Only from 1761-1860 we witness an internal competition against *par inadvertance*: all paradigm members are seemingly leagued against this particular form (Fig. 5.18). However, can we really consider that the whole paradigm is competing against one of its own members? This would be an absurd instance of bootstrapping: the paradigm is pumping frequency for itself out of one of its own members. The two best explanations for this fact might be that one member of the paradigm is sacrificed, so as to strengthen the other members; however, given the low frequency of *par inadvertance*, this would seem a bizarre choice. It is more likely that *par inadvertance* has simply be left aside from the overall reappraisal of the paradigm, for unknown reasons.

Let us linger slightly longer on this matter. From a pattern extraction of *par mégarde*, a very nice sigmoidal frequency increase is found for the period 1821-1890. The question we could ask facing such an empirical observation would be: if the



(a)



(b)

Figure 5.16: Giant component of the correlation network of the adverbial [par {N}] construction. (a) 1621-1720; (b) 1801-1900.

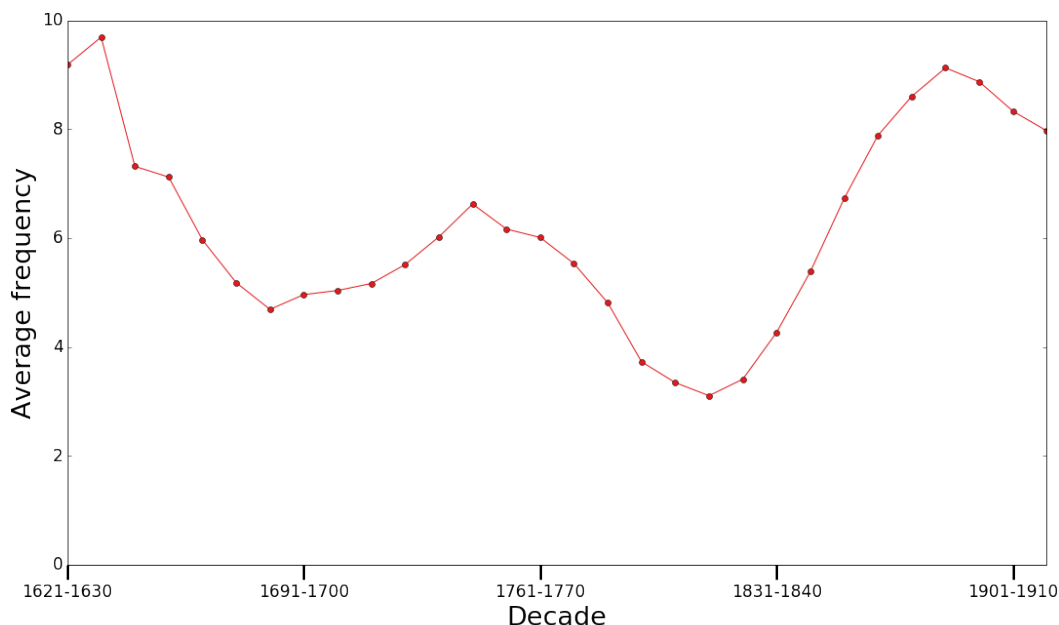


Figure 5.17: Average frequency of the whole [par {N}] paradigm over the period 1621-1920.

frequency rise corresponds to a winning competition, then against which form is *par mégarde* competing. It might be tempting to think of *par inadvertance*. Indeed, they both have, more or less, the same meaning; yet *inadvertance* is a complex, opaque word, especially as *advertir* has lost its *d* and shifted meaning. It meant originally ‘to shift oneself attention towards’, and was a near synonym of ‘to focus on’. Then it was expanded to a more transitive use, which is the only one retained today, and could be translated by ‘to warn’. Yet, *inadvertance* refers to the original, lost meaning. It is thus both phonetically opaque, and semantically opaque. On the other hand, *mégarde* is not much clearer, and exists only as a member of the [par { N}] construction in Present Day French. Still, a competition could have happened. This assumption would be confirmed by a look at the rescaled frequencies (Fig. 5.19). The replacement seems then very clear. Without any contextual cues, it would be legitimate to conclude that we have successfully identified the true competitor of *par mégarde*.

Yet we know better now. We have seen that both forms are part of the paradigm. For two centuries, they have co-evolved together, their frequency rose and fell in unison, and they even appeared around the same time (Fig. 5.20). But the crucial fact that rule out this scenario of a one-to-one competition is the anti-correlation network of Fig. 5.18. As I said, all the paradigm members cannot have risen together by competing with a single member of their kind, as it would be an instance of bootstrapping. Therefore, the rise of *par mégarde* is only to be understood as the reappraisal of the paradigm at this time; the sigmoid increase does not reflect the evolution of a single, individual form, but is part of wider process. In this regard, the anti-correlation with *par inadvertance* cannot be the sign of a semantic replacement.

We shall conclude with this disappointing observation that our diachronic network

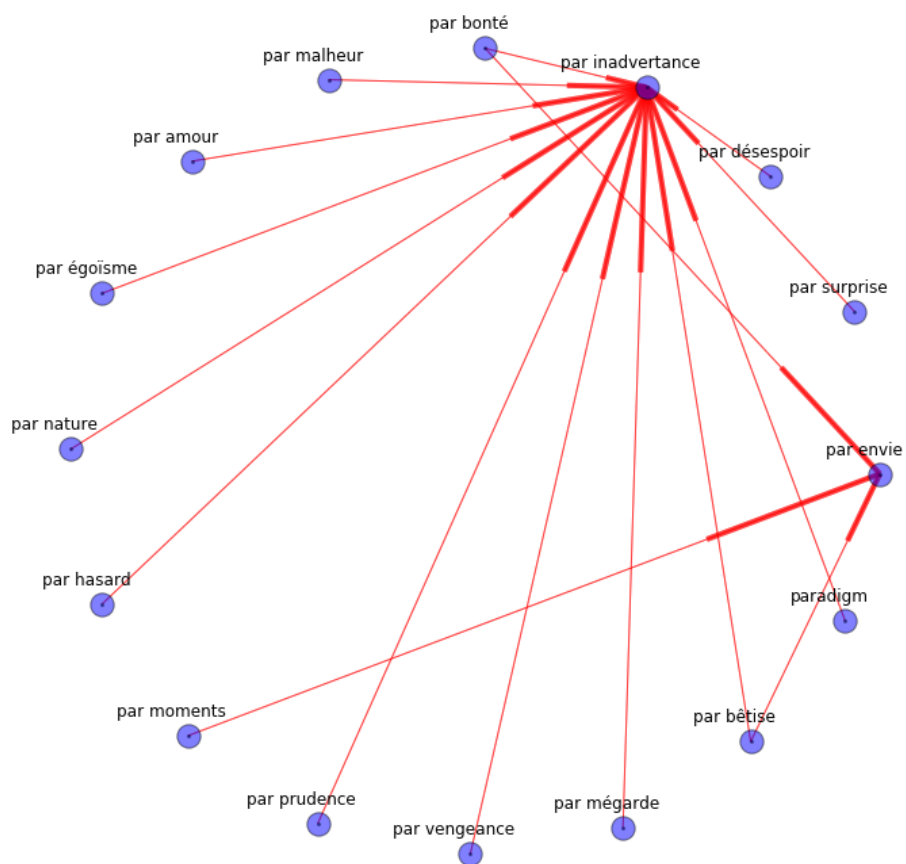


Figure 5.18: Anti-correlation network of the [par {un}] paradigm over the period 1801-1910.

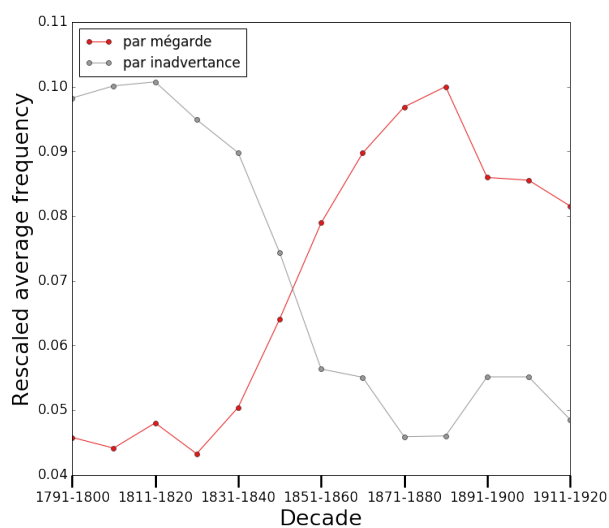


Figure 5.19: Rescaled average frequency profiles of *par mégarde* and *par inadvertance* over the period 1621-1920.

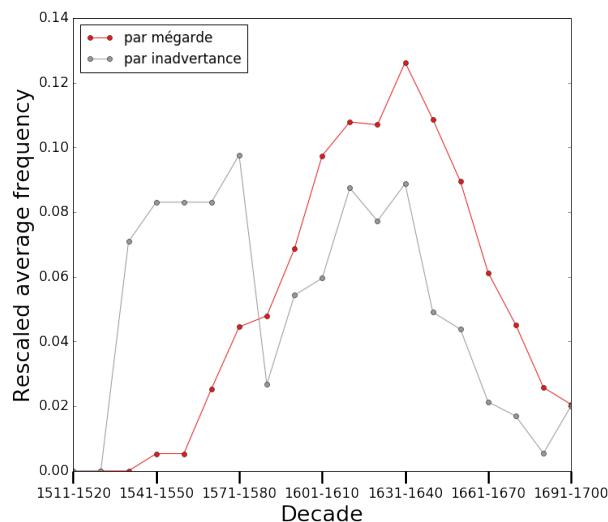


Figure 5.20: Rescaled average frequency profiles of *par mégarde* and *par inadvertance* over the period 1511-1700.

of anti-correlations do not necessarily help to identify the true competitor of a frequency rise. However, we could include all cluster members as one single linguistic item, and compare it with other linguistic items, not belonging to the same construction. To do so, these ‘other linguistic items’ must first be identified by a prior linguistic analysis. Once more, it appears that these networks can help to refine hypotheses and even allow to test some, but they critically need the input of a true linguistic expertise.

5.3 The Constant Rate Hypothesis revisited

The Constant Rate Hypothesis, proposed by Kroch (1989a,b), states that a change will follow, in different contexts, different S-curves, all characterized by the same rate of change (i.e. the same logit slope). We already saw in chapter 3 that this hypothesis was a tricky one, since the logit slope is not a reliable parameter. It depends heavily on the bounds of the change, and can vary hugely according to how these bounds are actually chosen and treated. Furthermore, an alternative hypothesis was proposed by Bailey (1973), and supported with actual linguistic data by Shen (1997), according to which the later the context is affected by the change, the higher the rate. In other terms, contexts affected later on by the change ‘catch up’ the process. It was proposed by Wang et al. (2004) that this acceleration was in fact due to an analogical pressure from already converted contexts to the newly affected one.

As we have developed an efficient method to retrieve S-curves, and especially to select optimally the boundaries of the change, we are now in good position to test the Constant Rate Hypothesis and its rival, the rate acceleration hypothesis, as we shall coin it. I propose to test these hypotheses on two selected examples, the first one being associated with the grammaticalization of the [en plein {N, Prep}] construction, the second to the development of idiomatic, non-compositional uses of the [dans {definite article} {N}] construction.

Form	Slope	t_0	Rise	Total frequency
<i>en plein</i> (total)	0.96 ± 0.05	1851-1860	1801-1910	43.75
<i>en plein air</i>	1.70 ± 0.13	1851-1860	1831-1890	2.73
<i>en plein jour</i>	1.27 ± 0.11	1851-1860	1821-1890	2.13
<i>en plein lumière</i>	1.03 ± 0.04	1861-1870	1821-1930	1.94
<i>en pleine nuit</i>	0.46 ± 0.01	1891-1900	1841-1950	1.45
<i>en plein soleil</i>	1.05 ± 0.05	1851-1860	1821-1900	1.02
<i>en plein vent</i>	1.05 ± 0.10	1851-1860	1831-1890	0.71
<i>en pleine campagne</i>	2.17 ± 0.38	1871-1880	1821-1860	0.55
<i>en pleine rue</i>	1.56 ± 0.11	1861-1870	1841-1910	0.52
<i>en pleine poitrine</i>	1.03 ± 0.10	1861-1870	1841-1900	0.47
<i>en plein dans</i>	0.88 ± 0.05	1851-1860	1821-1890	0.47
<i>en plein hiver</i>	1.01 ± 0.02	1851-1860	1821-1900	0.43
<i>en plein sur</i>	0.65 ± 0.09	1861-1870	1821-1880	0.19

Table 5.5: Quantitative features of the S-curves of the [en plein {N, Prep}] paradigm. This table provides, from left to right: the slope of the linear fit of the logit transform of the S-curve, the time t_0 during which the S-curve reaches its maximal intensity, as computed from the intercept of the linear fit of the logit transform, the total period during the frequency rises according to an S-curve, and the total frequency of the form during this period.

5.3.1 En plein

The construction [en plein {N, Prep}] appears as soon as the XIVth century with the rough meaning ‘in the middle of’, ‘in the midst of’, ‘in the thick of’, complemented by ‘right’ in some occasions (e.g. *en plein milieu*, ‘right in the middle’). The first occurrences of the construction already conform to its ensuing meaning: *en pleine compagnie* (1372), *en plein été* (1400), *en pleine meslée* (1508). After a first frequency rise in the sixteenth century, the frequency of the form stays remarkably latent, until the XIXth century which sees a nice S-shaped increase of frequency (Fig. 5.21).

We can search for its most frequent collocates. These can be interpreted as different contexts into which the construction is expanding. Therefore, we can consider the frequency profiles of all these constructs, extract the S-curve pattern from them, and compare the slopes. In doing so, collocates with less than 50 co-occurrences over the whole nineteenth century were not included in the study. Also, the criterion $r^2 > 0.98$ has been occasionally adapted so as to focus on the period of interest: for *en pleine nuit*, *en pleine poitrine*, *en plein sur*, *en pleine campagne*, for which it was set to 0.99, 0.979, 0.96 and 0.94, respectively. Two constructs, *en pleine mer* and *en plein midi*, did not lead to an S-curve. It may be explained the fact that they were among the leading members of the collocates paradigm in the earlier centuries, so the construction did not expand further with respect to these two. The results of this procedure are given in Table 5.5.

It appears that the Constant Rate Hypothesis is mildly supported by this survey. Two of the most deviating members, *en plein dans* and *en plein sur* could be discarded

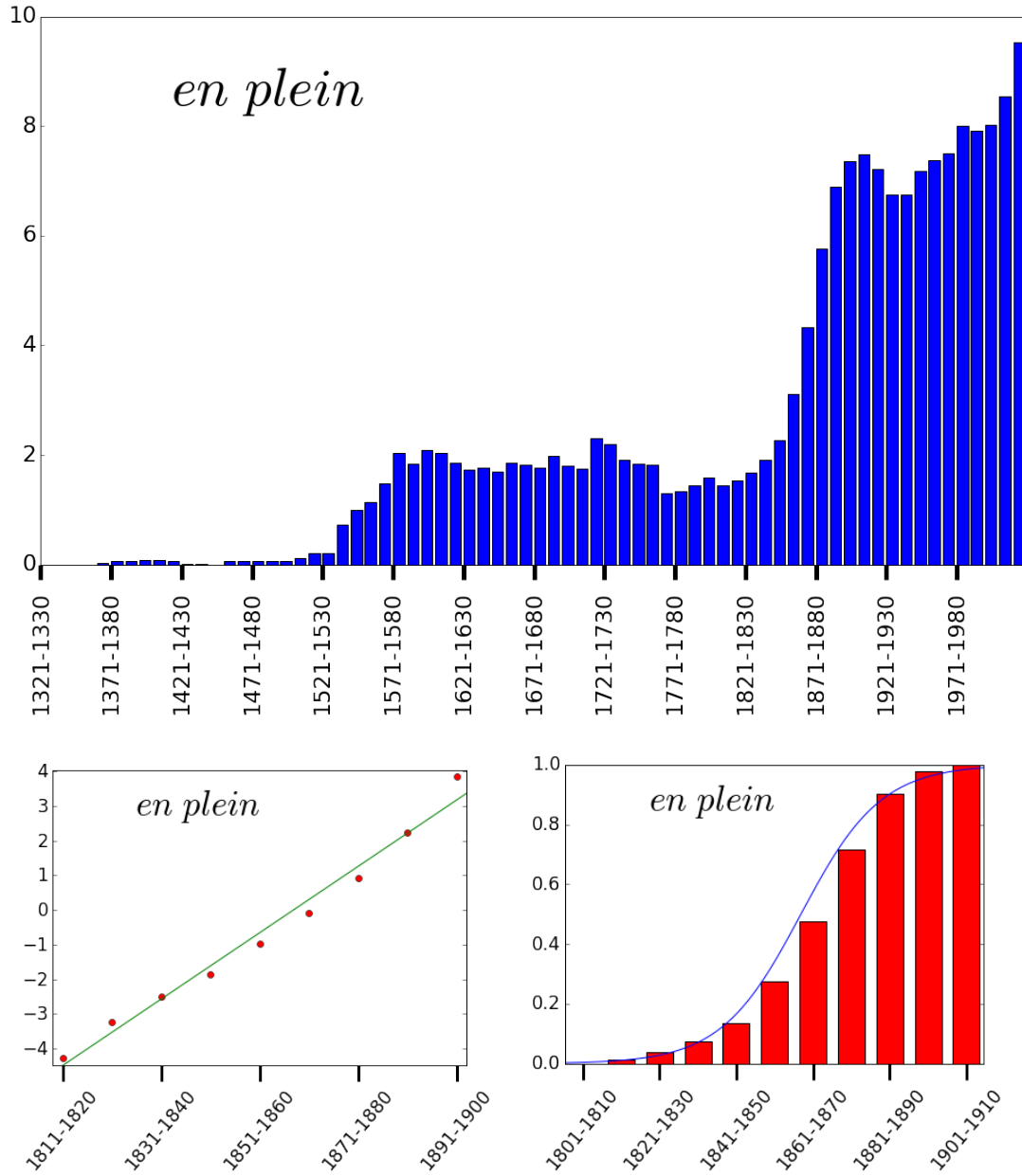


Figure 5.21: Top: Frequency profile of *en plein*. Bottom left: Logit transform of the extracted growth part of the curve. Bottom right: Sigmoidal fit of the extracted growth part.

on the ground that they are not the same construction. That is, [en plein {N}] is one construction, and [en plein {Prep}] is another (and it may also be that there would be as many different of these constructions as there are different prepositions). If the Constant Rate Hypothesis were assumed to be true, it would be a nice way to distinguish the two otherwise very similar constructions. As it stands now, it would be the other way around: we posit that the two constructions are different, and so we save the Constant Rate Hypothesis.

There are other deviations to the Constant Rate Hypothesis though. The form *en pleine campagne* could be excluded, because it was required to tweak the criterion in order to obtain it (the tweak for *en pleine poitrine* was so mild that it admittedly does not count as such. And yet many other forms deviate.

Actually, there might be a way to convincingly save Kroch's hypothesis. Here the S-curve of the rise is supposed to be associated with a semantic expansion. However, the construction was already covering a wide paradigm of collocates. When the form expands, the frequency of these collocates is expected to rise, by analogical strengthening from the frequency rise of the construction, but only the new collocate members are real contexts of the expansion, and therefore only them should be considered as reflecting the change. We can thus consider the paradigm of the construction during the eighteenth century, and keep only the members newly arising with the S-curve change of the nineteenth.

Here are the twenty most frequent collocates paradigm members of the construction in the XVIIth century:

- | | | |
|-------------|--------------|-----------------|
| • jour | • campagne | • théâtre |
| • air | • sénat | • terre |
| • mer | • santé | • rue |
| • liberté | • conseil | • parlement |
| • midi | • possession | • convalescence |
| • paix | • champ | • vent |
| • assemblée | • eau | |

Keeping only the members not belonging to this list shows a clear confirmation of the Constant Rate Hypothesis (Table 5.6), with the blatant exception of *en pleine nuit*, which seems anyway to have developed later (for unclear reasons, since it is attested since 1769 at least). I would also like to point out that not only these contexts share a same rate of change, they also pretty much share the same time of highest intensity of change. In the Constant Rate Hypothesis, it was assumed that the intercepts should distinguish one change from another. This is not the case here.

As for the accelerating rate hypothesis, given how close the intensity peak are from one change to another (apart from two they all fall into two adjacent decades), it seems ill-advised to try to test it. The only form which has developed much later, *en pleine nuit*, is anyway associated with a much lower rate, contrary to the hypothesis.

Before concluding on this matter of the Constant Rate Hypothesis, let us consider another example.

Form	Slope	t_0	Rise	Total frequency
<i>en plein</i> (total)	0.96 ± 0.05	1851-1860	1801-1910	43.75
<i>en plein lumière</i>	1.03 ± 0.04	1861-1870	1821-1930	1.94
<i>en pleine nuit</i>	0.46 ± 0.01	1891-1900	1841-1950	1.45
<i>en plein soleil</i>	1.05 ± 0.05	1851-1860	1821-1900	1.02
<i>en pleine poitrine</i>	1.03 ± 0.10	1861-1870	1841-1900	0.47
<i>en plein hiver</i>	1.01 ± 0.02	1851-1860	1821-1900	0.43

Table 5.6: Quantitative features of the S-curves of the [en plein {N, Prep}] paradigm. This table provides, from left to right: the slope of the linear fit of the logit transform of the S-curve, the time t_0 during which the S-curve reaches its maximal intensity, as computed from the intercept of the linear fit of the logit transform, the total period during the frequency rises according to an S-curve, and the total frequency of the form during this period.

Table 5.7: Next page: Quantitative characterization of the S-curve frequency rise of various forms from the *dans* paradigm. We provided the total number of occurrences of the form in the whole corpus, the slope of linear fit of the logit transform of the S-curve, the time at which the S-curve reaches its peak of intensity, as computed from the intercept of the linear fit of the logit transform, and finally the time period covered by the S-curve according to our extracting pattern procedure. A form may appear several times in the table if several non-overlapping S-curves have been extracted from the frequency profile.

5.3.2 Idiosyncratic uses of *dans*

In French, the preposition *dans*, appeared during the XIVth century, gradually replaced the former *en* as the generic locative preposition ('in/into'). We can therefore consider individual contexts of the lexical diffusion which resulted from this replacement.

Here, we have selected but a very few of them, either atypical (e.g. *dans la douleur*, *dans l'immédiat*), semantically idiosyncratic (*dans l'histoire*, *dans le coup*, *dans l'attente*), functional (*dans la mesure*, *dans l'intention*, *dans le but*), idiomatic, i.e., recurrent as a unit (*dans l'avenir*, *dans l'esprit*, *dans l'ambiance*), or even associated with pragmatic uses (*dans l'absolu*, *dans l'ensemble*, *dans le pire des cas*), and searched for possible S-curves (Table 5.7).

The typical Frantext query associated with these forms was:

`dans {article} &q(0,1) form`

where the paradigm of definite articles was manually defined, and where the wildcard allows to account for the possibility to be accompanied with an adjective (e.g. *dans la ferme intention*). Such queries are likely to encompass spurious occurrences, but this disturbance would be marginal. When we targeted a specific idiomatic use of the form (e.g. *c'est dans la boîte*), we did not use the wildcard and favored a more constrained request (i.e. `dans la boîte`). Also, some forms were included in the study but did not lead to an S-curve, so they do not appear in Table 5.7:

Form	Occurrences	Slope	t_0	Rise
<i>dans l'histoire</i>	5737	1.09 ± 0.07	1631-1640	1591-1670
<i>dans l'intérêt</i> (i)	1601	1.43 ± 0.08	1631-1640	1611-1670
<i>dans l'avenir</i> (i)	1653	1.80 ± 0.09	1631-1640	1611-1670
<i>dans l'esprit</i>	10117	1.03 ± 0.04	1641-1650	1581-1700
<i>dans la douleur</i>	404	1.13 ± 0.09	1641-1650	1611-1670
<i>dans l'attente</i> (i)	1352	0.90 ± 0.05	1641-1650	1611-1690
<i>dans le sens</i> (i)	6324	1.36 ± 0.08	1651-1660	1611-1700
<i>dans l'ordre</i> (i)	5584	0.72 ± 0.03	1661-1670	1611-1720
<i>dans l'affaire</i>	2370	0.79 ± 0.03	1671-1680	1611-1720
<i>dans les faits</i>	442	0.99 ± 0.07	1671-1680	1641-1720
<i>dans le détail</i> (i)	994	0.76 ± 0.04	1681-1690	1641-1740
<i>dans l'intention</i>	1035	0.54 ± 0.02	1701-1710	1611-1750
<i>dans les détails</i>	682	1.11 ± 0.04	1751-1760	1711-1700
<i>dans l'espoir</i>	1732	0.82 ± 0.03	1761-1770	1721-1800
<i>dans l'ensemble</i> (i)	1310	0.85 ± 0.03	1781-1790	1741-1830
<i>dans l'avenir</i> (ii)	1653	0.53 ± 0.02	1781-1790	1741-1840
<i>dans l'hypothèse</i>	673	1.45 ± 0.10	1791-1800	1761-1840
<i>dans le sens</i> (ii)	6324	1.25 ± 0.09	1801-1810	1781-1840
<i>dans l'intérêt</i> (ii)	1601	1.30 ± 0.08	1821-1830	1781-1870
<i>dans le but</i>	1812	0.87 ± 0.04	1821-1830	1781-1880
<i>dans le coup</i> (i)	332	1.00 ± 0.06	1831-1840	1801-1860
<i>dans l'absolu</i>	264	0.42 ± 0.04	1841-1850	1811-1870
<i>dans l'air de</i>	375	0.59 ± 0.03	1851-1860	1791-1900
<i>dans la boîte</i>	533	1.14 ± 0.05	1861-1870	1831-1900
<i>dans l'air</i>	4917	1.04 ± 0.04	1871-1880	1831-1910
<i>dans les conditions</i>	3721	0.78 ± 0.03	1871-1880	1821-1930
<i>dans l'attente</i> (ii)	1352	0.89 ± 0.07	1891-1900	1871-1930
<i>dans le cadre</i>	2433	0.32 ± 0.01	1901-1910	1841-1970
<i>dans la direction</i>	2679	0.80 ± 0.03	1901-1910	1861-1960
<i>dans le sens</i> (iii)	6324	0.98 ± 0.02	1901-1910	1881-1960
<i>dans le détail</i> (ii)	994	1.04 ± 0.05	1911-1920	1881-1950
<i>dans la mesure</i>	4753	0.56 ± 0.02	1921-1930	1841-1970
<i>dans l'ordre</i> (ii)	5584	1.57 ± 0.01	1921-1930	1901-1960
<i>dans l'ambiance</i>	56	0.76 ± 0.04	1931-1940	1881-1990
<i>dans la ligne</i>	543	1.00 ± 0.07	1931-1940	1901-1980
<i>dans l'ensemble</i> (ii)	1310	0.75 ± 0.03	1941-1950	1881-1990
<i>dans les délais</i>	148	1.58 ± 0.15	1941-1950	1921-1980
<i>dans la perspective</i>	676	1.28 ± 0.06	1951-1960	1911-1990
<i>dans le contexte</i>	295	1.26 ± 0.08	1951-1960	1911-1990
<i>dans le coup</i> (ii)	332	1.04 ± 0.04	1951-1960	1931-2000
<i>dans l'immédiat</i>	338	1.08 ± 0.05	1961-1970	1911-2000
<i>dans la lignée</i>	80	0.43 ± 0.02	1971-1980	1921-2010
<i>dans le pire des cas</i>	94	0.80 ± 0.04	1971-1980	1921-2010

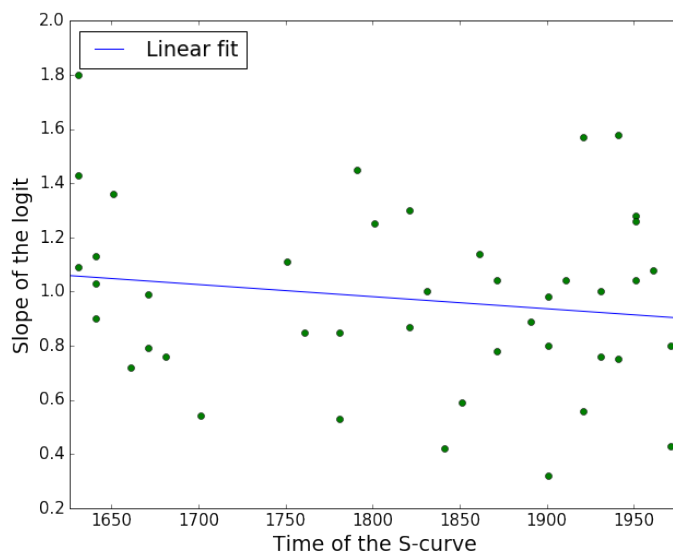


Figure 5.22: From Table 5.7: Slope of the logit transform of the S-curve versus time at which the S-curve is found to occur, defined by the moment when it reaches its peak of intensity.

- dans le sac
- dans la possibilité
- dans l'impossibilité
- dans la peur
- dans l'angoisse
- dans l'urgence
- dans l'optique
- dans la visée
- dans les temps
- dans la vue
- dans les cordes
- dans l'affirmative
- dans l'idéal
- dans l'idée
- dans la limite
- dans la continuité
- dans la foulée
- dans l'éventualité
- dans les termes

It is easy to rule out the lexical diffusion law of rate acceleration (Fig. 5.22). The associated Pearson correlation coefficient is equal to -0.16. This is too weak to be significant, and anyway goes in the opposite sense of the hypothesis, according to which the rate must be higher for contexts affected later.

The Constant Rate Hypothesis seems irrelevant as well to account for the data. Yet, one can see that some changes occur almost at the same time and have very close slopes. This is the case of *dans l'affaire*, and *dans le détail* (i), *dans l'ambiance* and *dans l'ensemble* (ii), *dans la perspective* and *dans le contexte*. Though these changes are too diverse for the Constant Rate Hypothesis to be broadly applicable, it may still work in some selected instances. In this case, I would favor a construction-based interpretation, according to which there are no 'contexts', but the development of one particular construction, i.e. [dans { la, le } {contexte, perspective}], with at least two paradigm members. This construction stems out of a larger [dans {location}] construction, but adds probably a layer of semantic, functional, or syntactic specificity; in other terms, it probably presents constructional features that the [dans {location}] construction lacks of, or at least fails to put the emphasis on.

Now, one could argue that the present case study is irrelevant in the sense that

those are independent changes, divergences from the regular uses of the construction, so that they should not be held as contexts of diffusion. This would be a legitimate objection. Still, there is some consistency in the evolution of the paradigm as a whole (Fig. 5.23), especially on the period 1561-1710, where it exhibits a clear S-curve.

What is more, and this is a crucial empirical observation, the evolution of *dans* itself is cohesive and remarkably obeys the pattern (Fig. 5.24). Arguably, the S-curve of growth of the paradigm coincides roughly with that of *dans*. Yet, these idiosyncratic constructs are better interpreted as branching constructions, all the more numerous that *dans* is frequent, and their behavior, their propagation, their rise, are completely irrelevant for the evolution of *dans*. In fact, they are extremely infrequent when compared to all uses of *dans* (totalizing 65k occurrences, to be compared with the 2M occurrences of *dans*, so that the paradigm represents about 3.3% of the occurrences of *dans*); and indeed, once subtracted to the evolution of *dans*, we find the exact same curve, with the exact same slope. Therefore, these marginal, speciated constructs, should not be seen as lexical contexts in the overall diffusion of *dans*. They are to be understood, and studied, rather independently.

Therefore, to study the contexts associated with the lexical diffusion of *dans*, it could have been more advised to consider either the different determiners with which it combines (*ma, cet, chaque*), as was done in (Fagard and Combettes, 2013) — without any quantitative measures of the S-curve —, or to search for the main nouns of the collocates paradigm of [*dans* {determiner}]. The relevant lexical contexts of diffusion could also be the syntactic positions in which *dans* occur, or more precisely, the constructions in which it is used. We actually tried the first approach, but the results were inconclusive.

5.3.3 What shall we conclude?

It may very well be that, in the case of *dans*, there are no clear contexts of lexical diffusion on which to test the Constant Rate Hypothesis or the rate acceleration law. In that sense, I would like to re-assess the idea that language is composed of a multiplicity of levels. As a result, language change can occur on all levels, and what happens on one level can influence what happens on another one. The spread of *dans*, for instance, has served as a catalyst for multiple, small constructions, to appear in its wake. The global process of the semantic expansion of *dans* happened on a higher level, and was certainly fed by what was happening on lower levels, as well as it was affecting the behavior and the frequency of many different constructions. As these levels are not on the same plane, or so to speak, the diffusion of *dans* was not horizontal, so we cannot find contexts that were converted, either in parallel or one at a time. Such a scenario would only work for a simpler process, as was almost the case with the [*en plein* {N}] construction.

Also, I would like to defend the idea that even if *dans* is probably not a construction in itself, it is a recognizable, individual, and independent linguistic item, which followed its own specific evolution. Now, should we consider that its remarkably cohesive evolution results from smaller changes, or that the smaller changes are splintered reflects of a bigger change? Obviously, the two perspectives are complementary. It illustrates well that linguistic changes, and semantic replacements, are not isolated events, and corre-

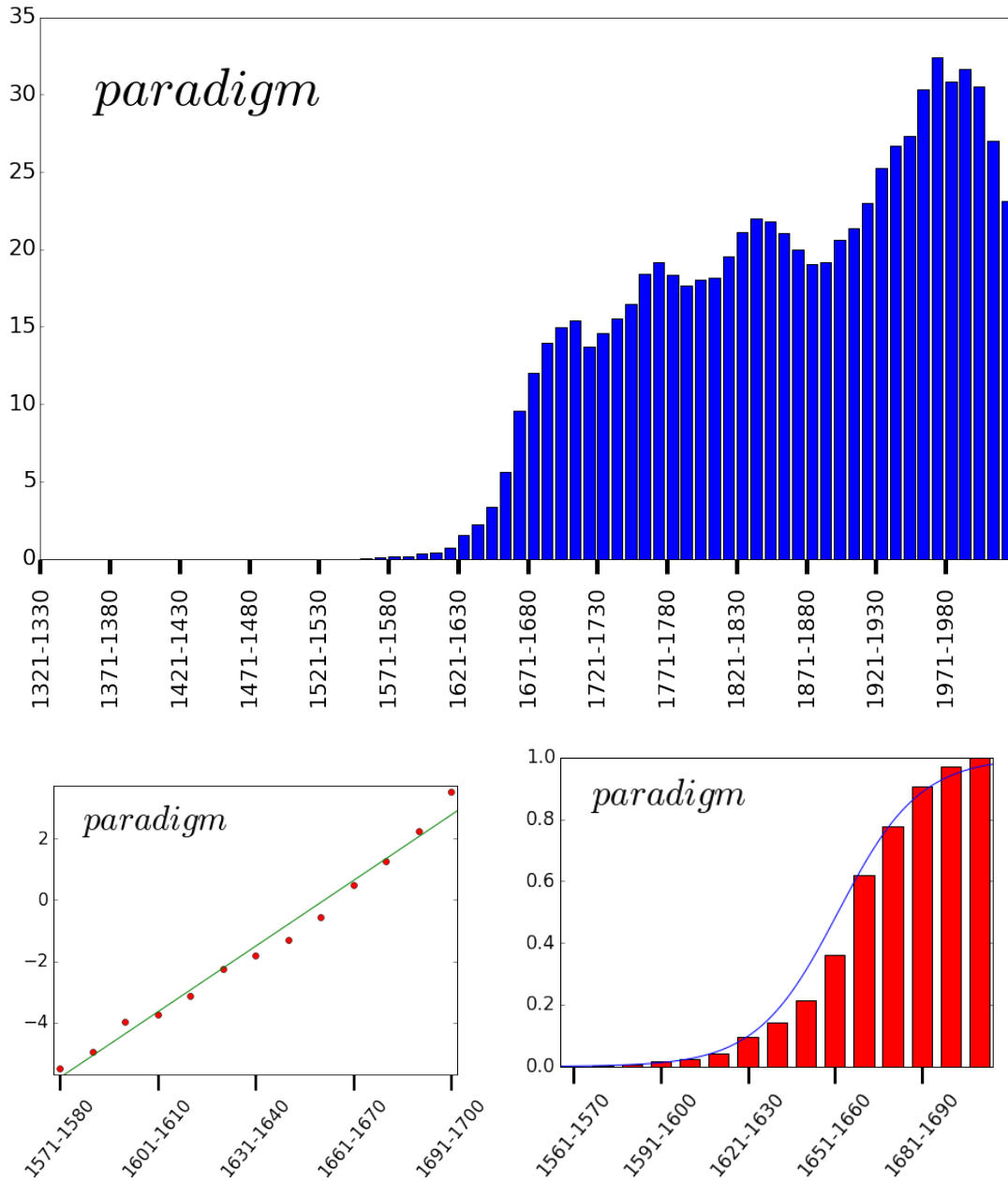


Figure 5.23: Top: Frequency profile of the paradigm of idiosyncratic forms of the [dans {definite article} {name}] pattern. Bottom left: Logit transform of the extracted growth part of the curve. Bottom right: Sigmoidal fit of the extracted growth part.

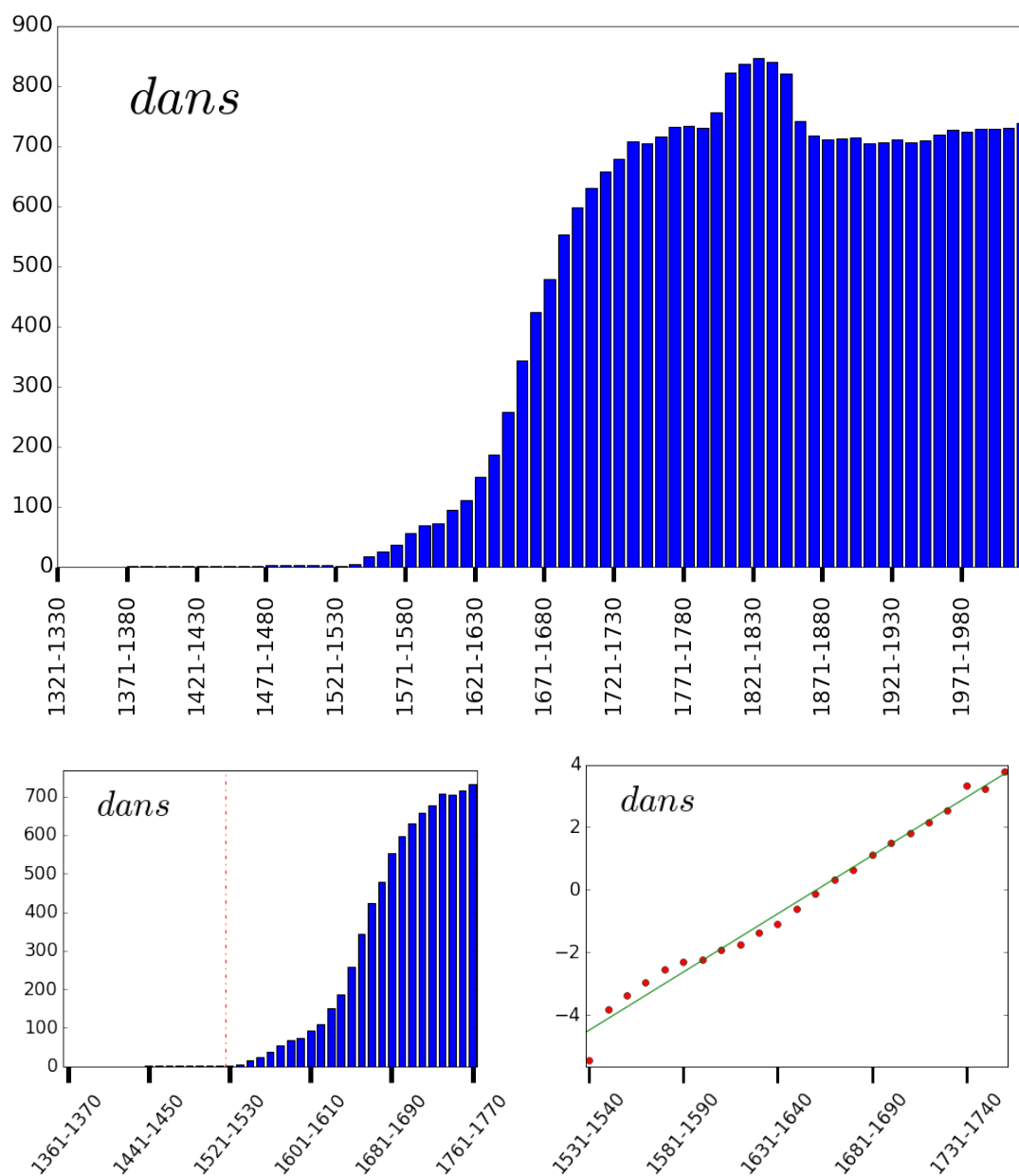


Figure 5.24: Top: Frequency profile of *dans*. Bottom left: Extracted rise pattern, with both latency and S-growth. Bottom right: Logit transform of the extracted growth part of the curve.

spond, only in the rarest, most marginal cases, to one-to-one competitions. It also may evoke the feeling that language changes by switch of abstract, underlying elaborated rules; more likely, language change would be, as ecological change, a mixture of processes happening on multiple, entangled, hierarchical scales. Changes on larger scales may ripple and reverberate all down the system, but they only happen through smaller, more local changes, both in time, and in the functional, semantic space. Therefore, it appears once more that the most delicate problem which an empirical account of language change is facing is the question of what, exactly, is changing.

The most reliable empirical sign that something is happening would be the S-curve. If a frequency rise fits so well a mathematical function, it must be that a specific process of change is at work. Yet, is it isolated, or is it just one aspect of a larger, broader change? As we saw throughout this chapter, this is a delicate, but relevant and challenging question. The two alternative hypotheses, the Constant Rate Hypothesis and the rate acceleration law, only make sense once we have already identified what exactly is the occurring change, and which contexts does it encompass, that is, once we have answered this crucial question. In the case of a phonetic change, it may be easier to handle, and this would be why Shen (1997) managed to find clear lexical contexts (different words), and a very narrowly identified change (a phonetical one), making it possible to test the rate acceleration law. Its relevance in semantic change remains to be assessed.

The Constant Rate Hypothesis might be more directly relevant to constructional change. We can indeed posit that, if two changes share the same S-curve, both in terms of rate and of peak intensity time (and we could add as well that they must have roughly the same boundaries), then these two changes are actually one and the same, or at least, reflect the same constructional shift of feature. This would be useful, for instance, to argue that a change makes sense only with reference to an over-arching construction. Yet this is not the Constant Rate Hypothesis, which states that one given change should affect the different contexts with the same rate. On the contrary, we propose to use the Constant Rate Hypothesis to detect a relationship between two changes which was not entirely expected. This is, basically, the same idea as the correlation approach we developed throughout this chapter.

It amounts to the following: empirical correlations in frequency are meaningful. They can be taken as conclusive evidence that two processes are related. The actual frame in which they are, however, needs not always be entirely clear. On the other hand, and this is most crucial, we saw that the emergence of an S-curve, no matter how robust, does not necessarily correspond to a single change. In the case of *dans*, unless we posit a peculiar Universal Grammar parameter to explain its most wondrous semantic expansion, it is likely that the S-curve encompasses many processes which, for a reason, concurs on a larger timescale. This would raise many questions: If the broad-scale S-curve of *dans* is made of several processes, which may end and begin at different times, separated by decades somehow, then how they relate within the bigger picture, across time and across entire generations of speakers, is at the very least intriguing.

In this chapter, we proposed, and made use of, several empirical treatments that may allow us to analyze in rich details specific instances of language change. We showed that it was possible to distinguish if a change was to be chiefly understood as

an internal paradigm reorganization, or as driven by linguistic pressures external to the pressure. At the very last, it was made clear that most changes are best understood within a larger scale, such as the paradigm to which the involved forms belong. We also argued that it was possible to extract information from the distribution of the collocates of a given construction. A further perspective of research would be to relate network features (connectivity distribution of nodes, mean shortest path, etc.) with linguistic interpretations. A longer term aim would be to integrate these different pieces of information in a more systematic way.

Yet, to understand something such as the evolution of *dans* and how it relates with smaller changes, I believe additional tools would be needed. Indeed, the study of the collocates, of the correlation between paradigm members, are chiefly horizontal relationships: they make sense on a single given level of the language structure. In chapter 1, we recalled the difference between collocation and colligation. Colligation describes the attraction of a construction to the slots of other, more complex constructions. This is typically a vertical relationship. Therefore, an empirical way of dealing with colligations would be a useful addition to this array of tools. Not all vertical relationships are subsumed within colligations though. The shepherding power of individual items such as *dans*, *par* or even *way* or *coup*, is also a vertical relationship, yet transverse with the respect to the hierarchy of constructions. How to handle empirically such matters is a challenge for future linguistic data analysis.

Part III

Modeling semantic expansion

Chapter 6

Previous works in language change modeling

Before turning to the presentation of our model of semantic change, we shall provide firstly a short survey of modeling works on the question of language change. Here, we strictly confine to analytical models, either mathematical or numerical, and leave aside other valuable approaches to the question such as, for instance, AI-based synthetic experimentations on language change and the emergence of a linguistic consensus (Steels, 1995; Bleys et al., 2009; Steels, 2015).

Excellent reviews on language change modeling have already been provided by various authors: Castellano et al. (2009) are more broadly concerned with the modeling of social phenomena from the point of view of statistical physics (though they encompass different approaches as well), yet specifically devote an entire section on language dynamics which despite of its brevity sums up efficiently the main tendencies of the field and gathers a great deal of relevant references; Vogt and De Boer (2010), introducing a special issue of the journal *Adaptive Behavior* devoted on language evolution, focus more exclusively on numerical approaches; Steels (2011) also includes works on synthetic experiments with artificial intelligence, alongside with numerous other interesting references; Loreto et al. (2011) offer a smaller scope as the other reviews, but provide a detailed account on recent works regarding the main Language Games (Naming Game, Category Game), an important family of models within the agent-based approach; Gong et al. (2014) survey more recent works and open the perspective towards rationalizations of the process of language change which are not necessarily grounded on any analytical formulation.

Given the richness of the existing literature, we shall therefore not aim to present a large-scale review on language change and focus instead on a few selected models which are of particular relevance for the modeling of the S-curve and of semantic change. We attempted to classify these works into different categories, according to the approach they favor to explain and reconstruct the S-curve. The first kind of models can be qualified as ‘macroscopic models’ (Stauffer and Schulze, 2005) or ‘equation-based models’ (Gong et al., 2014), as they do not rely on any specific details. In these models, language evolves deterministically according to a general equation usually describing a competition between two linguistic alternatives. Different explanations

may be provided to argue for the relevance of the different terms of the equation, but they are not bound to any interpretations and chiefly aim to generate a particular shape of diffusion. The second kind of models relates language change to language learning, and makes use of an explicit learning rule between generation of speakers. The third approach accounts for the S-curve through sociological diffusion. Usually, they are agent-based models and a specific network structure is usually provided. Then we shall focus on the presentation of the Utterance Selection Model and its variants, for it specifically deals with the S-curve problem, and from which we drew a lot of inspiration for the devising of our own model. Finally, a last model will be discussed, which stands apart from the four previous categories and does not aim at modeling specifically a situation replacement, but presents a special interest for the understand of semantic change.

6.1 Macroscopic models

Macroscopic models, or equation-based models, are deterministic model of language change. They typically set up a competition between two mutually exclusive variants. The frequency of a variant, say the new one, is then described by an evolution equation, whose solution gives the diachronic profile of frequency for this variant. As these models are deterministic, the outcome of the evolution only depends on the initial conditions. The terms of the equation are usually given an interpretation which describes the mechanisms of the change, from a sociolinguistic perspective most often, as we shall see below. This family of models is further discussed by Solé et al. (2010).

6.1.1 The Abrams-Strogatz model of language competition

One of the first macroscopic approach to language competition is due to Abrams and Strogatz (2003), and has led to an extensive number of further studies and expansions (Castellano et al., 2009). The initial idea is that two languages X and Y are competing against each other for monolingual individuals, who can switch from language X to language Y and conversely. The frequency x associated with language X is interpreted as the fraction of individuals speaking X in the population. This frequency evolves with time with the following equation:

$$\dot{x} = yP_{Y \rightarrow X}(x, s) - xP_{X \rightarrow Y}(x, s), \quad (6.1)$$

where $P_{Y \rightarrow X}(x, s)$ (resp. $P_{X \rightarrow Y}(x, s)$) is the switching probability from Y to X (resp. from X to Y). These switching probabilities depend on x , the frequency, and s , the social prestige associated with X . Necessarily, we have $y = 1 - x$ (individuals are monolingual), and furthermore, the social prestiges of the two forms sum up to one ($s_Y = 1 - s_X$), hence the notation s . The probability to switch from Y to X depends solely on the features of variant X . We can therefore note $P_{Y \rightarrow X}(x, s) = P(x, s)$ and $P_{X \rightarrow Y}(x, s) = P(1 - x, 1 - s)$ for the sake of clarity.

A specific choice for the switch probability is proposed:

$$P(x, s) = csx^a, \quad (6.2)$$

where c and a are two numerical parameters.

In this paper, the authors provide data from language in decline and fit the empirical curves with this model. They discover that the a exponent is the same over all studied cases, and extract the social prestige s of the declining language.

What is of interest for us would be how this model applies to a semantic change. The authors note that there are three fixed points to this equation, two stable ones, corresponding to the fixation of either one variant or the other ($x = 0$ and $x = 1$), and an unstable one, which can be computed:

$$x^*(s) = \frac{(1-s)^b}{s^b + (1-s)^b}, \quad (6.3)$$

with $b = (a - 1)^{-1}$.

If $s = 0$ (the variant X has no prestige) then we get $x^* = 1$, so that $x = 1$ is no longer stable anymore. X cannot prevail unless every one in the population is already speaking X , and as soon as Y manifests itself, X is bound to get extinct. If $s = 1$, then the first occurrence of X is enough to trigger a replacement, since $x = 0$ becomes unstable. The intermediate case is more interesting. If we consider X to be the new variant, then it says that, unless X achieves the threshold frequency $x^*(s)$, it cannot replace Y . Below this threshold, the system would otherwise be driven towards $x = 0$. The greater the prestige of the new variant, the lower the threshold. However, this model cannot explain how the new variant, no matter how prestigious it can be, is able to reach the frequency threshold in the first place. Thus, for this model to explain a semantic replacement, there must be a non-negligible frequency of people switching towards X as a prelude of the process; only then can the equation describe the remaining of the replacement.

That a frequency threshold needs to be reached for the replacement to occur seems perfectly reasonable, since otherwise, any new variant would take over the entire system, as soon as it is more prestigious than its predecessor. In the cases studied by the authors, the frequency threshold can be overcome through a language contact situation between two populations speaking a different language, especially in colonial or imperialist settings. Then, it is indeed conceivable that the frequency of the prestige language reaches a threshold independently of the evolution equation. However, in cases of semantic replacement, which occur spontaneously without attested language contact, this frequency threshold poses a theoretical difficulty.

6.1.2 Macroscopic models in lexical diffusion

The same kind of macroscopic models has been developed in the framework of lexical diffusion (Wang et al., 2004; Wang and Minett, 2005). Interestingly, these works are led by William S.-Y. Wang, who proposed the idea of lexical diffusion initially in the late 1960s (Wang, 1969).

The set-up is quite similar and aims to describe phonetic change. Each word i , or lexical context, exists in two forms, the unchanged one, and the changed one, reflected by frequencies u_i and c_i which represent the fraction of speakers in the population adopting either the unchanged form or the changed one. Thus we have, as before, $u_i = 1 - c_i$. For a given word, there exists a switching probability to go from u_i to

c_i during time interval dt , this switching being unilateral. This switching probability grows with the number of speakers having adopted the changed form, and the simplest possible form, $\alpha c_i dt$, is assumed. Therefore we get for the evolution equation:

$$c_i(t + dt) = c_i(t) + \alpha c_i(t) dt u_i(t), \quad (6.4)$$

or, in a continuous time limit:

$$\dot{c}_i = \alpha c_i(1 - c_i), \quad (6.5)$$

which is the usual dynamic equation associated with a sigmoid (Pearl and Reed, 1920). This equation has two fixed points 0 and 1, the former being unstable and the latter, stable. This entails that, as soon as the changed form appears in the system, if only marginally, it will eventually take over according to an S-curve of growth.

The novelty of the approach is to couple this traditional picture with lexical diffusion. There is not only one form undergoing the change, but several. There exists furthermore a linguistic pressure to harmonize the system. Henceforth, as there are more and more forms changing, the other forms, still unchanged, feel a pressure to change themselves. Therefore, the rate of change now depends on all the forms, through the couplings α_{ij} between words:

$$c_i(t + dt) = c_i(t) + \sum_j \alpha_{ij} c_j(t) dt u_i(t), \quad (6.6)$$

leading to:

$$\dot{c}_i = \sum_j [\alpha_{ij} c_j] (1 - c_i) \quad (6.7)$$

This equation yields two interesting features. The first one is that if a lexical context is initially associated with a zero frequency of the changed form, then it can still develop this new variant, driven by the pressure of the other lexical contexts. The second is that, if we assume that $\alpha_{ij} = \beta$ if $i = j$ and $\alpha_{ij} = \gamma$ otherwise, then the slope at mid-way will be equal to:

$$\left. \frac{dc_i}{dt} \right|_{c_i=1/2} = \frac{1}{2} \left[\frac{\beta}{2} + \gamma \sum_{j \neq i} c_j(t) | c_i(t) = 1/2 \right] \quad (6.8)$$

Therefore, the later the form develops, the higher the rate, since the other $c_j(t)$ will be bigger by this time. In that sense, the later developing forms catch up their delay — a hypothesis fundamentally at odds with the Constant Rate Hypothesis of Kroch (1989b). We can also imagine that a particularly influential lexical context, once it starts to change, will fast forward the replacement process in all other contexts.

6.1.3 Internal and external factors of change

Another model has recently been proposed to go beyond the S-curve process, and to distinguish internal and external factors of change (Ghanbarnejad et al., 2014). They study the evolution of the frequency ρ of a new variant (once again interpreted as the

fraction of adopters in the speakers' population). They propose the following adoption rate, given by:

$$P_{adoption} = a + b\rho. \quad (6.9)$$

Note that the new variant can only be further adopted, and is never cast off; since the rate is non negative, for any value of ρ the novelty is bound to be adopted by the whole population, even before it appeared once.

This adoption rate is made of two terms. The second one, $b\rho$, is the usual part of the logistic equation; it grows as the number of speakers, and reflects social imitation. The first one instantiates a constant push over for the new variant, independently of its frequency. It leads to the following evolution equation:

$$\dot{\rho} = (a + b\rho)(1 - \rho). \quad (6.10)$$

The case $a = 0$ is the usual logistic one. In this framework, the S-curve is therefore a limiting case. Conversely, if $b = 0$, the evolution would be exponential, and its momentum would wear off as more and more speakers adopt the new variant.

This paper also provides a way to weight the impact G^f of a given factor f , through the formula (slightly adapted here):

$$G^f = \int_0^{+\infty} \frac{g_f}{\sum_{f'} g_{f'}} \dot{\rho} dt, \quad (6.11)$$

where g_f is the contribution of factor f in the adoption rate. The returned value lies between 0 and 1, and is a measure of how much the factor shapes the change (1 would mean that the change only depends on this particular factor, 0 that the factor did not intervene in the change). This formula is a very general one and its scope does not restrict to the simple evolution equation (6.10). In this case, the contribution of the exogenous factor is:

$$G^{exo} = \frac{a}{b} \log \left(\frac{a+b}{a} \right). \quad (6.12)$$

Interestingly, (6.12) shows that a comparatively small exogenous parameter can have a high impact of change; if $a = 0.25$ and $b = 0.75$, then the exogenous mechanism still accounts for about one half of the change.

This exogenous mechanism is especially interesting because a can be extremely small, and provide a way to start off the process even in case of non-zero initial frequency. This term is called by the authors an 'exogenous factor', in contrast with $b\rho$, 'endogenous', because it reflects how the community of speakers acts upon itself. a is therefore the product of an external pressure to change. In the empirical cases investigated by the authors, only one case shows a a parameter significantly different from zero, which is the actuation of a spelling change in German, following a decree by the German Language Academy.

However, in a interpretation which would not be a sociolinguistic one, the exact nature of the a influence would be unclear. The $b\rho$ mechanism is, in any case, a self-reinforcement loop. In a situation of semantic change, the exogenous mechanism could correspond to an invitation of the linguistic system for a particular innovation.

Even if it were very small (say, $a = 0.01b$), it would still account for some part of the change (5% in that particular example), and could especially serve as a trigger to initiate the change. Then, the ‘exogenous’ linguistic factor can be nearly invisible from an empirical point of view, and yet play a crucial role in setting off the change. This, it would seem, is a simple, straightforward and convenient way to solve the Threshold Problem, as such a barely noticeable primer can be posited in nearly all models. However, it cannot explain why the initial frequency, in most data-based accounts of the S-curve, is significantly different from zero.

6.2 S-curve and learning

In this part we will briefly present a few works on language change where the learning of language by new generation of speakers is the main mechanism for change. Numerous models exist along this line, including the Evolutionary Language Game of Nowak et al. (1999) and the Iterated Learning Model by Kirby and Hurford (2002), both of which have stirred up an important number of subsequent works. Here we shall only focus on two works, for they directly address the question of the S-curve.

6.2.1 An early model of language change

The model of Niyogi and Berwick (1995) is perhaps one of the earliest analytical models of language change. The framework of the model is Universal Grammar. There is a space of possible sentences and each grammar constrains the corresponding language in such a way that only some of the sentences are acceptable. More precisely, to each language i corresponds a probability distribution P_i over the space of possible sentences. This probability distribution entirely characterizes the associated language.

The probability distribution for a language learner to hear the different possible sentences is given by a barycentric sum of all languages co-existing in the population. Therefore, if the population of speakers comprise N individuals and each language i is adopted by n_i speakers, the probability of utterance of sentence ω associated with the whole population is given by:

$$P(\omega) = \frac{1}{N} \sum_i n_i P_i(\omega). \quad (6.13)$$

The authors then define a ‘learning algorithm’ which is a mapping between a set of n sentences d_n and the hypothesis, from the language learner, that this set of sentences has been generated by a grammar h . The probability of the set d_n is given by the quantities $P(\omega)$. Thus, the probability $p_n(h)$, for a language learner to adopt grammar h after being exposed to n sentences as a learner, depends on the population structure, $\{n_i\}$, on the grammars $\{P_i(\omega)\}$ of the different languages, and on the specifics of the learning algorithm. This probability $p_n(h)$ is crucial as it defines a transition probability between two population states.

There is then an iterated dynamics over the generations. At each step of time, there are speakers and learners. A learner receives input from the whole population of speakers, and chooses a grammar on the basis of this input at the end of the critical learning period. Then the speakers are replaced by the learners and a new generation

of learners is introduced. Therefore, the model describes the evolution of the set $\{n_i\}$ of the adopters of each language (or grammar).

Further details are given in a later study by Niyogi and Berwick (1997). A grammar is described by a set of n binary parameters. The associated language is a uniform distribution over a subset of all possible sentences. The learner begins with an arbitrary grammar. Through the learning period, a total of m sentences is presented, and each one is processed at a time. As long as the sentences belong to the language associated with the current grammar, the learner will stick to it. Whenever a sentence is not sanctioned by this grammar, a new one, able to process the sentence, is chosen by the switching of a Boolean parameter. If no compatible grammar can be reached by the switching of a single parameter, then no change of grammar is made.

In the case of a one-parameter grammar, Niyogi and Berwick (1997) show that the dynamics reduces to:

$$p_{n+1} = Ap_n^2 + Bp_n + C, \quad (6.14)$$

where p_n is the fraction of adopters of one of the two grammars at the n -th generation. This is more or less in line with the previous macroscopic models, with the addition of an unusual constant term. However, their model is much more general. They focus on a three-parameter case (with eight different grammars), and compare it with actual language change scenarios (Niyogi and Berwick, 1995). They nonetheless do not provide details on the set of all possible sentences, nor on the specific sub-sets of each language. However, starting with homogeneous populations, they obtain S-curves describing the transition of the population from one language to another.

This model is extremely rich, and allows for many variants and possible extensions. However, its complexity forces one to provide a lot of input elements, such as the sets of possible sentences associated with each language. Actually, for the population to switch entirely from one language to another, there must be very specific learning biases, and specific language sets over the space of possible sentences. If it were to be applied to a case of a semantic change, with two linguistic competitors such that each one represents a distinct ‘language’, the two sets would probably be non-overlapping, so that all sentences produced by a homogeneous population of speakers will belong to any one subset, leaving no choice to the learner but to make the choice of the already entrenched variant.

As a conclusion, this model does not seem to adapt well to a simple semantic change scenario.

6.2.2 Bayesian learning of language

Reali and Griffiths (2010) propose a model which is similar to the previous one, but does not require one to specify any grammar or any set of possible sentences. The idea is that a given linguistic form can exist in K different variants. A learner is provided with N occurrences of the form, which can be described by a corresponding vector $\{x_k\}$ of the frequencies of all variants in this sample. From this stimulus, the learner infers the statistical distribution θ over the variants through a Bayesian learning rule. At the next time iteration, the learner is now the speaker and is producing occurrences for a new learner, which will infer a distribution in its turn. What differs from the previous model is the learning rule (now Bayesian), and the probability of sentences

(interpreted as variants of a linguistic form) associated with a given hypothesis, which is no longer uniform over a subset, but given by a Dirichlet distribution. Moreover, there is no longer any population dynamics.

To present further the model we will take a simpler example where there are only two possible variants, X and Y . We shall write x the frequency of the new variant, so that the frequency of the old one is $1 - x$. At each time step, the speaker produces N occurrences of the form, choosing variant X with probability v_t and Y otherwise. The corresponding production is then described by x_t , which is the total number of occurrences of X . The learner is then given this input x_t and must infer from it a probability distribution $P(\theta|x_t)$, where the θ corresponds to the estimate probability to produce X in an utterance.

This probability $P(\theta|x_t)$ is given by the Bayesian learning rule:

$$P(\theta|x_t) = \frac{1}{Z} P(x_t|\theta) P_0(\theta), \quad (6.15)$$

where Z is the integral of $P(x_t|\theta) P_0(\theta)$ over all values of θ .

The likelihood $P(x_t|\theta)$ is the probability that, given that there is a θ chance to produce x occurrences of X among the N . This is just a binomial probability:

$$P(x_t|\theta) = \frac{N!}{x_t!(N-x_t)!} \theta^{x_t} (1-\theta)^{N-x_t}. \quad (6.16)$$

The prior probability $P_0(\theta)$ reflects the initial belief of the learner concerning the probability to produce either variant. As the learner has no reason to favor any, this prior probability must be symmetric ($P_0(\theta) + P_0(1-\theta) = 1$). The authors have chosen a Dirichlet distribution of parameters (γ_X, γ_Y) :

$$P_0(\theta) = \frac{1}{Z_0} \theta^{\gamma_X-1} (1-\theta)^{\gamma_Y-1}, \quad (6.17)$$

with the normalization constant Z_0 given by:

$$Z_0 = \int_0^1 d\theta \theta^{\gamma_X-1} (1-\theta)^{\gamma_Y-1}. \quad (6.18)$$

Furthermore, the symmetry condition between variants imposes $\gamma_X = \gamma_Y$, which shall be noted γ_0 .

The Bayesian rule ensures that the Dirichlet distribution is conserved through the update rule, but the γ parameters are modified according to:

$$\begin{cases} \gamma_X = \gamma + x \\ \gamma_Y = \gamma + (1-x) \end{cases}, \quad (6.19)$$

which, as expected, breaks the symmetry between the two variants.

For the model to be complete, the learner, at the end of the learning period (which is processed as a whole), becomes the speaker and produces in its turn N occurrences

of either X or Y for a new learner. Occurrences of X are produced with probability v_{t+1} , which is itself given by the mean of the distribution $P(\theta|x_t)$, actually equal to:

$$v_{t+1} = \frac{\gamma_X}{\gamma_X + \gamma_Y} = \frac{\gamma + x}{N + 2\gamma}. \quad (6.20)$$

The iterated learning process is therefore stochastic can be described as:

$$v_0 \rightarrow x_0 \rightarrow v_1 \rightarrow \dots v_t \rightarrow x_t \rightarrow v_{t+1} \rightarrow x_{t+1} \dots, \quad (6.21)$$

where the regular arrow \rightarrow denotes a deterministic relation and the harpoon arrow \rightarrow a probabilistic sample of a distribution (hence, the stochastic nature of the process). The inferential process is entirely deterministic and all stochasticity comes from the speaker. The whole inference mechanism is not directly involved and reduces to formula (6.20) to relate v_{t+1} and x_t . Note that the prior distribution P_0 is the same at each timestep.

This prior distribution is actually crucial. Indeed, depending on the value of γ , it can reflect two different a priori beliefs. If γ is greater than 1, then it peaks at $\theta = 1/2$, which means that the learner expects a mixture of variants. On the other hand, if γ is smaller than 1, then it diverges for $\theta = 0$ and $\theta = 1$, which means that the learner expects that the input should be dominated by either one or the other variant. That means that the latter prior will favor situations where either one or the other variant is significantly used more than the other and will be reluctant towards a statu quo situation.

Now we can turn to the S-curve. If we start with $v_0 = 0$ (and hence $x_0 = 0$), it is possible that the X variant becomes eventually dominant. Indeed, even if Y only is produced, the prior distribution always acknowledge for the existence of X , so that the hypothesis may weight the odds in favor of X even if it has not been produced. Indeed, if we suppose that $x_t = Nv_t$ at each time step (deterministic approximation of the process), then we get:

$$\dot{x} = \frac{\gamma}{N + 2\gamma} (N - 2x), \quad (6.22)$$

with $x = 1/2$ as the only fixed point. Also, $v_{t+1} > x_t/N$ for $x_t < N/2$, so that the learning rule always over-estimates the new variant as long as it is in minority.

Over many runs of the process in the configuration which favors the domination of either one or the other variant (γ smaller than 1), and with the initial condition corresponding to a domination of the Y variant, the authors have selected all processes that lead to a domination of X (which is only a small proportion of the total). Note that the domination of one variant over the other is a finite size effect of the stimulus. In the deterministic approximation, corresponding to an infinite stimulus, this could never happen, and these processes are all the rare when the stimulus becomes bigger. From this subset of processes, it appears that the average trajectory has the shape of an S-curve. Interestingly, if the prior favors mixture ($\gamma > 1$), then the shape that is recovered is the reciprocal of the S-curve, with high slopes at the beginning and at the end, and a slowing down in the middle.

Blythe and Croft (2012) have discussed the relevance of this result for the modeling of the S-curve because it requires an average. Actually, the average makes sense, since

this describes only one lineage of learner-speaker successive pairs, while the S-curve is supposed to describe the whole population. But then, we would forget that this is not an average over all processes, but only over the processes leading to the replacement of Y by X . If the average were performed over really all processes, then only a very small minority of lineages would lead to the adoption of X (Reali and Griffiths, 2010), so that there would be no S-curve. Therefore the average cannot be interpreted as an average over the whole population of speakers and the criticism of Blythe and Croft (2012) is to be taken seriously.

We would have two additional concerns as for the applicability of this model to semantic change. The first one is that the new variant X must be known innately by all new learners of the language. But why would that be the case? How could new learners know, at birth, constructions which have not yet appeared in the language, and take them into account while learning the language for the input of their elders? The second one relates to the timescale of the change. With a very low number of occurrences on which to base the inference ($N = 50$), the authors find that the change would take, according to the average S-curve, fifty generations. We can assume that it would take more with a richer, more realistic stimulus. If we associate about thirty years for each generation, then the change would make approximately 1500 years to happen. However, a real language would be considerably renewed over this time period, and a real lot of changes would have happened in the meanwhile. Therefore, this process takes far too many generations of speakers to unfold so as to be able to account for actual instances of semantic change.

6.3 Sociologic accounts of the S-curve

The sociolinguistics-inspired models treat the S-curve phenomenon as the diffusion of the novelty among the community of speakers. In order to describe this diffusion, they usually present agent-based models on a network structure, much reminiscent of the general spin models in statistical physics. Different models of propagation have also been proposed, relying on diffusion equations (Patriarca and Leppänen, 2004), but they will not be treated here.

6.3.1 The Threshold Problem

One of the first works on this question is due to Nettle (1999), and openly relies on an even earlier work by Nowak et al. (1990), which, however, did not make particular reference to language change, while it is the chief aim of Nettle (1999). The ultimate goal of the paper is to overcome what is coined the ‘Threshold Problem’. A new variant, in language, would be completely drawn into the massive use of the older variant entrenched in the whole community. Contrary to a biological situation where the limited number of parents ensure reasonable chances of transmission of the mutation, a language learner has very many ‘parents’ (the linguistic input comes from all sorts of speakers in its personal network), so that a new variant has all chances to remain inconspicuous and should soon be forgotten.

What Nettle proposes is that this neutral view of language evolution should be replaced by another, more in line with the sociolinguistics realization that ‘Speakers do

not simply copy the most frequent items around them; they rather model themselves on those with whom they wish to associate and identify, and with whom their aspirations are bound up' (Nettle, 1999, p.100). Therefore, variants are weighted by a 'social impact' and a speaker shall adopt not the most frequent of variant, but the one which is perceived to have the higher social impact.

This social impact of a variant is subjective (i.e. it differs from one speaker to another) and depends on three things: the social status of the users of this particular variant, their social distance from the individual weighting the variant, and their numbers. Also, a further, constant bias is introduced, which reflects the ease to learn this particular variant. A precise mathematical formula for the social impact \hat{i}_X of variant X is given by:

$$\hat{i}_X = b_X n_X^a \left(\frac{1}{n_X} \sum_{k=1}^{n_X} \frac{s_k}{d_k^2} \right), \quad (6.23)$$

where b_X is the learning bias towards X , n_X the number of speakers of X , a a numerical parameter, s_k the social status of individual k and d_k its social distance from the individual under consideration. The term between parentheses is the mean of the quantity (s/d^2) over all speakers of X in the community.

The simulation considers 400 speakers on a 20×20 grid network, with periodic boundary conditions. Initially, all speakers have adopted variant Y . There is no learning weight towards one or the other variant. Furthermore, individuals are associated with an age, ranging from 1 to 5. Speakers of age 1 have no social status, so their way of speech does not impact anyone. All other speakers have an equal status of 1. Only children and adolescents (speakers of age 1 and 2) update their language during a 'lifestage', by selecting the variant which they perceive as having the highest social impact. This is a way to implement a 'critical period' after which the language of the individual becomes frozen. At each of these 'lifestages', the grid is swapped over left from right, up from down, so that each user of age 1 or 2 updates its choice of a variant sequentially. Then, the age of all individuals is increased by 1, and the individuals who were 5 at the previous lifestage are replaced by new individuals with age 1. Social distance d between speakers is given by the distance over the network.

Furthermore, Nettle adds another ingredient, which is spontaneous mutation. That is, there is a 0.05 probability that the learners would adopt the variant with the lowest perceived social impact, due to imperfect learning. That way, the new variant can spontaneously appear in the society. This mechanism plays a role alike the 'external factor' of Ghanbarnejad et al. (2014).

Nettle also discusses the role of parameter a , which weights the imitation mechanism. Most social diffusion models, he argues, take an a of 0.5, which favors a co-existence of the two variants. As we expect the community to reach consensus, a higher imitation weight is required. A value of 1, however, muffles the new variant, which can never take over. Therefore a value of 0.8 is chosen. However, no S-curve is found in that case; nor any complete replacement from one homogeneous linguistic convention to another. Either the new variant never spreads out, or the two variants come to co-exist in a final and stable statu quo.

To observe a true replacement, one has to add differences of social status. To each individual a status uniformly drawn between 1 and 20 is given. Then, the status of each

individual is changed to 100 with probability of $1/40$. To skew the social status inequalities further more, those statuses are then squared. Then complete replacement can occur. Actually, one can easily understand what is happening. The ‘hyper-influential’ individuals have such a social status that, if they choose the new variant through mutation, this variant has good chances to spread over. Social diffusion is then chiefly governed by the linguistic choices of a reduced number of individuals. That is, if one adopts the new variant, then the threshold is immediately crossed and the variant can spread. As there is no directionality, the consensus constantly goes back and forth between the two variants, with short transitions and long period of stability.

This mechanism is variant-neutral. Nettle also considers a case where the two variants are not made equal, by changing the learning biases. Interestingly, in the very skewed community presented before, a negative bias, even small, against the new variant, prevents it from taking over. Conversely, a favoring bias allows the new variant, once the transition has occurred, to remain undisputed. In a community where there is no status, those biases can still make a replacement happen, yet it never reaches full completion (the older variant still lingers in this case), and the consensus is going back and forth once again. Both social stratification and functional selection are therefore required to have replacement, and subsequent stable consensus.

Thus, the threshold problem can be overcome by adding different types of ingredients. One is a learning bias in favor of the new variant, the other is the existence of hyper-influential individuals whose influence is such that their individual choice alone is sufficient to pass the threshold and allows a further spread of the novelty.

However, this model does not lead to S-curves for the replacement. Transitions are sharp and abrupt, or they proceed by successive bursts and plateaus, depending on which mechanisms are activated. Therefore, if the S-curve can be trivially caught by a macroscopic, deterministic equation, implementing it on an agent-based model is far from being granted.

6.3.2 Network structure

There is actually one thing that Nettle (1999) did not change in his model. He varied the sheer weight of imitation, he tried to assess the optimal degree for the distance cut-off, he changed the statuses, the biases... Yet he did not investigate any other community structure than a grid-based network.

Acknowledging this limitation, Ke et al. (2008) propose to investigate social propagation on different types of networks. They also point out that, in Nettle’s model, the innovation can only come from highly influential individuals, while in actual instances of the change, novelties can be propagated ‘from below’. The framework follows closely that of Nettle (1999), especially on the learning perspective (1 are learners, 2 are learners and teachers, 3 to 5 are all only teachers). Instead of a social impact, the authors speak of a variant ‘fitness’, F , which is defined for each individual agent i :

$$F_X^i = f_X n_x^i \quad (6.24)$$

where n_x^i is the number of nearest neighbors of agent i having adopted variant X , and f_X is a ‘functional value’, independent of the agent. Note that, contrary to Nettle (1999), the frequency is not weighted anymore, corresponding to the case $a = 1$. There

is also no social status (the functional value corresponds therefore to the learning bias), and no social distance, but a cut-off beyond the nearest neighbors in the network. As in the previous model, an agent would adopt the variant associated with the highest fitness. Last, the mechanism of mutation has been discarded. Therefore, the initial state must account for a non-zero frequency of the new variant, which otherwise cannot appear in the system.

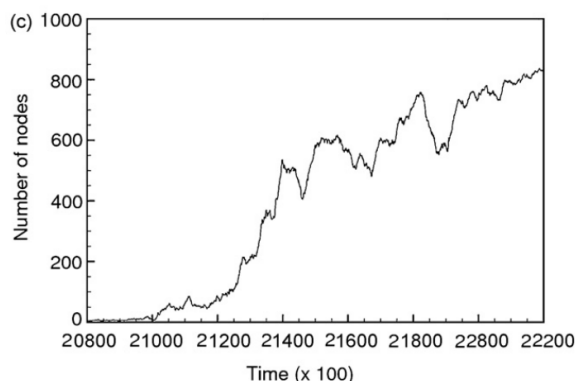
The authors investigate the dynamics of replacement for four kinds of networks: regular (all agents have the same connectivity), random, small-world, and scale-free. All of these networks have the same average connectivity, set to 20. The total numbers of agents is 500. Initially, one randomly chosen adult (age 3-5) is provided with X while all others are set to Y . In this setting, the propagation of the new variant is favored because it benefits from a functional bias compared to the entrenched variant ($f_X = 20$, $f_Y = 1$). Indeed, the authors are not chiefly interested in the conditions which can set the replacement, but focus on the shape of the diffusion curve.

They obtain S-curves for all types of networks save the regular one. They also investigate different functional biases ($f_X = 10$, $f_X = 2$), adjusting correspondingly the initial number of adopters of the new variant (respectively 10 and 100). In the regular and small-world networks, the propagation is slower, but the fixation probability is higher. Interestingly, the small-world network is the one presenting the widest variation on fixation times, ranging from 15 to 30 timesteps (which would be equivalent to a century and a half -three centuries if we take each timestep to represent ten years). This is still a bit long (we found an average growth time of ninety years in our statistical survey of semantic changes, given that we discarded all changes shorter than sixty years). Yet, it is remarkable that non-regular network structure alone is sufficient to generate an S-curve in an agent-based model.

6.3.3 Social networks of influence

Combining this network structure with Nettle's idea of polarized social influences, Fagyal et al. (2010) also propose a network model of the social propagation of language change, where the links are now directional (i.e. agent i way of speaking can influence that of agent j while the reverse is not necessarily). The network is scale-free, and is composed of 900 agents. The total connectivity (inward + outward) is equal to about 8 in average. Furthermore, they consider a total of 8 variants, instead of 2 as was the case in the two previous models. The population is initialized as heterogeneous.

The dynamics is also a bit different, as there is no longer any consideration of different ages, critical learning period, and a distinction between non-influent, learning children and influential but still learning adolescents. The theoretical linguistic framework of the model is indeed a bit different for it is now more exclusively oriented towards Sociolinguistics. The 'update' rule has also little in common with the two previous cases, and has been much simplified: at each time step, one agent is randomly chosen, and, among the subset of agents influencing him, a neighbor is selected with a probability proportional to its in-degree (i.e. the number of people under his influence in the convention chosen by the authors). That is, the higher the influence of the neighbor, the more likely the agent will listen to him so as to decide which variant pick up. Once the pair has been drawn, the first agent conforms to the way of speaking of



(c) middle portion of an S-like diffusion path of one variant on an extended time scale.

Figure 6.1: Figure 5c of Fagyal et al. (2010), displaying an S-curve of replacement resulting from a single run of their model.

the chosen neighbor, i.e., adopts the same variant. All agents are monolingual (they use one variant at a time).

The simulation shows a rapid take-over of one variant. After this, the dominating variant is repeatedly replaced by one of its competitor. The marginal persistence of all variants is ensured by the existence of ‘loners’, that is, agents which are influencing a few others, but are not influenced by any one. Therefore, stable periods alternate with comparatively abrupt transitions from one variant to another. The authors argue that those transition periods are S-like (Fig. 6.1), though they admit that it is not ‘prototypically S-like’ (Fagyal et al., 2010, p.2071).

The authors then show that the presence, in the network, of both loners (who store the rare variants away from the social consensus) and leaders (who are able to efficiently propagate around them even the rarest variants) are both necessary for the replacement and the fixation to happen. Therefore, the S-curve picture of replacement relies on a specific social structure, which can be argued for and is certainly not irrelevant to describe real linguistic communities.

These three models, both exploring different aspects of the sociolinguistics aspects of change through agent-based models, show unanimously one thing: despite what the macroscopic models could have led us to believe, the S-curve is not granted as a picture of language diffusion through a social community, and depends on very characteristic ingredients, such as a high functional bias and/or initial number of innovators in (Ke et al., 2008), a very hierarchical structure in terms of social influence in (Fagyal et al., 2010), and a specific network structure in both models.

6.4 The Utterance Selection Model

The Utterance Selection Model draws from the evolutionary theory of change by Croft (2000). This theory makes a sharp distinction between ‘innovation’ and ‘propagation’. Innovation is the appearing of a new variant, sanctioned by the language organization. Propagation is the spread of the said variant within the linguistic community. The

Utterance Selection Model deals exclusively with the latter. It should thus rightly belong to the Sociolinguistics category of models. However, this model introduces a further layer of description in the sense that speakers are no longer making binary choices (as spins in an Ising model, or a Potts model in presence of more than two variants), but are represented as reservoirs of different linguistic variants. Therefore, besides the community level where agents are interacting with each other, this model adds a further, cognitive level. This new level is the theater of another population dynamics, where all variants (species) are represented by a number of individuated occurrences (individuals). The different speakers are seen as a parallel of different islands in a population dynamics framework (Baxter et al., 2009).

As our model draws entirely from this picture of multiple individual occurrences of different species/linguistic forms competing over a cognitive territory, we felt obliged to present this model, and the several extensions it received over the years, with a particular focus. Furthermore, it may be the only model which specifically investigated the S-curve as a template for change (Blythe and Croft, 2012).

6.4.1 The original model

The model was first developed by Baxter et al. (2006). Here we shall present it in a simplified way, by assuming that there are only two competing linguistic variants, X and Y , so that the comparison with the previous models should be clearer. The notations have also been slightly altered.

Ingredients

The speech community is described by a network between nodes, where each node represent to a speaker. The link between speaker i and speaker j is associated with a weight G_{ij} . These weights are such that their total sum is equal to 1. Also, the influences between speakers are implemented through a matrix H whose coefficients h_{ij} denotes the influence of speaker j on speaker i . Those coefficients can be asymmetric (i.e. h_{ij} is not necessarily equal to h_{ji}).

Each speaker is described, at any time, by the frequency x_i of occurrences of X in its memory, called ‘grammar’ by Baxter et al. (2006). The frequency of the competing variant Y is therefore given by $1 - x_i$. The total frequency of variant X in the speech community can be described by:

$$x = \frac{1}{N} \sum_{i=1}^N x_i, \quad (6.25)$$

in which N would be the number of speakers, assumed to be constant. Therefore, the population frequency is the average frequency acknowledged by each speaker.

Also, those frequencies x_i , which can be interpreted as an occupation number of variant X in speaker i , must be distinguished from the probability p_i for speaker i to produce variant X . This accounts for production bias in favor of or against the variants. They are obtained from the frequencies through a linear combination of

frequencies of both variants:

$$\begin{pmatrix} p_i \\ 1 - p_i \end{pmatrix} = \begin{pmatrix} 1 - \alpha & \beta \\ \alpha & 1 - \beta \end{pmatrix} \begin{pmatrix} x_i \\ 1 - x_i \end{pmatrix}. \quad (6.26)$$

Thus, if $\beta > 0$, the probability to produce variant X can be non-zero, even if there is a zero frequency of X in the speaker's grammar. A cut-off must however be introduced so as to prevent the production probabilities p_i and $1 - p_i$ from leaving the $[0, 1]$ interval.

Interestingly, the authors stress that the frequencies x_i are not observable: what is observable are the numbers of occurrences produced in a given time window. This is entirely in accordance with diachronic corpus data, though they are more likely referring to sociolinguistics observations.

Dynamics and results

The dynamics follows three steps: first, one link of the network between agent i and agent j is chosen with probability G_{ij} . The two speakers enter a communication session: they both produce T utterances, each representing either variant X or Y . Speaker i produces n_i occurrences of X among the T and speaker j , n_j . Those two numbers are drawn from the respective binomial distributions $\mathcal{B}(p_i, T)$ and $\mathcal{B}(p_j, T)$. Finally, the frequency x_i of speaker i (and the same goes for the frequency x_j of speaker j) is updated according to:

$$x_i(t+1) = \frac{1}{1 + \lambda(1 + H_{ij})} \left[x_i(t) + \frac{\lambda}{T} (n_i + H_{ij}n_j) \right], \quad (6.27)$$

where λ is a parameter stating how much a single communication event weight in the speaker's representation of the frequency.

A possible approximation of the model is to focus on the cognitive level by ruling out all sociolinguistics details. This is equivalent to look at only one agent. In this case, a simple, deterministic approximation can be obtained, which is a balance of the inward and outward fluxes:

$$\dot{x} = \frac{\lambda}{1 + \lambda} [\beta(1 - x) - \alpha x] \quad (6.28)$$

The solution of this equation is an exponential which converges to $\beta/(\alpha + \beta)$. Recall that those two parameters, α and β , are the probabilities for an utterance to 'mutate', respectively from Y to X and from X To Y . Therefore, in this single speaker approximation, the new variant is always assured to rise, but can never evict the old one. Complete replacement is therefore, as in Reali and Griffiths (2010), an effect of finite size effects (manifested in drawing of T occurrences in a communicative event).

In Baxter et al. (2006), the complete dynamics of the model is thoroughly analyzed in a Fokker-Planck framework, in the context of a complete network (i.e. all speakers are related with each other) and symmetric, uniform influences, setting $h_{ij} = h$, $\forall(i, j)$ (all speakers are mutually influencing each other equally). Especially, the question of the fixation probability of either variant through finite size effects is extensively discussed. The deterministic approximation leads once more to a linear system, associated with exponential curves. Therefore, no S-curve is found at this point. Indeed, this

work addresses more the emergence of a linguistic consensus over a community (akin to the Naming Game objectives) than the possibility, for a new variant, to replace an established consensus.

6.4.2 Application to a linguistic case

Before turning to the S-curve problem, we shall first briefly present an application of the Utterance Selection Model to an actual linguistic case of change (Baxter et al., 2009). Few models that we have reported here have been tested on actual datasets, and those were macroscopic models, with a few parameters to be fit. It is therefore of particular interest that a model as rich and multifaceted as the Utterance Selection Model can be used to understand a specific language phenomenon.

The linguistic situation investigated in (Baxter et al., 2009) is the well-documented emergence of a New Zealand English dialect. In the second half of the XIXth, waves of immigrants came into New Zealand. They were speaking different varieties of English and a lot of linguistic variability was acknowledged, both from one speaker to another, and from one utterance to another for a given speaker. However, in fifty years, the language ‘focused’, and only a few variants were retained. Trudgill (2004) has argued that this focusing was the result of ‘accommodation’, i.e., speakers were accommodating to the variants most frequent around them. Therefore, the winning variants were the most frequent ones to begin with, and the formation of the dialect was, in that sense, ‘deterministic’. We note that this is a very specific phenomenon, not akin to semantic change and semantic replacement. It is the emerging of a consensus at the scale of the linguistic community. Semantic change, on the contrary, takes place in a setting where a full consensus has already been reached.

Baxter et al. (2009) propose to ‘evaluate’ the proposal of Trudgill in the framework of the Utterance Selection Model. They first recall that there are three ways through which the model can deviate from the ‘mechanical’ logic of Trudgill’s dialect emergence: first, the fact that each speaker has a limited number of interlocutors (the social network is not a complete graph and the G_{ij} weights can be equal to 0 if two speakers i and j do not interact); second, a speaker could be sensitive to the identity of its interlocutor (i.e. the coefficients h_{ij} of the influence matrix would not be uniform and can possibly be asymmetric, with $h_{ij} \neq h_{ji}$); third, there can be an intrinsically better variant, which will be reflected by a mutation rate in the utterance production step of the dynamics. Trudgill’s account would still be compatible with the first one of these mechanisms, and is chiefly opposed to the second.

What Baxter et al. (2009) show is that, in the simplest possible Utterance Selection Model, with no mutations and no polar influences between speakers, and independently of the network structure and the different parameters, the fixation probability of a variant is proportional to its initial frequency. Therefore, Trudgill’s theory is supported by the Utterance Selection Model. They also note that some of the exceptions to Trudgill ‘deterministic’ view (the initially most frequent variant wins), in which variants in slight minority achieved fixation, are also compatible with this neutral setting, which predicts that they had low, but still reasonable chances to occur.

However, the authors contest this apparent agreement by discussing the timescales at which the focusing of the New Zealand Dialect occurred. From the massive affluence

of immigrants to the dialect fixation, fifteen years elapsed. Assuming (an assumption grounded on solid arguments as well as some empirical support) that speakers produce and hear, in their lifetime, 13 M of utterances, then the cognitive time window of speakers (related to the λ parameter of the update rule (6.27)) must be of two days in order to get a mean fixation time of fifteen years. This cognitive time window is the time after which the reinforcement caused by an occurrence of the new variant fades away, so that the new variants must occur several times within that time window to make any progression in the speakers' grammar. According to psychological experiments, this value of two days is too little to be reasonable. On this ground, Baxter et al. (2009) reject Trudgill's deterministic account of dialect formation, since according to the Utterance Selection model, it would not be compatible with a focusing time of fifteen years. Thus, either some variants have an edge, or there are sociolinguistics influences factors at work in the process.

This result is of no relevance for our own concerns, yet the approach is of uttermost interest. Indeed, the authors have striven to substantiate the different parameters of the model, using results from multiple fields, so as to make the quantitative, analytical model, comparable with actual sociolinguistics data. This is truly remarkable and has remained an isolated feat in the literature. It shows that the mathematical models of language can help to aggregate knowledge from different research traditions in order to test the plausibility of a given mechanism as an explanation for a linguistic phenomenon.

6.4.3 The search for the S-curve

The paper that we now turn towards (Blythe and Croft, 2012) is of direct importance for the present work, as it discusses how the S-curve pattern can emerge from the Utterance Selection Model.

First, a slight modification is introduced. There are no longer mutations in producing; that is, we always have $p_i = x_i$: the probability for speaker i to utter variant X is equal to the frequency of X as represented in his memory. As before, two speakers i and j randomly selected will utter n_i and n_j occurrences of X respectively, among T uttered occurrences. However, the update rule (6.27) becomes more general:

$$x_i(t + dt) = \frac{1}{1 + \lambda} [x_i(t) + \lambda \tilde{x}_i(t)] , \quad (6.29)$$

where \tilde{x}_i is a 'perceived' frequency of variant X on the basis of the exchange, and is computed from both n_i and n_j . In the previous model we had:

$$\tilde{x}_i = \frac{n_i}{T} + h_{ij} \frac{n_j}{T} , \quad (6.30)$$

h_{ij} being the influence of speaker j on speaker i . Now we have:

$$\tilde{x}_i = h_{ii} f \left(\frac{n_i}{T} \right) + h_{ij} f \left(\frac{n_j}{T} \right) . \quad (6.31)$$

The perceived frequency is still a linear combination of the speaker contribution and its interlocutor one, but this contribution is no longer bound to be the produced frequency

of variant X . Also, the diagonal elements of the influence matrix H can be different from 1. To obtain an S-curve, it is possible either to change the function f which relates the produced frequency to the perceived one, or to try a specific influence matrix.

As the authors now investigate the S-curve, the initial situation is such that one variant (Y) is established in a population of $N = 80$ speakers. However, as there are no longer mutations, the variant X must be initially present in the system, with an initial frequency $x(0)$ (apparently the same for all speakers), where x is the average frequency of variant X over all speakers ($x = (\sum_i x_i) / N$). Furthermore, the network is a random one, where all agents have an equal connectivity of 8. All links are chosen with equal probability $G = 1/320$.

In the simplest case (uniform symmetric influences and $f(u) = u$), which the two previous works already treated, the probability of fixation of new variant X is equal to its initial frequency, and the trajectory, in a fixation case, is not an S-curve, but a noisy, diffusive random walk. To correct this picture of semantic replacement, the authors consider then three possible variations to the initial model.

Non-linear perceived frequency

The first one is to assume a new form for f , which is now:

$$f(u) = u + au(1 - u)(2u - 1) \quad (6.32)$$

In the single speaker, deterministic case, we would have $h_{ij} = 0$ and $n_i/T = x_i$, leading to:

$$\dot{x} = ax(1 - x)(2x - 1), \quad (6.33)$$

That is, provided $a > 0$, we obtain two stable fixed points for $x = 0$ and $x = 1$ and an unstable one for $x = 1/2$ (in the usual logistic picture we get only two fixed points, 0 and 1, the former being unstable). This means that the system will favor fixation of the variant whose initial frequency is the highest. If a is negative, then the system leads to mixing and stable coexistence of the variants.

Therefore no S-curve is obtained. One can observe something close to an S-curve if the new variant starts with an initial frequency greater than 1/2, but this would not describe replacement, only eviction of a weaker competitor.

Leaders and followers

Another ingredient would be the splitting of the society between influential leaders and influenced followers, much alike the partition found in Nettle (1999). The influence of a leader on a follower is α -times bigger than the reverse. The model investigates the case where all leaders start with the new variant, and all followers with the old one. α is chosen equal to 20 and the number n of leaders to 5. Then, the frequency of the new variant will shortly reach a characteristic frequency, from which the fluctuation can make the speech community deviate to fixation of either one variant or the other. Therefore, the situation is the same as in the initial Utterance Selection Model, only that the probability of fixation, given by the characteristic frequency which almost acts as a new initial frequency, is now heavily boosted, equal to:

$$x^* = \frac{\alpha n}{\alpha n + N - n}, \quad (6.34)$$

Then the fixation probability is greater than 1/2 if:

$$n > \frac{N}{\alpha - 1}, \quad (6.35)$$

which is indeed the case for the specific choice of parameters. Still no S-curve, then.

To obtain an S-curve, another influence structure is considered. It is now concentric. The community is divided into groups of growing sizes, and each group influences the next bigger one, but does not influence the next smaller one (nor the bigger groups beyond the immediate next one). Therefore, there is a pyramidal hierarchy, where the influence flows from the top to the base. In this ‘Illuminati’ model, if the speaker at the top of the pyramid adopts the new variant, then it can spread downwards the scale over the whole society. In this particular case, the frequency rise of the new variant is a clean and neat S-curve, even in individual cases. One can also consider an ‘ovoid’ layered society, with few people at the top, a lot in the middle, and few again at the bottom, with the same idea that a layer is influenced by the one immediately above it and influences the one immediately below. Then again, an S-curve is obtained. If, however, there is one hyper-influential individual at the top, and each societal layer contains the same number of individuals, then the frequency rises more exponentially-like, not S-like. Also, a reduced asymmetry (i.e. the members of one layer exert a low influence on the layer above them) breaks down the S-curve pattern of frequency rise.

These results show that the S-curve is far from granted, and may depend on a very specific structure to arise. It would have been interesting to consider the case of a ‘failed’ innovation, where the new variant arises among a group which is not at the top of the hierarchy. Then, the innovation would be corrected from above, but would have had time to spread below. Quickly, only the lowest layers would still speak the novelty, which will be slowly cleansed from top to bottom. However, the authors do not report any results on such a possibility.

Stronger variants

Finally the case where the novel variant is favored by an intrinsic bias is investigated. In this case, $f(u) = (1 + b)u$, with a cut-off at 1 for $u > 1/(1 + b)$ and b a small, positive parameter, which quantifies the bias. Thus, the new variant is always perceived as more frequent than it really is. Those biases are interpreted as the social weight associated with the early adopters of the new variant (which are responsible for the initial frequency of the form). The prestige of the initial adopters of the variant will therefore be attached to it throughout its diffusion over society. Of course, such a bias can receive many different interpretations, and the result does not hinge upon this particular sociolinguistics association.

As it happens, a very small bias ($b = 0.01$) suffices for the new variant to take over and becomes fixed in the population. Furthermore, the associated frequency rise is S-shaped, so that the desired pattern is recovered. One can try to get a quick idea of how the S-curve pattern would appear, with a simple deterministic approximation ($T \rightarrow \infty$). With a proper time rescaling, we would have, without any approximation:

$$\dot{x} = f(u) - x \quad (6.36)$$

Therefore, if $T \rightarrow \infty$, $u = x$ and we get:

$$\begin{cases} \dot{x} = bx, & \text{for } 0 < x < \frac{1}{1+b} \\ \dot{x} = 1 - x, & \text{for } \frac{1}{1+b} < x < 1 \end{cases} . \quad (6.37)$$

The solution of this equation will be given by:

$$\begin{cases} x(t) = x_0 e^{bt}, & \text{for } 0 < t < t^* \\ x(t) = 1 - \frac{b}{1+b} e^{-(t-t^*)}, & \text{for } t > t^* \end{cases} , \quad (6.38)$$

with:

$$t^* = \frac{1}{b} \log \left(\frac{1}{x_0(1+b)} \right) . \quad (6.39)$$

This matching of these two exponentials indeed produces an S-curve, though an asymmetric one. Its logit transform should therefore not be a straight line. It would be interesting to check in empirical data if the S-curves obtained for actual changes are symmetric or not with respect to the inflexion point. In the former case, the present account of the S-curve should be ruled out as non-conforming to the phenomenology of language change.

This asymmetry effect may be tampered for low T (the number of occurrences produced during an exchange). If $T = 1$, the new variant does not propagate on average; yet if $T = 2$, $f(u) = 1$ with probability x^2 , $f(u) = (1+b)/2$ with probability $2x(1-x)$ and $f(u) = 0$ with probability $(1-x)^2$, which leads on average to:

$$\dot{x} = bx(1-x), \quad (6.40)$$

which is the usual logistic equation, giving rise to a symmetric S-curve. With $T = 3$, we get:

$$\dot{x} = bx(1-x)(1+x), \quad (6.41)$$

with an inflexion point for $x = 1/\sqrt{3} \approx 0.58$.

For $T > 2$, the inflexion point occurs when the new variant has reached a frequency greater than one half. This property could be checked for actual S-curves. In this case, the Utterance Selection model would be a great tool to relate empirical features to model mechanisms. Combining the slope and the frequency of the inflexion point, it could be feasible to get an estimate of some of the parameters involved in the model (T , b , x_0).

Interestingly, the obtained S-curve also holds when some fraction of speakers (up to one third) are negatively disposed to the novelty, which is translated by the following, speaker-specific function:

$$f(u) = (1+b)u - p \quad (6.42)$$

with the appropriate cut-offs.

The S-curve can be thus captured relatively simply by introducing a small bias in favor of the novel variant. As this result also holds in a deterministic approximation independent on any details of the network, the Utterance Selection Model, with biases,

is an efficient way to generate the S-curve pattern of semantic replacement. Otherwise, if we want to consider a situation in which the variants are neutral, only a specific social structure of influences is compatible with the S-shaped frequency rise. The ultimate conclusion of Blythe and Croft (2012) is that neutral evolution is not compatible with an S-curve, and that language must either be speaker-biased or variant-biased for its change to follow the expected S-pattern.

6.4.4 Another way to the S-curve

Building up on these results, Stadler et al. (2016) have tried to give a more satisfying account of these variant-based biases, necessary to recover the S-curve in the Utterance Selection Model. The main idea is that these biases are not a priori and intrinsic, but come from a conformity to what is trending rather than what is frequent.

The set-up is as before; there are two variants, X and Y ; N speakers, fully connected. The influence matrix H is a bit different in the sense that we have now $h_{ij} = 1$ if $i \neq j$ and zeros on the diagonal (speaker i bases the update on the sole basis of the output n_j of its interlocutor j). The dynamics is unchanged: a pair of speakers is chosen randomly; they both utter T exemplars, choosing either variant X and Y with Bernoulli trials of parameter x_i and x_j , for speaker i and speaker j respectively; then they both update their frequency using the following rule:

$$x_i(t+1) = (1 - \alpha)x_i + \alpha f_i(n_j/T), \quad (6.43)$$

where α is a parameter which plays the same role as the previous λ (with a slight alteration in its definition). In the following we shall drop the i and j indices, keeping in mind that the argument u of function f is given by the interlocutor output n_j while everything else refer to speaker i .

Two crucial mechanisms are now added. First, the speaker does not only track the frequency x_α , where the α index indicates the associated learning rate, but also the short-term frequency of the term, x_γ , whose update rule follows similarly:

$$x_\gamma(t+1) = (1 - \gamma)x_\gamma + \gamma f(u). \quad (6.44)$$

As γ is defined to be greater than α , this short-term frequency associates a higher weight to the most recent events, and the memory of the older ones fades away more quickly, with a rate $\gamma > \alpha$. Therefore, the frequency x_α has more inertia whereas the frequency x_γ is dominated by recent exchanges.

Keeping track of this short-term frequency is actually used to compute the bias which comes into the perceived frequency f so as to alter the received output u :

$$f(u) = u + b \frac{x_\gamma(t) - x_\alpha(t)}{m(\alpha, \gamma)}, \quad (6.45)$$

where b is the strength of the bias, independent of the linguistic variant. Indeed, the bias is not variant specific, but depends on the recent history of the system: it is given by the ‘momentum’ of variant X , that is, the difference between its short-term frequency and its long-term one. If the variant appears all fresh and new (i.e. it has been more used recently than it has been in the further past), then it will be associated

with a high momentum and be associated with a higher bias. Finally, the quantity $m(\alpha, \gamma)$ is introduced for normalization purposes and is computed as the maximal possible momentum:

$$m(\alpha, \gamma) = \left(\frac{\gamma}{\alpha}\right)^{\frac{\gamma}{\gamma-\alpha}} - \left(\frac{\gamma}{\alpha}\right)^{\frac{\alpha}{\gamma-\alpha}} \quad (6.46)$$

This momentum-based bias allows to obtain an S-curve growth of the frequency for the rising variant. This is particularly interesting, since the dynamics never favor it explicitly, only as a result from its momentum. Indeed, the old established frequent, at the beginning of the process, is as frequent in the short-term as it is in the long-term. But the new variant, as soon as it comes to be produced, gains momentum, since its short-term frequency will rise more than its long-term one, which has more inertia. Therefore, the bias in favor of the novelty will come to favor its entrenchment by boosting its perceived frequency, leading to a reinforcement loop between momentum and produced frequency. Eventually, the momentum will wear off (as the new variant becomes almost equally frequent in the short-term and the long-term), and the growth will slow down, the bias becoming weaker and weaker. This latter waning of the growth accounts for the upper part of the S-curve. A slow beginning, a fast rise, a slowing down at the end: here is the S-shaped pattern, nicely explained.

This model also accounts for rare aborted changes, where the frequency of the new variant rises, and then falls down at mid-point due to an early loss of momentum, never reaching completion. It is unclear whether this is happening in actual language changes, but it is nonetheless an interesting prediction, worth investigating. It also shows that the S-curve is not inevitable, and that the phenomenology of change can occasionally bring about an unexpected turn of events.

The fact that the bias favoring the new variant is not posited a priori, but results from the dynamics itself, is a very interesting idea; that it proves to efficiently capture the main phenomenology of language change is all the more satisfying. And yet, we would have two concerns regarding these results. The first one is that, even if the slow-fast-slow pace of the S-curve is accounted for, the obtained S-shape is less pronounced and less clear than what we have found in actual linguistic data. The other concern is more serious. Indeed, this simulation relies on curious initial conditions, in which each speaker is provided with a non-zero initial frequency of the new variant, so that, no matter the chosen pair, each agent has a non-zero probability to hear the new variant from its interlocutor. There is no serious nor specific objection against this initial state of things, yet if the aim of an agent-based model is to account for the synchronization of speakers as the change moves further on, it is a bit strange to postulate that they have initially agreed on the initial frequency of a variant which has not yet been produced. Even though the authors state that their work does not deal with the emergence of the new variant as a linguistic receivable alternative, the emergence of a community-scale initial consensus regarding this variant should fall within their concerns and should receive a due explanation in the framework of the model. To summarize, if there is a mechanism powerful enough to impose, even residually, the new variant to all speakers, then this mechanism might also have enough strength to handle as well the remaining of the process. This is actually the line we will follow in our own modeling attempt to catch the S-curve. It would, however, be interesting to see if different initial conditions (e.g. one speaker only starts with the new variant) can lead to the same results.

6.5 Another model of semantic change

To complement this short survey of the attempts to model the S-curve in a situation of linguistic replacement, we include the work of Victorri (2004), which addresses neither replacement nor the S-curve. It nonetheless presents an interesting, dynamical, out-of-equilibrium picture of semantic change, which is analytically simple, phenomenologically rich, and theoretically innovative.

The model builds up from the Saussurian realization that both sounds and meanings are arbitrary closed domains in the otherwise indistinct continua that are the phonemic space and the semantic space, respectively (de Saussure, 1995). Therefore, both are represented as a continuous infinite line. Then, a phoneme, a meaning, act as probability distributions on these spaces. That is, the lexical unit u_i is not a closed region of the semantic continuum, but rather a density on this continuum, centered around a prototype. Though the model allows to represent signs as correspondences between a phonemic distribution and a semantic one, we will discard all the phonetic part and focus on meaning only. Indeed, it does not change the main dynamics, but it leads to a simpler picture.

6.5.1 Communication event

Let us consider that each lexical unit u_i (which would be a linguistic form) is represented by a Gaussian (μ_i, σ_i) . The mean μ_i would be the prototypical meaning of lexical unit u_i and the variance σ_i how semantically acute and specific this lexical meaning is (the greater the variance, the more broad and general the meaning of u_i). When the speaker wants to express a meaning m , it does not have a clear, sharp, infinitely precise idea of the meaning. Therefore, we can represent the associative cognitive representation of this intended meaning by another Gaussian (m, σ) where σ indicates how much precise is the intention of the speaker. We can consider that there is a lower bound for σ , which would be a cognitive limitation, but there is probably no upper bound: the speaker does not always aim to be specific.

To communicate this meaning, the speaker has to choose a linguistic form. This choice is made by computing how much the different lexical units distribution overlap with the Gaussian intended meaning. The probability to choose a lexical unit is then directly proportional to this overlap. Once a lexical unit has been selected, it is transmitted to the hearer. The ‘phonemic’ part of the model would then introduce here the possibility that a different lexical unit is recovered, but we assume here that this never happens. From the distribution associated with the received lexical unit, the hearer selects a meaning m' , which is, once more, associated with a Gaussian (m', σ') .

It is unclear whether the σ can vary or not from one utterance situation to another, and how the σ' is actually recovered. If we associate σ to a general, constant cognitive limitation, then we would have $\sigma = \sigma'$, and a σ independent of the utterance situation. We can also consider that σ' can be computed from the variance σ_i of the received lexical unit. If the lexical unit has a very broad and general meaning, then the hearer will not consider that the speaker’s intention was very precise either. Then we can assume, for instance, that $\sigma' = \max(\sigma_b, \lambda\sigma_i)$, where σ_b would be a cognitive limitation (the minimal semantic precision achievable by the human mind), and λ a constant

parameter, smaller than 1, which conveys the idea that in the context of an utterance, the recovered meaning is more precise than the general, isolated meaning of the lexical form. All this, however, is not discussed by Victorri (2004).

6.5.2 Semantic change of a lexical unit

So far we have described how a basic situation of language use could be represented in the model's framework. Now we must look at how language change, and more specifically in this situation, semantic change, can happen through language use. After the communication event, the speaker and the hearer compare their meanings and compute the overlap θ between their two distributions. This is a very unrealistic feature, and it leads one to wonder what is the point of communicating if meanings can be directly compared, but it is nonetheless an extremely widespread and common mechanism in modeling, for instance in a Naming Game simulation (Baronchelli et al., 2008). The lexical unit u_i will be therefore modified to account for the fact that it conveyed meaning s , and it will be modified all the more that the overlap between the two meanings (intended and inferred) is great. Therefore:

$$u_i \leftarrow \frac{u_i + \theta \mathcal{N}(m, \sigma)}{1 + \theta}, \quad (6.47)$$

where $\mathcal{N}(m, \sigma)$ is the Gaussian distribution associated with the intended meaning.

In the original paper, it is only said that the original lexical distribution is replaced by a weighted sum of itself and the intended meaning sum distribution, the latter being all the more important in case of a success. Here we have proposed the weight to be proportional to the overlap. Note that the resultant distribution is not necessarily a Gaussian, so that the semantic distributions associated with lexical units will not remain Gaussian after the initial step.

Unfortunately, no further analysis of the model is provided, for it was chiefly designed as a conceptual picture of the Saussurean semiotics. Only a few cases are qualitatively discussed; for instance, how can the simulation bring up a split of the meaning of a lexical unit, so that it comes to cover two separate meaning regions of the semantic continuum — a phenomenon that we refer to as a ‘speciation’ event. This model could have been analyzed further, yet it completely escaped notice. The fact that it has been written in French, in a little known overspecialized journal (*Cahiers de praxématique*) devoted to a rather endemic linguistic theory (‘la praxématique’), certainly did not help this work to achieve recognition.

6.5.3 Semantic replacement

This model is however very general, and much less specific to a given framework than many other models we reviewed so far. Furthermore, we believe it could be applied to describe semantic replacement, given the appropriate setting. In the following, we will remain on a qualitative level, though numerical simulations could be easily performed.

If we consider only one linguistic form, then its associative meaning distribution will spread over the semantic continuum in a diffusive way until we get an ultimately flat, uniform distribution. Yet, semantic change takes place in a fully formed language,

embracing very many entrenched lexical units, each one associated with their own semantic distribution. In this model, we are forced to think of the semantic expansion as happening within an organized, structured system, not as an isolated event.

Let us imagine that, at some point in a language, a lexical unit u is associated with a semantic distribution. To the left of this unit lies other lexical units, characterized by clear, sharp distributions. The distribution of the form cannot really expand this way, for any intended meaning falling into that zone will have a very high overlap with the sharp and precise distributions which cover this area. Now, to the right of the unit u lies another one, v , which has cleared its domain from the interferences of the other lexical units. Yet, this domain is large, and the distribution is widespread.

We then assume that the unit u has bleached a little bit, so that its distribution starts to leak out in the big dominion of the v unit. When the speaker wants to express a meaning in this leaking part, there is still a bigger probability to choose the established form. Yet, as this form is semantically bleached, i.e. it corresponds to a flattened distribution, there is a good chance that the hearer will not select the rightful form, so that the overlap between the intended and the inferred meaning will be small. As this overlap will be small, so will be the reinforcement. Henceforth, the distribution of the widespread form will change only very weakly through this particular event.

Now, there is also a small probability to choose the first semantic form. As this semantic form has a much more tight distribution, the inferred meaning will be closer to the intended one, and so the overlap will be bigger, and the reinforcement, stronger. This means that the foothold of the new variant has assured will become a little bit firmer. As time goes by, this will be repeated. The probability to choose the new variant will grow as the distribution shifts further within the dominion of the established variant. Yet the reinforcement will also become smaller. If we consider one point of the semantic continuum belonging to the dominion of the old variant, and we consider the relative height of the two semantic distributions — that of the old and of the new variant —, is there a possibility to observe an S-curve? Only numerical simulations can tell. In the scenario we have suggested, there is at least a possibility of a replacement, whose cause would be the bleaching of a lexical unit which semantic scope has become too large and too imprecise.

Maybe we can go a little bit further, with particular reference to the grammaticalization and its well-known unidirectionality. It may be that the level of precision associated with the lexical realm is greater than the one usually attained in the grammatical one. This may be because we do not need that precision; e.g., when we want to say that something happened in the past, we do not need to be much precise usually. For instance vague temporal indications can be very widespread, such as French *l'autre jour* ('the other day'), which is used to set the narrated scene in a not very distant past, or *tout à l'heure*, which roughly means 'at some other moment of the day' which can be either past or future. It may also be because, cognitively, grammatical and abstract meanings are harder to focus on cognitively, which would explain why there are in fact two unidirectionalitys in language change, which overlap but do not equate: from concrete to abstract, and from lexical to grammatical. Anyway, we can associate with grammatical intended meanings a greater variance than it would be the case for lexical, or concrete intended meanings. Therefore, the associated distribution will have a longer tail, and will be sensitive to a wider array of lexical units. A neighboring,

unusual lexical unit would consequently have greater chances to be chosen and uttered for a grammatical meaning, than for a lexical meaning.

The picture put forward by Victorri (2004) is thus both elegant and powerful. Properly adapted, it could efficiently relate bleaching and meaning change, and provide original explanations for the unidirectionality in semantic change, rooted in the most basic features of language use. We will not explore it further in this work, but it is nonetheless a useful conceptual sandbox to toy a little bit with the different ideas on semantic change and grammaticalization.

6.6 Summary

We reviewed here different models of language change relevant to understand, describe and explain the linguistic phenomenon of replacement and its associated S-curve signature. It turned out that the S-curve, though extremely easy to obtain through a macroscopic logistic equation ($\dot{x} = \alpha x(1 - x)$), does not arise trivially from more detailed models of language. Especially, relating the microscopic dynamics of language use and the macroscopic pattern of the S-curve has proven to be a subtle task. Indeed, only three of these usage-based models lead to an S-curve: the model of Ke et al. (2008), the Utterance Selection Model with intrinsic biases (Blythe and Croft, 2012), and the Utterance Selection Model with momentum selection (Stadler et al., 2016).

The first one requires two ingredients: a huge bias in favor of the new variant, and a social network with long-range ties. The Utterance Selection Model needs either a highly specific influence network, or a small intrinsic bias associated with the new variant. Finally, the Utterance Selection Model with momentum selection relies only on the specificity of its dynamics, the bias being equated to the momentum gathered by the new variant.

The models also deal differently with the Threshold Problem, which is a necessity, for the new variant to be propagated, to first reach a minimal frequency. The macroscopic model of Ghanbarnejad et al. (2014) shows that a barely noticeable push-up in favor of the new variant is enough to start off the process. In Nettle (1999), which specifically addresses this issue, the threshold is either overcome by an intrinsic bias, or by the presence of hyper-influential individuals, whose adoption of the variant automatically raises its frequency above the threshold. A proposal by Wang et al. (2004) would be that the new variant first reaches some non-zero frequency in one lexical context, which will then come to feed related contexts. This does not entirely solve the issue, but offers a convenient way to lower the threshold. The Utterance Selection Model, in some of its variants, makes room for mutations through which the new variant can start to appear (Baxter et al., 2006). Yet, none of these proposals could account for the latency phenomenon, according to which the frequency of the new variant stabilizes first at a very low level before the S-curve can take place.

All these models — except the macroscopic ones — are furthermore stochastic at some level, which is a good thing, since we have seen in chapter 4 that different empirical observables of actual linguistic changes (the logit slope, the growth time, the latency time) were obeying statistical distributions. However, none of these models actually describe which distribution is expected for the two main quantities related to

the S-curve, the growth time and the slope. Moreover, none of these models, save the macroscopic ones, try to relate the analytical prediction with actual quantitative data, with the outstanding exception of Baxter et al. (2009).

There is thus room left for a new attempt to tackle the modeling of the linguistic phenomenology associated with the replacement of an entrenched convention, and more specifically semantic replacement. As suggested before, our idea is that the mechanism causing the new variant to appear in the first place might also be responsible for the entirety of the process. Therefore, it seems important to deal with the ‘actuation problem’, i.e. how a new variant comes to appear within the pool of linguistic possibilities, as it is part of the whole process of change. This process has most often be relegated outside the scope of the models, e.g.:

That is, we simply state that an innovation has reached some (usually small) frequency by some time $t = 0$, but do not explicitly model the process by which this occurs. This is reasonable for studying the propagation of a new convention that has already been adopted by some minority group within the community.
(Blythe and Croft, 2012, p.285)

But if there is a mechanism which can be held responsible for the adoption of a variant by a minority group, why could it not account as well for the subsequent propagation?

It is also interesting that most of these models ascribe the change to sociolinguistic mechanisms. Admittedly, the mechanisms displayed in those models are not bound to a single and particular interpretation, yet the interpretation put forward falls most often within the Sociolinguistics framework. There are three exceptions to this: the two learning-based models, in which change arises through the transmission of language across generations, and the model of Victorri (2004), where change is a structural effect of the semantic organization itself. This sociolinguistic tropism of language change modeling is nonetheless a bit surprising, since most documented changes, and most attested S-curves, come from diachronic corpus data. Therefore, it does not compare well with models whose account of change builds on intrinsic preferences for the novelty, or specific social structure. We shall therefore, in our modeling attempt, favor a structural, language internal vision of language change, which shall nonetheless remain usage-based and cognitively grounded, as detailed at length in the next chapter.

Before turning to the presentation of the linguistic and cognitive framework of our model, we shall first briefly mention the modeling components that our model has in common with those of this survey. First of all, the idea that occurrences of different species compete over a cognitive territory, as put forward by the Utterance Selection Model (Baxter et al., 2006), is at the very core of our modeling attempt. The conception that language changes because of its internal, slippery structure, that we find in (Victorri, 2004), is also in line with our own, as well as the lexical diffusion idea, according to which what is happening in one lexical context can influence other contexts (Wang et al., 2004). Finally, macroscopic models, for they offer a clear guideline and are easily compared with each other, offer a previous ground of discussion and understanding. For this reason, we shall first provide a deterministic, equation-based approximation of the model, which should already display the main phenomenological features of semantic replacement. Then, a stochastic version of the model will be dealt with, as the stochastic nature is required to obtain statistical distributions akin to the ones resulting from our empirical analysis of semantic change.

Chapter 7

Hypotheses of the model

Before turning to the presentation of our own model to describe and to understand language change, we would like to set the key concepts which underline the design of our model. Indeed, a model is not only a descriptive tool, and should also serve as a guide to analyze the data by setting up a narrative in which the empirical facts make sense. In that perspective, a very special attention will be devoted to the interpretation of the different terms and mechanisms of our model. This is why I will first list the main hypotheses behind our model, so that the analogical matching between its mathematical terms and the process of language change may become entirely clear.

7.1 Language change as a product of language use

We first consider that language change is not a phenomenon separated from language use; the two are one and the same. In this view, language change is not a perturbation of language use, but a mere corollary. This does not mean that language change cannot happen through external influences, which are pretty much independent of the ordinary, daily use of language; it only implies that our model does not address such changes. As Ghanbarnejad et al. (2014) showed, it is possible to distinguish between externally-driven instances of change and internally-driven ones (even if external and internal may have different meanings in our own approach and theirs). As their study shows, most changes can be seen as internally-driven (i.e. resulting from the social imitation of peers). The only clear example of an externally-driven change is a consequence of an orthographic reform. It shows that this kind of language changes can occur, but it also allows to put them aside of our own concerns as most changes can be considered spontaneous and independent of specific external circumstances.

To state that language change is nothing but a mere consequence of language use, might seem a bit too strong a claim. Yet, we have good reasons to believe so. One is that there is no moment in time when the language is not changing. Of course, deep structural changes, as the rise of a new paradigm of determiners in French (Marchello-Nizia, 2006), only occurs from now and then; and this example shows that a given change can stir up new ones. Yet, even during hypothetical stable periods, we can always find some functional change going on. If language change were separated from language use, one would then have to explain why this is so, and how some external,

exceptional circumstances are always and constantly acting on language while not being part of it. This is far from impossible, yet it seems very unlikely.

There are, however, two external perturbations which would be good candidates to explain this constant messing with the course of language. The first one is the speaking apparatus; people are prone to make errors whenever they speak. Syntactic, semantic, phonetic errors are highly frequent in speech. Yet this is as much part of language use as anything else and cannot be rightly counted as external; we will henceforth deal later with this hypothesis. The second one is more crucial, and has been invoked quite often to account for language change: It is the passing of generations, and the learning of language by the new ones. As much as there always is some change going on in the language, there always is some child learning it. Learning is not an error-proof process, to be sure, and so it can stand as a perturbation of language use, a perturbation which can be assumed to be responsible for language change.

There has in fact been a long-lasting debate between the tenants of this hypothesis and its opponents; a debate which has not been settled to this day. It still survives, though it has been crippled by numerous arguments (Croft, 2000): The kind of language errors made by children do not correspond to usual language changes; children are compelled to conform to a sociolinguistic norm and themselves have no social influence, so they cannot change the norm; there may be no 'critical age' beyond which language ceases to be learned, so that learning language would be a part of using it; errors are individual, and they could not affect language as a social convention unless children are coordinating their language and adopting the innovations made by their fellow comrades. This explanation would then fall into the category of sociolinguistics, which is also a theory of language use: different social weights of the linguistic variants leads to different probabilities of uttering them. That is, social influences or linguistic identification to a social group are taken into account in the process of speech production, so that it is indeed through language use that sociolinguistic factors induce language change.

Another weakness of enthroning imperfect learning as the core mechanism of language change is that it would not explain why a change can span several centuries, unless there were some gradual sequence in the imperfection of learning. It would thus assume that there exists some learning bias which lasts over generations and drives the language in some new direction, inch by inch. Consequently, the learning bias would be more faithfully transmitted than language itself. This is reasonable in a Generative Grammar framework, as children are born with a language learning apparatus rather than a language proper. Genetic mutations affecting the language learning apparatus could then create new biases in the learning process, and so lead to new language changes across the ages. However, this would suppose a correlate with language changes and genetic lineage, a correlate which has not been supported by empirical evidence so far.

For all these reasons, it seems legitimate to posit that language change is rooted in language use. In physical terms, we would say that language is a far-from-equilibrium system. This is a powerful idea: it states that language is the main and most important agent of its own gradual transformation. Whenever language is used, it is as well perpetuated and altered; there is no way to speak in a way which would preserve language in the state it is. Language is perpetually changing, yet slowly, and most

often locally. Why it is so is another matter. There are several ways by which language could change through language use. Sociolinguistics account of language change, as we saw, is one of them. When speaking, individuals try to conform to the way some other individuals are speaking, because they want to show that they belong to the same community, or because this community has some higher prestige than their own. The relationships between communities are thus expected to shape the way language is changing. Though this is a serious possibility, which has yielded undebatable results, I prefer to favor another picture of language change, that appears more fitting to address specific linguistic problems.

Indeed, in a sociolinguistic account, all changes are pretty much equivalent. It is not the change which matters, but the communities in which it occurs, and the social ties it strengthens or weakens. If a community chooses to adopt a new variant *B* instead of an old one *A*, then it could as well have adopted another new variant *C*. In that sense, language change is completely arbitrary from the point of the view of the language as an organized structure and a coherent system of signs and meanings. The phenomenon of grammaticalization would thus be relegated entirely out of the study of language change; it is no more relevant than a phonetic change, and a semantic shift from grammar to lexis would be as likely and as unimportant than a shift from lexis to grammar. Language change would be no different than a change of clothing, for instance. The only thing that matters is how variants are ‘colored’ in a social, community sense.

The influence of the speech apparatus has also been discussed, especially through the phenomenon of phonetic reduction (Geurts, 2000). The more frequently linguistic forms are uttered, the more they wear off, until they get unrecognizable. At this point it becomes necessary to introduce a new, fresher variant, which still has a decent amount of phonetic substance. Yet this explanation does not work every time. For instance, the word *quoi*, extremely frequent, phonetically reduced (it is a monosyllabic word), has come to be used, by the middle of the XIXth century, as an extracausal pragmatic marker, used at the end of an utterance:

GERPIVRAC [...] Voilà le cas que vous faites de mon bouquet.

MADemoiselle EURYDICE Eh bien, il est tombé, quoi.

GERPIVRAC: This is all the care you treat my bouquet with.

MISS EURYDICE: Well, it’s fallen, that’s it.

HUGO, Victor, *L’Intervention*, 1866 (Frantext)

In this case, *quoi*, while already phonetically much reduced, has acquired a complex pragmatic meaning, whose use is pervasive in Present Day Parisian colloquial speech. Thus if phonetically reduced forms can semantically expand, there is need for a complementary mechanism of change.

As for pronunciation errors (or spontaneous and unconscious innovations), the same remarks hold for them as in the case of learning errors. An error cannot change the language, unless it is transmitted and repeated by other members of the speech community. Furthermore, pronunciation errors are hardly able to account for semantic expansion, in which there is very seldom attested phonetic changes of the form under-

going the expansion. Ease of pronunciation can probably play a role; but this role is, most certainly, a limited and occasional one.

This is why there is some room left for another explanation of language change through language use.

7.2 Language as memory

As has been pointed out in chapter 3, the S-curve manifests a phenomenon of saturation, which has been explained in three ways. First, the saturation is caused because there is a limited amount of speakers: once the new variant is used by the whole community, it reaches a point beyond which it cannot spread more, hence the saturation (Osgood and Sebeok, 1954). One cannot deny this; but if evidence was provided for the language change within a single individual, and if the course of change was displaying an S-curve as well, then one could conclude that this evidence should have another explanation. Disregarding this particular assumption, it is nonetheless established that individuals alternate between variants at a given point in time, and that this probability distribution for a given individual matches those of the people surrounding him (Labov, 1994; Denison, 2003). This would mean that the competition between the two variants is slower than the social coordination over the community, as was assumed in the wave-based picture of Bailey (1973). Hence, the S-curve we observe in time is more likely to represent this internal competition than the social diffusion of the current state of the competition. In that sense, the saturation mechanism is best not be ascribed to the finite number of speakers.

Second, from a lexical diffusion perspective (Wang, 1969; Chen, 1972; Wang and Cheng, 1977), saturation occurs because there is a limited number of compatible linguistic contexts in which the new variant can be used. The compatibility seems to be set a priori, which makes sense in a sound change (very roughly, if sound a is replaced by sound b , then all words with sound a constitute a compatible context for sound b), much less so in cases of grammaticalization where the compatible contexts co-evolve with the meaning of the form. Also, it would not explain why an S-curve is observed within individual contexts as well, as exemplified in Kroch (1989b).

Third, there can be a saturation because change is but a competition between an old and a new variant for each environment, or context. Saturation, in a given environment, is then achieved when the old variant dies out, though this position has often been left implicit (Kroch, 1989a; Denison, 2003). It entails that those linguistic environments must have some limiting capacity, and that they can carry a varying number of instances of the same variant — they are not binary occupied by one or another.

This can be interpreted straightforwardly in terms of an organized memory. Each compartment C of the memory can store a given number $M(C)$ of exemplars. If two variants are competing for this compartment, then the exemplars are of two different types. A variant of this view, in a more neural perspective, would be to consider that variants correspond to attractors that the memory can retrieve, through a Hopfield-like mechanism (Hopfield, 1982; Ruppin and Yeshurun, 1991). A linguistic form (output of the retrieval task from the neural memory network) can only be associated with a

limited number of inputs (stimuli of the memory). The saturation would come from the fact that, if all inputs come to be associated with the same output, we would always utter the exact same linguistic form, no matter the circumstances, so that saturation could result from the necessity to keep numerous and separate memories.

It should be noted that the saturation is also a mnemonic phenomenon in the Utterance Selection Model (Baxter et al., 2006). Indeed, even though there is a limited number of speakers, what causes the frequency to saturate is the entrenchment of the new variant at the cognitive level. The rise of the average frequency of the new variant over the whole community not only reflects the propagation from one speaker to another, but also the fact that the variant progresses in all speakers' memories.

The saturation observed on the S-curve would then be a mere consequence of the fact that language does not exist per se, but only through speakers who carry, in their minds, the records of their past linguistic experiences.

7.3 The semantic territory

This immediately raises the question of the nature of those compartments. They should reflect, at the same time, the linguistic environments for which different forms can compete, and the cognitive organization of language memories by the speakers. They must fulfill the two very different requirements of empirical applicability to diachronic linguistics, and psychological plausibility on the cognitive structure of linguistic knowledge. This mismatch is particularly salient in the case of grammaticalization. Functions are psychologically salient because they are instantiated by a linguistic form; reciprocally, functional differences within the range of use of this form are not easily accessed through linguistic analysis. It is nevertheless this functional gradience that a form explores and reveals when it comes to express a new function. As a consequence, the new function (i.e. the function of the new variant) does not coincide exactly with the function assumed by the old variant, even when this old variant has been entirely evicted.

7.3.1 Non-overlap of meanings across a semantic replacement

An example of this is the replacement of *moult* in Middle French by the tandem *beaucoup/très* (Marchello-Nizia, 2006). *Moult* possesses features that its two successors lack, such as scope on the whole utterance:

Mout leur avint fiere aventure

Moult happened to them extraordinary adventure → Many great adventures happened to them
Roman de Thèbes, ca. 1150

Moult samble bien qu'il fust estorés et compassés por envoisie gent et pour jolie et pour mener geus et fentes et deduis.

Truly it seems well that [this place] has been built and set for merry people, to the pleasures of life and to hold games and parties and leisures.

La suite du Roman de Merlin, c.1235-1240 (Frantext)

On the other hand, *beaucoup* and *très* have uses which are not attested for *moult*, such as litotes and partitives (*non (moult|mout|mult|mut)*): 2 occurrences; *pas très*: 6575 occurrences; **moult pain vs beaucoup de pain*). Indeed, the query:

```
(moult|mout|mult|mut|molt|moulte)
(pain|pan|pein|paen|pains|pans|peins|paens)
```

does not return any occurrences, while we find occurrences such as:

```
pain et vin a grant foison
meat and bread with profusion           Le roman d'Eneas, c.1160 (Frantext)
```

```
[B]ien comanda as penetiers et as queuz et aus botelliers qu'il livrassent a grant
planté, chascun selonc sa volanté, et pain et vin et veneison[.]
```

```
He ordered to the bread basket carriers, to the cooks and the wine pourers, that
they offered with profusion, to any one as many as one might want, and the bread
and the wine and the meat.           DE TROYES, Chrétien, Erec, 1170 (Frantext)
```

Hence, *moult* could not occupy all intensifier functions, and was complemented by *à pleinté* and *à foison*, who also disappeared from common use. This shows that the divisions of the semantic territory evolve as new linguistic forms come to claim it.

An efficient solution is to consider that the semantic territory is made of several small fine-grained domains, and that the domain occupied by a given form usually encompasses much more than only one such fine-grained semantic domains. In other terms, all linguistic forms are assumed to be polysemic — monosemy being a very rare but theoretically possible exception. Describing these micro-meanings is usually pointless and tedious, since the linguistic forms are too blurry tools for such a task. Similarly, the specific meaning of a given linguistic form is evasive, since it has many sides and facets, which it is not always relevant to subsume under a broad and general meaning. In that sense, we are favoring a representation akin to the semantic network proposed by Brugman (1988) to account for the polysemy of the English preposition *over*, and to Talmy (2000) schemas for functional words.

7.3.2 An epistemologic perspective on the semantic territory

Are all these fine-grained domains really contained in the memory of an individual? And how could they be learned if they are not specifically expressed by a linguistic form? To these questions I have no satisfying answer; the most I can provide would be a possible point of view. Knowledge is a carving of reality; linguistic knowledge, similarly, is a carving of the semantic territory. The broader the carving, the cruder the knowledge. For instance, the knowledge of the natural surroundings of a community is reflected in the diversity of names that an individual is capable to provide. A great number of cases are discussed by Lévi-Strauss (1962). To give but one, the Tewa people distinguish about fifteen different parts of the maize plant, and give a different name to each.

In some sense, the meaning of a form is like the political carving of the geographical territory. We are more aware of the geographical territory details of South America than we are of Russia, while both have comparable sizes, because the former is composed of more geographical units than the latter (the geopolitical country division of

the World ‘expresses’ South America more finely than it does with Russia). The meaning of a form plays the same role than the borders of a political country; it makes the bounded territory salient as a unit.

This is somehow a semi-platonic view of the semantic territory — as platonic as would be the idea of a geographical territory beneath the geopolitical carving. It exists, in a sense, independently of language, independently of the individuals, independently of learning. What is learned is the carving. The smaller, the more precise the carving is, the more salient a specific meaning becomes. In that sense, all languages express the same semantic territory, but they do not devote the same amount of efforts in the same parts of this territory. As is well known, English has a large array of modalities compared to French. This does not mean that French language is incapable of expressing those modalities; it only means that it cannot express them specifically — which, to be sure, does not make much difference practically. Such cross-linguistic variation in semantic contrast is not limited to grammar, but is actually pervasive to all lexical domains. Among other studied examples are color categories, body parts, locative prepositions (Majid et al., 2015), kinship relations (Nerlove and Romney, 1967), or landscape features (?). For instance, in Mandarin Chinese, there are four words for *aunt*, depending on the sex of the parent and whether the aunt is younger or older than the parent. We can therefore say that a language does not lack meanings; it lacks contrast.

If we consider that the semantic territory is finite and contains a huge number K of contexts, all of which have the same capacity M , then all languages are associated with a memory of KM units. There is no one language better than the others in terms of this capacity: only the carving of this memory differs from one language to another. This would mean that French people associate a greater memory to the very few modals they have than English people do with their own, as the former modals encompass a greater number of micro-meanings than the latter. The same holds within a language: *can* is less precise and more general than *ought*, because it covers a greater number of contexts. Hence, all languages would be associated with the same overall mnemonic capacity, a claim which seems quite reasonable. Their difference lies in the internal organization of this capacity. Thus, we do not learn the fine-grained meanings. We discover them as the linguistics forms focus on them.

There are two main problems associated with this notion of completeness of the semantic territory. The first one is that there are new semantic units corresponding to new objects in the natural and technological environment of man. There are many specific lexical words to describe all parts of a ship, for instance, and it would be foolish to think that they can be expressed in the language of a people living far from major rivers and seas. However, they can certainly be described, or pointed at. If we consider gestures as parts of the language, then clearly, almost anything can be expressed in any language, even if the means are very rough compared to a precise and clearly delineated word. This means that there are parts of the semantic territory that our language can virtually express, but does not. If a new kind of aerial transportation is built in the next five decades, it will correspond to a new semantic unit, and a specific word can be proposed to name it. But it is already contained in the semantic territory, and all languages can already speak of it (e.g. ‘that new flying thing’).

The second problem is less trivial. It may be expected that all languages do not

express the same semantic units with the same frequency. Similarly, the mnemonic capacity associated with a specific semantic unit may not be the same in all languages. This would mean that there may be a feedback loop between the typical productions of a language, and the allocation of memory over the semantic territory. Though this idea seems quite natural, it poses some problems. First, is there a limit to the capacity of a semantic unit? Or can a semantic unit be populated by indefinitely many occurrences, as long as the language favors it? Second, this process must be highly constrained and the memory allocation over the semantic territory must answer certain needs. But what would they be? Why would a language need to allocate more memory to, say, the expression of temporal relations between events, or why would it need to express these more frequently?

This memory weighting seems furthermore redundant with the fact that languages already specialize their expression means in the course of their history through the carving of differentially delineated meaning domains. Such an idea would also make it difficult to relate the overall frequency of use of a given word to the broadness of its semantic meaning. It is clear that the most frequent words of a language have a very broad and loose meaning, especially if we consider that all semantic units have the same capacity. Hence, a word which occupies many of these units would present a high frequency. We shall revisit this idea later on chapter 8 as it may prove rewarding.

7.3.3 Empirical cartography of the semantic territory

Of course, one has not to ascribe to these epistemological lengths to accept our hypothesis that language can be conveniently represented as population of occurrences of different species, competing in language users memory for domains of the semantic territory. A more convincing argument in favor of this view is that it is possible to build empirically these micro-meanings; this is the aim of the Semantic Atlas (Ploux and Victorri, 1998; Ji et al., 2003; Ploux et al., 2010). The Semantic Atlas exists in two alternative variants, each of them providing a different way to build micro-meanings.

The first one relies on synonymic relations between words. Several different synonyms dictionary have been compiled so as to build lists of synonyms as extensive as possible. Then, a graph is constructed whose nodes represent the different words, and where links correspond to the synonymic relations between these words. A micro-meaning (called *clique* in the Semantic Atlas) is then defined as a complete subgraph of this whole network which is not included in any other complete subgraph. Those micro-meanings are much more numerous than words and, usually, a word is associated with many micro-meanings. For instance, the word *shuffle* is associated with 12 cliques, e.g. {mix, ruffle, shuffle} and {rearrange, shuffle}.

While the first approach reflects paradigmatic alternations between words, the second one focuses more on syntagmatic relationship. A second way to build up micro-meanings is to make use of collocations. The idea of *cliques* remains the same, but now the network differs: A link is added between two words if they are contonyms of each other, that is, if they collocate in some training corpus (with the possibility to introduce a collocation range beyond immediate neighboring, and to fix a threshold of frequency above which the collocation is considered significant). In this case, *shuffle* belongs to only one clique, {shuffle, cards}, but *iron* belongs to 30 (including {chain,

gold, iron} and {coal, iron, wood}). In the following, I will call *S-clique* a clique of synonyms and *C-clique* a clique of contonyms.

The problem is that, by construction, all those cliques are inhabited by several variants, while we should expect that competition concerns only a minority of micro-meanings. The cliques are thus located at the fringe of word's meaning domains, but these domains remain uncharted. This is, however, not quite true I believe. While we can expect that synonyms do compete in spite of some division of labor, it seems particularly odd to consider that contonyms compete for the same context. I think more useful to hold that, contrary to S-cliques, which reveal the boundaries between word meanings domains, C-cliques flesh out those domains. The more contonyms a word present, the wider his dominion over the semantic territory. A change of word meaning corresponds accordingly to a change of contonyms (Boussidan and Ploux, 2011).

However, this idea encounters a few difficulties. The main one is that a C-clique is shared by all the words in that clique. The respective domains of these words should overlap in this region, so the corresponding micro-meanings would be boundary meanings too, and within them a competition would take place between all contonyms. Clearly, this cannot be the case: *shuffle* does not compete with *cards*. We can, therefore, not assume that the C-cliques can be straightforwardly interpreted as regions of the semantic territory.

We might actually doubt as well that S-cliques are actual sub-domains of this territory. Though their definition is less sensitive than C-cliques (which heavily relies on the collocational range), they also raise the question of whether or not they could indicate a competition that is going on between its different members. An S-clique can easily contain 3, 4, 5, or even more words. A competition between so many words over a single sub-domain is unlikely. We can assume, however, that the bigger the clique is, the closer to a meaning unit it focuses. A two-word clique is already more localized than a word — as a boundary compared to a country. Then, if an S-clique contains more words, it becomes much more precise. For instance, the meaning corresponding to the clique heighten, sharpen is more vague than the meaning of the clique grind, hone, sharpen, strop, whet. In the second case, the margin of interpretation is much more constrained; the area of the semantic territory corresponding to the intersection of so many word meaning domains is bounded to be restricted.

Another problem is that cliques are word-dependant. If the meaning of a word evolves, then the cliques it belongs to will be different. The idea of the semantic territory, on the contrary, is to provide a stable underlining plane of meaning where this kind of word meaning changes can be mapped and tracked down. In conclusion, the cliques cannot be used, as such, as an empirical representation of the semantic territory. Yet, the Semantic Atlas shows that we can empirically work out sub-units of meanings, more fine-grained than word meanings. It also provides interesting new insights on the semantic organization of language, which are far from being exhausted by the short preceding discussion. These cliques provide precious information on language; as we proposed, the size of the S-cliques reflect psychological salience of the corresponding part of the semantic territory for a language users, while the number of C-cliques a word belongs to is a hint on the wideness of its dominion over the semantic territory.

7.3.4 Vectorial representation of semantic units

There is a simple, convenient way to represent the semantic units. We can postulate that the semantic territory is a high-dimensional vectorial space (say L) and describe a semantic unit by listing its coordinates. This has actually been performed for words (Mikolov et al., 2013) and even allowed to account for their semantic evolution (Dubossarsky et al., 2015). This, however, would allow to build semantic units. In the following section, I suggest two ways to build such semantic units, one which relies on the Semantic Atlas S-cliques, the other on diachronic competition. The purpose of the following proposal is not to provide a right-away workable method, but to illustrate the possibility to represent the semantic territory through available information on language.

Let us note θ^μ the vectorial representation of word μ with $\theta^\mu = \{\theta_1^\mu, \dots, \theta_i^\mu, \dots, \theta_L^\mu\}$. In Mikolov et al. (2013) this representation is built out of the collocation frequencies over a given corpus. A diachronic variant of this representation is possible (Dubossarsky et al., 2015, 2016) if one divides the corpus into successive time windows. Then $\theta^\mu(t)$ represents the meaning of word μ as computed for the time window t .

Now, we can assume that the semantic units can also be represented by a K -dimensional vector. We thus note θ^s the vectorial representation of semantic unit s . If we note x_s^μ the frequency of word μ in semantic unit s , then we expect the word meaning θ^μ to be related to semantic units by the relation:

$$\theta_i^\mu = \sum_s x_s^\mu \theta_i^s. \quad (7.1)$$

Variations of this formula are possible; for instance, if the coordinates are binary ($\theta_i = \pm 1$), then one should add a sign rule. If they are bounded, one has to add a normalization rule. In any case, we can rewrite 7.1 by defining a matrix X of elements x_s^{mu} such that:

$$\theta_i^W = X \theta_i^S, \quad (7.2)$$

where θ_i^W (resp. θ_i^S) is a vertical vector of the i -semantic component for all words (resp. for all semantic units). Then, knowing the matrix X and the vectorial representation of all words, we can compute the i -semantic component of all semantic units by inverting the non-square matrix X :

$$\theta_i^S = X^{-1} \theta_i^W. \quad (7.3)$$

However, we do not know the matrix X . Yet, we can have an approximate idea of it using the Semantic Atlas, by equating the S-cliques to the semantic units. It is indeed not possible to access the exact frequencies x_s^{mu} : considering a word in a given utterance, it seems a difficult task to decide which clique the word expresses in this particular context. It is all the more intractable if we want to do it for all words of all utterances of the corpus. There is nonetheless the possibility to consider that the x_s^{mu} are binary variables, equal to 1 if word μ belongs to clique s . This task is still to be performed, but it is, at least, empirically feasible: both the Word2vec model of Mikolov et al. (2013) and the Semantic Atlas are currently available.

Let us now turn to the second method. The first one was entirely synchronic and global, because the synonymy relations between words reflect a temporally located state of the language, and because the inversion of the X matrix, unless it has internal

block-structure, involves all the elements of language simultaneously. The second one is diachronic and local (at least partially, because it makes use of vectorial representation which are computed using data on the whole language).

We consider two words, μ_1 and μ_2 , which compete for a given semantic unit s . We note t_i the time at which the competition initiates (assuming word μ_2 dominates s) and t_f the time at which it ends (i.e. word μ_1 has evicted μ_2 out of s). We will then have:

$$\begin{cases} \theta^{\mu_1}(t_f) &= \theta^{\mu_1}(t_i) + \eta_1 + \theta^s \\ \theta^{\mu_2}(t_f) &= \theta^{\mu_2}(t_i) + \eta_2 - \theta^s \end{cases}, \quad (7.4)$$

where η_1 and η_2 are two perturbation terms accounting for the fact that during the time $t_f - t_i$, other processes may interfere. Of course, there may be some technical problems if a normalization is required so that the contribution of the newly conquered semantic unit does not influence $\theta^{\mu_1}(t_f)$ through a mere summation. We remind that this is but a sketch of a method.

Combining these two equations, we obtain:

$$\theta^s = \frac{1}{2} [\theta^{\mu_1}(t_f) - \theta^{\mu_1}(t_i) - (\theta^{\mu_2}(t_f) - \theta^{\mu_2}(t_i))] + \frac{\eta_2 - \eta_1}{2}. \quad (7.5)$$

Obviously, only one of the two equations (7.4) would be sufficient. Yet, we do not know the η terms and cannot extract them empirically. We can hope, however, that they somehow cancel out, so that accounting for two contexts is more reliable than one. Assuming thus that the spurious term $\frac{\eta_2 - \eta_1}{2}$ is negligible compared to the two differential terms, we only need, to achieve the computation of θ^s , to find two variants which are indeed competing. Yet, we already provided in chapter 5 a method to identify competing variants from diachronic data. Combining the diachronic vectorial representation of words of Dubossarsky et al. (2015) with this information, we should be able to empirically extract θ^s .

This should enable us to detect cases of repeated replacements, and maybe, if we find that the θ^s are clustered in the vectorial space, to provide a typology of changes. It could allow, if the word μ_1 is successively competing with different forms, to represent the discrete steps of meaning corresponding to each of these competitions, and to account for the whole meaning shift by decomposing it into several invasions of related semantic units. It would also be very tempting to compare the coordinates of the θ^s between parallel cases of grammaticalization across two different languages. However, the vectorial space is built out of a given corpus, so that the dimensions are fitted to the description of one language only, but not the other. If the training corpus is not the same, there is no match between the dimensions. This mismatch is attenuated in the case of the diachronic discretization of the corpus, because the authors take, as an input for each time window, the output of the preceding one. The different corpus are thus not trained independently, which may, as a consequence, produce some kind of inertia in the algorithm computing the $\theta^\mu(t)$'s.

7.3.5 Summary

From all the preceding attempts at giving the semantic territory a less abstract, hypothetical existence, it appears that it is a notion which is, expectedly, hard to sustain

empirically (as it is independent from linguistic forms), but meet the two requirements of psychological plausibility (meanings being made salient through their linguistic expressions, which is not the same from one language to another) and efficient reflection of diachronic competition between forms.

Of course, this representation can still seem rather crude compared to the formidable complexity of language. It is not obvious to see how this semantic territory populated by different species of linguistic forms can combine with the triadic picture of constructions, paradigms and collocations that we developed in chapter 1. However, I strongly believe that empirical approaches such as diachronic frequency tracking, word vector representation, or the Semantic Atlas, can considerably help to refine and clarify theoretical insights on the organization, structure and components of language, as well as provide an accurate description of meaning.

7.4 The semantic territory as a conceptual network

Earlier on I insisted on the fact that the semantic territory was best conceived as composed of discrete semantic units instead of a continuum of meanings. There are two reasons for this preference. First, it allows us to grasp more easily the competition between two words for a given unit; second, we can enrich the semantic territory with a network structure by drawing links between the different semantic units.

This is consistent with many previous claims. Hudson (1984) famously postulated that ‘Language is a conceptual network’. In Construction Grammar literature, networks have also received a special focus (see (Traugott and Trousdale, 2013) for a summary). However, there are many ways to conceptualize language as a network, and the one which we will focus on presents little in common with networks as defined in Construction Grammar literature.

In the previous subsection, we encountered two kind of networks, based on synonyms and contonyms respectively. There are of course other ways to build lexical networks: Gaume (2004), for instance, using a dictionary, draws a link between two words when one is found in the dictionary entry of the other. Those lexical networks usually present a characteristic small-world structure, independently on the specific method to draw links between words: Ferrer i Cancho and Solé (2001) found this structure for co-occurrences links, Motter et al. (2002) arrived at the same conclusion using synonymic links, and the method used in (Gaume, 2004) led to a small-world structure as well. We can thus safely assume that the semantic territory exhibits small-world properties.

There is, however, a difference which I ought to emphasize once again: Lexical networks cannot map right away to the semantic territory, for a lexical network lies on links between words, while a word can cover several semantic units, just as a semantic unit can be covered by several competing linguistic forms. A possible way to achieve this mapping would be to use the fine-grained S-cliques defined by Ploux and Victorri (1998), and to draw links between two cliques iff they share a common member. This method would certainly be a crude one to represent the links over the semantic territory (especially because 1 - the S-cliques do not chart the whole semantic territory and 2 - linguistic forms are far from being limited to words); yet it would give us a reliable

insight on its general structure, so that it would be a way to test whether the semantic network is a small-world or not.

If those links are empirically out of reach, one might wonder why they would be interesting at all. As we shall see, they play in fact a crucial role in the model. However, instead of focusing on the global network structure, I will report two related views of change which rely on a local account of links between contexts. They have been developed initially to describe grammaticalization more exclusively, yet they can certainly be applied to semantic change in general. Interestingly, both have been initially published in the same volume collecting various contributions on grammaticalization. After a presentation of these context-based accounts of change, the relation between linguistic contexts and the semantic territory shall briefly be discussed.

7.4.1 A context-based view of grammaticalization

Heine (2002) distinguishes four steps occurring in any grammaticalization process, each one associated with a particular type of context. The first step is the initial situation, where a linguistic form carries a given meaning, called the source meaning. Then, this form comes to be used in a ‘bridging context’ in which a particular inference is possible, leading to the target meaning. At this point, both interpretations of the form are available, and the source meaning is still dominant. The target meaning becomes only prominent in the third step when the form comes to be used in a ‘switching context’. Here the source meaning becomes less appropriate, though it remains in the background. However the form can now be used with its new target meaning in additional contexts, where the source meaning is no longer relevant. This is the fourth step called ‘conventionalization’.

We see that the process of meaning change can be seen under two different perspectives: the first one is semantic (there are two meanings involved, a source and a target, so that if we consider a semantic network, the form goes from one site to another), the second one is syntagmatic, and depends on the particular utterances with which a form is compatible. When the form comes to be used with its new meaning, it can be involved in a higher number of utterances. In the following, I will argue that those two views are one and the same, and that each context corresponds to a particular meaning, but that the source and the target meaning correspond to large clusters, or domains of the semantic network, while contexts are associated with one or a small number of semantic units. The ambiguity felt in the intermediate contexts reflects thus the fact that the corresponding semantic units are too fine-grained compared to the specific decoupage of the semantic territory operated by linguistic expression. In other terms, we can only conceive ‘big’ meanings. However, a linguistic form, in order to pass from one of these big meanings to another, has to take the small bridge which exists between the two big clusters, and so manifests a very precise meaning which feels ambiguous because we cannot conceptually achieve this particular level of precision. This would explain why, while these intermediate contexts are always found by diachronic linguists, change is seldom acknowledged by speakers, because their cognitive apparatus is not acute enough to grasp such detailed meaning, so that they are aware of the change only once the bridge has been crossed -which is the fourth ‘conventionalization’ step of Heine: the form now occupies a domain large enough of

the big target cluster so that the target meaning could finally be perceived.

Diewald (2002) has also proposed a context-based account of semantic change in a grammaticalization-oriented perspective. Her theory is similar to Heine's, but there are some differences between the two that are worth being highlighted. There are also four kind of contexts: Isolating (i.e. non-ambiguous) contexts for the lexical meaning, isolating contexts for the grammatical meaning, untypical contexts of use for the linguistic item with its lexical meaning, and critical contexts. Untypical contexts arise through an extension of the lexical meaning, through which the grammatical meaning might be hinted at. They do not necessarily lead to a grammaticalization, which is only 'triggered' with the emergence of the critical context, but they condition it. Without this step, the linguistic item cannot develop out a grammatical meaning. Untypical contexts are furthermore numerous, and show a wider array of uses of the form, leading to its involvement in more diverse syntactic and semantic situations. In this process, a critical context can emerge, where the usual semantic and structural reading of the form will become highly unusual, leading to a conflict which can only be resolved with the emergence of the grammatical meaning. In this scenario, the new meaning arises out of an inferential process or a 'conversational implicature' in the hearer's mind, confronted to a 'clash' characterizing the use of the form in this critical context. Finally, the two meanings -the lexical and the grammatical ones- will part away as isolating contexts, which are compatible with either one or the other meaning, start to appear simultaneously.

As Diewald herself stresses, one of the main differences between Heine's account and her own is that in Heine's view, all contexts which have been gradually accessed in diachrony coexist in synchrony. On the contrary, Diewald's critical context serves as a 'catalyst' (Diewald, 2006), and is bound to disappear after having played its role in the grammaticalization process. Another difference is that there is no clear correspondence between the types of context they distinguish. Even if the general view is the same — the linguistic form proceeds from context to context through inferences or conversational implicatures in its way to acquire a new meaning —, and if it could be tempting to assimilate bridging contexts with untypical ones and the switching context with Diewald's critical context, this does not hold in the detail. In the switching context, the new meaning is already there, while the same meaning appears as a solution of the conflicts posed by the critical context. In other terms, in Heine's account, the main semantic leap occurs between the bridging context and the switching context, while in Diewald's, it occurs from the critical context. Also, in (Heine, 2002), there is no necessity of jumping from the bridging context to the switching context, whereas the critical context compels to find a new interpretation of the linguistic form. What is contingent, in Diewald's account, is the emergence of such a critical context. It would thus seem that the best way to resolve this mismatch between these two parallel views is to postulate that the critical context lies in between the bridging context and the switching context. Because of its transitory nature, it should be expected that it is not to be found in most examples.

There may be no need, however, to offer such a conciliation between the two views. Indeed, there is a further difference which sets them apart. Heine (2002) focuses on meaning; one context brings forward one particular meaning, so that if the form comes to be used in new contexts, it gets associated with new meanings in some

systematic way. Diewald (2002) on the contrary puts the emphasis on the structural aspect of a context. This is why she tries to describe contexts as constructions in the Construction Grammar sense of the term (Diewald, 2006). However, as constructions are form-meanings pairs, contexts should be systematically associated with meaning as well. This is not the case: a critical context, as we saw, is not associated with any compatible meaning — hence the necessity to work out a new meaning through conversational implicature. Indeed, it is not recognized as corresponding to any stored construction (Diewald, 2006). This scenario is actually most compatible with the *constructionalization* process (Traugott and Trousdale, 2013), the process by which new constructions occur in the language. We can then understand the disappearing of the critical context as the fact that, once the new construction has emerged, this context is no longer critical, for it matches with the structure of this new construction. This proposal, however, would be at odds with Diewald’s view of the development isolating contexts which are necessary to separate and keep well apart the new meaning from the old.

In the following, I will try to make my own proposal in line with those two, so as to tentatively explain how constructions, contexts, meanings and forms can be understood together in an instance of meaning change, and how this can go along with a competition scenario.

7.4.2 Contexts and semantic units

The simplest picture that we can have of a meaning change with competition is the following. A linguistic form X is used with a meaning μ_0 and a second form Y is used with meaning μ_1 . Each of these two meanings is associated with a bag of words of size M : meaning μ_0 , respectively μ_1 , contains only occurrences of form X , respectively Y . However, there is some link between the two meanings. This link is assumed to be unidirectional and asymmetric. As we saw, this kind of links have long been postulated in grammaticalization literature, and has been called invited inference (Traugott, 1987), subjectification (Traugott, 1989), pragmatic inference (Traugott and König, 1991), conventionalization of an implicature (Nicolle, 1998), conversational implicature (Diewald, 2002), inference (Heine, 2002), subjectivation (Marchello-Nizia, 2006), etc. Through this link, form X can also come to be associated with meaning μ_1 ; hence, a competition then occurs between the two forms X and Y over meaning C_1 .

Crossing the bridge

We can outline the exact same scenario in the case of contexts, to follow Heine’s proposal, whose application is more straightforward than the one of Diewald (2002). First, form X is used in a context c_0 ; then it comes to be used in a new context c_1 which gives rise to inference leading to a new context c_2 where the form receives a different meaning, and thus becomes able to invade new contexts. In other terms, the form initially occupies a whole cluster of closely related sub-meanings, and, through a long-range inferential link between one of these sub-meanings and a sub-meaning belonging to another semantic cluster, which are respectively identified as the ‘bridging’ and the ‘switching’ contexts, the form is able to express the meaning of this new semantic cluster.

The problem is, if we add this layer of description, a discrepancy starts to appear. Indeed, in a Construction Grammar framework, contexts can be interpreted as constructions (Diewald, 2006). For instance, there is, in French, a construction [$\{V_{\text{participate}}\}$ de $\{N\}$], where the noun expresses the cause of the V (V occurred because of N). Some instances of this construction are: *tombé de faiblesse*, *tombé d'effroi*, *crié de peur*, *mort de peur*, *tombé du coup*. This construction seems to be the context from which the construction [du coup] with the meaning 'as a consequence', emerged in the nineteenth century. In this case, the linguistic form *du coup* is filling a slot of a construction in one context, defined by the construction, and becomes a construction by itself in the second context (which is no longer a construction, and would be rather defined from specific semantic, discursive, and syntactic features).

The situation is pretty much the same for *way too* if we believe in the specific scenario I proposed in chapter 2. The specific following context :

$$[[[\text{give } \{\text{pronoun recipient}\} \text{ way}]_{C_1}]_{C_2} [\text{too } \{P(C_2)\}]_{C_3}]_{C_4} \quad (7.6)$$

with :

- C_1 : give way, a lexical item which still remains analyzable as the two words can be split apart,
- C_2 : specific transitive verbal construction with pronoun,
- C_3 : intensification through *too* of a qualifier whose nature is constrained by the nature of what is qualified (this is a specific evaluation of the more general intensifying construction which presents a whole paradigm of quantifiers very, too, far too, ...),
- C_4 : qualification construction : C_3 qualifies C_2 ,

allowed for the following reanalysis:

$$[[[\text{give}]_{C_1} \{\text{pronoun recipient}\}]_{C_2} [\text{way too } \{P(C_2)\}]_{C_3}]_{C_4} \quad (7.7)$$

which is almost the same as 7.6, except that C_1 is now the lexical *give* alone, while the intensifier in C_3 is now *way too* instead of *too*. In other terms, *way too* has been added to the paradigm of intensifiers through the specific context made explicit above. However, in 7.6, the whole context is required for the specific inference to give rise to the new meaning of *way too* while in the second one, the construction C_3 alone is sufficient to process this particular linguistic form.

This means that, after the passing of the bridge, the form which is actually recorded might not be the same. In the first context, the form to which *way* belongs is clearly *give way* while in the second one, it must be *way too*. Thus, *way too* occurrences are fed by the *give way* occurrences, but they are not of the same form. It implies that, though in the model we assume that the form X stays the same on both sides of the bridge, it may not be so in reality, but it will not make any significant difference during the competition process.

The competitors

Another important question is the relationship between contexts and meanings. It may happen that, after a reanalysis of the constructional organization of the utterance, the linguistic form will be the only one compatible with this particular context. Then, it would seem that we should consider that the new meaning is granted for the form, because there will be no competitors. There are two major ways for this to not be the case. The first one would be to consider that the picture of two competing forms over a semantic sub-domain is inappropriate. The competition would take place over the reanalysis, struggling to be or not to be. There would be no invasion anymore, only an existentialist struggle of certain new interpretations, a fight against their own nothingness. The prevalence of such a view would explain why there are so few instances of meaning change where it is possible to identify the actual competitor of the form. I will, however, favor an alternative view, emphasizing the fact that such a competitor can be found sometimes, and while it might not be always salient, it can exist in every case.

I presented earlier the rather strong and debatable assumption that all languages cover the entirety of the semantic territory, at all times. The linguistic forms share this territory into multiple domains. The wider the domain, the comparatively more obscure the units within this domain. Nonetheless, it is possible to focus on a single unit (or close to it) by adding context features so as to constrain further the processing of the linguistic form. It means that all semantic units can be expressed by a given language because they are carved out through specific forms and contexts. However, and this is the key point, it may happen that a same semantic unit becomes carved out by two different contexts. In that case, the two contexts are themselves competing for the semantic sub-domain, each one occupying the semantic unit through a bounded linguistic form or a bounded paradigm of linguistic forms. An interesting question then would be the nature of the recorded occurrence in those cases. Is the whole elaborated context recorded, or is a specific construction perceived as conveying the meaning more than the others, in that particular context?

Anyway, it implies right away that competition can occur at different levels; at the context level, where both contexts struggle to carve out a semantic unit according to their own structure; at the linguistic form context, where two forms can compete for a given context. It means that the one-to-one struggle between two forms needs to be refined: Forms can cooperate as a paradigm within a hosting context against another context, or they can compete within the context. Obviously, this raises new questions: What should decide between cooperation and competition for different forms within a given context? How can this be modeled? For simplicity, we will consider, in our model, that the competing 'forms' can equally stand for a single form, or a set of cooperating forms representative of a context, so that the two cases will not be distinguished in any way.

Also, this competition has a good chance to escape notice. Indeed, if the semantic unit was expressed by the older form through an elaborated context, then losing it does not change dramatically the uses of this older form. A new context, by the spread of its supporting forms, can thus carve out a piece of the semantic territory within the dominion of an already existing context without altering much the uses of this former

context, if its dominion is large enough and the disputed units are marginal enough.

The same would be true between forms. For instance, *way too*, in the paradigmatic context of the intensifying construction, can compete with *too* and grab some of its former semantic units without changing significantly the meaning and the uses of *too*. In this particular case, the prototypical meaning of *too* would be to express intensification in terms of going beyond a certain reasonable threshold. Certainly, it is possible to get beyond this threshold at different degrees that all fall under the semantic scope of *too*, each of these degrees corresponding to a distinct semantic unit. Yet, because *too* is too vague to express one of this degree specifically, it can be refined (or substituted, depending on the point of view) by different forms such as *way too* or *a little too*. In that case, these two latter constructions would express a sub-domain of the *too* dominion, formerly obscured by the extent of this dominion.

Hence, it seems plausible enough that the competition between two forms, or two contexts, appears as if one of these two forms (or contexts) meets some invisible resistance in its efforts to extricate into existence. The account proposed here might be a simplification, and obviously leaves crucial questions unsolved (e.g. what is actually stored during the processing of an utterance?); yet it relates, in a cohesive way, meaning, contexts, and forms, taking into account both the competition, and the context-wise spreading of a given linguistic form. From a modeling point of view, it also appears that the simplistic picture brought forth at the beginning of this section — two related meanings/contextes occupied each one by a different form/species and asymmetrically related so that the form occupying one can invade the other through a competition process — is actually much richer in its interpretation, and remains applicable in more sophisticated situations — e.g. if the form invading the target-meaning is not the same as the form occupying the source meaning, or if a constructionalization process gives rise to a new context, moving up the competition on the context-level rather than of the form-level.

Unfortunately, if our current account of the role of contexts in semantic change is close to be correct, this would suggest that understanding fully a semantic change is bound to be difficult. Indeed, it would mean that the switch responsible for the semantic change is very precisely localized over the semantic territory and therefore easy to be missed. In a diachronic corpus, available data is limited, and it may well be that the very specific context in which the link is drawn is not represented among the evidence (especially so if this critical context corresponds to a speech situation which has not be represented in any written account). Indeed, in many cases, the new meaning seems to pop out of nowhere, and while we can still encounter many instances of Diewald's 'untypical contexts', there seems to be an abrupt shift between the old and the meaning of a form and the exact path is not be retraced. This certainly limits the reconstruction of the detailed process, and would place the exact nature of the critical link out of our reach.

7.5 The links of the network

We have already stressed the network nature of the semantic territory, and review a few works establishing its small-world structure. However, in the competition scenario

that we sketched in the previous section, we mentioned the possibility that the links could be asymmetric, even unidirectional. As this point is crucial in our model so as to allow for the change to proceed, it must be discussed further. We shall therefore discuss a few cognitive interpretations of these links, in order to understand better why they could be unidirectional.

7.5.1 Pragmatic inferences and implicatures

First of all, the nature of the semantic associations between the units of meaning appears to be very diverse. They may be motivated by structural factors, analogical processes, metaphorical extensions, etc. They can also result from a conversational implicature. This kind of inference is a process by which the hearer draws a specific meaning from the speaker's utterance, while this meaning is not associated with this specific utterance. Rather than relying on the direct semantic content of the utterance, the hearer instead makes a guess about the speaker's communicative intent underlying the utterance. Actual conversations are full of implicatures, and they are actively used by the speakers to convey meanings more or less overtly. A humorous instance of such an inference is to be found in *The Simpsons*, ep. 329. The Simpson family is sharing a meal with a young girl from a juvenile detention center in her own cell, and the director of the center, who has just brought back the girl to her cell, witnesses the beginning of the meal.

The director: Well, my shift's over. I guess it's back to my bachelor apartment. Make a tuna sandwich, turn on *Will and grace* and cry myself to sleep.

Marge, the Simpson mother: Would you like to join us?

The director: Didn't you hear me? I've got an evening planned.

From the complaints of the prison director, Marge Simpson infers that he would like to join them, which is indeed an obvious inference to draw (why would he complain otherwise?). The humorous effect of the scene comes from the fact that this inference is completely ruled out by the director, who unexpectedly sticks to the literal meaning of his own utterance.

Conversational implicatures therefore reflect how the hearer rely on the speaker's plausible intentions to process the meaning of an utterance, and are clearly directional. They allow one to always understand a little bit more than what has actually been said. Whether they do or do not correspond to actual pathways of meaning change is an entirely different question. The point is, there are many more conversational implicatures than there are instances of meaning changes. That is, most conversational implicatures do not lead to a semantic expansion and remain instead at the level of pragmatics. They provide general, highly directional lines of reasoning which certainly guide the change and favor some innovative semantic interpretations of linguistic forms. In other words, we can rely on them to find the general path of a diachronic change and assess its psychological plausibility, but they cannot serve as a proper explanation for this particular change.

7.5.2 Metaphors

Metaphorical extension also prove to be a powerful way to derive new meanings. This has been proposed first in Osgood and Sebeok (1954), and then, independently as

it appears, by Lakoff and Johnson (1980), a famous book which enriches this idea and fleshes it out with numerous examples. The cross-linguistic importance of these metaphorical links as well as their relevance in meaning expansion phenomena have been further established in Heine (1997), and, in the specific case of grammaticalization, in the crucial *World lexicon of grammaticalization* (Heine and Kuteva, 2002). Some metaphorical links are entrenched in a specific culture (Kövecses, 2005), e.g. the importance of the network metaphors in our contemporary connected societies, while some others may appear universal (e.g. the mapping of the time domain onto the space domain). Even if metaphors are culture-specific, they go along some universal lines: There are constant mappings between the psychological mood, for instance, and more salient, external features of the natural, technical, or numerical environment of the speakers.

Metaphorical mappings allow us to carve out a semantic target domain with constructions and conceptual patterns initially developed for a source domain. An important feature of these mappings is that they have directionality. Indeed, we, as humans, conceive more easily what we can immediately grasp through our senses than what we need to reconstruct on the basis of this perceptual input. Hence, time is more difficult to conceive than space. It would make no sense to express our knowledge of a familiar domain, through the prism of a less apprehensible one. For instance, language discourse is a domain less concrete, less phenomenal, than space and movement, even less than time. As a result, discursive markers are often expressed through metaphors from these domains (e.g. *by the way*, *in the same time*, *actually*, *from* as in ‘from what you said’, etc.).

The importance of metaphorical reasoning in the shaping of the linguistic means is therefore not to be dismissed. They certainly play an active role in semantic expansion. However, I tend to believe (though this is purely conjectural) that they might not be at the core of processes such as constructionalization, though they might initiate a grammaticalization (e.g. the French *par ailleurs* had a literal spatial/movement meaning from which a metaphorical expansion led to its contemporary discursive function). Indeed, it seems that the triggering of grammaticalization usually relies on a very specific context (or on a specific link between a bridging and a switching context in Heine’s account), while metaphorical mappings are pervasive in the organization of the semantic territory. In that sense, a context-specific conversational implicature or an analogical reanalysis might be necessary for the process to start off.

7.5.3 Analogy

Metaphorical inferences and conversational implicatures both suffer from the fact that they lack explanatory power because, if they would be responsible for a change to occur, there would be more changes than what we actually observe. I suggest that in order to trigger a change, they need to be combined with another mechanism, related to the structural features of the contexts and the forms, which is analogy. The relevance of this mechanism has been stressed in particular in (Fischer, 1999, 2008, 2011).

One of the main cases where analogical reasoning would be at work is the phenomenon of paradigm attachment. We saw the importance of paradigms in constructions as a way to regroup and organize constructions, but also as a set of entrenched

forms that are commonly used to fill the free slots of a schematic construction. This set or paradigm contributes in turn to the meaning of the construction it partakes in and restricts its semantic scope. Attachment to a paradigm would be an efficient way to give rise to a semantic expansion; indeed, the linguistic form, as it would join the paradigm, would be given access to many contexts where the paradigm is already used in. For instance, from the moment *par ailleurs* joins the paradigm of discourse markers, it acquires its syntactic freedom (as soon as the first unambiguous use of *par ailleurs* as a concessive marker appears in Frantext, in 1797).

However, analogical attachment to a paradigm is, most of the time, unnecessary, since forms already belong to a paradigm. It may nonetheless happen, in very specific and idiosyncratic contexts, that the usual parsing of the form is inhibited by contextual factors. Analogy then becomes necessary to map the unparsed form to a suitable construction, or a suitable paradigm. In doing so, the analogical reasoning is guided by possible conversational implicatures and/or metaphorical mappings. Such a context would be the ‘critical context’ of Diewald, or Heine’s ‘bridging context’. Then, analogy is the answer triggered by the use of the form in a specific context.

A crucial caveat must be expressed at this point. The previous account leaves the impression that analogical reasoning is essentially the doing of the hearer. There is, however, no reason for the speaker to be more careless about language than the hearer. There are probably as many efforts devoted to the on-line process of uttering, than to the one of parsing and understanding. It seems thus unreasonable to assume that the speaker can utter ambiguous sentences, leaving all the pain to make sense of it to the hearer. (If some part of this manuscript seems unclear, it should not be necessarily assumed that I was somehow negligent in the process of writing it.) On the contrary, we should see them as equals in the coping of communication. Thus, if the speaker uses a form in a context which seems to prevent its usual parsing, it is probably because his own analogical reasoning supports him in doing so.

Analogy may also be responsible for paradigm creation. If several forms are felt to behave in a similar way, then a paradigm of these forms can emerge. Thus, idiosyncratic use of one or two forms can become a template for a new construction. French determiners, English modals, may have emerged in this way. In that regard, a given grammaticalization instance can favor the triggering of multiple subsequent grammaticalizations based on analogy, as suggested by Marchello-Nizia (2006). However, how this idiosyncratic behavior of the forms has developed in the first place remains an open question worth investigating.

Two important points must be stated on analogy. First, they have no reason to be directional, and their directionality is more likely to be the result of the implicatures and metaphorical reasonings who support its unfolding. Second, they are necessarily language-specific. They depend on the actual organization of the linguistic forms into paradigms, on the available constructions, on the subtleties of contexts. This is a fact of great consequence for the theoretical picture we are trying to draw: It entails that the channels of meaning change vary from one language to another. Hence, they also vary from two different times of a single language. Links over the semantic territory are therefore subject to change. This change, however, is expected to be slow, at least slower than the meaning change of a form. This would also be the case for metaphors. For instance, the metaphorical organization of emotions in French still relies partly

on the Aristotelian theory of humors (Kövecses, 2005), while this theory has lost all relevance centuries ago and has ceased to be common knowledge. In our model of semantic change through competition, we will therefore make the assumption that the links are constant throughout the process, as the timescale of their change is expected to be longer than the timescale of a meaning change.

A last point to be made on links is that they may be probabilistic. A complex analogical remapping as the one we suggested to be responsible for the shift of meaning is certainly far from being automatic and granted, and it may actually fail. We can thus weight the link with the probability of giving rise to the expected inference. Once again, we will suppose, in our modeling attempt, that those probabilities are constant throughout the process, though they may actually evolve in time. For instance, they could be reinforced as the form becomes present in the new context — past recorded utterances playing then the role of precedents on which to rely in the devising of the analogy. Yet, it may also not be the case. If the analogical reasoning is supported by an elaborated conversational implicature, then the implicature does not need to become simpler once it has succeeded. Some reasonings, even if they have been pursued out, remain difficult to follow and to outline; the understanding of a chapter of a Kantian treaty, of a subtle mathematical demonstration, the computation of the critical temperature in the 2D Ising model, are all achievements of the mind whose complexity and difficulty are not quite diminished by a former successful attempt of sorting them out. At a more modest level, it is possible that conversational implicatures do not benefit for their past successes, especially if they have to be performed on-line, in the middle of a speech process, with very little time available. It thus seems a safe assumption to posit the constancy of those link probabilities.

7.5.4 Priming

Priming is a psychological effect first studied in the 1970s (Meyer and Schvaneveldt, 1971; Meyer et al., 1972), according to which some stimuli can provoke a spread of neural activity, and henceforth facilitate the subsequent retrieval of related memories. Semantic priming (Neely, 1976), in particular, addresses how the retrieval of words, and how it is facilitated or inhibited by previous lexical stimuli. For instance, synonyms, antonyms, category members, tend to prime each other (Hutchison, 2003). This semantic priming can furthermore be asymmetric (Ferrand and Alario, 1998); for instance, in French, *alligator* triggers *crocodile*, while *crocodile* does not trigger *alligator*. Furthermore, these semantic priming relations between words are language-specific (Ferrand and Alario, 1998).

Priming and grammaticalization

Priming has been recently proposed to account for the directionality in grammaticalization (Jäger and Rosenbach, 2008). The idea is that, due to an asymmetric priming between meanings A and B , a word expressing A can come to be felt as if expressing B as well. Because the priming is unidirectional, so would be the change. Furthermore, as priming primarily occurs from lexical meanings to grammatical meanings, then grammaticalization would be a unidirectional change. Hence, semantic expansion

would be due to a repeated neural co-activation of asymmetrically related meanings. This proposal is most consistent with the view we defended throughout this chapter.

However, this scenario of change has been inquired through a psycholinguistic perspective (Hilpert and Saavedra, 2016), which spreads serious doubts on its actual relevance. Hilpert and Saavedra (2016) have tried to answer the following question: given an utterance as a primer, where a given form *X* is used with a lexical meaning, how much does it facilitate the processing of a second utterance, where *X* has now a grammatical meaning? If the lexical meaning primes the grammatical one, then the primer should make easier to process the utterance in which *X* is grammatical. They contrast, for instance, the processing of ‘I’ve rented out this movie I can’t wait to see.’ (no lexical primer) with the processing of ‘I’m waiting for the movie I can’t wait to see.’ (lexical primer). In the latter case, the occurrence of *wait* with a lexical meaning should favor the processing of *wait* with a grammatical meaning.

As it happens, their results show that a lexical primer tends not to facilitate the processing of the grammatical meaning, but to inhibit it. Moreover, they also investigate the reverse case, i.e., the effect of a grammatical priming for a lexical processing, and find that there is a positive effect on the processing time: the grammatical meaning primes the lexical meaning. Those two effects are therefore completely at odds with the picture proposed by Jäger and Rosenbach (2008).

Actually, they are not completely unexpected, and can be understood at the light of a work by Sweetser (1988). According to her, grammaticalization is associated with a phenomenon of semantic bleaching, which she describes by associating cognitive schemas to the different meanings. As a linguistic form is grammaticalized, it comes to be associated with a more abstract schema, fleshed out of its lexical features. For instance, the preposition *à côté de* (‘next to’), in French, comes from a body part (the ribs). In this case, the lexical source meaning is fuller, richer, than the grammatical target meaning, which only keeps the more skeletal and essential features of the original schema.

Then, in a sentence, if the form is uttered with the grammatical meaning, the skeletal, abstract schema is activated; when the form is uttered for the second time, with the lexical meaning intended, it suffices to complement the schema by enriching it with the corresponding features. If we proceed the other way around, the utterance of a form with its lexical meaning activates the richer, fuller schema. Then, to retrieve the grammatical meaning in the second utterance of the form, the hearer has to remove the features which have just been activated. Yet this is not an easy task, for those features are blended in the skeletal schema, and the latter cannot be easily abstracted out. For instance, from the observation that planets revolve elliptically around the sun, Newton abstracted out the concept of a gravity force. Then this concept was almost immediately used to account for the fall of physical bodies. Once a concept is extracted out of an empirical embodiment, it can be dressed with new phenomenological garments. Apparently, the same goes for semantic reasoning.

Priming in our model

Even if those results can be understood, they still pose a threat for an account of grammaticalization based on a cognitive asymmetry between meanings. In particular, our

model relies on an asymmetric relationship between a lexical meaning and a grammatical meaning. We can contest the relevance of this general picture by arguing, as we stressed throughout this chapter, and in the line of Heine (2002) and Diewald (2002), that grammaticalization happens only in very specific contexts, precisely in contexts in which the lexical meaning is neither obvious nor fully appropriate. The contexts investigated in the psycholinguistic experiment of Hilpert and Saavedra (2016) would correspond to what Diewald (2002) called ‘isolating contexts’, i.e. contexts in which either the grammatical or the lexical meaning is overtly and non-ambiguously favored. They would therefore not correspond to the kind of bridging context or critical context involved in the outset of grammaticalization. It should also be insisted upon that grammaticalization was a highly specific phenomenon, which requires specific links, and specific cognitive efforts concurring for it to be triggered. As the elliptic revolution of celestial bodies was a specific empirical context which favored the abstracting out of the gravity force, grammaticalization results from the abstraction of a schema in highly favorable linguistic circumstances. Of course, the comparison between scientific discovery and grammaticalization is limited, and should not be continued much further.

Nonetheless, we have to acknowledge the result of Hilpert and Saavedra (2016), and face it as it stands. In our model, we shall propose two alternative mechanisms, each one mimicking a cognitive process of the language user. One favors the perspective speaker, the other that of the hearer. In both cases, the asymmetry of the link between the source and the target meanings is crucial. The first one consists in the retrieval of a linguistic form in memory so as to express a given, selected meaning. However, the memory contents of the related meanings interfere with the retrieval task, so that the speaker may happen to select a linguistic form from one of the related meanings instead of the selected meaning.

The priming directionality evidenced by Hilpert and Saavedra (2016), from grammatical meaning to lexical meaning, supports well this picture. Indeed, the priming shows that the grammatical meaning is able to activate the lexical meaning — as we suggested, by adding features on a more abstract and simpler semantic schema. Thus, during the retrieval task, the activation of the grammatical meaning also activates the related, more lexical meanings, so that a form expressing a lexical meaning can be recruited so as to produce the grammatical meaning. As a result, the unilateral direction of the occurrences’ invasion is expected to go in the opposite way of the cognitive activation. Therefore, the proposed mechanism is compatible with the specific features of priming.

Things do not go along so well in the hearer perspective. This second mechanism consists in the elicitation of a supplementary meaning once a first meaning has been recovered from the utterance of a linguistic form. Hence, an occurrence of the linguistic form is recorded in the memories associated with both the intended meaning, and the additional, elicited meaning. For the semantic replacement to have a change to happen, the elicitation must be unidirectional. Therefore, if there is no priming (i.e. co-activation) of the lexical meaning towards the grammatical meaning, this scenario does not hold. If, furthermore, the priming goes the other way around, then we would expect the form expressing the grammatical meaning to invade the lexical meaning, and not the converse.

However, in certain cases, there is an asymmetric relationship between a lexical meaning and a less concrete, usually more pragmatic meaning. The inference mechanism (or conversational implicature) would not relate an elaborated, abstract meaning, difficult to coin with a proper word, to a concrete meaning, which can be more easily expressed. In this case, the asymmetry clearly goes towards the more grammatical. Hilpert and Saavedra (2016) show that priming and inference are two very distinct cognitive processes, which can work in opposite ways. In the perspective of our modeling attempt, this would suggest that some changes may be supported by priming association, other by inference. In the first case, the speaker mechanism would be responsible for the change while, in the second, the change would be carried on by the hearer mechanism. Whether these two mechanisms can pair up and work together — i.e., whether there are links whose unidirectionality is compatible with both — is another question. It can be that an inference relationship between two meanings and a priming co-exist. As they correspond to two different cognitive processes (priming is an automatized co-activation of concepts, inference is the cognitive effort devoted to decipher the intentions of the speaker), they can work in opposite ways without suffering from contradiction.

Regarding why inferences do not manifest themselves in the experiments of Hilpert and Saavedra (2016), it might be that the sentences to be processed do not correspond to any actual situations of use. If one has to process a sentence *per se*, there is no point in inferring the speaker's intentions, for there would be no speaker in such a case. Moreover, it might just be that inferences do manifest in the results of Hilpert and Saavedra (2016). Indeed, drawing inferences require a further processing time (Lenzner et al., 2010). Thus, because the lexical meaning has been activated, retrieving the grammatical meaning costs the inference processing time, which would not be the case if the grammatical meaning was retrieved directly, as a stored construction. Hence, the same experimental evidence reported in Hilpert and Saavedra (2016) can speak both in favor of a lexical > grammatical inference and of a grammatical > lexical priming. This would conveniently allow for both of our mechanisms to benefit from the two-sided associative asymmetry between the two meanings.

What this discussion highlights is that the modeling of language change can only benefit from further psycholinguistic experiments and results, which help to clarify and constrain the overall cognitive picture of language use. As these matters are still largely unclear, caution is advised, and the interpretation of our model, as it crucially relies on the assumption that asymmetrical associations do exist between lexical and grammatical meanings, would be severely crippled if those associations were proved to be dubious. We believe anyway that language change cannot receive a proper and satisfying explanation as long as it does not rely on the cognitive handling of language and all related psychological processes.

7.6 Communication is efficient and successful

The last hypothesis on which the model rests is certainly the boldest one, and the most debatable one. It can be stated as such: Communication is a perfect process, which does not suffer from any mistake. This is, most evidently, wrong. However,

communication is usually successful, and it is much more often so than it is not. Misunderstandings are many, to be sure, but their roots lie in differences between general ideas about the world, rather than in language errors. Language, for the most of the time, is a powerful, reliable, efficient tool.

7.6.1 Communication errors

If we adhere to this statement, then would it be reasonable to attribute language change to specific language errors? This would be equivalent to the assumption of a neutral evolution in biological evolutionary processes. However, the analogy between language evolution and biological evolution is limited. There are no such thing as parental lineages in language evolution; language users and the language itself do not belong to the same tight relationship than the gene carriers and the genome. Neutral evolution can have a crucial importance in biological evolution, because if an individual carries mutated genes, he has no choice but to live with them, and he has good probability to transmit them to its offspring. If a language speaker introduces inadvertently arbitrary ‘mutations’ in his speech, then he is not bound to stick to them and will probably correct himself at the next occurrence. The hearer, similarly, may not record this distorted occurrence, and might even correct it unconsciously — as a reader can do when reading a misspelled word. Furthermore, the hearer is likely to receive many inputs from other speakers, who have no particular reason to repeat the initial error. Among that many inputs, the distorted occurrence will have no weight. It will be foreshadowed by the amount of other occurrences. The speaker itself might not repeat this error again, if it is random as we expect from a neutral evolution. This feeble, infinitesimal influence of an isolated occurrence is what has been called the ‘Threshold Problem’ (Nettle, 1999). Quite often, as we have seen, modeling attempts tend to overcome this problem by postulating a non-zero initial frequency of the novelty.

It might be the case, however, that the error is not isolated, but repeated, by multiple speakers over multiple times. For this to happen, there must be a bias in language production which favors the occurrence of this error. In other terms, the error must be invited by a specific language use already entrenched in the community. Then, this variation stands as an error only from the point of view of an official, institutionalized language Academy; it however reflects the living, interconnected organization of language. For instance, in French, there is a recurring error in the usage of the ‘passé composé’ tense which consists in replacing the ‘participe passé’ (an adjectival form of the verb) by an infinitive. For instance:

Du coup, et pour être honnête avec toi j’ai piger ça ya pas longtemps en remontant une team de zero[.]

As a result, and to be honest with you, I got it not long ago by building up a new team from scratch. [https://www.wakfu.com/fr/forum/, 2016]

where the correct form would be *j’ai pigé*. A Google search, for what it is worth, sends back 5,500 results for “j’ai piger” (to be compared with 49,000 forms for the sanctioned form “j’ai pigé”). The pervasiveness of this mistake is all the more striking with the variation *je suis allé/je suis aller*, sending back respectively 612,000 and 576,000 results on Google, which is the same order of magnitude in both cases. Symptomatically, automatized websites of translation of language chunks, such as Reverso and

Linguee, offer an English equivalent for the string ‘je suis aller’, hinting that it has been recognized as a recurrent form in French.

It should be stressed that, phonetically, the two forms are identical. However, the frequency of this spelling error shows that an analogical reanalysis is going on, which may be indicative of a greater transformation. The past tense is indeed reinterpreted within the paradigm of verbal constructions $[V + V_{\text{infinitive}}]$, a pattern which is used for the future tense $[\text{aller} + \{V_{\text{infinitive}}\}]$ and for modalities $[\text{pouvoir} + \{V_{\text{infinitive}}\}]$. In this case, the past tense $[\{\text{avoir}, \text{être}\} + \{\text{participate}\}]$ is reinterpreted according to the former construction, a reinterpretation favored by the phonetical identity between the two.

This reinterpretation is further revealed as the pattern sanctions new occurrences phonetically incompatible with the $[\{\text{avoir}, \text{être}\} + V_{\text{adjectival}}]$ construction, for instance:

J’ai finir par comprendre que ce n’est pas l’avis des autres qui fait qu’on avance
mais plutôt qui fait qu’on s’enfoncé[.]

I have come to understood that the others’ opinions do not help us to go forward
but rather drag us down. [\[https://www.facebook.com, 2015\]](https://www.facebook.com)

In this occurrence, *finir* is phonetically distinct from *fini*. This kind of analogical reanalysis does not only belong to the wild realm of World Wide Web where language can reach unbelievable levels of distortion, but is also heard in the streets. I have deliberately chosen two examples where the remaining of the sentence is orthographically correct, and syntactically usual. The hypothesis that the *r* in *j’ai finir* comes from a typing mistake is ruled out by the fact that that *r* is placed far from *i*, *p* and the spacebar on a keyboard, so it probably does not result from a simple slip of the fingers.

This new schema $[V_{\text{mode/tense}} + V_{\text{main, infinitive}}]$ applied to past tense may be shocking from a language policy point of view, but it harmonizes quite well with the schema of the future tense, and makes a lot of sense within the overall structure of French mood and tense system. As a result, the parallel between the two tenses, past and future, becomes all the more striking — the letter *v* being the only difference between the two in most cases (all singular forms + third person plural form) when the *avoir* auxiliary is used to form the past tense. Interestingly, $[\text{être} + \{V_{\text{adjectival}}\}]$ seems to be slightly more resilient than $[\text{avoir} + \{V_{\text{adjectival}}\}]$ in the case where $V_{\text{adjectival}}$ is not phonetically identical with $V_{\text{infinitive}}$, which may be a consequence of the fact that the morphology of the verb *être* does not parallel the one of *aller*, so that the analogy would be less compelling. This latter remark belongs, of course, to the conjectural domain.

This example suggests, in any case, that what counts as a language ‘error’ might be the response of the speakers to a specific invitation of the language organization. It thus seems legitimate to picture those ‘errors’ in terms of the invasion, by a given linguistic form, of a semantic subunit of the semantic territory, this invasion being sanctioned and channeled by a link resulting from the structural organization of this semantic territory by a particular language. In many cases, such innovations may not be felt as errors at all (in case of a semantic change, for instance, which has no influence on the surface form of the linguistic item). I will therefore assume that the majority of language changes can be seen as resulting from language use, this use favoring innovations through the network organization of the semantic territory.

In short, the view of linguistic of communication which underlies the model may be summarized by these three maxims:

1. Communication is efficient and successful;
2. We mean more than we say;
3. We understand more than we are told.

Of course, the last two maxims depend on each other. If we could not mean more than we say, then we would not be able to understand more than we are told. Similarly, if our understanding of linguistic utterances was limited to their immediate meaning, it would be pointless to hint at some additional meaning when producing one. In both cases, we rely on the same inferences.

The main idea behind these hypotheses is that language is changing because of this constrained overflow of meaning beyond the raw meaning of the linguistic forms of an utterance. Pun intended, language changes not because of its flaws, but because of its flows. Meaning flows, specifically, are the source and the guideline of meaning change. Without these fluxes, communication would be near impossible, for the literal meaning is often too unspecific and could correspond to a wide array of unrelated interpretations. Meaning must thus be inflated with input stemming both from pragmatic inferences, the overall discourse content, and the general lines of cognitive reasoning.

7.6.2 Mean field hypothesis

If we assume that communication is, as a rule, successful, then it automatically implies that the majority of language users have comparable language knowledge. They are able to draw the same inferences, to work out the same associations, to grasp the same analogical innovations. This entails some uniformity in the language community, some common sharing of the network connecting together the semantic units. It has been shown, furthermore, that different speakers may use different linguistic variants with similar proportions (Labov, 1994). Thus, it makes sense to posit that all language users of a given linguistic community are characterized by the same language representation. This strong assumption is known as the representative agent hypothesis in Economical Sciences, and is akin to the mean field approximation in Physics. Albeit it cannot be held as an absolute truth, it seems indeed legitimate to assume that the variation between two different language representations (each one associated with a different language user) is, in most relevant cases, smaller than the variation in time of either one of these representations. This amounts to foreshadowing the sociolinguistic aspects of language change, whose importance is nonetheless well recognized and undeniable. Yet, I already provided arguments to support this choice and to justify such a strong approximation in the context of meaning change.

If one is reluctant to accept that all language users share a similar representation in average, we can also consider that the populations of occurrences over the semantic network reflect the language of the community as a whole, a sort of common, akashic memory. This view holds all the more if every member of the community is closely connected with all others, in other terms, if the shortest path between two individuals

is small in average. This is fortunately true for small-world networks, which is representative of social communities (Newman and Park, 2003). Yet this is exactly the view held by the representative agent hypothesis: the linguistic community is connected and homogeneous enough, so that no strong heterogeneities have to be taken into account.

Note also that the difference with a social diffusion picture is only a matter of timescales; there is no clear dichotomy between the cognitive and social diffusions, since both are, obviously, necessary to account for any language change. However, sociolinguistic models, such as most agent-based models, make the assumption that the cognitive diffusion is much faster than the social diffusion (speakers internalize the change in less time than it takes for this change to spread over the whole community), so that they consider the agents to flip binarily from one variant to another. On the contrary, we make the reverse assumption, according to which language users coordinate on a rather short timescale, while it takes a longer time for an individual to switch habits and to forget the old variant entirely so that only the novelty would be produced.

Both diffusions can also be encompassed at once. This would be the case in the Utterance Selection Model (Baxter et al., 2006), in which the agents interact on a social network, as well as they are described by slowly changing preferences towards one or the other variant (represented by a perceived frequency of each). Our model could also account for a social network of agents characterized by semantic networks — which would result in a network of networks —, especially considering that we have different rules for the hearer and the speaker roles of the language user. Since the aim was to keep the model simple and tractable, it seemed however better advised to stick to the approximation that the social diffusion timescale is much shorter than the cognitive diffusion timescale, so that the ‘mean field’ representative agent approach would be justified. This crucial question of timescales is nonetheless worth pursuing, yet it lies beyond the scope of the present work.

7.7 Summary

The model that shall be presented in the following section is thus supported by the following hypotheses:

- All language users of a community share a similar language representation, so that the evolution of the language of that community can be accounted for through the evolution of the language of a single, fictive, representative agent, whose own language representation is the average over all community members;
- Language is depicted as a semantic territory divided into many semantic units;
- Linguistic forms are populating these units through many occurrences, reflecting the past linguistic experiences of the representative agent;
- However, the capacity of these units is held at a constant level, so that new occurrences can only be recorded if past ones are erased;
- Those semantic units are related within a network structure;
- The links of this network can be weighted, and directional;

- All semantic units are always populated at maximum capacity;
- Occurrences of a form can use the edges of the network as channels to invade a new semantic unit.

The hypotheses that we rely on are crucial to relate our model to the phenomenology of language, but they are not bound to be true, if it needs to be said. It seemed important, however, to show that the specific ingredients of the model can be integrated into a consistent picture of language organization and language representation. We tried, especially, to show that it was possible to bring together a competition-based view of change, and a context-based account of the steps of this change. Thus, we can understand how the empirical S-curves resulting from such competitions relate to the subtle structural and semantic changes occurring in grammaticalization and constructionalization processes. Only through this effort of theorization can the empirical modeling of frequency data shed light over the processes at work in language change.

Chapter 8

A model of semantic expansion

In this chapter we propose a model of semantic expansion based on the hypotheses previously exposed. Its principal features are:

- A minimal semantic network of two sites (two different meanings).
- Two competing species of occurrences, each one populating one of either two sites initially.
- An asymmetric, unilateral influence of one semantic site on another.
- A usage-based dynamics based on memory retrieval and mnemonic limitations.

The goal of this model is to reproduce the main qualitative features of the semantic expansion brought forth in chapter 4, especially the S-curve, and to give a clear picture of the simplest competition possible.

8.1 Presentation of the model

We consider a minimal networks of two sites A and B , with a unidirectional link between C_0 and C_1 of weight γ . We also consider two different species or forms X and Y , whose occupation numbers in site S at a given time t in C_0 and C_1 are respectively noted $N_X^S(t)$ and $N_Y^S(t)$. The tokens of each species (occurrences) correspond to the recording of past linguistic experiences in the memory associated with each of the sites.

The sites have a limited capacity of occurrences, noted respectively M_{C_0} and M_{C_1} , and we shall assume that $M_{C_0} = M_{C_1} = M$. The *actual* frequency $x_F^S(t)$ of form F in site S at a given time t is thus given by:

$$x_F^S(t) = \frac{N_F^S}{M} \quad (8.1)$$

As the relation $N_X^S + N_Y^S = M$ holds for all times and for all sites S , we can focus on either one of those two quantities, say on N_X^S . In the following, we shall henceforth drop the X index.

Initially, we have $N^{C_0} = M$ and $N^{C_1} = 0$ (X is populating C_0 , Y is populating C_1). Since C_0 influences C_1 but not the reverse, the population of A is not expected

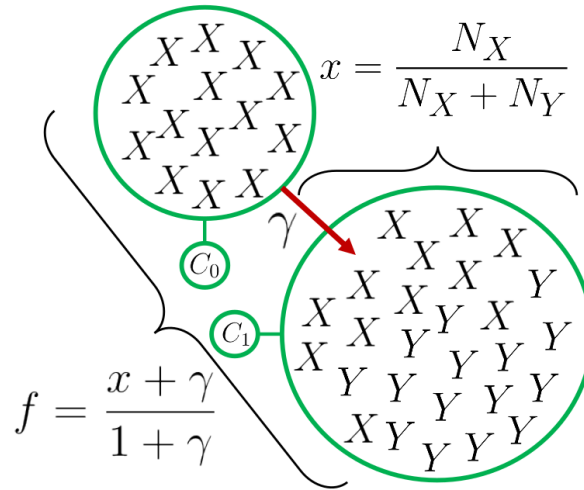


Figure 8.1: Schematic illustration of the difference between the actual frequency x and the effective frequency f of form X in site C_1 .

to change, so that $N^A(t) = M, \forall t$. We shall therefore focus on what is happening in site C_1 , we can also drop the site index, knowing that all quantities (X, N) refer to form X in site C_1 .

We define another quantity as well, which is the *effective* frequency, f (Fig. 8.1). This frequency accounts for the fact that the perceived occupation of X in site C_1 acknowledges that X already occupies C_0 and that C_0 influences C_1 with weight γ . Hence:

$$\begin{aligned} f(t) &= \frac{N^B(t) + \gamma N^A(t)}{M^B + \gamma M^A} \\ &= \frac{x + \gamma}{1 + \gamma} \end{aligned} \tag{8.2}$$

Initially, the actual frequency of X in C_1 is 0: the form X is never actually used to express meaning C_1 so there is no linguistic experience corresponding to it in memory. However, the fact that X is used to express C_0 , and that meaning C_0 elicits meaning C_1 , entails that X is felt as if to have expressed C_1 . Hence, the cognitive network-based organization of language can trigger the semantic expansion, as we shall see.

8.1.1 A random walk in frequency space

The dynamics of the system has been designed so as to mimick repeated language use by a community of speakers, hereby played by a representative agent. Each time step corresponds to the utterance of a linguistic form to express meaning C_1 and covers the following events:

1. A form, either X or Y , is randomly chosen according to its effective frequency of use, f , to express meaning C_1 .
2. An occurrence of this form is encoded in the memory so as to record this linguistic experience.

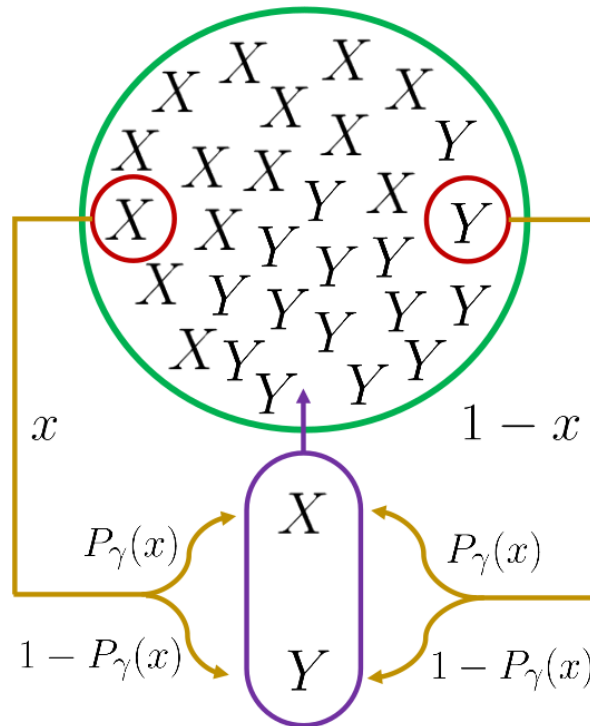


Figure 8.2: Schematic picture of the dynamics. Purple cartouche: one form, either X or Y , is chosen with respective probabilities $P_\gamma(x)$ and $1 - P_\gamma(x)$ and the corresponding occurrence is recorded in memory. Red circle: to keep a constant memory size, one occurrence, chosen with uniform probability, is erased from memory. An occurrence of X (resp. Y) is erased with probability x (resp. $1 - x$). The occupation numbers N_X and N_Y will change whenever there is a mismatch between the produced occurrence and the erased one.

3. As the memory capacity is finite and equal to M , a former occurrence is randomly erased.

This process is schematically summarized on figure 8.2.

Picking up a form

In the first step, the probability to choose either X or Y is not proportional to the effective frequency and is a non-linear function $P(f)$ instead. This function must have the following properties:

1. It must be bound between 0 and 1 (it is a probability).
2. It should be symmetric under the transformation $f \leftrightarrow 1 - f$.
3. It must be non-linear, so as to respect $P'(0) = 0$.

The second condition is debatable. At first sight, it seems reasonable: we can write down the dynamics from the perspective of X , so that $f = f_X$, or from the

perspective of Y , so that $f = f_Y$. In the second case, we thus have $1 - f = f_X$. Thus, the symmetry condition only states that the dynamics must be invariant under a change of perspective: what X is and what Y is has no importance, only their effective frequencies matter. In other terms, we must have:

$$P(f) + P(1 - f) = 1. \quad (8.3)$$

The symmetry would be for instance broken if speakers were sensitive to the past frequency history of the form (Stadler et al., 2016), or more generally if the choice of either variant was described by two different functions, P_X and P_Y (for instance if X is intrinsically advantageous over Y , be it for internally linguistic or sociolinguistic reasons). The symmetry condition actually amount to $P_X = P_Y = P$, in other terms, the choice depends on the forms through their frequencies only.

The non-linearity condition is important because we want the linguistic system to be protected against change. If the choice probability P was linear in f , the frequency of any new variant would rise instantaneously. Our two-sites minimal competition setting would lead to a sudden increase of X no matter the weight of the influence of C_0 on C_1 . Furthermore, the change would only slow down as the frequency gets higher and higher, and would take a very long time to reach completion. Hence, it is necessary to assume that P is non-linear. To stress this change-resilience, the further condition $P'(0)$ seems appropriate.

A simple third-order polynomial can meet the three conditions:

$$P(f) = -2f^3 + 3f^2, \quad (8.4)$$

however it leaves us with no free parameter to adjust the steepness of the non-linearity. The derivative $P'(1/2)$, which quantifies how much the system is tipped by an unbalance between the two forms, is indeed set to $3/2$, though this could be overcome by positing a five-order polynomial. Another possibility could be:

$$P(f) = \frac{f^2}{1 - 2f(1 - f)} \quad (8.5)$$

which has also one free parameter and a steepness fixed to $P'(1/2) = 2$.

We instead propose the following form for $P(f)$:

$$P(f) = \frac{1}{2} \left[1 + \tanh \left(\beta \frac{f - (1 - f)}{\sqrt{f(1 - f)}} \right) \right], \quad (8.6)$$

where β is a parameter which quantifies the non-linearity: the smaller it gets, the weaker is the non-linearity. It allows to tune directly the steepness of the curve, since the derivative $P'(1/2)$ is now equal to 2β . β should be no smaller than $1/2$, for the slope cannot be smaller than 1.

Note, however, that none of these choices is easily generalized to more than two competing forms. The symmetry criterion would become, for K different forms:

$$\sum_{k=1}^K P(f_k) = 1 \quad (8.7)$$

with the special case that if $f_k = 1/K$, $\forall k$, then we must also have $P(f_k) = 1/K$. If we further want to make room for a varying number K of competing species (the number of competitors has no reason to be always the same over time), then this should hold for any K . There is no other possibility, in that case, for P but to be linear. The only way out of this problem would be to accept that, in some cases, the speakers can fail to retrieve an occurrence in their memory (for instance if all competing forms are under the $1/2$ threshold), which is equivalent to release the condition that the production probabilities of all forms must sum up to 1. In such a case, nothing would happen and we would wait for the next timestep.

$P(f)$ depends on x through f . To make this dependence explicit, we will make use of the notation $P_\gamma(x)$ defined by:

$$P_\gamma(x) = P\left(\frac{x + \gamma}{1 + \gamma}\right) = P(f) \quad (8.8)$$

Understanding the occurrence

As we made a hypothesis of perfect communication, the occurrence is always understood as expressing the meaning it has been chosen for, and recorded with full reliability in the corresponding site. The form chosen in the first step is therefore recorded in C_1 , even though it may more commonly express C_0 . Note that this could represent either the speaker, recording what has just been uttered (much alike a Pólya urn mechanism), or the hearer, receiving the utterance and storing it in the memory. As we deal with a representative agent supporting both roles, it does not make any difference.

Keeping up memory size

To keep up memory size at its constant value M , an occurrence has to be erased any time a new one is recorded. As this deleting process concerns the C_1 site alone, it involves actual frequency instead of effective frequency. The symmetry principle (this time phrased in terms of x instead of f) is still to be respected.

The simplest choice would be an erasure probability proportional to the actual frequency, which reflects the picking up of an occurrence with a uniform probability:

$$E(x) = x \quad (8.9)$$

One can favor other functions, for instance to account for the fact that it may be advantageous to remove with higher probability occurrences existing in large number in the memory, for instance:

$$E(x) = \frac{e^{\alpha x} - 1}{e^{\alpha x} + e^{\alpha(1-x)} - 2}, \quad (8.10)$$

where α quantifies how much the frequent forms are punished and the infrequent forms encouraged (the limit $\alpha \rightarrow 0$ corresponds to the uniform proposal).

However, this ‘curiosity’ principle could also be replaced by a ‘semiotic’ principle (one unit of meaning, one linguistic form), so that it would be better to clean the

memory out of all parasite occurrences not corresponding to the dominant form. A possible choice would be:

$$E(x) = \frac{e^{-\alpha x} - 1}{e^{-\alpha x} + e^{-\alpha(1-x)} - 2}, \quad (8.11)$$

or the following:

$$E(x) = \frac{\sqrt{x}}{\sqrt{x} + \sqrt{1-x}}. \quad (8.12)$$

As both principles seem plausible, the neutral, uniform choice, seems as suitable as any and has the advantage of simplicity.

It would also be possible to erase the oldest occurrence, or to assign a probability of erasure depending on how long the occurrence has stayed in memory. However, if it were so, the model would become Markovian in the whole memory, instead of being Markovian in the sole frequency (Markovian in M variables instead of one). As this would significantly increase the complexity of the model, we restrict ourselves to the simplest choice.

The model therefore depends on three parameters: M , the size of the memory, γ , the strength of the influence C_0 exerts over C_1 , and β , the steepness of the production probability we have chosen.

Master equation

The dynamics previously described amounts to a random walk in memory occupation between $N = 0$ and $N = M$, with the following jump probabilities:

$$\begin{cases} \mathbb{W}(N \rightarrow N+1) = P_\gamma\left(\frac{N}{M}\right) \left(1 - E\left(\frac{N}{M}\right)\right) \\ \mathbb{W}(N \rightarrow N-1) = \left(1 - P_\gamma\left(\frac{N}{M}\right)\right) E\left(\frac{N}{M}\right) \\ \mathbb{W}(N \rightarrow N) = P_\gamma\left(\frac{N}{M}\right) E\left(\frac{N}{M}\right) + \left(1 - P_\gamma\left(\frac{N}{M}\right)\right) \left(1 - E\left(\frac{N}{M}\right)\right) \end{cases}, \quad (8.13)$$

where the symbol \mathbb{W} stands for a probability. Since $E(0)$ and $E(1)$ are respectively equal to 0 and 1, there are no special boundary conditions. Furthermore, since $P_\gamma(1) = 1$, the site $N = M$ of the walk plays the role of an absorbing boundary.

The whole process can be rewritten as a master equation:

$$\begin{aligned} \mathbb{P}(N, t+1) = & \mathbb{P}(N+1, t)\mathbb{W}(N+1 \rightarrow N) + \mathbb{P}(N-1, t)\mathbb{W}(N-1 \rightarrow N) \\ & + \mathbb{P}(N, t)\mathbb{W}(N \rightarrow N) \end{aligned}, \quad (8.14)$$

where $\mathbb{P}(N, t)$ stands for being at site N at time t , and with the initial condition:

$$\mathbb{P}(N, 0) = \delta(N, 0), \quad (8.15)$$

δ being the Kronecker δ .

If we now introduce $R_N = \mathbb{W}(N \rightarrow N + 1)$, the probability of a forward jump (to the right), and $L_N = \mathbb{W}(N \rightarrow N - 1)$ that of a backward jump (to the left), we can write:

$$\boxed{\mathbb{P}(N, t + 1) = R_{N-1}\mathbb{P}(N - 1, t) + L_{N+1}\mathbb{P}(N + 1, t) + (1 - R_N - L_N)\mathbb{P}(N, t)} \quad (8.16)$$

We can also define the net flux $\Delta\mathbb{P}(N, t)$ through site N at time t by:

$$\Delta\mathbb{P}(N, t) = \mathbb{P}(N, t + 1) - \mathbb{P}(N, t) \quad (8.17)$$

whose evolution is given by:

$$\begin{aligned} \Delta\mathbb{P}(N, t) = & L_{N+1}\mathbb{P}(N + 1, t) + R_{N-1}\mathbb{P}(N - 1, t) \\ & - (R_N + L_N)\mathbb{P}(N, t) \end{aligned} \quad (8.18)$$

This is actually a generic one-step stochastic process, as defined by Van Kampen (1992, p.139), with a reflexive boundary at $N = 0$ and an absorbing one for $N = M$, and has been widely used to describe competition scenarios between two populations (Kendall, 1949; Brauer et al., 2001; Blythe and McKane, 2007). The originality of our model therefore lies in the specific R_N and L_N coefficients.

8.1.2 Deterministic approximation

If we now write everything down in terms of x instead of N , we get, with the additional notation $\Delta x = 1/M$, :

$$\begin{aligned} \Delta\mathbb{P}(x, t) = & \mathbb{P}(x + \Delta x, t)L(x + \Delta x) + \mathbb{P}(x - \Delta x, t)R(x - \Delta x) \\ & - (L(x) + R(x))\mathbb{P}(x, t) \end{aligned} \quad (8.19)$$

Under the approximation that Δx is very small and stands as a unit of time ($dt = 1/M$), and defining $v(x) = R(x) - L(x)$, we can consider the first-order expansion of the previous equation:

$$\frac{\partial\mathbb{P}(x, t)}{\partial t} = -\frac{\partial}{\partial x} [v(x)\mathbb{P}(x, t)] \quad (8.20)$$

As $\mathbb{P}(x, t)$ is initially a Dirac delta, it will stay so throughout this first-order evolution, so that we can write $\mathbb{P}(x, t) = \delta(x - y(t))$ where $y(t)$ is the position of the Dirac delta peak, evolving with time. If we perform a Laplace transform on both sides of the equation, we get:

$$\int \frac{\partial}{\partial t} [\delta(x - y(t))] e^{-sx} dx = - \int \frac{\partial}{\partial x} [v(x)\delta(x - y(t))] e^{-sx} dx \quad (8.21)$$

We first derive with respect to time:

$$-\dot{y} \int \frac{\partial}{\partial x} [\delta(x - y(t))] e^{-sx} dx = - \int \frac{\partial}{\partial x} [v(x)\delta(x - y(t))] e^{-sx} dx \quad (8.22)$$

Then we perform the integration by parts on both sides:

$$-s\dot{y} \int \delta(x - y(t)) e^{-sx} dx = -s \int v(x)\delta(x - y(t)) e^{-sx} dx \quad (8.23)$$

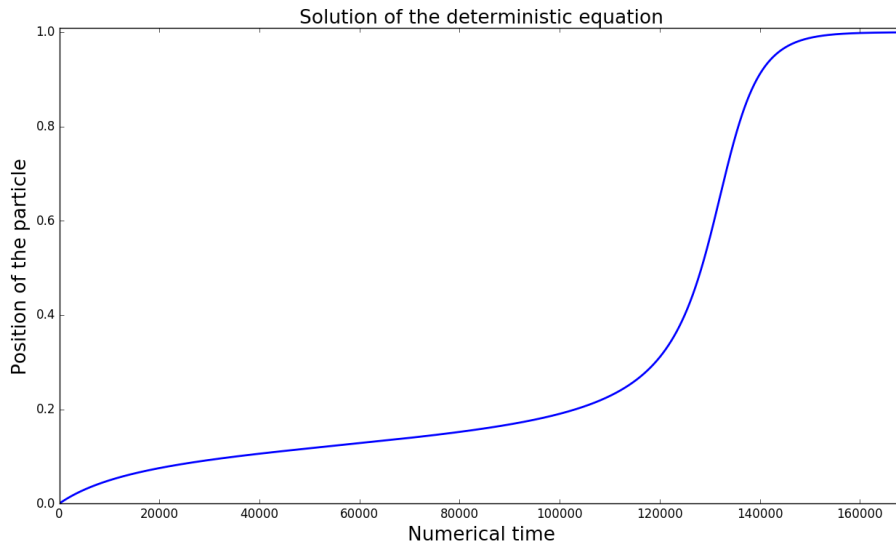


Figure 8.3: Time evolution of the walker's position according to the deterministic evolution equation. $M = 5000$, $\gamma = 0.225$, $\beta = 1.0$.

Then we integrate over the Dirac deltas:

$$-s\dot{y}e^{-sy} = -sv(y)e^{-sy}. \quad (8.24)$$

We thus obtain:

$$\dot{y} = R(y) - L(y). \quad (8.25)$$

Without any risk of confusion we can note x instead of y the deterministic position. Applying the definitions of $L(x)$ and $R(x)$, we finally get:

$$\boxed{\dot{x} = P_\gamma(x) - x}. \quad (8.26)$$

Solution

The differential equation for the walker position (8.26) can be solved in the following way:

$$\int_0^x \frac{du}{P_\gamma(u) - u} = t \quad (8.27)$$

This fully defines $t(x)$, the time needed to reach position x . By reverting the axis, we can therefore obtain a picture of $x(t)$, the time evolution of the position of the walker (Fig. 8.3).

We see that the solution meets our qualitative expectations: first, a sudden, short rise to a low value of frequency; some lingering of the system around this value; a sudden, S-shaped frequency rise.

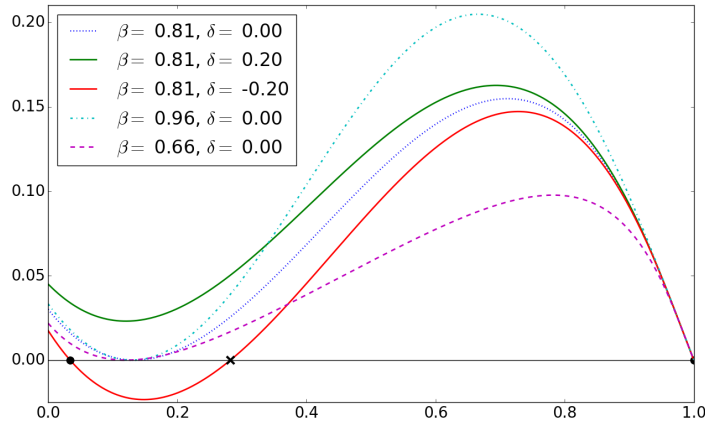


Figure 8.4: $P_\gamma(x) - x$ as a function of x , for different values of the steepness β and different distances δ to the critical threshold ($\delta = (\gamma - \gamma_c)/\gamma_c$). Intersections of the curve with the X-axis represent fixed points of the system. Stable fixed points (associated with a negative slope of the curve at this point) have been marked with a black dot, unstable ones (associated with a positive slope) with a black cross. The position $x = 1$ is a fixed point no matter the value of γ . A saddle-node bifurcation, with the creation of a pair of fixed points, occurs whenever γ is lower than its critical value.

Fixed points and bifurcation

The process is now reduced to a simple dynamical systems. We can now look for its fixed points, i.e. the set of values $\{x^*\}$ such that:

$$\left. \frac{dx}{dt} \right|_{x=x^*} = 0 \quad (8.28)$$

or equivalently such that:

$$P_\gamma(x^*) = x^*. \quad (8.29)$$

$x^* = 1$ is always a fixed point (remember that $N = M$ was an absorbing boundary). If γ is low enough, there can be two more fixed points, one unstable and the other stable (Fig. 8.4). The strength γ of the influence of C_0 over C_1 serves thus as a control parameter of a saddle-node bifurcation (Fig. 8.5).

The threshold value γ_c below which the bifurcation occurs is solution of the following system:

$$\begin{cases} P_{\gamma_c}(x_c) = x_c \\ \left. \frac{dP_{\gamma_c}(x)}{dx} \right|_{x=x_c} = 1 \end{cases} \quad (8.30)$$

By playing with the algebra, it can be shown that those two conditions amount to

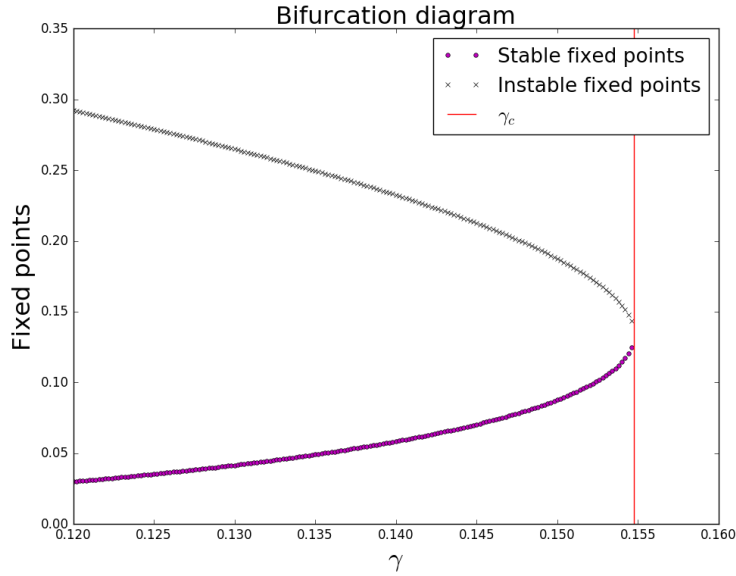


Figure 8.5: Bifurcation diagram of the random walk. Below a critical value γ_c of the γ parameter, a saddle-node bifurcation occurs and two fixed points, one stable and the other unstable, appear. The fixed point $x = 1$, always stable no matter the value of γ , has not been shown.

the following system:

$$\begin{cases} x_c^4 + (3\gamma_c - 1)x_c^3 + [\beta^2(1 + \gamma_c)^4 + 3\gamma_c(\gamma_c - 1)]x_c^2 + \gamma_c^2(\gamma_c - 3)x_c - \gamma_c^3 = 0 \\ \beta^2\gamma_c^2 + [2\beta^2(2x_c - 1) - (1 - x_c)\alpha(x_c)]\gamma_c + \beta^2(2x_c - 1)^2 - x_c(1 - x_c)\alpha(x_c) = 0 \end{cases}, \quad (8.31)$$

where $\alpha(x)$ is the function defined by:

$$\alpha(x) = \operatorname{arctanh}^2(2x - 1). \quad (8.32)$$

This can be solved numerically so that, for any given value of β , we can find the critical threshold γ_c as well as the position x_c at which the saddle-point appears (Fig. 8.6a).

Instead of choosing γ as a parameter, we can therefore choose directly the distance δ to the threshold, with δ defined by:

$$\delta = \frac{\gamma - \gamma_c(\beta)}{\gamma_c(\beta)}. \quad (8.33)$$

The corresponding value of the γ parameter will thus be given by:

$$\gamma = \gamma_c(\beta)(1 + \delta) \quad (8.34)$$

The critical threshold γ_c is a monotonic increasing function of β (Fig. 8.6a). However, the critical frequency x_c at which the saddle-node bifurcation occurs is non-monotonic and reaches a maximum value for $\beta = \beta^*$, roughly equal to 0.81 (Fig. 8.6b).

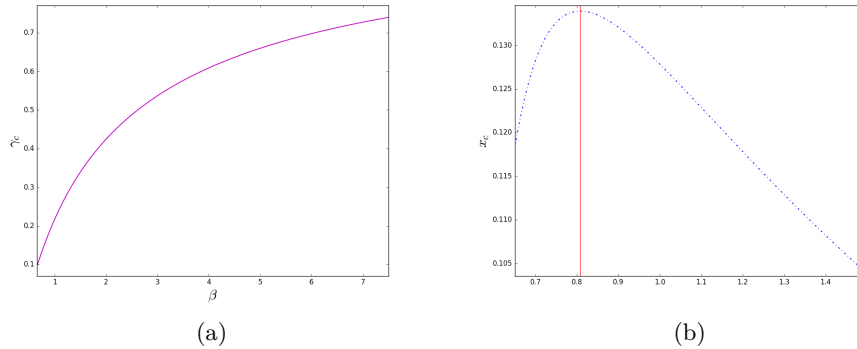


Figure 8.6: (a) Critical threshold γ_c as a function of the steepness parameter β . γ_c is monotonically increasing with β . (b) Critical frequency x_c as a function of the steepness parameter β . x_c displays a maximum for the value $\beta = \beta^*$ roughly equal to 0.81, marked with a red vertical line.

It is interesting to take this value as a reference value, for if the critical value x_c is as high as possible, it allows the new variant X to marginally co-exist with the old one Y even if X cannot take over the meaning. As this synonymy-margin seems compatible with language organization, we shall, as a rule, choose β equal to this value β^* .

Critical slowdown

The differential equation for the walker position (8.26) can be solved in the following way:

$$\int_0^x \frac{du}{P_\gamma(u) - u} = t \quad (8.35)$$

Hence, we can get the time Δt spent to get through an infinitesimal region $[x, x + dx]$:

$$\Delta t(x) = \int_x^{x+dx} \frac{du}{P_\gamma(u) - u} \quad (8.36)$$

From this, it is clear that when $P_\gamma(u) - u$ becomes small, the time spent in the region increases dramatically. If this quantity goes to 0, then the integral diverges and the walker becomes stuck in this region. Hence, according to the deterministic process, there is a critical slowdown when γ is slightly above the threshold. Below this threshold, the walker stops on the first fixed point, associated with a low frequency, and never reaches the fixed point $x = 1$. In this case, the transmission time, i.e. the time required to get through the whole system, becomes infinite.

We can thus consider the inverse transmission time as the order parameter associated with the transition. If we plot this quantity as a function of the control parameter δ (Fig. 8.7), then the bifurcation becomes apparent.

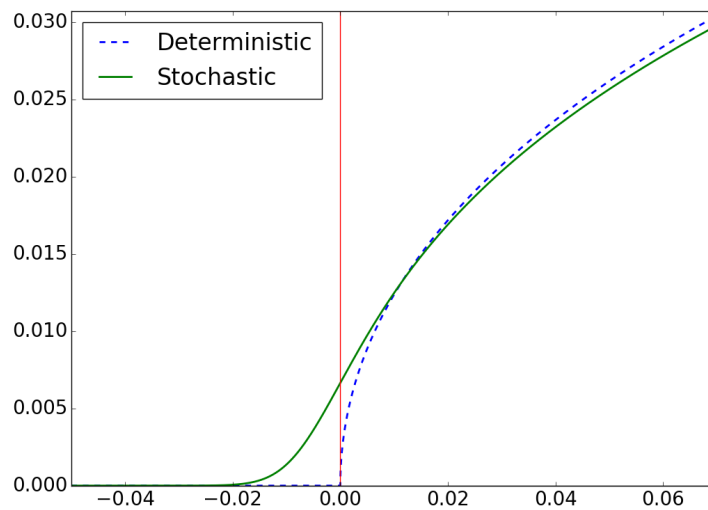


Figure 8.7: Inverse transmission time as a function of the distance δ to threshold, for both the deterministic process and the stochastic process (in the latter case, the transmission time is the average transmission time). Below the threshold, the inverse transmission time is zero: the walker never goes through the system and remains stuck at the first fixed point encountered. $M = 5000$, $\beta = \beta^*$.

Consequence of the bifurcation for the semantic expansion

This bifurcation phenomenon is quite important. Indeed, it implies that the change is not bound to happen, without certain conditions to be met. A model which would lead to change no matter what would overpredict the change. Here, we learn from the bifurcation that change depends on the cognitive strength of the semantic channel which is driving it. If the cognitive link between the two meanings C_0 and C_1 is too weak, then the form X cannot invade the site, even though it has uses which elicits meaning C_1 . The frequency of use in C_1 will therefore remain at a low, marginal value.

It also entails that, if the strength γ is to evolve, especially when the overall language organization comes to favor some reanalysis (as we saw it was the case for *du coup* and *way too*), then it can pass the threshold. Hence, a change which was nearly possible can suddenly meet the conditions to unfold towards completion. We saw for instances that the uses of the *du coup* construction (already slightly idiosyncratic in its semantic and syntactic behavior) were hinting as soon as the XIVth century at the possible semantic expansion which finally happened five centuries later, when the overall semantic bleaching of *coup* at last set the ecological conditions for the change to happen.

The existence of a control parameter for the occurrence of the semantic change is thus an interesting feature of this model, since it provides a better understanding on some features of semantic expansion. Especially, it highlights how a very small change in the language, strengthening the semantic association between two meanings, can trigger an expansion which was lingering for years before that. A remarkable frequency

rise can thus be caused by seemingly insignificant variations in the language. Obviously, this only speaks in favor of a more refined, more detailed, but also less local, linguistic analysis of the change.

Further modeling of the change should thus focus on how changes in frequencies can produce a feedback on the association weights between the different meanings. As those changes are probably non-local, it might be a complicated task. Furthermore, the semantic expansion from site C_0 to C_1 does not necessarily reinforce the link between the two: as Diewald (2002) argued, the link can even get lost in the process.

Anticipating with the next chapter, it would also imply that, if X , in our set-up, happens to expand further and further in the direction of C_1 , then it can more easily lose C_0 . If the links have been strengthened throughout the expanding of X , then its new competitor, coming from behind, could take over all its semantic territory in no time. And another competitor coming after would have yet a clearer trail to expand, so that the semantic evolution of a given language would gradually accelerate along some semantic lines. As this acceleration phenomenon has never been reported before, we can doubt that the feedback from occurrences population change to semantic association weights is as simple as a mere mutual reinforcement loop.

This scenario (that γ increases as occurrences of X comes to populate C_1) would further make the change to rapidly leave the critical region (the region where γ is close to the threshold and the walker stuck at a frequency around the critical value x_c). In that case, there would be no critical slowdown, and hence, no latency. As we saw the empirical importance of latency, this reinforcement is unlikely.

We shall now have a closer look at the behavior of the model in this critical region, so as to understand better this latency phenomenon.

8.1.3 Stochastic behavior in the critical region

As can be seen in Figure 8.7, the true behavior of the model is not as sharp in the critical region as the deterministic approximation may suggest. Hence, the walker has a chance to pass through this region even if γ is slightly below γ_c —though these chances tend to vanish as γ decreases.

We can rewrite the master equation (8.18) by expanding it to the second-order, in a continuous approximation, so as to obtain the corresponding Fokker-Planck equation:

$$\frac{\partial \mathbb{P}(x, t)}{\partial t} = -\frac{\partial}{\partial x} [v(x)\mathbb{P}(x, t)] + \frac{1}{2M} \frac{\partial^2}{\partial x^2} [D(x)\mathbb{P}(x, t)], \quad (8.37)$$

with:

$$v(x) = R(x) - L(x) = P_\gamma(x) - x \quad (8.38)$$

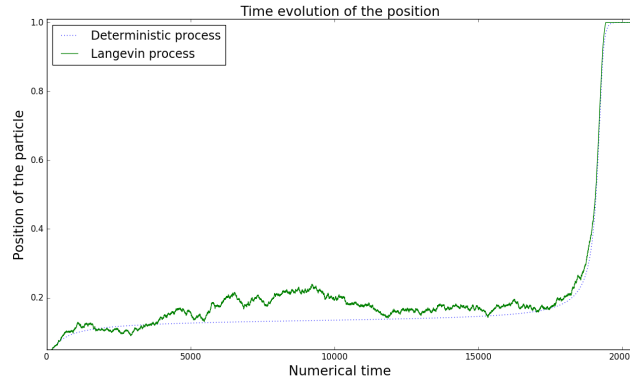
and:

$$D(x) = R(x) + L(x) = (1-x)P_\gamma(x) + x(1-P_\gamma(x)). \quad (8.39)$$

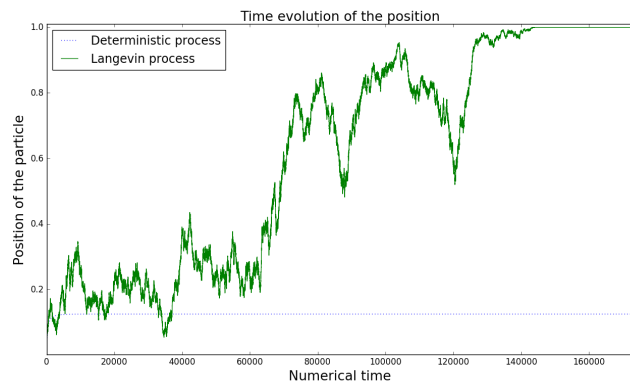
From this equation we can define a corresponding Langevin process (Van Kampen, 1992, p.244):

$$\dot{x} = v(x) + \sqrt{\frac{D(x)}{M}} \xi(t), \quad (8.40)$$

where $\xi(t)$ is a Gaussian white noise, of mean 0 and variance 1, uncorrelated in time.



(a)



(b)

Figure 8.8: (a) Simulation of the deterministic process and of one run of the Langevin process, with $M = 80$, $\beta = \beta^*$ and $\delta = 0.001$. The Langevin process is noisier than its deterministic counterpart, especially in the critical region. (b) The same, with $M = 50$, $\beta = \beta^*$ and $\delta = -0.001$. The Langevin process allows to escape the critical region while the deterministic process remains stuck at the first stable fixed point.

The idea is that the noise should enable the walker to get free of the low frequency stable fixed point so as to escape the critical region and eventually reach the final fixed point $x^* = 1$. We can now simulate, at the same time, the deterministic process, and the equivalent stochastic Langevin process (Fig. 8.8a).

The problem with the Langevin process is that the noise scales as $1/\sqrt{M}$. If M is too big, the noise becomes negligible and the walker cannot pass the critical region. To have the walker transmitted even in the $\delta < 0$ case, the memory size must therefore be very small. Yet, the noise becomes important in all parts of the process and the S-shape frequency rise is destroyed as well (Fig. 8.8b): the stochastic process has now little to do with the deterministic one.

We can therefore model the master equation (8.16) directly (which acts over the discretized space). As the time evolution of a probability distribution is not easy to represent, we focused on two quantities. First, the position $x_p(t)$ associated with the

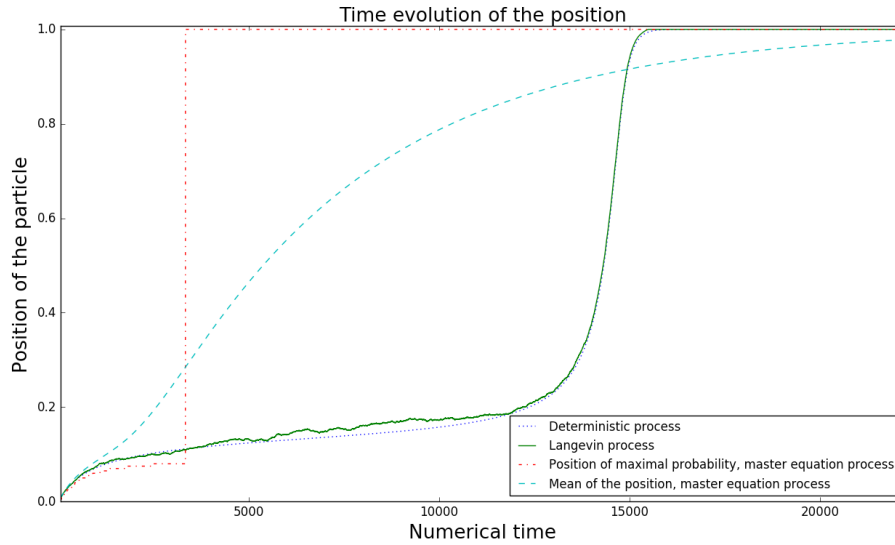


Figure 8.9: Time evolution of 1 - the position $x(t)$ according to the deterministic process (blue dotted line) and the Langevin process (green full line) 2 - the quantities $x_p(t)$ (red dash-dotted line) and $\langle x(t) \rangle$ (cyan dashed line) according to the master equation. $M = 200$, $\delta = 0.01$, $\beta = \beta^*$.

highest probability:

$$x_p(t) = \frac{1}{M} \operatorname{argmax}_N (\mathbb{P}(N, t)) \quad (8.41)$$

and second, the mean $\langle x(t) \rangle$ of the distribution:

$$\langle x(t) \rangle = \frac{1}{M} \sum_{N=0}^1 x \mathbb{P}(N, t). \quad (8.42)$$

While the quantity $\langle x(t) \rangle$ is smooth, $x_p(t)$ shows a discontinuity (Fig. 8.9). This discontinuity can be understood as follows: if we imagine many walkers following the process, in a short time most of them will be trapped in the critical region. Then, some of them can escape the critical region and will quickly move towards the end of the walk (the $x^* = 1$ fixed point). At some point, there would be more walkers to have escaped and have reached the endpoint, than walkers trapped in the critical region. Hence, the peak of the probability distribution will shift from somewhere in the trap region to the end of the walk. This escape mechanism is captured by neither the Langevin process nor the deterministic one.

This can be understood if we look at $R(x)$ and $L(x)$, the jump probabilities (respectively forward and backward). In the critical region, they are almost equal (Fig. 8.10); hence the quantity $v(x)$, used in both the deterministic and in the Langevin processes, vanishes. Yet, the jump probabilities themselves are still high enough for the walker to move around, and eventually allow him to escape.

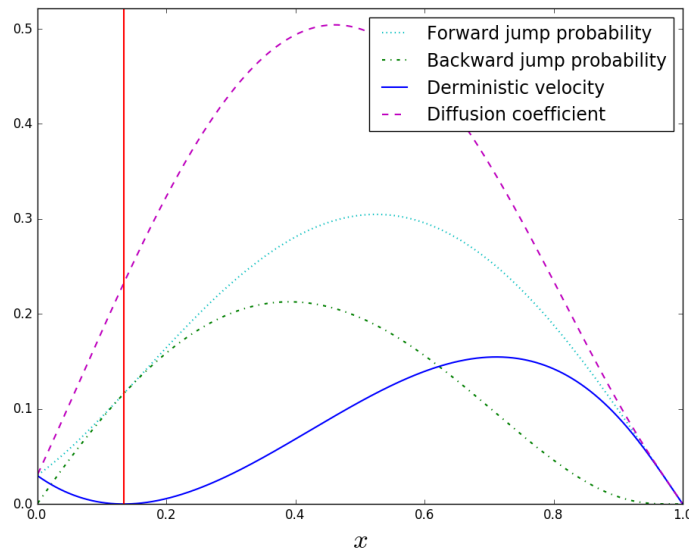


Figure 8.10: Forward and backward jump probabilities, deterministic speed, and diffusion coefficient as a function of the walker’s position. The critical point x_c is indicated with a red line. $M = 5000$, $\delta = 0.0$, $\beta = \beta^*$.

8.2 Theoretical and numerical analysis of the model

We established in chapter 4 how crucial the latency phenomenon was, as it could be a specific signature of a semantic change. This is the reason why we shall devote special efforts to understand latency on our model, so as to build its distribution and compare it to the empirical quantitative features of latency in semantic change that we previously brought to light.

Latency can be seen as resulting from the trapping of the walker in a critical region. As we saw, there is a probability of escape; and once the walker gets out, it will reach the end of the walk with certainty and achieve domination over the meaning at stake. The evolution of the mean position on figure 8.9 almost evokes an exponential law, as if, at each time, there was a given and constant probability to get out.

It follows that the latency time can be defined as the first passage time out of the trap region (minus the first passage time at the entrance of the said region). The problem is, there is no obvious way to define the boundaries of the trap. Before facing this issue, we shall first describe how the first passage times are actually computed.

8.2.1 First passage time

We define $T_{n \rightarrow m}$ the first-passage time at site m , starting from site n . $T_{0 \rightarrow M}$, for instance, corresponds to the transmission time and $T_{n \rightarrow n}$, the time to reach n starting from n , is 0.

Rather than computing the whole probability distribution of these quantities, we shall compute their two first moments only, though theoretically, we could obtain any.

The probability $\mathbb{P}(T_{n \rightarrow m} = t)$ that the first passage time at m starting from n is equal to t obeys the following relation:

$$\begin{aligned} \mathbb{P}(T_{n \rightarrow m} = t) &= R_n \mathbb{P}(T_{n+1 \rightarrow m} = t - 1) + L_n \mathbb{P}(T_{n-1 \rightarrow m} = t - 1) \\ &\quad + (1 - R_n - L_n) \mathbb{P}(T_{n \rightarrow m} = t - 1), \end{aligned} \quad (8.43)$$

which can be rewritten:

$$\begin{aligned} \mathbb{P}(T_{n \rightarrow m} = t) &= R_n \mathbb{P}(T_{n+1 \rightarrow m} + 1 = t) + L_n \mathbb{P}(T_{n-1 \rightarrow m} + 1 = t) \\ &\quad + (1 - R_n - L_n) \mathbb{P}(T_{n \rightarrow m} + 1 = t). \end{aligned} \quad (8.44)$$

The boundary conditions in $n = 0$ and $n = m$ is given by:

$$\mathbb{P}(T_{0 \rightarrow m} = t) = R_0 \mathbb{P}(T_{1 \rightarrow m} + 1 = t) + (1 - R_0) \mathbb{P}(T_{0 \rightarrow m} + 1 = t) \quad (8.45)$$

Now if we take the moment-generating function of $\mathbb{P}(T_{n \rightarrow m} = t)$, we get:

$$\begin{aligned} \sum_{t=0}^{+\infty} e^{\lambda t} \mathbb{P}(T_{n \rightarrow m} = t) &= R_n \sum_{t=0}^{+\infty} e^{\lambda t} \mathbb{P}(T_{n+1 \rightarrow m} + 1 = t) \\ &\quad + L_n \sum_{t=0}^{+\infty} e^{\lambda t} \mathbb{P}(T_{n-1 \rightarrow m} + 1 = t) \\ &\quad + (1 - R_n - L_n) \sum_{t=0}^{+\infty} e^{\lambda t} \mathbb{P}(T_{n \rightarrow m} + 1 = t) \end{aligned} \quad (8.46)$$

or equivalently:

$$\begin{aligned} \langle e^{\lambda T_{n \rightarrow m}} \rangle &= R_n \langle e^{\lambda(T_{n+1 \rightarrow m} + 1)} \rangle + L_n \langle e^{\lambda(T_{n-1 \rightarrow m} + 1)} \rangle \\ &\quad + (1 - R_n - L_n) \langle e^{\lambda(T_{n \rightarrow m} + 1)} \rangle. \end{aligned} \quad (8.47)$$

Average first passage time

A first-order development in λ of equation (8.47), once evaluated for $\lambda = 1$, leads to:

$$\langle T_{n \rightarrow m} \rangle = R_n \langle T_{n+1 \rightarrow m} \rangle + L_n \langle T_{n-1 \rightarrow m} \rangle + (1 - R_n - L_n) \langle T_{n \rightarrow m} \rangle + 1, \quad (8.48)$$

which we can rewrite:

$$R_n (\langle T_{n+1 \rightarrow m} \rangle - \langle T_{n \rightarrow m} \rangle) - L_n (\langle T_{n \rightarrow m} \rangle - \langle T_{n-1 \rightarrow m} \rangle) = -1 \quad (8.49)$$

with the two special conditions for $n = 0$ and $n = m$:

$$R_0 (\langle T_{1 \rightarrow m} \rangle - \langle T_{0 \rightarrow m} \rangle) = -1 \quad (8.50)$$

and:

$$\langle T_{m \rightarrow m} \rangle = 0. \quad (8.51)$$

At this point we define the quantity μ_n by:

$$\mu_n = \langle T_{n \rightarrow m} \rangle - \langle T_{n+1 \rightarrow m} \rangle, \quad (8.52)$$

which is positive, since the time required to reach m is greater when starting from n than from $n + 1$. It obeys the following recursion, according to (8.49):

$$\mu_n = \frac{1}{R_n} + \frac{L_n}{R_n} \mu_{n-1} \quad (8.53)$$

starting with $\mu_0 = 1/R_0$ and up to μ_{m-1} . Then, condition (8.51) entails that:

$$\langle T_{m-1 \rightarrow m} \rangle = \mu_{m-1}. \quad (8.54)$$

From this we can climb down the ladder, using:

$$\langle T_{n \rightarrow m} \rangle = \langle T_{n+1 \rightarrow m} \rangle + \mu_n. \quad (8.55)$$

Accordingly, we get:

$$\langle T_{n \rightarrow m} \rangle = \sum_{k=n}^{m-1} \mu_k. \quad (8.56)$$

Note that it respects the Markovian nature of the process, since it amounts to:

$$\langle T_{n \rightarrow m} \rangle = \sum_{k=n}^{m-1} \langle T_{k \rightarrow k+1} \rangle. \quad (8.57)$$

We can now plot, for each site k , the corresponding site-wise first passage time μ_k (Fig. 8.11a), which is equivalent to the average time required to jump on the next site, as well as the transmission time for all starting points from 0 to M (Fig. 8.11b). Interestingly, the peak in the average site-to-site transmission time is not reached for $x = x_c$, but slightly after that. Anyway, those plots confirm the existence of a trap region after which the time required to cross the whole system diminishes greatly.

Variance of first passage time

Similarly, we can obtain the second moment of the first passage time by a second order development of equation (8.47). We then obtain:

$$\begin{aligned} -1 - 2 \langle T_{n \rightarrow m} \rangle &= R_n \left(\langle T_{n+1 \rightarrow m}^2 \rangle - \langle T_{n \rightarrow m}^2 \rangle \right) - L_n \left(\langle T_{n \rightarrow m}^2 \rangle - \langle T_{n-1 \rightarrow m}^2 \rangle \right) \\ &+ 2R_n (\langle T_{n+1 \rightarrow m} \rangle - \langle T_{n \rightarrow m} \rangle) - 2L_n (\langle T_{n \rightarrow m} \rangle - \langle T_{n-1 \rightarrow m} \rangle) \end{aligned} \quad (8.58)$$

Similarly, we can define a quantity u_n such that:

$$u_n = \langle T_{n \rightarrow m}^2 \rangle - \langle T_{n+1 \rightarrow m}^2 \rangle, \quad (8.59)$$

we are left with:

$$-1 - 2 \langle T_{n \rightarrow m} \rangle = -R_n u_n + L_n u_{n-1} - 2R_n \mu_n + 2L_n \mu_{n-1}. \quad (8.60)$$

We then come up with the following recursion relation:

$$u_n = \frac{L_n}{R_n} u_{n-1} - 2 \left(\mu_n - \frac{L_n}{R_n} \mu_{n-1} \right) + \frac{1 + 2 \langle T_{n \rightarrow m} \rangle}{R_n}, \quad (8.61)$$

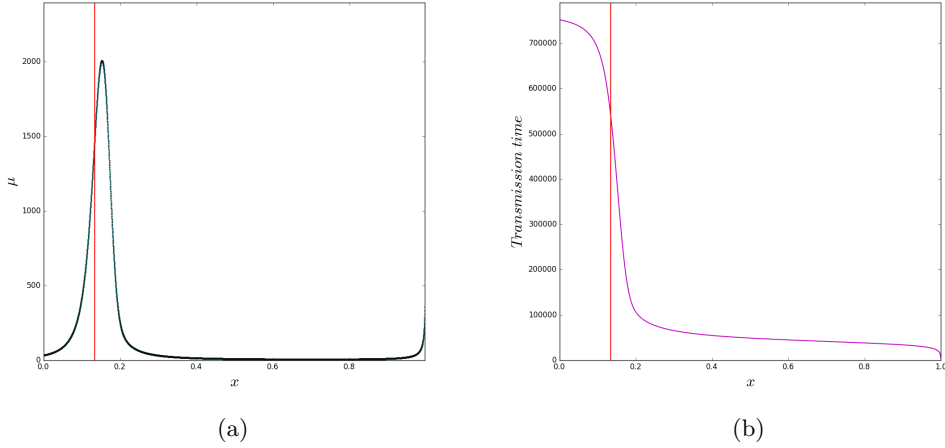


Figure 8.11: (a) Site-wise average first passage time. The red line indicates the critical value x_c . (b) Average transmission time to the site M , for all possible starting sites. The red line indicates the critical value x_c . $M = 5000$, $\delta = 0.0$, $\beta = \beta^*$.

which, according to (8.53), simplifies to:

$$u_n = \frac{L_n}{R_n} u_{n-1} + \frac{2 \langle T_{n \rightarrow m} \rangle - 1}{R_n}, \quad (8.62)$$

with the boundary condition:

$$u_0 = \frac{2 \langle T_{0 \rightarrow m} \rangle - 1}{R_0}, \quad (8.63)$$

and $\langle T_{m \rightarrow m}^2 \rangle = 0$. Then we can compute all μ_n 's up until μ_{m-1} , which is equal to $\langle T_{m-1 \rightarrow m}^2 \rangle$, so that we can get all $\langle T_{n \rightarrow m}^2 \rangle$ recursively as before. We shall therefore obtain:

$$\langle T_{n \rightarrow m}^2 \rangle = \sum_{k=n}^{m-1} u_k \quad (8.64)$$

This formula, however, is a bit tricky to handle, since the u_k 's, contrary to the μ_k 's, actually depend on m through the term $\langle T_{n \rightarrow m} \rangle$. We can explicit this dependence by adding a m exponent:

$$u_n^m = \frac{L_n}{R_n} u_{n-1}^m + \frac{2 \langle T_{n \rightarrow m} \rangle - 1}{R_n}, \quad (8.65)$$

Thus, if we want to compute $u_k^{k+1} = \langle T_{k \rightarrow k+1}^2 \rangle$, we face the problem that a term u_{k-1}^{k+1} appears instead of u_{k-1}^k in the recurrence relation:

$$u_k^{k+1} = \frac{L_k}{R_k} u_{k-1}^{k+1} + \frac{2\mu_k - 1}{R_k}. \quad (8.66)$$

However, since the process is markovian, we can write:

$$\langle T_{k-1 \rightarrow k+1}^2 \rangle = \langle (T_{k-1 \rightarrow k} + T_{k \rightarrow k+1})^2 \rangle, \quad (8.67)$$

which translates to:

$$u_{k-1}^{k+1} + u_k^{k+1} = u_{k-1}^k + u_k^{k+1} + 2\mu_{k-1}\mu_k \quad (8.68)$$

and thus:

$$u_{k-1}^{k+1} = u_{k-1}^k + 2\mu_{k-1}\mu_k. \quad (8.69)$$

Reinjecting this expression in the recurrence relation (8.66), we obtain:

$$u_k^{k+1} = \frac{L_k}{R_k} u_{k-1}^k + \frac{2(\mu_k(1 + L_k\mu_{k-1}) - 1)}{R_k}. \quad (8.70)$$

Invoking finally (8.53), we get:

$$u_k^{k+1} = 2\mu_k^2 - \frac{1}{R_k} + \frac{L_k}{R_k} u_{k-1}^k. \quad (8.71)$$

We therefore have a recurrence relations for the quantities $\langle T_{k \rightarrow k+1}^2 \rangle$. From this, we can obtain the variance σ^2 of any first passage time by using:

$$\begin{aligned} \sigma^2(T_{n \rightarrow m}) &= \sum_{k=n}^{m-1} \sigma^2(T_{k \rightarrow k+1}) \\ &= \sum_{k=n}^{m-1} (u_k^{k+1} - \mu_k^2). \end{aligned} \quad (8.72)$$

The behavior of the site-to-site transmission time variance is quite analogous to the one of the average site-to-site transmission time (Fig. 8.12a). As the variance of the transmission time can be interpreted as an uncertainty regarding this time, it is expected that this uncertainty follows a sudden decrease associated with the release of the walker outside of the trap: as long as the trap region has not been crossed, there is no possible way to know when the walker will be transmitted to the end of the system. This is indeed confirmed by the computation of the uncertainty over the global transmission time (Fig. 8.12b). In that case, the latency region could really be interpreted as the region of uncertainty decrease: as the trapping time is determinant to know the transmission time, before reaching that region, the uncertainty stays very high and constant, and after it, the transmission time is almost known since the process is nearly deterministic.

Note that the end process also introduces some uncertainty. Yet it is of no real interest, since from the point of view of which form is actually produced, when the walker is almost through it makes no difference whether or not a handful of the old variant's occurrences remain. Indeed, as the derivative of $P_\gamma(x)$ is null for $x = 1$, the function $P_\gamma(x)$ is insensitive to the lingering of very few spurious occurrences in memory, and the old form is, practically, never produced. Hence, the state of the competition is settled down at this point, and this last step is not actually reflected in any way in language use. We shall see later, in chapter 9, a much better explanation to account for the 'residue' phenomenon, which states on the contrary that the old variant may have a long persistence even after the competition is lost.

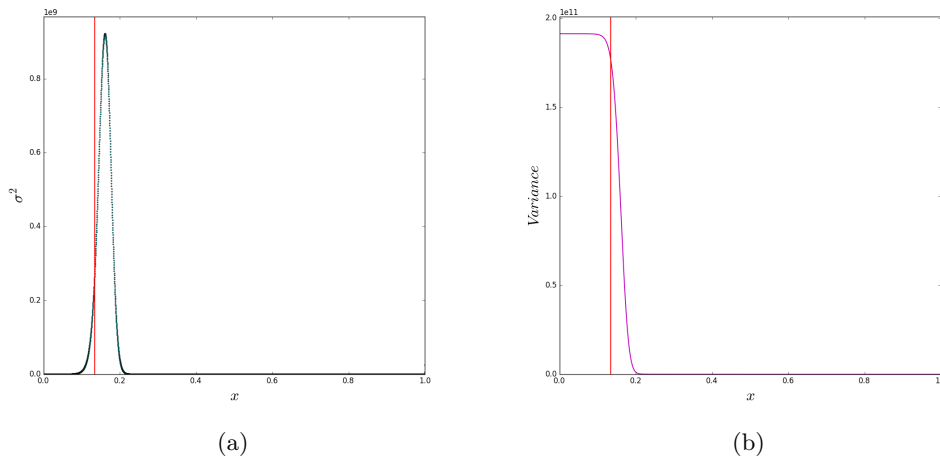


Figure 8.12: (a) Site-wise variance of transmission time. The red line indicates the critical value x_c . (b) Variance of the transmission time to the site M , for all possible starting sites. The red line indicates the critical value x_c . $M = 5000$, $\delta = 0.0$, $\beta = \beta^*$.

Last, we can discuss the relative uncertainty on the transmission time, σ/μ . Once more, we can consider it site-wise (transmission time from a site to the next, Fig. 8.13a), or globally (transmission time to the end of the system, Fig. 8.13b). However, the relation between the two is no longer that of a simple summation, therefore this quantity is less reliable. Anyway, the relative uncertainty shows a clear bump followed by a significant decrease, which is yet another hint that something peculiar happens in this region. Note that the peak of relative uncertainty goes beyond 1, which means that the standard deviation is bigger than the average, indicative of how uncertain the transmission time is due to the trap region.

Now that we can compute the two first moments of any first passage time in the system (i.e. from any starting point to any end point), we only need, to define the latency time, a robust definition of the trap boundaries.

8.2.2 Trap region

Though it is clear, from the previous analyses, that there is some sort of a ‘trap region’ in the system, associated with very high uncertainty (i.e. the uncertainty does not decrease before entering the trap and reaches a low, stable level once out of it), there seems to be no clear-cut position at which the trap can be said to begin or to end. In the following, we shall therefore explore the different possibilities of defining this crucial region.

Back and forth the trap region

The first idea to find the boundaries of the trap region is to exploit the fact that in the trap region, the walker does not stay still. If such a time is spent in this region, it is because there is no drift towards one direction or another, therefore the walker

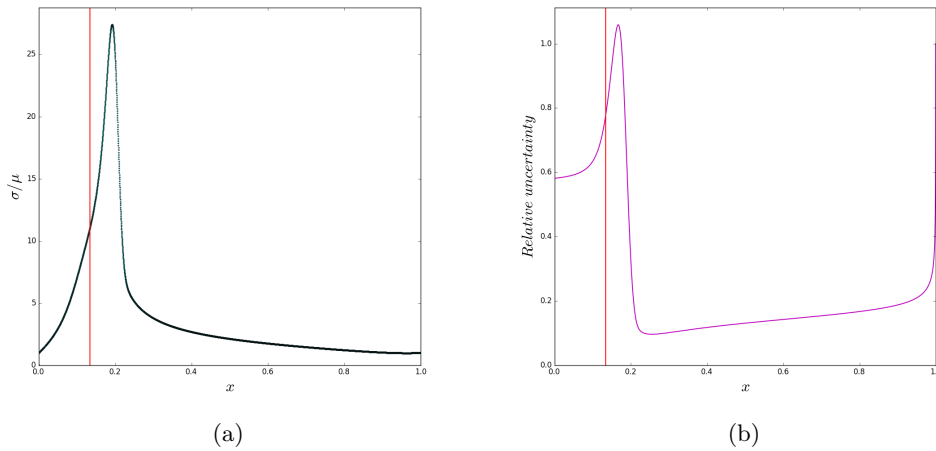


Figure 8.13: (a) Site-wise relative uncertainty on transmission time. The red line indicates the critical value x_c . (b) Relative uncertainty of the transmission time to the site M , for all possible starting sites. The red line indicates the critical value x_c . $M = 5000$, $\delta = 0.0$, $\beta = \beta^*$.

goes back and forth aimlessly. Hence, there shall be a lot of backtracking in the trap region, whereas outside of it, the walker will move forward without wander. A good way to characterize the trap region would thus be to search for the region of maximal exploration.

For each site, we can plot how far the particle explores, in average, both backward and forward, before its first passage and its last return to the site (Figs. 8.14a and 8.14b). One can observe a little nodule located around x_c which might correspond to the trap region (Fig. 8.14a), where the particle is wandering back and forth without being compelled to go much further in one direction or another. This corresponds closely to a case of pure diffusion, which is more or less the case since, in this region, the drift $v(x) = R(x) - L(x)$ is negligible.

Overall, the total range of the diffusion is maximal at the critical point (Fig. 8.15a), which concurs with the deterministic picture that we first drew. However, this diffusive behavior is asymmetric with respect to the critical point (Fig. 8.15b). Indeed, left to x_c , there is a small, positive drift (hence the behavior is diffusive with a tendency to explore forward), while right to x_c , the drift is comparable, but the diffusion coefficient is bigger, which makes the effect of the drift smaller in comparison (Fig. 8.10). Furthermore, there is a large-tail on the right of the critical point, which is somehow unexpected given the fast S-shaped rise expected after the passing of the trap.

From these quantities, an interesting way to define the trap could thus be to take, as an entrance, the first point whose average forward exploration range encompasses the critical point, and as an exit, the last point which reaches it thanks to backward diffusion. We would thus end up with a trap region defined without any threshold, focused on the critical point, and consistent with the diffusive behavior we exposed here.

There are two drawbacks with this choice. The first one is that we already saw that the stochastic latency was quite different from the deterministic one, so that if we

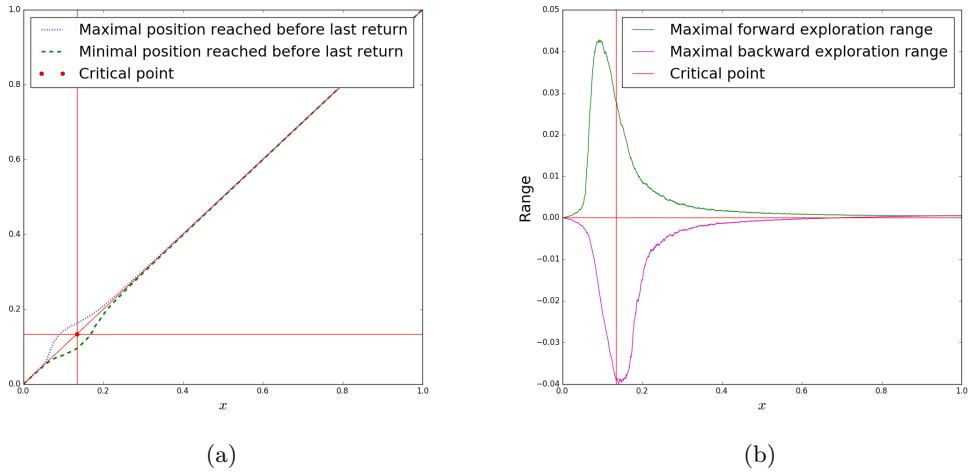


Figure 8.14: (a) Average extremal positions reached between first passage time and last return time, for each site. (b) Average extremal positions reached between first passage time and last return time, for each site. $M = 2000$, $\delta = 0.0$, $\beta = \beta^*$.

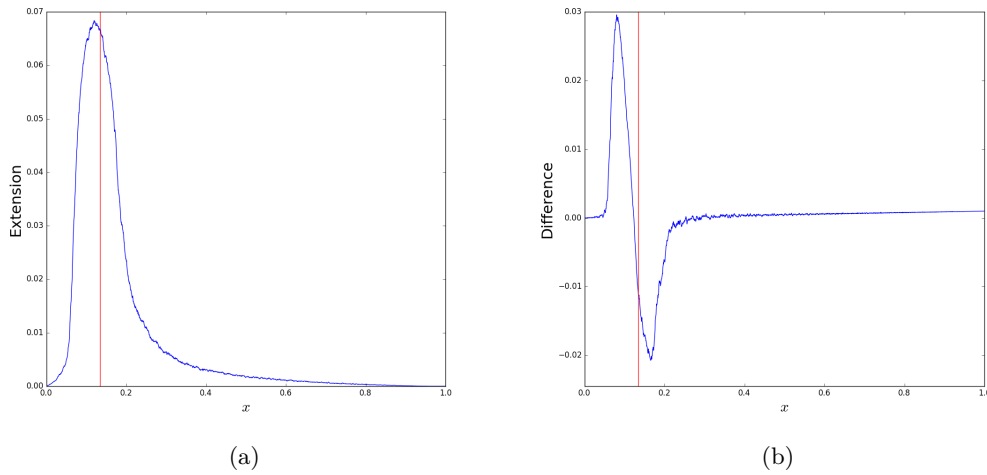


Figure 8.15: (a) Range of the exploration both forward and backward a given site, for each site. (b) $M = 2000$, $\delta = 0.0$, $\beta = \beta^*$.

want to catch latency based on quantities which agree too well with the deterministic picture, we risk being unable to catch the real, stochastic latency. The other thing is that all quantities relevant from a stochastic point of view (average transmission time, variance of the transmission time, etc.) are skewed on the right of the critical point. Hence, a better description of this trap region would not be focused and centered on the critical point, but shifted from it further into the system.

Trap region and uncertainty

Of the three quantities we considered before with respect to transmission time (average, variance, and relative error), the variance shows the clearest peak, and its interpretation is the most straightforward: as long as the walker has not crossed the trap, we know very little on the transmission time, which therefore shows a high variance. Getting closer to the trap region does not teach us anything on the transmission time, and it's only once the trap is crossed, that the uncertainty get finally reduced, if not vanishes entirely.

Hence we can look at it more closely and figure out whether it can help us to naturally define boundaries for the trap region. Invoking a threshold is always a possibility (i.e. threshold in average site-to-site transmission time), yet it would be nicer to put as few input as possible in the definition of this trap region.

It turns out that the variance of the transmission time has the nice advantage to show an interesting structure in its third-derivative (Fig. 8.16a) —equivalently, in the second derivative of the site-to-site transmission time variance (Fig. 8.16b)— with two minima corresponding roughly to the boundaries expected of the trap region. What is more, the first minimum coincides exactly with the deterministic critical point, which is an interesting feature, to be sure, since we already saw that the critical point was close to the place where the walker started to diffuse backward rather than forward (Fig. 8.15b). Hence, past the critical point, the diffusion starts to pull the walker backwards the critical point rather than to push it forward. It thus seems not inappropriate to regard the critical point as the entrance to a slowing down region.

There are several advantages to this choice. As its computation relies on the transmission time variance, which can be computed exactly through a simple recursion formula (8.71), we can find the boundaries of the trap region for any parameter without heavy simulations as those required to find the forward and backward exploration range for each site of the system. As an example, we have shown the behavior of the trap boundaries with respect to the normalized distance δ to the critical threshold γ_c (Fig. 8.17).

It appears that the entrance of the trap coincides exactly with the critical point $x = x_c$ for $\gamma = \gamma_c$ (i.e. for $\delta = 0$), which is already remarkable. Furthermore, the unstable fixed point is always included in the trap and both the entrance and the exit seem to follow it closely. This would indicate that the trap region actually corresponds, below the threshold at least, to the boundary between the two coexisting attraction basins of the system —one for each of the two stable fixed points.

The fact that the trap region widens for $\gamma > \gamma_c$ might seem surprising, since we do not expect latency to be important in this regime. However, above the threshold, this region no longer corresponds to a trap region, as is illustrated by the evolution of the

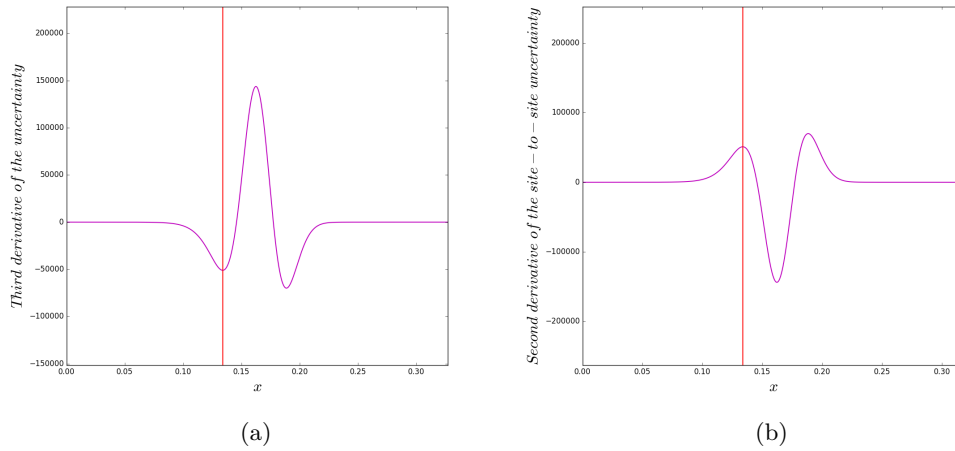


Figure 8.16: (a) Third derivative of the uncertainty over the transmission time. (b) Second derivative of the site-to-site transmission time variance. On both figures the red line indicates the position of the critical point x_c . $M = 5000$, $\delta = 0.0$, $\beta = \beta^*$.

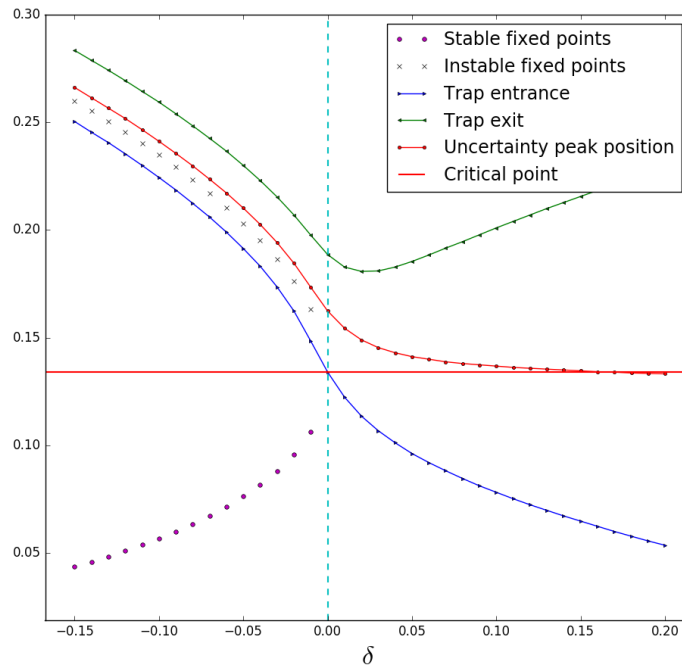


Figure 8.17: Entrance and exit of the trap as a function of the distance to threshold δ , as given by coordinates of the minima of the third derivative of the uncertainty. The uncertainty peak corresponds to the maximum in-between. The fixed points of the deterministic equation have also been plotted for comparison. The dashed cyan line indicates $\delta = 0.0$. The critical value x_c has been represented by a full red line. $M = 5000$, $\beta = \beta^*$.

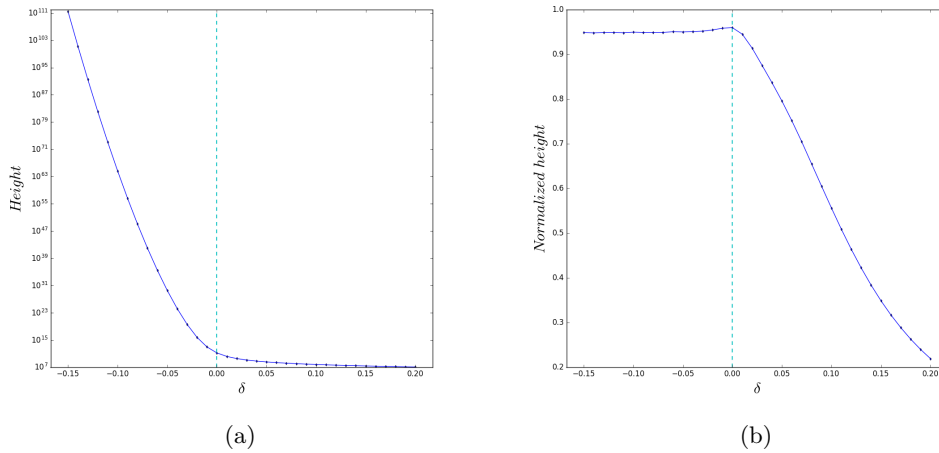


Figure 8.18: (a) Semilog plot of the height of the decrease of uncertainty regarding the transmission time before and after the trap, as a function of the distance δ to the threshold. (b) Same quantity as before, though divided by the uncertainty at the trap entrance. The threshold $\delta = 0.0$ is indicated by a dashed cyan line. $M = 5000$, $\beta = \beta^*$.

difference of uncertainty between the entrance into the trap and the exit. Above the threshold, crossing the trap region no longer leads to a significant drop in uncertainty (Fig. 8.18a): the trajectory of the walker within the trap becomes predictable, so that it cannot be considered as a trap anymore. As a result, the size of this region does not matter and its widening is irrelevant regarding the latency phenomenon.

We can also observe that this difference diverges (even on a logarithmical scale) as γ gets further away below the threshold. In this domain, the latency phenomenon increases to such a point that the system never reaches the end of the system practically, so that the uncertainty on transmission time at the trap entrance goes to infinity. More interestingly, we can consider the normalized height between the two uncertainty plateaus, i.e. the difference in uncertainty between the entrance and the exit of the trap, divided by the uncertainty at the entrance (Fig. 8.18b). It becomes clear then that the difference of uncertainty only diverges because of the uncertainty at the trap entrance. It also shows a remarkable connection between the deterministic bifurcation picture, and the stochastic uncertainty on the transmission time.

Another interesting feature is that the entrance of the trap, which happens to be tuned to x_c , does not depend on M (Fig. 8.19a) as it is merged with the critical point. The exit of the trap, however, does, which is not much of a surprise, since the stochastic latency is linked to the discrete dynamics, and is therefore expected to depend at least weakly on the discretization step, $1/M$. However, though the trap region becomes smaller, it does not mean that so does the latency, quite the contrary. If we plot the difference of the average transmission times right after and right before the trap (normalized by the average transmission time before the trap), it shows a roughly linear increase of that time with M (Fig. 8.19b), even if the region is narrowing (in terms of the $[0,1]$ continuum though, not in terms of the number of sites).

To conclude, it seems that the determination of the trap region by the coordinates

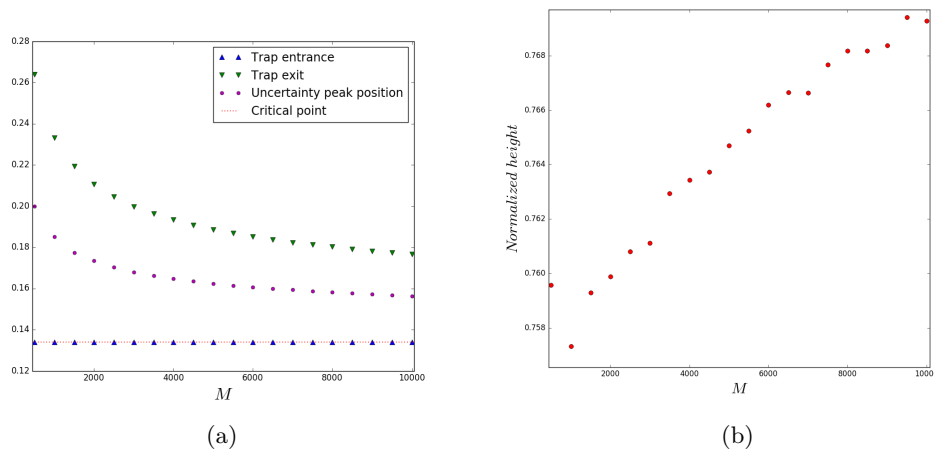


Figure 8.19: (a) Entrance and exit of the trap as a function of the system size M (which defines the discrete step), as given by coordinates of the minima of the third derivative of the uncertainty. The critical value x_c has been represented by a full red line. (b) Difference between the average transmission time at trap entrance and at trap exit, normalized by the average transmission time at trap entrance, with M , the number of sites of the system. $\delta = 0.0$, $\beta = \beta^*$.

of the maxima of the second derivative of the site-to-site transmission time uncertainty behaves correctly with respect to the parameters of the model, conforms to the expectations, shows nice connections with the deterministic picture of the system, and finally is easy to compute for any parameters and does not require any numerical simulations to be obtained.

8.2.3 Are the trapping time and the latency time really the same?

Though this choice seems therefore relevant and satisfying from an analytical point of view, it is not clear whether it captures latency as an empirical feature of a competition process, similar to the one we extracted from our examples of semantic expansion processes.

Numerical surrogate data for the frequency of use

To investigate this matter, we should first ask what exactly would be the equivalent of latency in this model. The model simulates a competition between two variants over memory. Hence, if we want to compare it with empirical data, we must consider how many times the new variant is produced over a given time window. As M is the natural timescale of the model, we choose it as the size of the time window over which the occurrences are counted. It would certainly be better and most enlightening to compare the memory-based timescale with the actual timescales of change, yet we found no obvious way to do so.

Thus, we can produce surrogate data for the frequency of use. Then, everything we performed over the empirical data can be done for the numerical data of our model

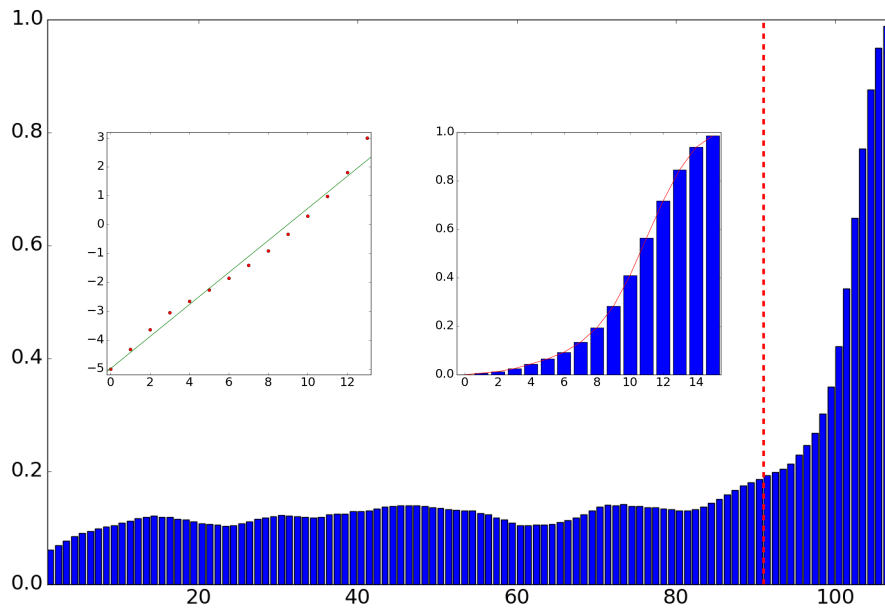


Figure 8.20: Surrogate frequency pattern of one run of the process. The red dashed line indicates the separation between the latency period and the growth period. The left inset displays the logit transform of the surrogate frequency for the growth part, with linear fit (red line). The left inset shows a sigmoid representation of the rescaled surrogate frequency, with its sigmoidal fit (red line). $M = 5000$, $\delta = 0.0$, $\beta = \beta^*$.

simulations: smoothing over a period of five occurrence-counting windows, extraction of an S-curve pattern, detection of the latency period, etc (Fig. 8.20). By setting the exact same parameters that we used to treat empirical data (with the exception of the threshold frequency in the determination of latency, set to 0.17 instead of 0.15, since the latency frequency is comparatively much higher for the numerical process, a minor issue which we shall overcome in the next chapter), we can obtain qualitatively similar results, with an S-curve pattern, preceded by a latency. Here the latency is much more prominent, since nothing interferes with the process, and also because we have tuned the parameters to be exactly at the critical point ($\delta = 0.0$).

Now we have what we shall call a ‘theoretical’ latency, given by the difference between the first passage times at the exit and the entrance of a trap region, and a ‘numerical’ latency, obtained through the numerical simulation of numerous runs of the process, to which are applied the same extracting procedure as the one underwent by our empirical, linguistic data.

Numerical latency-driven determination of trap boundaries

A first question would be: can we set the boundaries of the trap region so as to find the two first moments of both the growth times and the latency times, and if so,

how do they compare with our theoretical determination of those boundaries? To do so, we performed the following procedure. Given the mean and variance of both the numerical growth time and the numerical latency time (μ_{growth} and μ_{lat} , σ_{growth}^2 and σ_{lat}^2 , respectively), we try to adjust the trap boundaries in order to make the mean and the variance of the theoretical growth time and latency time to fit their numerical counterparts.

We actually tried to place three boundaries, and not only two, which would correspond to the trap region. Indeed, we already stressed that the end process was somehow problematic, and that it could take quite long for the walker to reach the end of the system (the dynamics slows down once more in this end region). Yet, the numerical process only cares about the occurrences produced, and once the walker has reached this end region, then only the new variant is produced in practice, so that the surrogate frequency of use would be truly equal to one. For instance, even though the walker lies at a distance of a hundred sites from the absorbing one ($x = 0.9798$), the new variant can already be produced with certainty ($P_\gamma(0.9798) = 0.999993$). Hence, it seems reasonable to define a cut-off boundary, such that we consider that the growth time ends once it reaches this cut-off boundary. In the following we shall note x_{in} , x_{out} and x_{off} the position of the three boundaries respectively corresponding to the trap entrance, the trap exit, and the cut-off.

We proceeded in the following way. First, we were to determine the pair (x_{out}, x_{off}) . To do so, we selected all candidate pairs over a reasonable range, and kept all pairs such that the relative error was less than 1%:

$$\left| \langle T_{n_{out} \rightarrow n_{off}} \rangle - \mu_{growth} \right| < 0.01 * \mu_{growth}. \quad (8.73)$$

Then, among this subset, we picked up the pair which was capturing best the variance σ_{growth}^2 . Hence, we could determine the trap exit, $x_{out} = n_{out}/M$.

To get the trap entrance, we are left with three possible choices: we can choose it so as to fit the mean numerical latency, its variance, or the mode of its distribution. The latter might seem a bit tricky, since we do not know the mode of the first passage time $T_{n_{in} \rightarrow n_{out}}$: we have only analytical formulae for the mean and the variance. We could run a lot of simulations with a lot of different couples of parameters to get an estimate of the mode for each, but it does not seem like a satisfying idea. What we can do is to assume an analytical formula for the mode, by positing that the distribution of the first passage time must be Inverse Gaussian (an hypothesis that we will justify shortly). As the mean and the variance entirely characterize the Inverse Gaussian, we then know anything about the first passage distribution, including the mode, given by the relation:

$$t_{mode} = \frac{3\mu^2}{2\lambda} \left[\sqrt{1 + \frac{1}{9}P_e^2} - 1 \right], \quad (8.74)$$

with λ , one of the two defining parameters of the Inverse Gaussian, equal to:

$$\lambda = \frac{\mu^3}{\sigma^2}, \quad (8.75)$$

and P_e , the Péclet number:

$$P_e = \frac{2\lambda}{\mu}, \quad (8.76)$$

Method	Trap entrance	Trap exit	Midpoint	Wideness
Surrogate data fit	0.1146	0.2164	0.1613	0.1018
Theoretical determination	0.1340	0.1886	0.1655	0.546

Table 8.1: Boundaries of the trap obtained through a fit of the distribution parameters associated with the surrogate data first, and through a theoretical determination second. Also provided are the middle and the wideness of the trap.

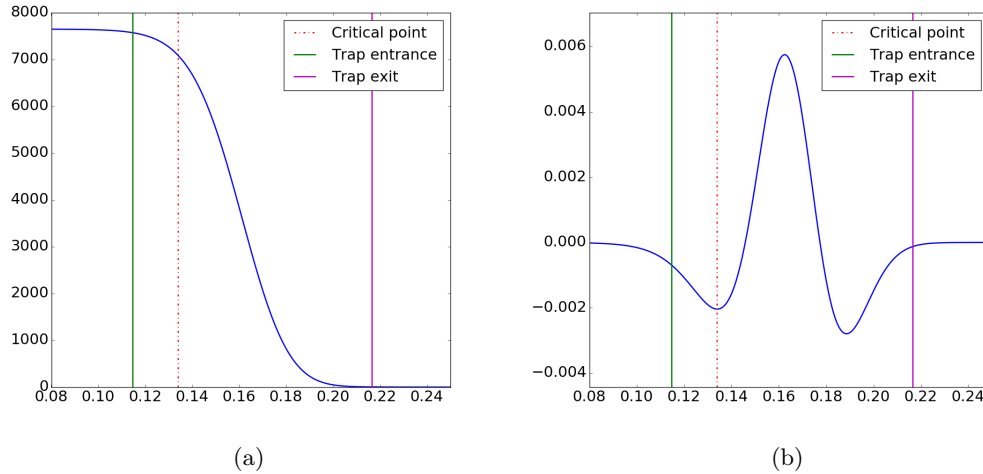


Figure 8.21: Position of the boundaries of the trap represented along (a) the variance of the transmission time for each different starting site; and (b) the third derivative of the previous quantity. $M = 5000$, $\delta = 0.0$, $\beta = \beta^*$.

whose importance we shall see later on.

With this assumption that the first passage time must be Inverse Gaussian, we can therefore compute the mode for any pair of points, and choose the entrance of the trap n_{in} such as to fit the numerical latency mode as best as possible. Fitting the mode proved the most efficient of those three choices, since the mode captures both the mean and the variance. Furthermore, as it associated with the peak of probability, missing it has great consequences on the goodness of fit. Doing so, we found coordinates for the trap boundaries which can be compared with the ones obtained theoretically (Table 8.1). The cut-off is set at 0.998.

We can now compare these new trap boundaries with the theoretical quantities which drove our previous investigations. The region of the trap effectively coincides with the region of main uncertainty reduction (Fig. 8.21a). It does not match the two third derivative minima (Fig. 8.21b), but encompasses them both. Hence, the trap region fitting our numerical latency is larger than the one we determined theoretically, yet it captures roughly the same region, indicative that the uncertainty on transmission time is indeed the best observable to describe and to understand latency. It may also be that, if the variation tolerance around the latency frequency was set smaller, so would be the trap region.

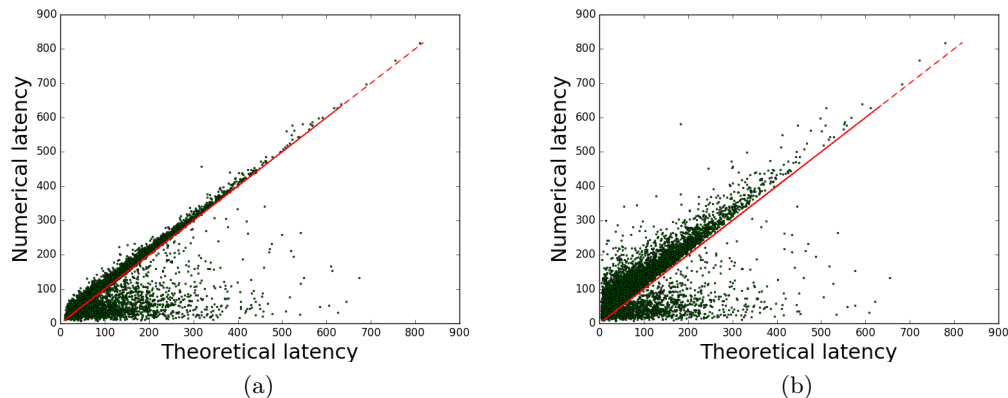


Figure 8.22: Correspondence between the numerical latency (extracted from the output pattern) and the theoretical one (computed as the difference between the first passage times at trap exit and trap entrance), for 10,000 different runs of the process. (a) Trap boundaries are optimized so as to fit the statistical features of the numerical observables (mean growth time, variance of the growth time, mode of the latency distribution). (b) Trap boundaries are theoretically set by the position of the third-derivative uncertainty minima. $M = 5000$, $\delta = 0.0$, $\beta = \beta^*$.

These findings are satisfying since they show that our theoretical understanding of the latency phenomenon compares well with the latency extracted by our empirical, output-driven procedure. Our empirical dealing of the linguistic data is therefore able to capture deep and well-grounded features of the latency phenomenon, while our theoretical understanding of our model features catches correctly the empirical phenomenology of latency in language change.

Latency match

We previously checked that, by optimizing the choice of the trap boundaries, we could catch the mode of the numerical latency distribution, and furthermore, that the trap boundaries obtained through this procedure were compatible, in their principle, with the trap region we theoretically identified. However, this correspondence only regards one overall statistical feature. To pretend that the latency phenomenon indeed corresponds to some entrapping in a delimited region of the frequency space, then it must be that the correspondence between the numerical latency and the theoretical latency is one-to-one for each process.

For 10,000 runs of the process, we thus computed the first passage times at trap exit and trap entrance, so as to obtain the value, for this process, of the latency as theoretically defined. We also obtained the latency value from our pattern-extraction procedure. Then we plotted a scatter plot showing all processes so as to compare the two values for the latency (Fig. 8.22a).

Furthermore, we also compared the numerical latency to trapping time associated with the theoretically determined trap boundaries (not the one which have been optimized to match the statistical features of the numerical observables). The disagreement

is, most expectedly, bigger (Fig. 8.22b), and consistently with the previous picture, we now have instances of numerical latency significantly greater than the theoretical one (the trap region is now much smaller). Yet, those two quantities still show a one-to-one rough correspondence, indicative of the robustness of the theoretical picture we proposed to account for the latency phenomenon.

Note that we did not try to match the numerical growth time and the theoretical one, because of the difficulties introduced by the end process, which even the cut-off does not help to solve (it only allows to match the mean and the variance of both growth times). Knowing that latencies match is enough: it shows that, indeed, once the walker reaches the trap exit, it has escaped the trap region for good, otherwise the numerical latency could be sometimes much higher than the theoretical one. Matching the growth times has proved to resist all our attempts, especially because they do not represent the growth of the same quantity (the numerical growth follows the production probability $P_\gamma(x)$ while the theoretical one focuses on the actual frequency in memory x).

As it appears from Fig. 8.22a, the two quantities match satisfyingly, at least enough so as to assess the relevance of this theoretical picture of a ‘trap region’ to account for latency, but there is still some disagreement between the two. Especially, there is a bulk of points below the line in which they should lie for the two quantities to be equal, corresponding thus to a numerical latency shorter than its theoretical counterpart. This result is actually most curious, since we could have expected the contrary. Indeed, the numerical latency can easily be greater than the theoretical latency, for instance if the walker spends more time outside the trap than what was expected. Or it could also be that the numerical latency would represent a last passage time at the exit boundary of the trap, hence getting bigger than the first passage time.

As this is not the case, it means that the growth can start within the trap region, so that the numerical latency would end even though the exit of the trap has not yet been reached. Indeed, the trap region is a region of uncertainty over the transmission time but it does not prevent to be crossed in a straightforward way. Hence the growth can blend in with the latency, even though the walker is still moving across the trap region. Since, in the procedure, the growth is computed first and the latency second (keep in mind that the latency is computed backwards from the start of the growth part), a part of the process can be considered to be either the mousy beginnings of the growth, or the last stage of the latency. The point is, as long as the growth does not take off, there is no possible way to distinguish the two, precisely because of the uncertainty which is the fundamental characteristic of this region.

Another viable explanation for this mismatch is that the numerical latency time can be cut earlier if the walker position in the trap fluctuates too wildly and goes beyond the boundaries set by the procedure (a data point is considered part of the latency pattern as long as it is comprised between two boundaries, symmetric with respect to frequency of the first data point in the growth pattern, and defined by a constant proportion, in this case 0.17, of the difference between the frequency at the end of the growth and at the start of it). Thus, the numerical latency can break, while the walker is still in the trap. In this case also, the numerical latency will be shorter than the theoretical one.

This explanation is partly corroborated by the observation that, if we narrow down

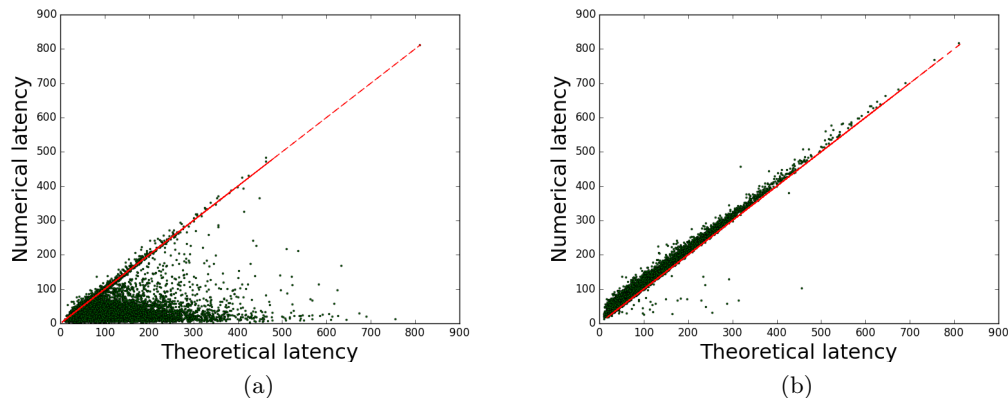


Figure 8.23: Correspondence between the numerical latency (extracted from the output pattern) and the theoretical one (computed as the difference between the first passage times at trap exit and trap entrance), for 10,000 different runs of the process. (a) With a variation tolerance equal to 0.11 instead of 0.17. (b) With a variation tolerance equal to 0.23 instead of 0.17. $M = 5000$, $\delta = 0.0$, $\beta = \beta^*$.

the acceptable spectrum of frequencies for the walker’s position to be considered latent in the numerical procedure, then we find more instances of processes such that the numerical time is smaller than the expected theoretical time (Fig. 8.23a). Contrary, if we widen this margin, then those mismatches situations become infrequent (Fig. 8.23b). If the latency were numerically shorter than the theoretical prediction because of a blending with the growth part, the number of mismatches would not be affected by the size of this threshold interval. Thus, the latter explanation of the mismatches seems more correct.

These observations echo with a previous remark, according to which a reduction of the frequency variation tolerance could lead to a better focus on the theoretically defined trap. Because of this spurious latency break, this hope is technically compromised. This suggests that there might be better ways to define latency than setting a range of acceptable frequencies with respect to the first frequency of the growth pattern. However, this latter way to proceed is appropriate for the empirical linguistic data, and as we wanted to set a comparison with the latter, there is little point in devising another way to extract latency from the output pattern of the process. What matters here is that we have now confirmed that the latency time can be closely equated to a first passage time at the exit of a trapping region.

8.2.4 Distributions

From one run of the process, we can consider two quantities: the number N_k of occurrences of the new linguistic variant produced during the k^{th} time window, which is a macroscopic quantity, and the position $x(t)$ of the walker at each numerical step, reflecting the occupation fraction of the linguistic memory. We saw that the numerical latency extracted from the first observable matches the one obtained from first-passage times of the microscopic walker. Similarly, we can extract a growth time, which is given

by the extracted S-curve pattern in the ‘macroscopic’ case, and by the first passage time at the cut-off in the ‘microscopic’ tracking of the process.

We can therefore generate the statistical distributions for all these quantities as we did for the corpus data, and try to compare both their qualitative and their quantitative features. But first of all, we shall briefly discuss which distributions are expected for these quantities.

The Inverse Gaussian

In a random walk with homogeneous jump probabilities (i.e. $R(x) = R$ and $L(x) = L$, $\forall x$), the distribution for the first passage time at a given threshold is given by the Inverse Gaussian distribution of parameters μ and λ (Redner, 2001, p.118):

$$IG(x) = \sqrt{\frac{\lambda}{2\pi x^3}} \exp\left\{-\frac{\lambda}{2x\mu^2}(x - \mu)^2\right\}. \quad (8.77)$$

The parameter μ grows as the inverse of the drift velocity $R - L$ (the larger the drift, the longer the first passage time in average), and λ grows as the inverse of the diffusion $R + L$. As the variance scales with λ^{-1} , this behavior is expected. When λ tends to infinity, the Inverse Gaussian becomes equivalent to a regular Gaussian distribution.

However, in our case, the jump probabilities are not constant, and neither is the drift nor the diffusion coefficient. Yet, we can try to approximate the first-passage time distributions with the Inverse Gaussian. It is at least a guideline for our expectations, and predicts for instance that, as the drift is weaker in the trap region than in the rest of the system, the parameter μ will be much bigger for the latency time than for the growth time. As for the λ parameter, predicting anything is trickier, since the diffusion varies a lot in the get-away region, and comparing it to the typical diffusion in the trap is not quite enlightening. Also, during the latency, since the walker is confined in a small region of system, the parameters vary only slightly, so that the Inverse Gaussian fit is expected to be much better than for the growth times.

If we choose the variation tolerance parameter so as to never cut latency because of a spurious fluctuation, and match the trap boundaries to get the mode of the latency, we obtain a very efficient match between numerical and theoretical latency (Fig. 8.24). Doing so, we make a dent in the ‘reduction of uncertainty region’ interpretative picture: the spectrum over which the frequency is still considered as latent being much bigger, latency starts roughly at the same time of the whole process, so that the entrance of the trap is really close to the starting point ($x_{in} = 0.0684$). This does not affect the growth, which does not depend on the variation tolerance parameter, and relies only on the trap exit x_{out} . The definition of the latter did not change and still conforms to our theoretical understanding of the phenomenon.

We can then compare the two distributions and fit them with an Inverse Gaussian (Fig. 8.25). The two distributions closely agree, for the two latencies are in a one-to-one, process-to-process correspondence. Both are nicely fit by an Inverse Gaussian (whose parameters are directly computed from the mean and the variance of the quantity it represents), with a D_{KL} score of 0.04.

The same comparison can be done for the growth times (Fig. 8.26a), as well as for the logit slopes (Fig. 8.26b), for which we do not have any theoretical counterpart.

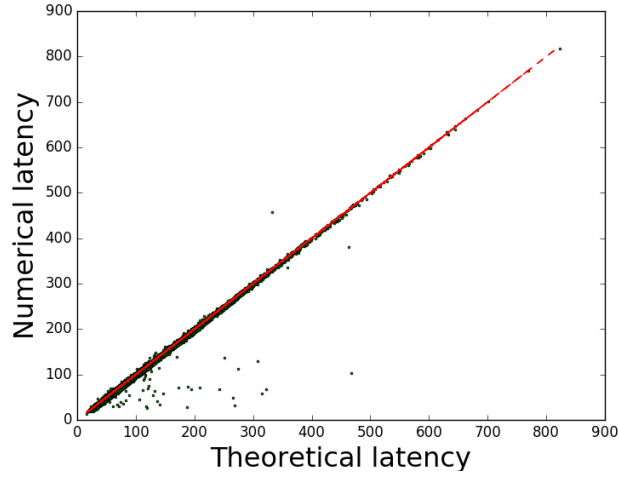


Figure 8.24: Correspondence between the numerical latency (extracted from the output pattern) and the theoretical one (computed as the difference between the first passage times at trap exit and trap entrance), for 10,000 different runs of the process, with a variation tolerance equal to 0.23 instead of 0.17. The trap entrance have been recomputed so as to fit the new mode of the latency distribution. $M = 5000$, $\delta = 0.0$, $\beta = \beta^*$.

D_{KL}	Growth time	Logit slope
Inverse Gaussian	0.091	0.013
Gaussian	0.122	0.009

Table 8.2: Kullback-Leibler divergence between the two numerical observables of the growth pattern and two possible fits, Inverse Gaussian and Gaussian.

The Inverse Gaussian fit is only approximative for both, and especially for the growth times. Also, given the parameters value (a very high λ), the distribution makes little difference with a regular Gaussian (Table 8.2).

The Péclet number

The Péclet number, that we already introduced:

$$P_e = \frac{2\lambda}{\mu} = \frac{2\mu^2}{\sigma^2}, \quad (8.78)$$

compares the mean and the variance of the Inverse Gaussian distribution. The factor 2 has been added here to match the definition of this number given by Redner (2001, p.119) and is entirely facultative. The interpretation of this number is all the clearer in relation with the mode of the Inverse Gaussian distribution:

$$t_{mode} = \frac{3\mu^2}{2\lambda} \left[\sqrt{1 + \frac{1}{9}P_e^2} - 1 \right], \quad (8.79)$$

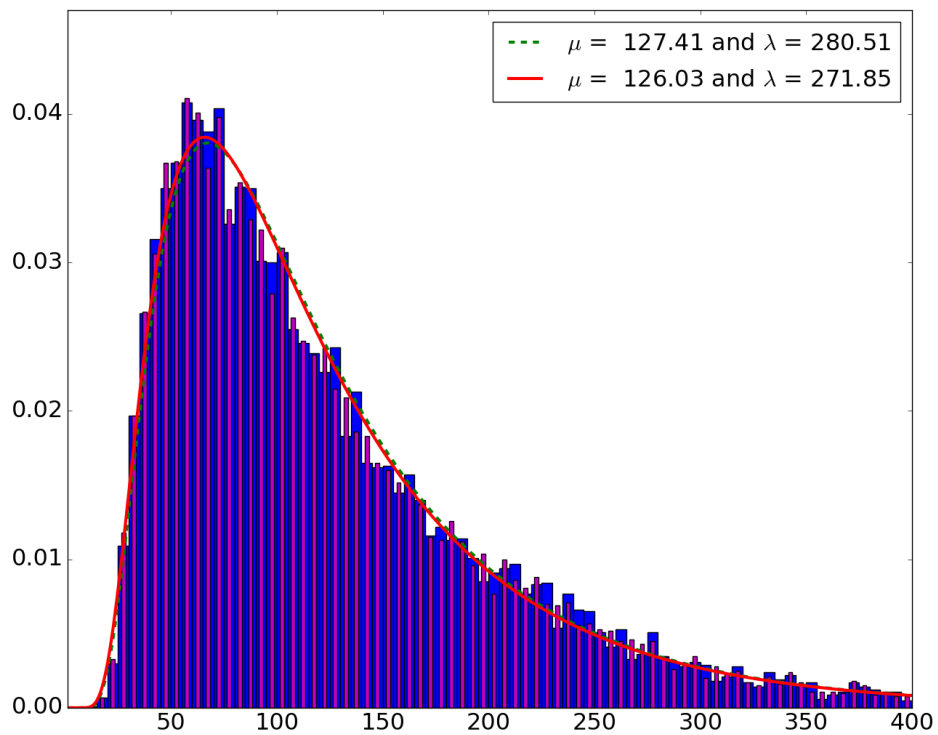


Figure 8.25: Distribution of latency times, both numerical (blue wide histogram) and theoretical (magenta thin histogram). The Inverse Gaussians of these two distributions are respectively shown by a red full line and a green dashed line. $M = 5000$, $\delta = 0.0$, $\beta = \beta^*$.

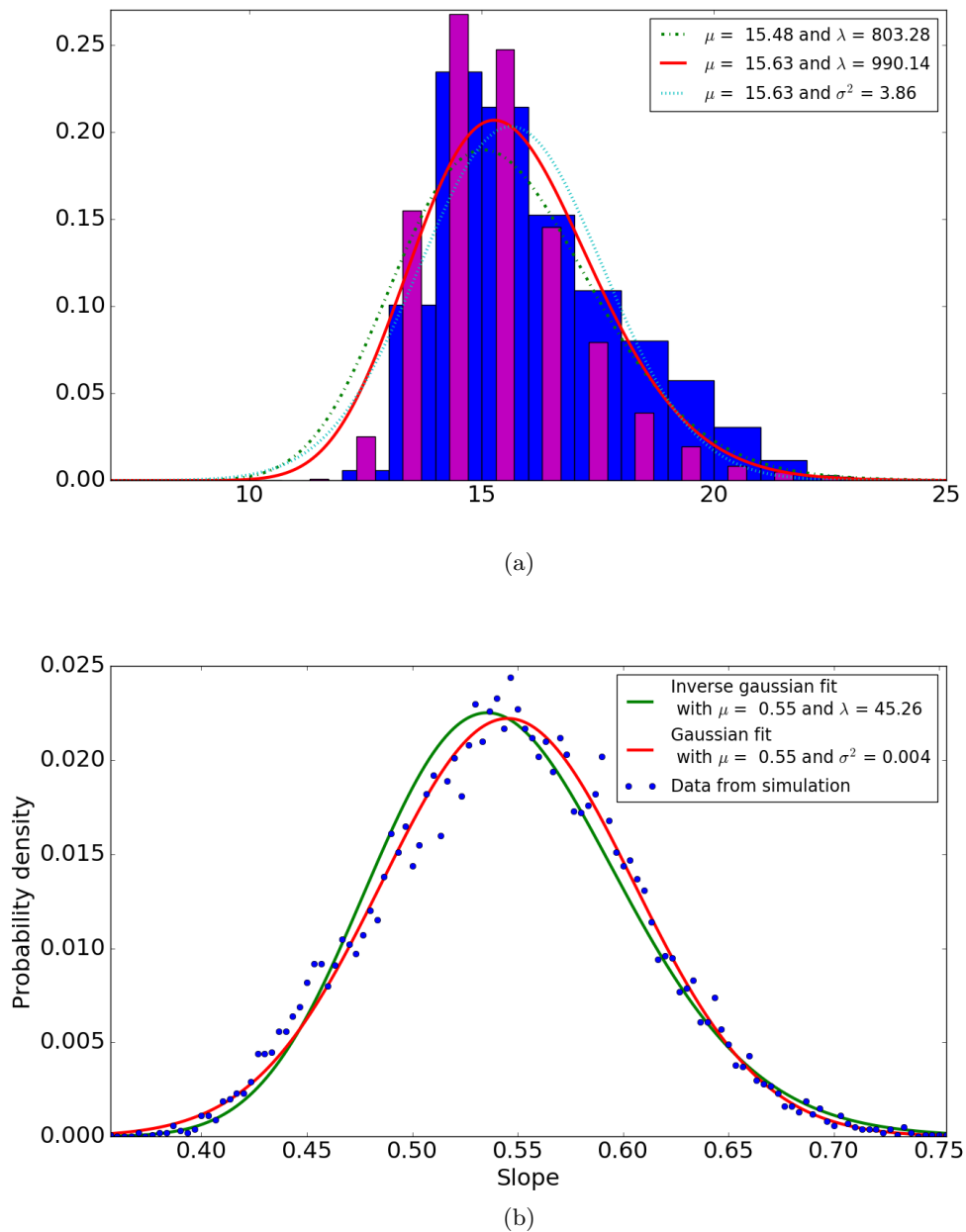


Figure 8.26: (a) Distribution of the growth times over 10,000 runs of the process. Surrogate data (blue wide histogram) and theoretical data (magenta thin histogram) are fitted with an Inverse Gaussian distribution (respectively the red full line and the green dashed line). Surrogate data distribution is also fitted with a Gaussian distribution (dotted cyan line). (b) Distribution of the logit slopes over 10,000 runs of the process. Inverse Gaussian fit is shown with a green full line, regular Gaussian fit with a red full one. $M = 5000$, $\delta = 0.0$, $\beta = \beta^*$.

Péclet numbers	Latency time	Growth time	Logit slope
Surrogate data	4.31	126.7	166.0
Corpus data	2.31	13.63	10.55

Table 8.3: Péclet numbers of the Inverse Gaussian fit of the latency times distribution, of the growth times distribution, and of the logit slopes distribution, for both the surrogate numerical data and the corpus data of chapter 4.

X	μ_{lat}/μ_{growth}	$\lambda_{lat}/\lambda_{growth}$
Surrogate data	8.06	0.27
Corpus data	0.88	0.15

Table 8.4: Comparison between corpus and surrogate data of two quantities which scales identically.

If P_e is small compared to one (i.e. λ is small compared to μ), then:

$$t_{mode} \approx \frac{1}{3}\lambda, \quad (8.80)$$

so that the mode is much smaller than the mean μ . If P_e is much greater than one we get:

$$t_{mode} \approx \mu. \quad (8.81)$$

In the latter case, the Inverse Gaussian is peaked at its mean value, and is shaped like a regular Gaussian distribution.

The Péclet number has the advantage that it does not depend on the scale of the process (since μ and λ have the same scaling behavior). This is a convenient property in our case, since we do not have any clear way to compare the scale of our surrogate data and that of our corpus data. Yet, we can compare their Péclet numbers (Table 8.3).

The agreement between the surrogate data and the corpus data is quite nice for the latency time (they are of the same order of magnitude), but really poor for the growth time and the slope. Whereas for the surrogate data, the distribution of these two quantities is almost Gaussian (as indicated by the high Péclet number), it is clearly not the case for the linguistic data.

We can also compare various scale independent quantities, such as the ratio between the mean of the latency time and that of the growth time, but the results are inconclusive (Table 8.4). As the slopes depend on a mixture of scales, it is not relevant to include their statistical features in this comparison.

Interestingly, if we trust the modeling of the latency, we can get an idea of the difference of the timescale between the surrogate and corpus data. The ratio of the λ parameters (λ numerical/ λ corpus) gives 28.8, the ratio of the μ parameters gives 15.4, so that both are of the same order of magnitude.

Now, if we divide the mean surrogate growth time (15.6) by this scale ratio, we obtain something between one decade and half a decade —between 5 years and 10 years. Thus, the growth pattern is too short in our model. Furthermore, the very

high Péclet number indicates that it varies too little. We shall come back to these discrepancies later on in chapter 9.

For now, we can already state that we do regard this estimate as a true, important result regarding language change. It shows that the minimal replacement (replacement in one context only) through a semantic expansion process lasts around five-ten years, which seems a plausible and appropriate timescale.

Correlations

We can also look for correlations between the different quantities, just as we did for the corpus data:

- Growth time and slope anti-correlate with a Pearson coefficient of -0.69
- Growth time and latency time are independent (Pearson coefficient of -0.01)
- Slope and latency time are independent (Pearson coefficient of 0.01)

Hence, we find back the negative correlation between the growth time and the slope; but this is an expected feature of an S-shaped curve, not specific of any process, and cannot be held as a significant trait. Furthermore, as the process is Markovian, it is expected that the latency time and the growth time are independent quantities, as are the latency time and the slope.

However, what is interesting is that, if we decrease the variation tolerance parameter for the latency pattern detection to 0.15, then we find a growth-latency correlation of 0.35 (and a latency-slope correlation of -0.36). This means that the procedure used to detect latency introduces a statistical correlation between the growth time and the latency time (expectedly, as the latency pattern is extracted on the basis of the growth time specific features).

Therefore, the latency-growth correlation that we found for corpus data (with a Pearson coefficient of about 0.2) might just have been spuriously introduced by the procedure that we performed to detect the latency pattern. It is thus legitimate to assume that, in the actual instances of semantic change, the latency time and the growth time are independent quantities, which is a strong prediction of our model.

8.2.5 Summary

Our model allows us to capture different quantitative and qualitative features of semantic change. Especially, it closely accounts for the non-trivial behavior of the latency: its distribution shape is both quantitatively and qualitatively predicted. Yet, something is missing for the S-shaped growth part, whose specific behavior is poorly reproduced by our model (in particular, we do not find the precise shape of the Inverse Gaussian distribution that fitted the data). Our model also predicts that the latency and the growth must be independent, and we showed that this prediction is compatible with our findings from corpus data.

8.3 Model variants

So far, we have described a mechanism of change which falls into the speaker's perspective entirely: semantic change occurs because the speaker innovates with the choice of a new, linguistically motivated choice of a variant (X) to express a meaning C_1 already covered by a former variant Y . As the innovation is motivated by the semantic organization of language, which is roughly shared by all members of the speech community, and because of contextual cues (be it situational, discursive, or semantic), we assumed that this innovation was immediately understood by the hearer. The role of the hearer is then reduced to acknowledge the innovation by recording it into memory. However, two points can be made against this simplistic picture. First, language change is not necessarily driven by speakers' innovation. Second, if the hearer makes use of the semantic channel to understand the new variant as it is meant to be, then this semantic channel can be used productively in the interpretation process.

The goal of this section is therefore to offer a different model, accounting for the hearer perspective.

8.3.1 Hearer perspective

The set-up and the notations are the same as before: two meanings, C_0 and C_1 , are populated with occurrences of two different, competing linguistic forms, X and Y , have a finite capacity of occurrences, M , and are related by a unilateral asymmetric influence of weight γ of C_0 over C_1 . The initial situation is also such that $N(0) = 0$, where N represents the number of occurrences of X in C_1 , which is the theater of the competition.

Dynamics

The dynamics proceeds along the following steps (Fig. 8.27):

1. A meaning is chosen to be expressed. C_1 is chosen with probability α , C_0 with probability $1 - \alpha$.
2. An occurrence is chosen to express the meaning corresponding to the chosen site S . X is chosen with probability $P_0(x^S)$ and Y otherwise, where x_S is the actual frequency of X in site S . There is no effective frequency in this framework.
3. An occurrence is recorded in memory and a former occurrence is deleted to keep the memory constant.
4. If the site C_0 has been chosen in step 1, then there is a probability γ that C_0 elicits meaning C_1 . If so, an additional occurrence of the chosen form is added in C_1 and a former occurrence is deleted.

The jump probabilities, $R^H(x)$ and $L^H(x)$ (the H superscript has been added to denote the 'hearer' perspective) are now given by:

$$\begin{cases} R^H(x) = \alpha P_0(x)(1-x) + (1-\alpha)\gamma(1-x) \\ L^H(x) = \alpha(1-P_0(x))x \end{cases} \quad (8.82)$$

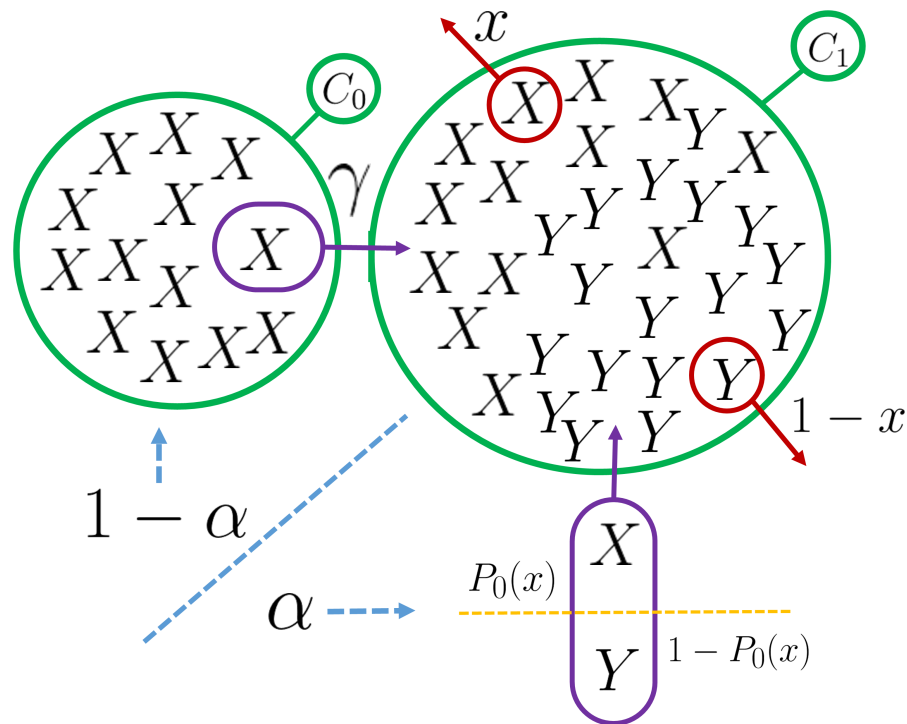


Figure 8.27: Schematic representation of the hearer process. First, a meaning is chosen: either C_1 with probability α or C_0 with probability $1 - \alpha$. Then a form is chosen to express the meaning. If C_0 is the chosen meaning, X is systematically chosen and then it can elicit meaning C_1 with probability γ , leading to the recording of an occurrence of X in C_1 . If C_1 is chosen, then X is produced with probability $P_0(x)$, where x is the proportion of occurrences of x in the memory associated with C_1 . Y is chosen otherwise. In all cases, whenever an occurrence is recorded in memory, a former occurrence is deleted with uniform probability; hence, the probability to delete an occurrence of a given form is equal to its frequency in the memory. Addition of an occurrence in C_1 is indicated by a purple cartridge, deletion is marked by a red circle.

A new occurrence of X appears in the C_1 memory after an iteration of the process if C_1 is chosen, X is chosen to express C_1 and an occurrence of Y is deleted, or if C_0 is chosen (X being systematically chosen to express it in that case), elicits meaning C_1 , and an occurrence of Y is deleted to make room for the newly recorded occurrence of X . A new occurrence of Y appears in C_1 only if C_1 is chosen to be expressed, Y is produced, and a former occurrence of X is deleted.

If we suppose that both sites C_0 and C_1 have equal probabilities to be chosen, so that $\alpha = 1 - \alpha = 1/2$, then the jump probabilities simplify into:

$$\begin{cases} R^H(x) = \frac{1}{2} (P_0(x) + \gamma) (1 - x) \\ L^H(x) = \frac{1}{2} (1 - P_0(x)) x \end{cases} . \quad (8.83)$$

We can remove the $1/2$ factor, since it only accounts for the fact that we are now conjointly considering the dynamics for the two sites C_0 and C_1 . It could have been added as well for the speaker jump probabilities (where if C_0 was chosen nothing could have happened so that it would not have changed the formulae). The associated drift is much higher at the beginning of the process than it was in the speaker perspective (Fig. 8.28).

Critical behavior

From (Fig. 8.28) it is clear that the hearer model also displays a critical behavior, for the same critical value of γ . We add a superscript to the critical value of the γ parameter (and to the critical frequency as well) to refer to the different variants of the model (S for ‘Speaker’, H for ‘Hearer’). It appears that we have $\gamma_c^S = \gamma_c^H$.

This can be proven easily. The fixed points are solutions of the equation:

$$v^H(x) = R^H(x) - L^H(x) = 0, \quad (8.84)$$

which translates to:

$$(P_0(x) + \gamma) (1 - x) - (1 - P_0(x)) x = 0 \quad (8.85)$$

and finally:

$$P_0(x) + \gamma - (1 + \gamma)x = 0 \quad (8.86)$$

If we now replace x by $y = (1 + \gamma)x - \gamma$, we get:

$$P_0\left(\frac{y + \gamma}{1 + \gamma}\right) = y. \quad (8.87)$$

According to the definition of P_γ , we have:

$$P_0\left(\frac{y + \gamma}{1 + \gamma}\right) = P_\gamma(y), \quad (8.88)$$

so that we obtain:

$$P_\gamma(y) = y. \quad (8.89)$$

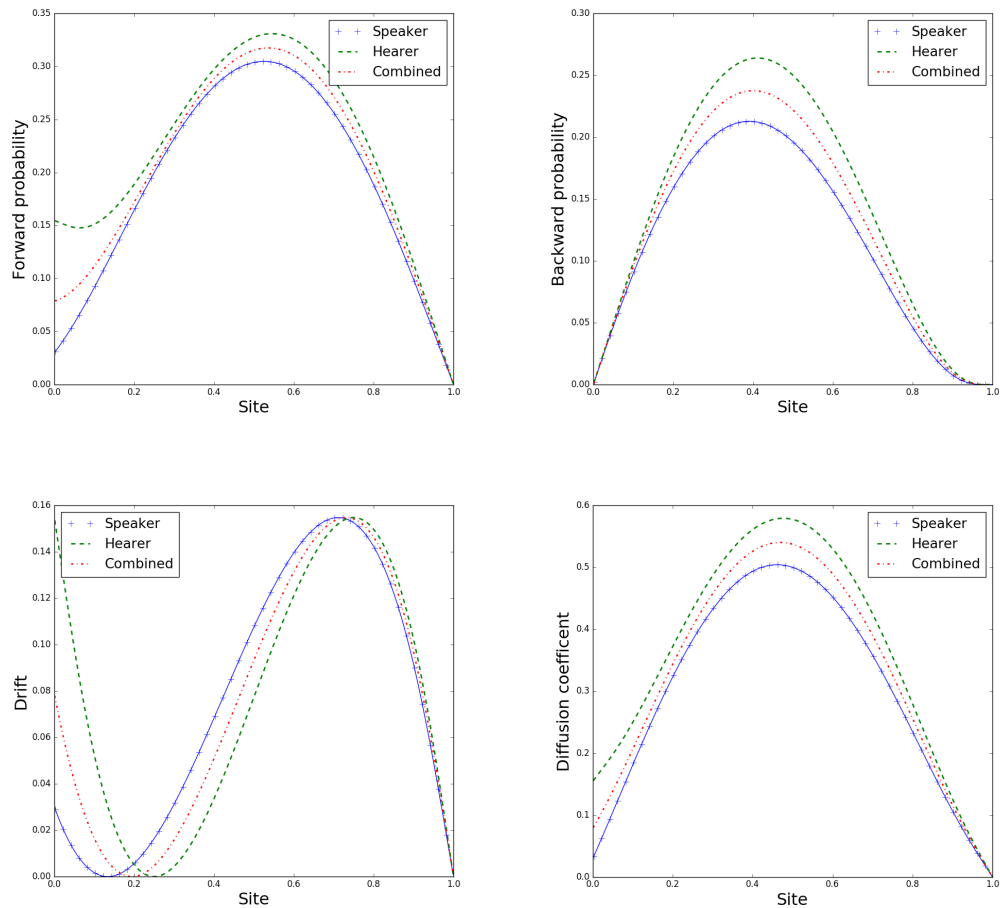


Figure 8.28: Hearer case. From left to right, top to bottom: forward jump probability, backward jump probability, deterministic velocity (difference between the jump probabilities forward and backward), diffusion coefficient (sum of the jump probabilities), for all sites of the system, for the three different model variants ('speaker' in blue crosses, 'hearer' in dashed green line, 'combined' in dotted red line). $M = 5000$, $\delta = 0.0$, $\beta = \beta^*$. The γ parameter has been tuned to criticality for each of the three models.

which is exactly equation (8.29).

Hence we have $y_c = x_c^S$ and $\gamma_c^H = \gamma_c^S$, from which we deduce:

$$\begin{cases} x_c^H = \frac{x_c^S + \gamma_c^S}{1 + \gamma_c^S} \\ \gamma_c^H = \gamma_c^S \end{cases} . \quad (8.90)$$

However x_c does not show anymore a non-monotonic behavior and there is no equivalent of β^* in this case. We shall therefore keep β at the same value as before.

Trap region

Similarly, we can consider the trap region and try to give a theoretical estimate with the coordinates of the third derivative of the uncertainty on transmission time (Fig. 8.29). The results obtained are comparable with those of the speaker model. Once more, the coordinate of the critical point matches with the first minimum of the third derivative of the uncertainty. The exit of the trap, in this determination, coincides nicely with the maximum of relative uncertainty (which immediately precedes a very sudden and abrupt decrease of this quantity).

We can also match the main features of the numerical latency and growth times as we did in the speaker case. Here, the things are made a bit trickier by the fact that the critical point x_c^H is much higher than its ‘speaker’ counterpart, so that the variation tolerance threshold for the latency must be set higher (0.30 in our simulations). Yet, the trap obtained by fitting the statistical features of the macroscopic quantities closely correspond to the theoretical picture (Fig. 8.30).

The results of the numerical simulation are overall very similar to the previous case. The correlation coefficients, the Péclet numbers, are all of the same order of magnitude. The comparison of the scales also predicts that the mean growth time would correspond to a real time of the order of a decade. There is thus no significant difference between the two variants of the model, except that the critical frequency is higher now. It also confirms the robustness of our findings, as they do not depend on the specific mathematical details of the model.

8.3.2 Combined perspective

As we have seen, in both cases, the innovation, be it due to the speaker or the hearer, is motivated by a language organization they both share to some extent (in the model, as a single agent plays both roles, the language organization is exactly the same in both cases). We can therefore consider that both speakers and hearers innovate by producing X to express C_1 , or by understanding C_1 when hearing X , and combine the previous two perspectives.

The dynamics change very little compared to the hearer case: when a site has been chosen to be expressed, the form to express it is chosen in terms of its effective

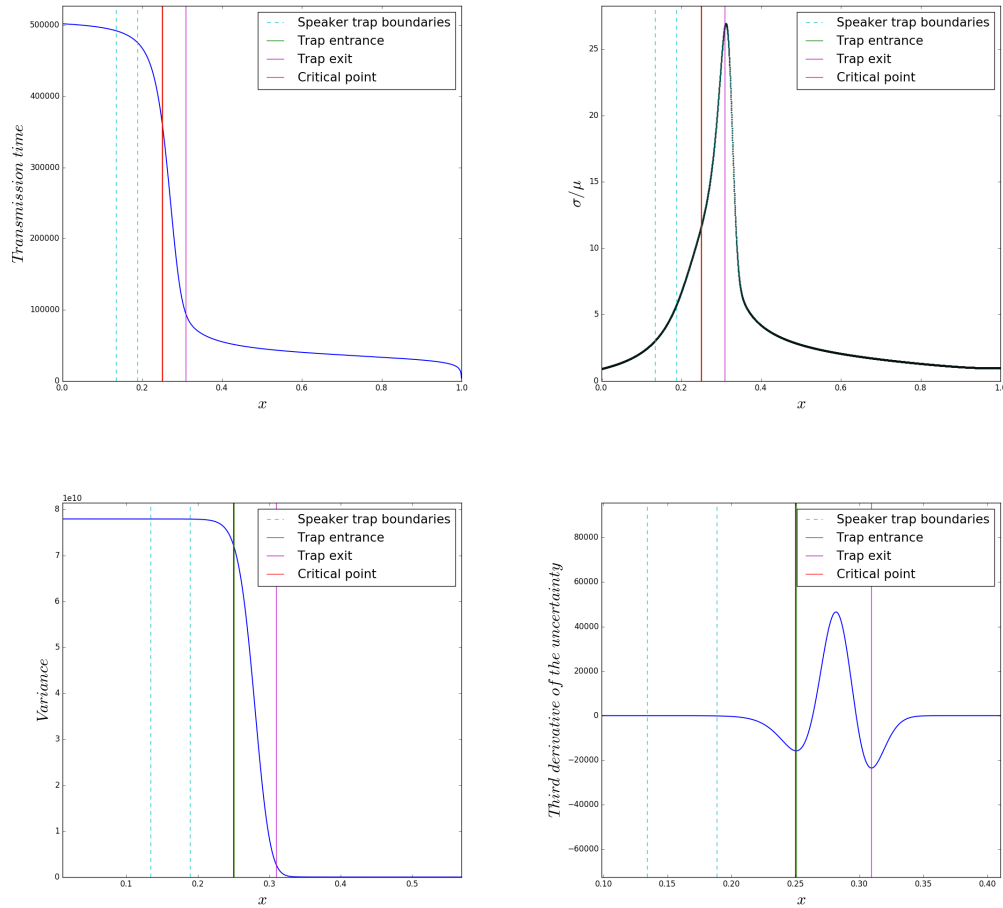


Figure 8.29: Hearer case. Trap boundaries, computed as the minima coordinates of the third derivative of the uncertainty, are shown along different quantities. From left to right, top to bottom: transmission time, relative uncertainty on transmission time, uncertainty (variance of the transmission time), and third derivative of the uncertainty. The trap boundaries obtained in the speaker case are shown in dashed cyan lines. The critical position x_c is shown in a red full line. The trap entrance is shown in a green full line (but overlaps with the critical point), the trap exit in a full magenta line. $M = 5000$, $\delta = 0.0$, $\beta = \beta^*$.

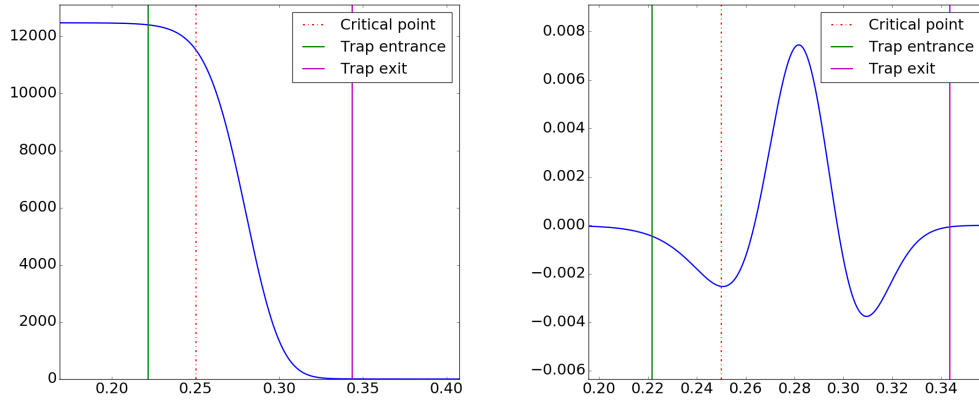


Figure 8.30: Hearer case. Limits of the trap so as to fit the main statistical features of the numerical latency and growth times, in the hearer perspective, shown along the uncertainty on transmission time (left) and the third derivative of this uncertainty (right). $M = 5000$, $\delta = 0.0$, $\beta = \beta^*$.

frequency instead of its actual frequency. Hence we have:

$$\begin{cases} R^B(x) = (P_\gamma(x) + \gamma)(1 - x) \\ L^B(x) = (1 - P_\gamma(x))x \end{cases}, \quad (8.91)$$

where the superscript B stands for ‘Blend’ since the speaker and hearer perspectives are now blended. These jump probabilities, once tuned at criticality, fall nicely between those of the speaker perspective and those of the hearer perspective (Fig. 8.28).

As before, we can compute the critical couple (x_c^B, γ_c^B) . The drift is now given by:

$$v^B(x) = P_\gamma(x) + \gamma - (1 + \gamma)x. \quad (8.92)$$

To find the fixed points we look for the zeros of this quantity. We define $y = (1 + \gamma)x - \gamma$ so that we get:

$$P_\gamma\left(\frac{y + \gamma}{1 + \gamma}\right) = y. \quad (8.93)$$

Using the definition of P_γ , we obtain:

$$P_0\left(\frac{y + \gamma(\gamma + 2)}{1 + \gamma(\gamma + 2)}\right) = y. \quad (8.94)$$

Hence, we can define $\tilde{\gamma} = \gamma(\gamma + 2)$ and find back the usual equation:

$$P_{\tilde{\gamma}}(y) = y. \quad (8.95)$$

From this, it follows that $(y_c, \tilde{\gamma}_c) = (x_c^S, \gamma_c^S)$ so that:

$$\begin{cases} x_c^B = \frac{x_c^S + \gamma_c^B}{1 + \gamma_c^B} \\ \gamma_c^B = \sqrt{1 + \gamma_c^S} - 1 \end{cases} . \quad (8.96)$$

From the equations of the fixed point obtained in the speaker case, we can therefore derive the fixed point for all model variants. In this blended variant, the critical threshold for the influence between meanings is much lowered, as the parameter $\tilde{\gamma} = \gamma(\gamma + 2)$ plays the role of an ‘effective’ weight. It means that the joint tendency to innovate as a hearer and as a speaker considerably reduce the required influence for the semantic replacement to happen. Whether the change is driven by the speaker innovation, or by that of the hearer, does not make the replacement easier or more likely; but if the two are combined, then a much weaker cognitive inference $C_0 \rightarrow C_1$ is sufficient to set the change at the benefit of the new variant.

As before, we can look at the theoretical trap boundaries (Fig. 8.31). Interestingly, the trap entrance almost coincides with the trap exit of the speaker’s case, while the trap exit closely overlaps the trap entrance of the hearer’s case.

We can also try to fit the numerical, macroscopic quantities by optimizing the trap boundaries. The numerical simulations ran with a variation tolerance of 0.25 (between those of the speaker case and the hearer case, as the critical point in this blended perspective lie between the critical point of the two previous ones). We obtained trap boundaries once more compatible with the theoretical analysis (Fig. 8.32). Interestingly, the trap entrances match almost exactly (site 972 in the theoretical case, 978 in the ‘fit’ case).

The theoretical simulations of this process are once more closely in line with the results of the speaker case. We find similar values of Péclet numbers, and a similar estimate for the real-time duration of the average growth (that is, a decade). Given the observables that we obtained from the corpus data, and how comparable those three variants happen to be, there seems to be no obvious way to assess if a real instance of semantic change is more speaker-driven, hearer-driven, or rely on a competition of both. The main difference between these three perspectives is that, in the combined, the strength of the semantic association between the two meanings required for the change to unfold up to completion is much weaker than in the two other cases. Hence, if we had a way to get an empirical estimate of this γ parameter for an actual instance of semantic expansion, we would be likely to distinguish between a case where the two mechanisms, hearer and speaker, concur to the change, and a case where one of the two has acted independently. Indeed, if a critical phenomenon is found associated with a high value of γ , it means that, certainly, only one mechanism was present. If the value of γ is found to be very low, then the two mechanisms are likely to have cooperated.

8.3.3 Different interpretations of the γ parameter

Different interpretations can be given to the γ parameter. So far we have considered it as the strength of an inference from one meaning to another, or as a probability

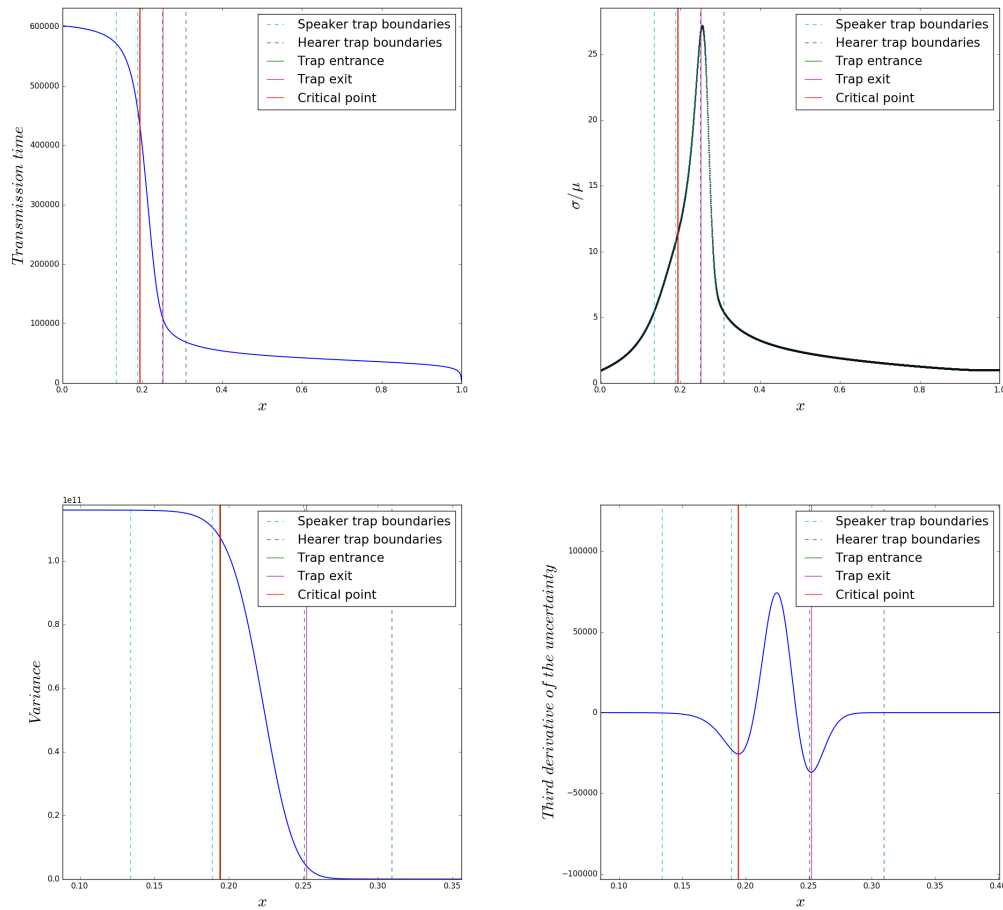


Figure 8.31: Combined case. Trap boundaries, computed as the minima coordinates of the third derivative of the uncertainty, are shown along different quantities. From left to right, top to bottom: transmission time, relative uncertainty on transmission time, uncertainty (variance of the transmission time), and third derivative of the uncertainty. The trap boundaries obtained in the speaker and the hearer cases are respectively shown in dashed cyan and dashed green lines. The critical position x_c is shown in a red full line. The trap entrance is shown in a green full line (but overlaps with the critical point), the trap exit in a full magenta line. $M = 5000$, $\delta = 0.0$, $\beta = \beta^*$.

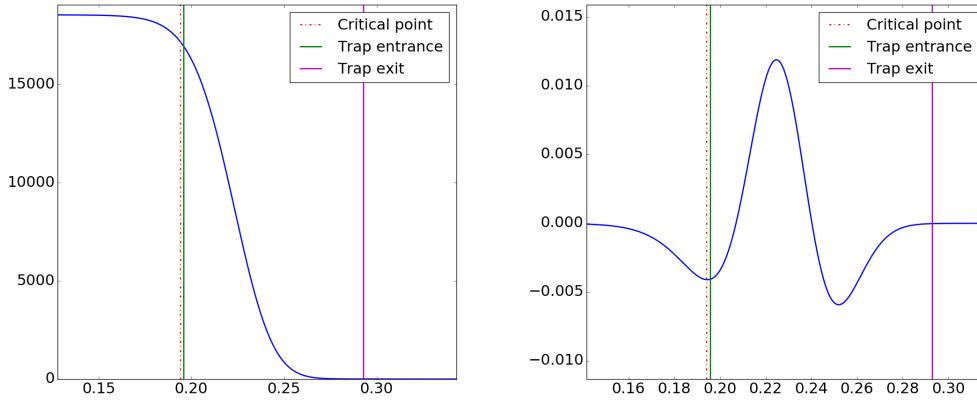


Figure 8.32: Combined case. Limits of the trap so as to fit the main statistical features of the numerical latency and growth times, in the hearer perspective, shown along the uncertainty on transmission time (left) and the third derivative of this uncertainty (right). $M = 5000$, $\delta = 0.0$, $\beta = \beta^*$.

of elicitation. It was also the weight associated with the unidirectional edge relating meaning C_0 to C_1 in the semantic network. We can release this interpretation and consider that the semantic network still has directional edges, but now unweighted (i.e. they all have a weight of 1 as long as they exist). We shall see how to make a γ value appear in this case, for the speaker perspective as well as for the hearer one. Then we shall discuss how to reconcile these two interpretations in the blended case.

γ as a difference of memory capacities

In the speaker model, the γ parameter appeared through the effective frequency:

$$f = \frac{N + \gamma M^0}{M^1 + \gamma M^0}, \quad (8.97)$$

where we had $M^0 = M^1 = M$. Now that the network is no longer weighted, the effective frequency would become:

$$f = \frac{N + M^0}{M^1 + M^0}, \quad (8.98)$$

so that if we release the assumption that all memory capacities are equal, we get:

$$f = \frac{N/M^1 + M^0/M^1}{1 + M^0/M^1}. \quad (8.99)$$

By setting $\gamma = M^0/M^1$, we find the exact same definition of the effective frequency as before. The main difference is that, whenever γ was the weight of the link, we could safely consider, by definition, that γ was lower than 1 (under the reasonable assumption that a meaning does not affect another meaning more than this other meaning affect

itself). This is no longer the case and γ can be greater than 1. If this were the case, we would not observe latency. The growth would be very fast at the beginning and would slow down by the end; it would no longer follow an S-curve.

We shall discuss in the next section why it could be the case that γ is very low in the contexts of semantic change.

γ as a difference of calling probabilities

In the hearer model, γ is the probability to elicit meaning C_1 when C_0 has been chosen. If the network becomes unweighted, then the elicitation probability becomes equal to one. Yet, the choice of the context to be expressed was governed by a set of ‘call’ probabilities α and $1 - \alpha$ (the probability to call either context C_1 or C_0 , respectively) so that the forward jump probability was given by:

$$R^H(x) = \alpha P_0(x)(1 - x) + (1 - \alpha)(1 - x). \quad (8.100)$$

At that point we assumed that the calling probabilities were equal. If we release this assumption, then we can still factorize by the calling probability α and get:

$$R^H(x) = \alpha \left[P_0(x)(1 - x) + \frac{1 - \alpha}{\alpha}(1 - x) \right]. \quad (8.101)$$

The α factor will only affect the timescale (and can be added as well in the speaker case) and is equivalent to the $1/2$ we got before. In this equation, it is clear that we can define a γ parameter equal to:

$$\gamma = \frac{1 - \alpha}{\alpha}, \quad (8.102)$$

so as to get back exactly the same jump probabilities as before.

Now, γ can be greater than 1 if the calling probability of C_0 is greater than that of C_1 . Conversely, the less probable it gets to call C_0 , the smaller the γ is. Then, we can have two alternative, equivalent pictures: one in which the two contexts have equal calling probabilities, and where the elicitation is itself probabilistic, or one where the elicitation is systematic, and the two contexts have uneven calling probabilities.

Combining the two interpretations

In the blended perspective, there is no need for the two γ (the one coming from the speaker mechanism, the other one from the hearer) to be equal. By adding superscripts to explicit to which mechanism each γ corresponds, we would have:

$$\gamma^B = \gamma^S + \gamma^H + \gamma^S \gamma^H, \quad (8.103)$$

where γ^B is the effective γ of the blended process.

However we can consider that there would be an equivalence between the two perspectives, so that we can try to set:

$$\gamma^S = \gamma^H. \quad (8.104)$$

It follows that:

$$\alpha = \frac{M^1}{M^0 + M^1}. \quad (8.105)$$

Hence the calling probability of C_1 , with respect to C^0 , is equal to the relative size of its affected memory. Conversely, the more frequently a given meaning is called, the bigger memory capacity it would be given.

The γ parameter can therefore be given an interpretation along two different lines: one would correspond to the strength of semantic associations in the cognitive organization of language, another to the frequency of use of the different meanings (not to be mistaken with the frequency of use of the different forms) and to the importance that they are given in memory. Those two interpretations have no need to be opposed and can actually complement each other. The γ parameter can actually be seen as a mixture of a semantic association, and of the hierarchical importance given to the different meanings. In both cases, it reflects deep features of the cognitive organization of language.

8.4 Discussion of the model

We have laid down the model, exposed its main theoretical and statistical features, and explored some of its possible variants. The model proved to compare well with the linguistic data of chapter 4, especially the latency phenomenon, capturing non-trivial qualitative and quantitative features. This comparison has furthermore proved to be robust, as the different variants of the model gave rise to comparable results.

Now that the relevance of the model regarding the phenomenology of semantic change has been assessed, we can discuss in a larger extent what exactly does the model explain, and which consequences for language change it can allow us to draw.

8.4.1 Understanding the phenomenon of semantic change

The model help us to understand semantic change in three ways: first, by proposing a narrative of semantic change, from start to completion, as it would manifest in corpus data; second, by solving the Threshold Problem efficiently, and last, by pointing out the difficulties that a linguistic account of the change has to face. We shall address them successively.

Setting things up

At the beginning of a semantic change lies a semantic association, γ . In our model, we set an initial situation which could either lead to a complete replacement ($\gamma > \gamma_c$) or to a marginal use of the new variant in the disputed meaning ($\gamma < \gamma_c$). The first question is, how did this initial situation come to be? There are two possible explanations: either X was not here before, or the semantic association between the source meaning and the target meaning was too weak.

The first explanation is unsatisfactory. If X was not here, it meant (at least according to our rather extreme hypothesis that the semantic territory is fully populated in its entirety at all times) that another linguistic variant was occupying its place, so

that it could have taken over the meaning C_1 as well. We shall see later on that it may be that another variant, let us call it Z , was in C_0 before X , and could not evict Y while X can, but this would remain an unusual situation.

The second explanation is much easier. As the parameter γ is a control parameter governing the occurrence or not of the change, a small change in γ could have led in a stalemate to a situation allowing for semantic replacement. We already stressed that the bifurcation, governed by the strength of the semantic association, was an interesting feature of our model, as a small variation in this quantity could change the odds entirely. Thus, at the beginning of the change, we can posit a small increase of a semantic association, which would pass unnoticed for the speakers, but large enough to set the replacement in motion. As for the question of how such an increase could have happened, we leave it open, having already suggested several possibilities (especially, the influence of other language changes, and possibly some analogical reasoning).

However, the fact that, prior to some point in time, no occurrence of X are found in C_1 , must mean that either X had not reached C_0 yet or that the semantic association was non-existent. Otherwise, even if γ is below γ_c , we would still find sporadic occurrences of X with the meaning C_1 , with a frequency $P_\gamma(x^*)$ corresponding to the coordinate of the stable fixed point (Fig. 8.5), which admittedly can be so low that it would be equivalent to a complete absence of X in C_1 . Still, it is not impossible for a semantic association to be created from scratch, though the conditions, the modalities, and the mechanisms of such a creation are unknown.

Note that, from the sole point of view of the semantics, at this point of the process we can already consider that the semantic expansion has happened. Indeed, the conditions of the cognitive organization of language are set so as to allow for X to be used with meaning C_1 . The semantic expansion is therefore enacted very quickly by corresponding occurrences incompatible with the pairing $[X \leftrightarrow C_0]$.

This little observation underlines the crucial difference between the semantic expansion as a cognitive association, and the semantic expansion in language use. The former can be effective, whereas the latter would only correspond to marginal linguistic practices. Once more, it reminds us that frequencies are an indispensable part of linguistic knowledge. The semantic organization is one thing; the actual population of this structured territory by occurrences encoding memories of past linguistic experiences is definitely another.

A scenario of semantic change

Now that we have settled down this matter of cognitive semantic expansion and of the initial situation, we can recap the main steps a semantic change is expected to follow.

- The semantic association between the two meanings C_0 and C_1 allows occurrences of X to be produced with meaning C_1 , and subsequently recorded in memory.
- This production will very shortly increase up to a point where the frequency production would be stuck at a low value. Then, X is sporadically produced with meaning C_1 and can be found occasionally in corpora.
- At some point the frequency of X will increase slightly, due to chance, so that the following rise will be triggered.

- The frequency of use of X grows according to an S-curve.
- As X comes to saturate the memory capacity associated with C_1 , the change ends. The former variant Y can no longer be used to produce C_1 at this point.

Hence, the main steps of the semantic change, as evidenced in diachronic linguistics, are accounted for in this scenario; first, the form shows up in a context (C_0) close to its new meaning (C_1) and will remain there as long as the conditions of invasion are not met (γ being too low); then the semantic association is strengthened and first, sporadic occurrences of the form with its new meaning can be found; then the frequency rise really begins and follows an S-curve to completion.

This scenario takes place in the linguistic memory of the speakers' community. It describes a feedback between language use, influenced by the semantic organization of language, and its encoding in memory. It does not involve pragmatic considerations, specific communication maxims (contrary to Keller (1994); Haspelmath (2000b)), nor an inherent superiority of the new variant. In this account, change is therefore a mere side-product of regular language use.

A slightly alternative scenario can be imagined in which γ would increase gradually. Then, the frequency of the new variant would follow the position of the stable fixed point, which itself moves forward as γ evolves. Then γ would pass the threshold γ_c and the frequency of the form will get free of the fixed point (which won't exist anymore once passed the threshold), and rise according to the S-curve. The difference would be that instead of the frequency to increase suddenly from zero to a non-zero value and then remain roughly stable for the long latency time, it would increase gradually and regularly. The specific evolution dynamics of the γ parameter would be directly reflected in this early frequency evolution of the form.

As the phenomenology of language change indeed exhibits a latency, it speaks in favor of a stable γ . This means that, for the change to be triggered the way it is, γ must not increase slowly and gradually, but by a small, nearly instantaneous jump. These two views (γ as constant for the duration of the change and γ as a trigger of the change) would lead to posit that the semantic association strength evolves through discrete and seldom increments, so that between two of such step changes, it remains stable and still for a long period. Here again, we are left with the question of how semantic associations are created and strengthened.

In the following chapter we will propose a view which can help to solve this puzzle, introducing a non-local component in this picture by associating with linguistic forms a likeliness to express the different meanings. Then the network structure and the weight of its edges can be assumed stable, and the trigger of the change would rely on the population dynamics happening in other, non-related sites. In this case, language could be seen as an out-of-equilibrium, ever-changing ecological system of linguistic forms competing over the semantic territory.

Overcoming the Threshold Problem

A great advantage of our model is that it solves the question of the initial frequency. As we saw in chapter 6, most models of the S-curve posit a non-zero initial frequency of the new form. However, how the new form can reach such an initial frequency is not a

trivial question, as Nettle (1999) has highlighted. Here, we proposed two mechanisms, grounded on attested linguistic features, that could lead a linguistic form to be used and accepted to express a meaning to which is already attached a whole population of occurrences of an established concurrent variant. Both rely on the idea that a ‘new’ form may be new from the point of view of this particular use, but is not linguistically new, and can rely on its own history to be supported in this competition process.

The first of these two mechanisms is effective frequency. Effective frequency allows to search in neighboring sites when one wants to select a linguistic form to express a given meaning. Thus, linguistic forms have an effective frequency, reflecting the influence of these neighboring meanings’ occurrences contents, which can be non-zero even if their actual frequency (the record of their past occurrences with this particular meaning) is zero. Thus, a linguistic form could have never been used to express a meaning, and yet be considered apt to convey it, because it is already associated with closely related meanings.

The other mechanism is elicitation. Indeed, a given meaning can elicit another meaning. Whenever a form is used to express its conventional meaning, it can be felt as if it would suggest another, additional, related meaning, and be recorded as an occurrence in the latter meaning associated memory. Thus, this memory site is constantly fed with occurrences of a new linguistic form. This feeding can be weak and irrelevant from the point of view of language use, but it can be sufficient to trigger the replacement.

In both cases, a zero initial frequency of the form does not prevent the change to happen. Yet indeed, the S-curve will generally start at a non-zero frequency, which is the frequency corresponding to the latency trap. The explanatory power of our model not only lies in the S-curve, which had been efficiently modeled in the past, but in the reunification, into a single picture, of the two parts of the change: the rise from zero frequency to a threshold frequency, and the subsequent S-shaped growth.

The latency drawback

The latency phenomenon, as we described it, has several consequences for the investigation of language change.

First, we cannot predict the length of the latency. It is a stochastic quantity, drawn from a probability distribution. It can last half a decade, or a century. According to our corpus data analysis, the most probable time of latency is equal to 13 years. However, one must not forget that the Inverse Gaussian (especially at low Péclet number, which is the case here) has a very large tail, so that the latency time can easily be way longer. If we are looking at ongoing change, and observe that this frequency stays still, we cannot know if and when it eventually rises. Hence, the time of rise cannot be predicted, which is a severe drawback of the latency phenomenon.

Second, it introduces a discrepancy between the time at which ‘something happened’, so as to set the conditions for the change to eventually happen, and the time at which the change finally unfolded into an S-curve. According to our scenario, there is no need to look for explanations of why the frequency would start to rise, why the rise would accelerate, and why this impetus would wane at some point. Indeed, all these features are already encoded in the change.

What crucially matters is the very start of the process, before the latency even starts. The linguistic, historical, cultural, sociological or political factors which may have triggered the change are to be investigated for as soon as the first innovative occurrences show up in the corpus. In the case of *du coup*, it would be in the first decades of the XIXth century.

This is a very strong prediction of our account of semantic change, which would drive the study of specific instances of language change. We shall hint in the next chapter that the later parts of the change also benefit considerably from a linguistic enquiry, but still, our theoretical picture invites us to focus on the linguistic settings allowing for the very first occurrences of a novel use. There lies the spark which causes the whole change to later unfold.

8.4.2 A near-critical behavior

We shall now address a puzzling question arising out of our model. The latency phenomenon, as we saw, is associated with a critical behavior. For it to occur, the γ parameter must be tuned to the critical value, or at least near to it. In our empirical account of language change, we observed that latency was not an infrequent phenomenon —quite the contrary in fact, as we had enough instances of it to get its statistical distribution. This would mean that the γ parameter, in real instances of change, is often tuned to criticality. How would it be so?

In the following, we will give two possible lines of answer. The first one is related to the alternative interpretations of the γ parameter that we suggested previously. The last one anticipates somehow the contents of the next chapter as it discusses criticality in relation with the network structure of the semantic territory.

A protective network

The γ parameter can be interpreted as a ratio of the memory capacities associated with C_0 and C_1 , or alternatively, as the ratio of their calling probabilities. In this interpretation, γ can easily be greater than 1. If it were so, then change would proceed chiefly from big memory capacity meanings, frequently called in language use, from meanings of lower capacity meaning and more sparsely used. If this were to be the case, then language would be invaded by the linguistic forms expressing the more frequent meanings, whatever they may be.

However, not all sites are expected to be related to all other. Even though the edges are not weighted in the frame of these two interpretations (γ as memory sizes ratio and γ as call probabilities ratio), the graph need not be a complete one (and is not, for the semantic network shows a small-world structure). Furthermore, the edges are directed. The existence of many edges with a γ close to criticality, as suggested by the latency phenomenon evidenced in actual changes, would therefore entail that many of this directional edges are directed from a low-capacity, seldom called site, to a big-capacity site, much more frequently used. Otherwise, γ cannot be equal to its critical value, which is smaller than 1. Thus, it would be as if the directed network structure were set so as to counteract the tendency for widely used meanings to have their contents invade the smaller, rarer one.

This has direct theoretical implications in the case of grammaticalization, or more broadly, semantic expansion towards more functionality. Indeed, in this picture, the pervasive criticality suggested by the latency behavior suggests that the grammatical sites are bigger and more frequently called than the lexical ones. This does seem reasonable: we use procedural, operative functions to structure all sorts of utterances, while the more lexical, referential functions have more limited use. Thus, if the network was not protective, the linguistic forms expressing grammatical meanings, because of the very large call frequency of the latter, would spread very quickly all over the network and take on the weaker lexical meanings. Hence, language would become more uniform, and would be supported by a much reduced diversity of linguistic forms.

In this line of thought, the unidirectionality of grammaticalization appears in a very different light. It does not reflect a specific phenomenon or mechanism, but a specific network structure, whose edges are directed so as to prevent a tremendous uniformization of language. Put it in another way, if grammaticalization were not unidirectional, then all meanings would soon come to be expressed by the more grammatical of the forms. The unidirectionality of grammaticalization is therefore a necessary feature of the semantic network to keep a reasonably wide array of linguistic means.

This picture combines well with Sweetser (1988) account of grammatical meaning as fleshed out semantic schemas. Grammatical semantic schemas are more shallow and present less features than lexical schemas. But one does not always need all the features of the schema and the simpler, more abstract schema would be suitable in more cases. This would explain why grammatical meanings would be associated with much higher calling probabilities. Why would they be associated with a greater memory though? This may be because grammatical functions are crucial to shape the meaning of an utterance. A bigger memory lowers the likeliness of a replacement; and if replacement is yet to occur, it furthermore slows down drastically the whole process. Henceforth, a bigger memory allows for a more robust, more stable linguistic expression over time, which indeed seems a desirable feature of grammatical meanings.

Now, this does not explain why γ would be specifically tuned to criticality; it only explains why it is profitable to be much smaller than 1. We can assume that the language network is protective enough so as to draw directional associations only if this does not lead to any invasion of the contents of a meaning towards another. However, specific circumstances could produce a weak increase of the calling probability of the source meaning (which, as a reminder, is not the same as an increase in the frequency of any linguistic, observable form), leading the γ associated with an already existing link to exceed slightly the threshold value. In this view, cultural changes, setting up differently the importance of things, could be the true trigger of linguistic change. For instance, it might not be a coincidence if the rise of terms such as *coup d'état*, or any occurrence of *coup* in the political sense, precedes of a very few decades the grammaticalization of *du coup*. And, as it happens, the former may have been more frequently called in the long period of political renewals and instabilities that followed in the wake of the Revolution.

I would not favor such a view, however, for it requires too much of external motivation for a change to occur. It seems better advised to prefer a view of language change in which it results from the simple course of things. Indeed, linguistic change is so pervasive, so deeply intertwined with the diachronic flow of language, that invoking

much specific and seemingly clever circumstances for any single instance of change appears to require too profuse an array of exclusive, singular explanations.

Unremarkable changes

We now offer an alternative view to explain why language change would be driven to criticality. It does not rule out the previous one, but is somehow more flexible. The main idea is that language change is not tuned to criticality, but that changes which are not at criticality are not remarkable, and therefore not really noticed.

Let us recall that the semantic territory is very likely to be organized according to a small-world structure. In such a network, there are clusters of closely inter-connected meanings, and a limited number of edges between the different clusters, acting as shortcuts. This structure is, as such, unrelated with our concerns, since criticality depends on the weights between the different sites, which are not entitled to any value in a small world network.

However, we can make a further hypothesis. As shall be detailed in next chapter, we can attach to each site a semantic representation, for instance a vectorial one. It makes sense to build these semantic representations so that neighbors are semantically close. As a result, if two sites share similar sets of neighbors, their semantic representations will be most alike. Given these semantic representations, it becomes easy to define a semantic distance between two neighboring sites of the network. Now we can use these distances to compute the weight attached to each link, with the idea that the smaller the distance, the stronger the weight. If we consider then a ‘shortcut’ edge between two clusters, because it relates two sites with none or very few neighbors in common, it is to be expected that they will be semantically afar from each other. Thus such an edge between otherwise unrelated clusters would be associated with a small value of the weight.

For the semantic network, it means that the lower γ will be found between the most distant sites. Semantic expansion can occur between meanings belonging to the same semantic cluster; then the γ parameter can be rather high. The change will be fast, and rather unremarkable from a semantic point of view. However, if semantic change involves an expansion from a meaning belonging to one cluster to a meaning belonging to another, then it will proceed along the far reaching, weakly weighted edge, associated with a much lower γ . The latency phenomenon is then likely to occur. Moreover, the semantic distance between the two meanings will be high enough for the semantic change to be quite noticeable. Typically, this is the sort of large semantic leap that we expect between a lexical and a grammatical meaning.

Furthermore, once the linguistic form has set a firm foot in one site of a new cluster, then it will pass quickly to the other, closely related meanings of the cluster. Hence, the first latency will be followed by a short succession of further semantic expansions, much less noticeable from a semantic point of view, and very fast compared to the first cluster-to-cluster expansion. Then, the first, near-critical semantic expansion can trigger a cascade of further semantic expansions related to the gaining control over the whole semantic cluster. From a frequency of use perspective, the whole process will appear as a single movement, as we shall see in the next chapter.

To sum up, we can roughly distinguish three categories of relations between mean-

ings. The first one is a distant, thin relation between two loosely associated meanings: the corresponding γ weight will be too small for a change to occur along this particular channel. This is the case below the critical threshold. The second type of relation would be the intra-cluster semantic association. The two meanings are close from a semantic point of view, the weight of the edge which relates one to the other is high, and if a form occupies one, it will most likely quickly come to occupy the second (with the caveat that there might be directionality, so that it can work only one-way). This would be the case above criticality.

Then, there is the case at criticality. The two meanings, belonging to two different semantic clusters, are related, but the semantic distance between them is typically much larger than the distance between meanings of a same cluster. Yet, the link between them is high enough for a linguistic form to proceed by this narrow path. Then, the associated γ parameter will be at threshold; a latency phenomenon is expected; the gap between the source meaning and the target meaning is noticeable.

Thus, remarkable changes, and especially grammaticalization instances, are associated with a critical behavior which is the latency phenomenon not because they are tuned to criticality, but because otherwise, they would not be as remarkable.

Sapir's drift

An unexpected consequence of this near-critical behavior would be the Sapir's 'drift' phenomenon, the idea that a language carries, in its structure, its own possible changes. The effect would be that sister languages, once individuated and separated from each other, will keep showing up the same changes, in Sapir's own words: 'drift is often such that languages long disconnected will pass through the same or strikingly similar phases' (Sapir, 1949, p.172). It has been proposed, in the framework of a simple birth-death process, that an asymmetric mutation rate, constant throughout the language evolution, and in favor of a new, unattested variant, could eventually favor the fixation of this novelty, due to finite population effects (Yanovich, 2016). Furthermore this fixation would be fast and abrupt, and the established conventional variant would remain unchallenged for a long time, until the transition occurs, at a random, unpredictable time. Therefore, we would witness different fixation times for different, phylogenetically related languages, even after their speciation occurred.

The explanation of Sapir's drift in our model is not much different. One could relate it to the random latency time, but it would be erroneous; first, because latency times are usually shorter than the time during which the change is lurking in a situation of Sapir's drift, second, and more crucially, because during the latency period, the new variant is already attested, and marginally used, which is not the case in Sapir's drift. However, we have posited that the language network was constant across generations (though we did not provide any explanation for this faithful cultural transmission of the conceptual structure). Under this assumption, Sapir's drift is easily explained by the bifurcation phenomenon.

It can be that a parent language is associated with a conceptual network where some asymmetric link between two sites is too low for the invading variant to be ever produced (in other terms, for the semantic expansion to occur). Yet, small changes in the language can lead to a small increase of this link and make it pass the threshold,

which will trigger the change. If all children languages share the same conceptual network, then they all share as well a fair probability to develop the change, because of the existence of this link below the bifurcation threshold. In this regard, Sapir's drift would be a hint of the robustness of the conceptual network across time and generations. We would then not only learn language, but also a complex structure of semantic and pragmatic associations, historically and patiently elaborated; as Tolkien (1992, p.40) put it, 'language goes back by a continuous tradition into the past'. The cultural transmission of language would be therefore akin to that of myths, which similarly exhibit a remarkable persistence over time (da Silva and Tehrani, 2016).

8.4.3 Sociologic interpretation

Though we have favored throughout the presentation of this model a purely cognitive interpretation of the process, the random walk can also be given very different interpretations, in particular, sociolinguistic ones.

In a sociolinguistic interpretation, the two meanings C_0 and C_1 now stand for two speech communities. C_0 is a prestigious one, and exerts an influence γ on community C_1 . As only C_0 has prestige, this influence is unilateral. Individual occurrences correspond now to individuals belonging to either one or the other community. They can choose binary between two linguistic variants, X or Y .

At the beginning of the process, community C_1 homogeneously speaks variant Y and community C_0 uses uniformly variant X . At each time step, we select randomly, with uniform probability, a given speaker of community C_1 . The linguistic variant of the speaker is updated without considering the variant that the speaker was using before. To make this choice, the speaker considers what the other members of the community are using, and also feels the influence of the prestigious community. Therefore, the choice favored by the latter can lead the speaker to use X even if very few members of the C_1 community make use of this variant. This would be the equivalent of the 'speaker' perspective. In the hearer perspective, there would be a probability γ that a speaker of the C_0 community is heard by a speaker of the C_1 community, which will then mimick the speaker of the prestigious community.

As before, the γ parameter can be interpreted with respect to the community features (size and frequency). Differences of memory size become now differences in community size. Let us consider that the prestigious community C_0 is also smaller. Then, the speaker of the C_1 community, when it comes to make a probabilistic choice in favor of either one or the other linguistic variant, will base this choice according to the behavior of all speakers, from both communities. In the meantime, members of community C_0 only try to match the linguistic practice of other members of C_0 and do not care about the way of speaking of the C_1 members. Hence, C_1 members base their choice on the overall frequency of the form, the γ parameter arising out of a difference between the two community sizes, while C_0 members base their choice on the frequency of the form in C_0 only.

We can also consider the associated frequencies of the community, which are now probability that a member of a community hears a member of another community. Members of the C_0 community only stay with other members of the C_0 community and never hear the language of the C_1 community. Yet, they have what we call in

French ‘voix au chapitre’: they can speak in the media, they are given a voice on many public occasions. Hence, members of the C_1 community have a chance $1 - \alpha$ to hear members of the C_0 community through the media, but a greater chance α to hear other members of their low-prestige community.

Of course, the two perspectives can be combined as easily as it was done in the cognitive point of view. Matching the influence γ parameter of both perspectives would lead to equate the probability, for members of the C_1 community, to hear members of the C_0 , with the relative size of the latter. This is not completely relevant, and a mixture of those various interpretations would probably be preferable. Also, compared to the complexity of the sociolinguistic factors, this picture might seem a bit crude.

Anyway, the three variants of our model are compatible with a sociolinguistic framework. It would nevertheless leave open the question of how members of the C_0 community have changed their language in the first place, if they had no model speakers to mimick. One would need to add a striving, for members of the C_0 community, to avoid the variants used by the C_1 community. However, it would imply that language is never stable — C_0 members would be constantly running away from the imitation of their low-prestige C_1 peers and there would be no end to this headlong rush. An explanation of the overall stability of language could nonetheless lie in the difficulty, for members of the C_0 community, to work out new linguistic variants as to distinguish itself from the C_1 community. Since innovations would be few and far between, there would be pretty long period of times in which the same variant is shared by both communities.

We personally favor the cognitive view of semantic change as it stresses that language change is not arbitrary, and that semantic expansions are themselves meaningful. It takes the existence of semantic chains and semantic inferences as an input for the change to happen. However, this brief overview of alternative, sociolinguistic interpretations, show that this model is relevant and efficient even if one does not ascribe to the particular, semantically focused view of change that we argued for in chapter 7. Our results are therefore entirely independent of the specific linguistic framework we proposed to understand and explain semantic change.

8.4.4 Limitations

We shall now end this discussion by pointing out the main limitations of our model.

First of all, there are obvious limitations, related to the fact that we included only a limited number of ingredients. The social structure has been ignored, and the hypothesis of a perfect communication is certainly far too strong. Yet, the ingredients that we have included (cognitive influences between meanings, probabilistic choice of linguistic variants, finite memory capacity) have proven to be sufficient to account for most of the phenomenology of semantic change. If we do not have further quantitative and qualitative observations to compare the model with, it is of little relevance to add some more features for the sake of realism.

What is of greater concern is the discrepancy between our theoretical modeling and the actual empirical features of the growth part of the change. As it stands, our model clearly lacks key components so as to account for the phenomenology of frequency rise. More particularly, we should account for four stylized facts which have emerged

out of our statistical survey of semantic change. One, the growth is typically longer in real instances of semantic change than it was in our model, once the timescales are properly compared. There is a factor 10 between the mean growth time found in our surrogate data and the mean growth time in the empirical distribution of actual semantic changes. Two, the growth time presents a comparatively wider variation in the empirical data than it does in our model. This, in particular, is reflected in the value of the Péclet number, which governs the shape of the probability distribution. Three, in our model, there is no saturating scaling law which can compare with the one exhibited by the empirical data. Last, in our model, the latency frequency is typically between one fourth and one fifth (20% - 25 %) of the final frequency. In the empirical data, the latency frequency is usually much smaller compared to the final frequency reached at the end of the rise. The next chapter will be especially devoted to tackle with these limitations, and shall offer some promising exploratory research in this direction.

The third kind of limitation that the model is facing is the most serious one, for it seems unclear if they can be overcome and how. The first one is the fact that the S-curves that we generate in our model are nice enough to pass the criterion, but they are not as beautiful as the best exemplars we got from actual linguistic data. This is concerning, since in our model, the competition occurs without any interference, and therefore should lead to the cleanest possible curves. Yet, corpus data leads to cleaner curves still. There might be a mechanism which forces these curves to be more sigmoidal, which our model would miss. The second limitation is that a lot of analytical details are arbitrary, while they should not be. This is especially the case for the production probability function $P(f)$ and the erasure mechanism (uniform draw of a past occurrence in a crude ‘bag-of-words’ memory model). Those mechanisms, at least in the cognitive interpretation that we put forth, have clear psychological reality, and they should be refined on this basis. Without adding any new ingredients, our model could thus benefit from a greater psychological realism by letting the neuro-cognitive realities of the mind guide the choice of the mathematical quantities involved in the model. For instance, we could make use of a neuron-based decision-making process (Kirillov et al., 1993; Wang, 2002).

Last, we acknowledge that the explanation of language change somehow evades this model. Our model shows that, if the linguistic conditions are met, a change can happen according to the characteristic phenomenology of semantic change that we highlighted in chapter 4. However, it fails to shed light on how those conditions could happen to be met. In some sense, it only buries the trigger of change into a specific network structure, of which it is not said how it came to be in the first place. It is therefore unsatisfying to relocate all trigger of changes into a set of weights and edges which seems even less accessible to the empirical investigation than language itself.

Chapter 9

Semantic bleaching and other network effects

The aim of this last, brief chapter is to propose some research perspectives, rather than delivering actual results. The model that we presented in the previous chapter can be expanded in two ways. We can consider what could happen if there were several competing linguistic forms, or we can add further structure to the semantic territory, and investigate what would happen once the new variant has finally set foot into one site of a larger cluster of related meanings.

We shall not investigate the first branch, for several reasons. The chief one is that, as we remarked in the previous chapter, the model cannot straightforwardly account for a competition between more than two forms. The symmetry condition of the production function ($P(f) + P(1 - f) = 1$) imposes a linear function if it is to be generalized to an arbitrary number of forms. There could be two ways to deal with this technical issue. The first one would be to authorize uttering failure; i.e., when all competing forms would have a frequency below one half, it would become likely, for the speaker, not to be able to retrieve a suitable linguistic form to express the intended meaning.

The second would be to describe the process of retrieval as a succession of duels. The forms would be presented in a queue. The first two of the queue would then compete, where their frequencies would be computed with respect to the memory occupied by their union, leaving aside all other forms. By doing so, the usual $P(f)$ function could be used. Then, the winner would compete against the next form in the queue, and so on until there is no longer any member to compete. Of course, if there is one form in majority, and all other forms are marginal, it would be advantageous, for these forms, to be the first to compete against the strong one. Therefore, the order of the forms in the queue would have a crucial importance. And how to decide this order? Should the forms be sorted randomly? In decreasing order of their frequencies? If so, what would matter most, the effective, or the actual frequency? All these questions could be answered and it is perfectly conceivable to account for a situation where more than two forms are competing, but anyhow, it would require to add some rules.

There are other reasons to neglect this part. As we saw in chapter 5, one-to-one competitions are the rule more than the exception, as long as we are aware that

‘one-to-one’ does not mean ‘one linguistic form against another’. Clusters of paradigm members, more complex constructions, can also compete one against another. Yet, these differences, crucial for the linguistic understanding of the process, matter absolutely not when it comes to the modeling. Most competitions with several members can be interpreted as a succession of one-to-one competitions. For instance, in Old and Middle French, there were a lot of words to express negation, such as *pas*, *point*, *mie*, or more marginal forms such as *goutte*, *bouton*, *grain*. It is unlikely that all forms would have started to compete with each other. More plausibly, the paradigm emerged as a whole (against an unknown and possibly impossible to identify competitor), different members playing the role of different contexts of diffusion for the new construction. Only then, once it was entrenched, it got reorganized. Some members of the paradigm have certainly entered a competition at some point (e.g. *pas* and *point* in the eighteenth century), but it was a later process, within the frame of an established paradigm, between two entrenched and fully grammaticalized forms.

The linguistic cases where different forms are competing at the same time, and independently, against a third one, are arguably extremely rare. One example which falls with all likelihood in this category would be the replacement, in Middle French, of *moult* by *très* and *beaucoup*, as studied in great detail by Marchello-Nizia (2006). It may well be that the two new variants, *très* and *beaucoup*, competed independently against *moult*, which — much like Harold Godwinson facing the independent and simultaneous invasions of Harold Hardrada and the Duke Willelm II — succumbed to the conjugated weakenings of its functional dominion, while there was no actual collusion between the two adversaries. It is not unlikely that the triadic *moult/très/beaucoup* situation could be described, without *très* and *beaucoup* ever occupying a same semantic site.

Anyway, the situation with more than two competitors seems marginal in language change. Therefore, there is no need to focus on it, when there are already possible model expansions which would echo the aspects of change which already received empirical evidence. This is why we shall focus on model expansions related to the semantic territory and its network structure. Note that it would also be interesting to extend the model towards considerations of vertical, inter-levels influences, for language is multi-layered, and what happens on same plane has consequences on others. Yet, the ‘horizontal’ semantic diffusion (i.e. within one language level) seems an easier task to begin with, and would anyway be needed to understand the ‘vertical’ one.

9.1 Pervading the network

Confronted with empirical data, our model proved inadequate in four major ways, all related to the growth part:

- The frequency gap between the latency frequency and the final frequency was too small. In actual instances of change, this gap can be huge, and the frequency can increase a hundred fold between the two, while it was limited to less than tenfold in our model.
- The obtained S-curves were not smooth enough, and at least less so than the best S-curves found from actual corpus data. In contrast, we should expect theoretical surrogate data to be cleaner than its empirical counterpart.

- The statistics of the growth times did not correspond to those of linguistic data. Indeed, the variance was much too small compared to the mean value, leading to a high Péclet number, approximately one order of magnitude bigger than what was expected.
- The mean surrogate growth time was too small compared to the mean surrogate latency time (μ_{growth}/μ_{lat} was equal to 0.12 instead of the 1.14 expected from corpus data).

We believe that these four limitations are related to the fact that we considered a minimal competition, where the new variant invaded only one site. Yet, we stressed in several occasions that the semantic network is likely to be small-world, with a few long-range links relating densely connected clusters. It should then be expected that semantic expansion is a phenomenon associated with the crossing of such a long-range link (hence, the usually large semantic distance between the source meaning and the target meaning). Now, once the link has been crossed, the new variant can spread all over the cluster with much more ease than it took to set foot upon it. Since all links within the cluster are tight, no further latency is expected and the spread should happen as a whole. As a result, the growth would be longer, and the frequency rise, wider.

9.1.1 Basic setup

To investigate this question, we need to provide a network structure. To begin with, we shall consider a simple, random network of K sites. The first one is not related to any but the second site, with weight $\gamma_{2,1} = \gamma_c$, which it influences unilaterally. Then a given site k is related to all sites j such that $j > k$ with weight γ_{jk} . Therefore the structure is heavily asymmetric and the change can only move forward. Actually, making room for backward links, for instance from site 7 to site 2, the latter being the site of first invasion, would dramatically increase the threshold value required for the new variant to start spreading within the cluster.

We also set $\gamma_{jj} = 1$. We shall at first assume the simplest possible choice of weights: $\gamma_{jk}, \forall k > 1, j > k$, is drawn from a uniform probability on the $[0, 1]$ segment. We shall change this condition later on. If K is large enough, then the probability for the new variant to take over the whole cluster is almost one: as long as the sum of the weights γ leading to a given site is greater than γ_c , and provided all previous sites have been successfully invaded, then the new variant shall conquer this site within a finite time.

For simplicity, we restricted the investigation to the speaker mechanism. The dynamics is the same as before, except that, in the first step, a site j is chosen with uniform probability. Then, the effective frequency of the new variant is computed with the formula:

$$f_j = \frac{\sum_{k=1}^K \gamma_{jk} x_k}{\sum_{k=1}^K \gamma_{jk}}, \quad (9.1)$$

where x_k is the actual frequency of the new variant in site k .

The new variant is then uttered with probability $P(f_j)$, recorded in memory of site j , and finally a former occurrence is randomly chosen within this memory to be erased. The erased occurrence is a token of variant X with probability x_j . The only differences with the simple competition process are the addition of a step to select a site, and the adapted definition of the effective frequency. Note that the β parameter appearing in P could be different from one site to another, hence forcing a notation $P_j(f_j) = P(f_j, \beta_j)$, but for the sake of simplicity, we will stick with the value $\beta = \beta^*$, defined so as to maximize the frequency of the variant during the latency.

Latency time in the network

The latency phenomenon should not be affected by this generalization. Indeed, the latency phenomenon is associated with the neighborhood of a critical value of the γ parameter characterizing the competition. For site j , this parameter, that we shall note $\tilde{\gamma}_j$, is an effective one. It appears from the definition of the effective frequency f_j ; assuming all sites $k < j$ have been dominated, we would get:

$$f_j = \frac{x_j + \tilde{\gamma}_j}{1 + \tilde{\gamma}_j}, \quad (9.2)$$

and so the effective $\tilde{\gamma}_j$ is given by:

$$\tilde{\gamma}_j = \sum_{k=1}^K \gamma_{jk}. \quad (9.3)$$

Note that for $j > 2$ the sum could equivalently run from 2 to $j - 1$, since all other terms are 0. Now this effective $\tilde{\gamma}_j$ can be either below the threshold γ_c , so the variant X can only be marginally present in this site, or above, in which case the new variant will promptly take over the site according to an S-curve. Only if $\tilde{\gamma}_j$ were to fall precisely within a small neighborhood around γ_c would we observe a latency phenomenon. Indeed, below the threshold, the latency increases dramatically (Fig. 9.1a); above, the latency time drops to approximately 5 units of M (Fig. 9.1b), which is about four times less than the growth time. The falldown to zero, on Fig. 9.1b, is due to the disappearance of the two wells motif of the third derivative of the uncertainty whose minima was used to compute the average latency time (actually the whole motif is shifted to the left as γ increases, so that one of the well is reached for a value of N inferior 0, which is not physical).

Therefore, to get a significant value of the latency, that is, which is neither exceedingly long nor shorter than the typical growth time for a single site, the parameter $\tilde{\gamma}_j$ must belong to a very restricted interval, of size 10^{-2} approximately. This has a very limited chance to happen, especially given that, if the network favors the spreading, $\tilde{\gamma}_j$ should on average be of order:

$$\langle \tilde{\gamma}_j \rangle \leq \sum_{k=2}^{j-1} \frac{1}{2}, \quad (9.4)$$

where $\langle . \rangle$ denotes an average over many possible realizations of the random network.

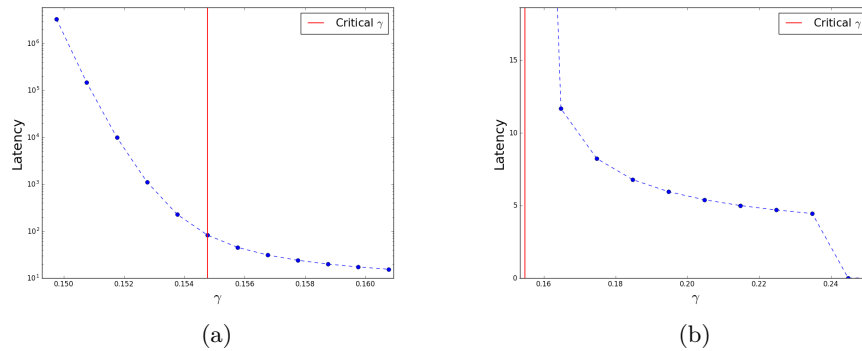


Figure 9.1: Average latency time (in units of M) as a function of the γ parameter. The critical value γ_c is indicated with a red line. (a) In the neighborhood of the critical threshold, in semilog scale. (b) Above the critical threshold, in normal scale.

We can thus safely consider that latency happens only during the invasion of the first site. Things will be different however for other kinds of networks, and different probability distribution of the weights.

9.1.2 A greater leap in frequency

We can now easily fix the first issue of the model. Indeed, if the new variant conquers now K sites instead of only one, and shows a latency for the first site only, then we can expect the difference between the latency frequency and the final frequency to be arbitrarily big, depending on the number K of sites in the cluster. If we assume that K_W of them are conquered by the end of the process, then the increase of frequency through the S-rise will be approximately given by K_W/x_c .

Empirical application

An interesting matter would be to relate these results with the empirical data. Given an instance of linguistic change, we can compute the gap between the frequency at the end of the S-curve and the frequency at the beginning, which should give us an estimate of K_W . Thus we can get an idea of how broad the semantic expansion is, and how many semantic units it actually encompasses.

There are, however, several problems. The first one would be the actual definition of the gap. Noting x_f , the frequency achieved at the end of the S-growth, and x_i that at the beginning, a simple definition for the gap g would be $g = x_f/x_i$. Then, the number of semantic units K_W covered by the expansion would be related to the gap through $g = K_W/x_c$, where x_c is the latency frequency of the model.

This, however, is suitable under two conditions, which are not necessarily fulfilled in language change. Firstly, the frequency must be ‘pure’, i.e., it must not encompass former uses of the form. Yet, obtaining this kind of ‘pure’ frequencies would be, in most cases, only possible through a manual sorting of the relevant occurrences. A second condition, provided the first one has been fulfilled, is that the frequency x_i can

be regarded as a ‘latency’ frequency, i.e., a non-zero initial frequency at which the system lingered due to the latency phenomenon. If there is no latency phenomenon (and this is a true possibility), then $x_i = 0$ and the gap cannot be defined this way.

This simple estimate of the cluster size would be independent of the network structure. Indeed, as long as the new variant takes over K_W contexts, then $g = K_W/x_c$ should always hold. In every case, we would get an empirical characterization of the scope of the expansion, in terms of the number of semantic units involved.

The problem is, this would force us to rely on the specific value x_c . Without any way to confirm the relevance of this value, it should be regarded as an arbitrary choice. Indeed, the value of x_c depends on the β parameter, and this β could be different from the value we used in this numerical and theoretical analysis. What is more, β could vary from one change to another — and so would x_c .

An alternative would be to define the gap as $x_f - x_i$, which holds even if the initial frequency is zero. This seems reasonable in the framework of our model: each site being called with equal probability, when the variant occupies one more site, its frequency should increase of a frequency quantum Δ_x . Therefore, the absolute difference is, in a sense, more relevant than the relative one. We would have then $x_f = K_W\Delta_x$, and thus, a direct relation between K and x_f . Furthermore, we would have $x_i = x_c\Delta_x$, and, as we know x_c from the model, we can infer Δ_x , and from it, K_W . Note that the value of K_W obtained through the two approaches would be the same. In cases where $x_i = 0$, since Δ_x is assumed to be a constant of language organization, we could use an estimate of Δ_x from another semantic replacement with latency to get K_W in a case without. However, the assumption that each semantic unit is uttered with equal probability is, at the very least, debatable.

Finally, we assumed so far that only one site was subject to latency. However, in real instances of change, it is almost certain that several sites can be affected by the new variant during the latency time. This was the case for *du coup* where the change immediately showed up in several contexts of use, despite waiting a few decades before taking off after that. This situation would be expected for other types of network. In this case, denoting K_L , the number of contexts affected by latency, we would have:

$$g = \frac{K_W}{x_c K_L}. \quad (9.5)$$

In such a case there would be no way to get the scale. Furthermore, it would not be possible to get K_W and K_L , only their ratio, which alas presents very little interest. An eventual solution would be to find the smallest initial frequency x_i^* among a big set of changes with latency, to posit that this smallest value corresponds to $K_L = 1$, and from that, to infer the scale $\Delta_x = x_i^*/x_c$ between empirical and numerical frequency.

Anyway, the empirical gap between the final and the initial frequency in a semantic change and the number K of sites covered by the expansion are expected to be correlated. How to exploit properly this correlation in the study of a specific change remains however a matter of speculation.

9.1.3 A smoother S-curve

Increasing the size of the frequency leap is, as we saw, a trivial matter. What is much less so is how to ensure that the overall replacement process remains sigmoidal. In

chapter 4, it could be seen that some processes of language replacement were very neatly described by an S-curve. It may be that this would not be the case for all instances of change, but at least our model must account for this possibility.

Let us address a preliminary question: is the S-curve associated with the shape of the competition within one single site (as we saw was the case), or does it arise from the spreading over a large network? In the latter case, the particular shape of the competition within a site should not matter much. It should be reminded here that a semantic expansion needs not be large (we saw examples involving a very limited number of occurrences), and yet the growth is still S-shaped. Therefore, the S-curve must already be captured in a one-site competition — a feature nonetheless almost granted by the few requirements that we assumed the production probability function had to meet.

This one-site S-curve allows one to trivially capture S-curves of higher leap. If we assume a cluster structure such that the change is almost immediately transmitted to all other sites, for instance a star network, with the first invaded site at its core, and asymmetric links pointing towards all other sites (Fig. 9.2). In Network 1, all weights are drawn with a uniform probability and we indeed find a beautiful S-curve, spanning over 25 sites of the cluster out of the 29 (Fig. 9.3). Four leaves have not been invaded by the new variant, as the weight of the link leading to them was below the critical threshold. In Network 2, we also added links between the peripheral nodes such that each leaf exerts an influence to the next one and the previous one (which makes the star a wheel). These weights are drawn with a uniform probability. This little trick helps us to harmonize the process over the network, and to increase the probability for a leaf to be affected by the new variant, even if its relation to the core is too weak. The S-curve is then a bit altered (Fig. 9.4) since there is now an asymmetry between the beginning of the process (only the core spreads the new variant) and the end, where the crown of leaves becomes a channel of diffusion as well.

Yet, such a network suffers from several problems. First, it is unrealistic and designed such as to give the expected result. Second, we shall expect the change to come from a marginal member of the cluster, rather than from a prototypical member. Last, the overall growth time will be sensibly the same as the growth time in the first site. Therefore, all changes would present the same growth time, since we already know that the growth time in the first site does not vary much. This eventuality has, nonetheless, already been ruled out by our empirical observations.

Consequently, we should answer our preliminary question by stating that the S-curve must be associated with both a specific shape in the micro-event of a one-site competition, and also with the overall spreading into the network. Only a specific structure, with appropriate features, can lead to an S-curve for the diffusion on the whole network. The question of the coherence of the S-curve over the entirety of the process of replacement is therefore not a trivial one. In the following, we shall investigate this matter on several network structures.

The shape of the network

We considered the following network structures:

- Random network. Links are placed randomly and successively. The first link

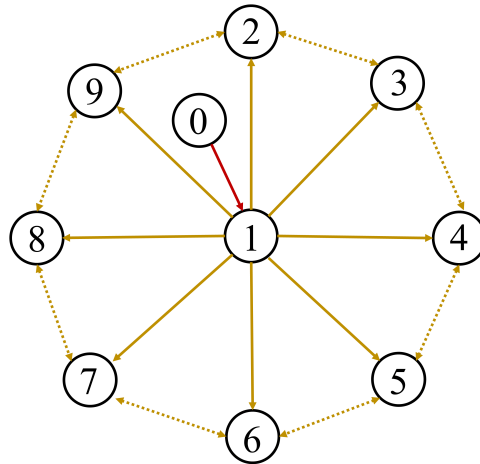


Figure 9.2: Star network with $K = 10$ sites. The reservoir site (site 0) feeds the first invaded site (site 1) with occurrences of the new variant X . Then, site 1 feeds unilaterally all other sites (the leaves) with occurrences of this variant, allowing it to spread over the whole network. We can also add the possibility for neighboring leaves to be related (dashed arrows), making it a wheel network. The links between the leaves have two directions, but a different, random weight is associated with each direction.

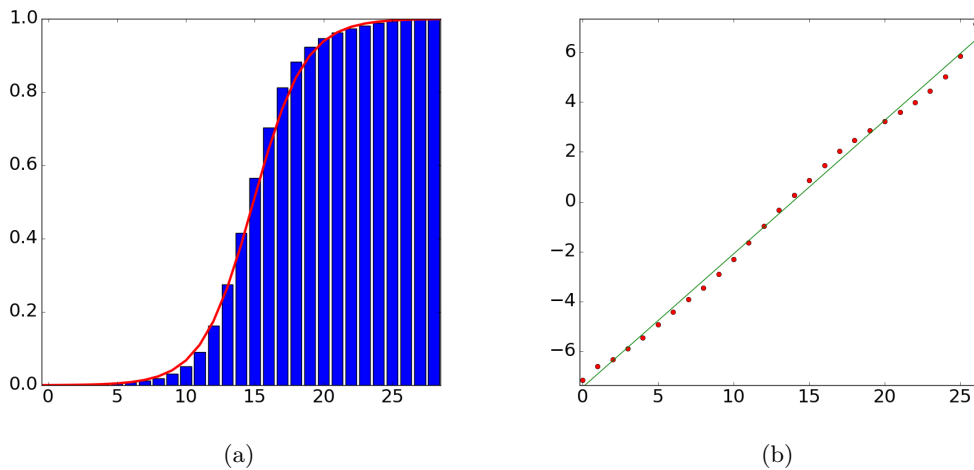


Figure 9.3: Network 1: Spreading of the new variant on a star network radiating from the first invaded site, with unilateral links, randomly drawn out of a uniform probability, pointing to all the leaves. (a) Rescaled frequency profile of the overall replacement (blue histogram) its numerical fit by a sigmoid (red line). $K_W = 25$. (b) Logit transform of the overall frequency (red points), and its linear fit (green line). $K = 30$, $M = 5000$, $\delta = 0.0$, $\beta = \beta^*$.

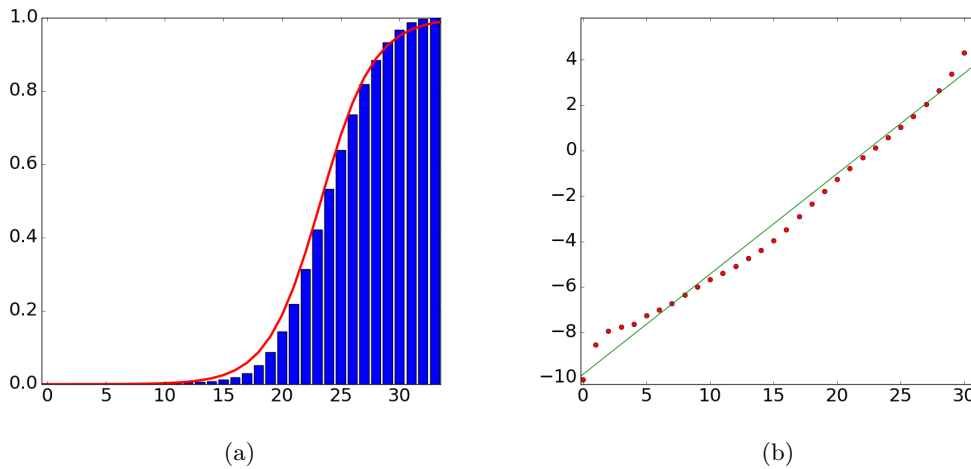


Figure 9.4: Network 2: Spreading of the new variant on a star network radiating from the first invaded site, with unilateral links, randomly drawn out of a uniform probability, pointing to all the leaves. The star structure is now altered by the addition of links between all neighboring leaves. (a) Rescaled frequency profile of the overall replacement (blue histogram) and its numerical fit by a sigmoid (red line). $K_W = 29$ (b) Logit transform of the overall frequency (red points), and its linear fit (green line). $K = 30$, $M = 5000$, $\delta = 0.0$, $\beta = \beta^*$.

starts from the first invaded site and joins a uniformly chosen site. All subsequent links start from the last chosen site and join any other site, save the reservoir. The building process stops when no isolated site remains (Fig. 9.5). If an oriented pair of nodes is chosen twice, the new value of the weight replaces the previous one.

At the end of the building process, there exists a directed path from the first invaded site to any other site. This process also accounts for the possibility that two sites mutually influence each other, though with different weights. We added a special option in the building rules: if the first invaded site is chosen as the attachment point of a link, the weight of this link is set to weight 0 and the path continues from the first invaded site. In that way, the first invaded site has a probability to influence several sites in the network, without being influenced itself.

- **Radiant network.** The network is built as previously, except that the first invaded site cannot be chosen to attach a link. Contrary to the previous one, there are now \sqrt{K} different paths of length $n_c\sqrt{K}$ which are drawn, all starting from the first invaded site (Fig. 9.6). Therefore, the number of added links is fixed, the average connectivity is given by n_c (set to 2 in the simulations), and there is a chance for a site to not be a part of the cluster (i.e. to stand apart from all other sites with neither inward nor outward links reaching it).
- **Grid network.** In this network, all sites except the reservoir site are disposed on a grid of length L . The size K of the network is thus given by $K = L^2 + 1$.

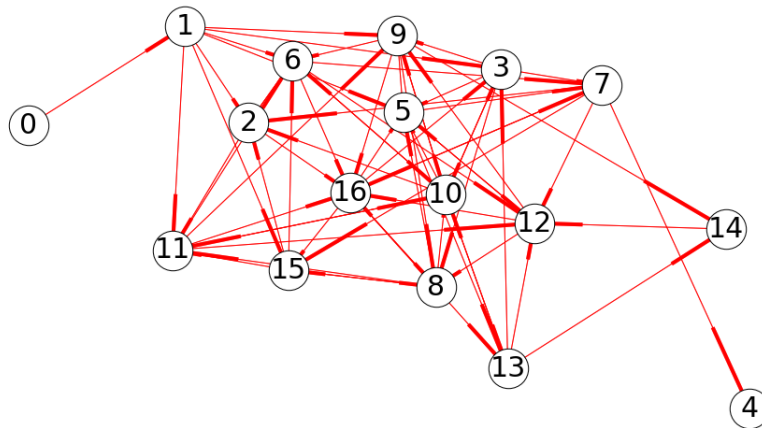


Figure 9.5: Random network sample with $K = 17$ sites. The reservoir site (site 0) feeds the first invaded site (site 1) with occurrences of the new variant X . Links are added randomly from site 1 until all sites have been reached. Incoming links to site 1 are removed. The network obtained presents a high connectivity on average, and many cycles. The thicker extremity of an edge indicates the direction of the link. Site 4 is the last site reached by the random path, and therefore a dead end.

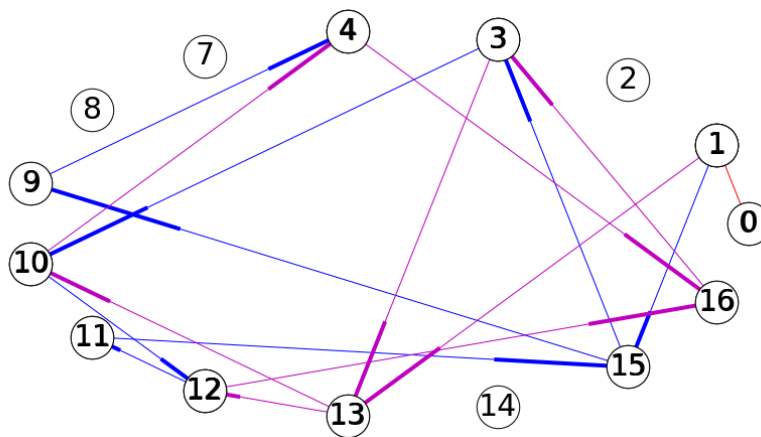


Figure 9.6: Radiant network sample with $K = 17$ sites. The reservoir site (site 0) feeds the first invaded site (site 1) with occurrences of the new variant X (red link). Then, $\text{int}(\sqrt{17}) = 4$ paths of $2 * 4 = 8$ steps are drawn between the $K - 2$ remaining sites, radiating from the first invaded site. Here, only two paths out of four have been drawn, in blue and magenta directed links respectively (blue path: 1-15-3-10-12-11-15-9-4; magenta path: 1-13-10-4-16-3-13-12-16). The thicker extremity of an edge indicates the direction of the link. Navigating the network involves typically a large amount of back and forth, as the two paths illustrate.

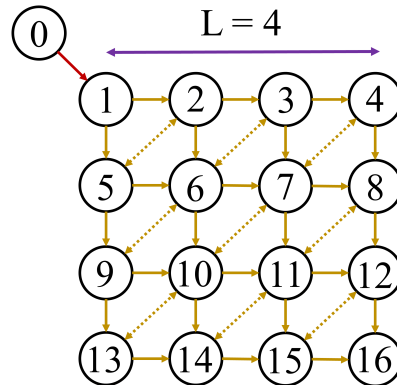


Figure 9.7: Grid network with $K = 17$ sites. The reservoir site (site 0) feeds the first invaded site (site 1) with occurrences of the new variant X (red link). Then, all sites are disposed on a $L \times L$ grid (here $L = 4$) so that a site can only influence its down and right neighbors (full arrows). One can also add diagonal, non-directional links (dashed arrows), so as to homogenize the contents of the cluster sites. The flow of occurrences goes from the upper-left corner (site 1) to the lower-down corner (site 16).

The links are such that, if the first invaded site lies at the upper-left corner of the grid, each site influences its right and down neighbor, except on the edges (Fig. 9.7). The dimensionality of the grid here is 2 but the network could be generalized for higher dimensions.

We added the possibility to have long-range links so that the new variant can take a shortcut — but it can also hinder its progression if the link is opposite to its flow. When we activated this option, we added D new directed links between randomly chosen pairs of nodes, and tried several values of D (e.g. $D = K$, $D = L$). New links can radiate from the first invaded site, but they cannot be attached to it.

We also considered the case in which some links of the grid are complemented by opposite links (flowing from right to left or from down to up), with a random weight drawn as all the other weights of the network. In the simulations, this break of unilaterality could happen with probability 0.25, which is the inverse of the connectivity of the sites inside the grid.

- Scale-free network (Fig. 9.8). To build this network, we choose a parameter α so that:

$$n_k = \text{int} \left[\left(\frac{K}{k} \right)^\alpha \right], \quad (9.6)$$

where n_k is the number of nodes that site k is influencing. These nodes are uniformly chosen with the only rule that the reservoir and the first invaded site cannot be chosen. In the simulations we shall set $\alpha = 1$, so that n_k is comprised between $K/2$ and 1. What is interesting with this network is that we can investigate the impact of the rank k of the entry point of the new variant.

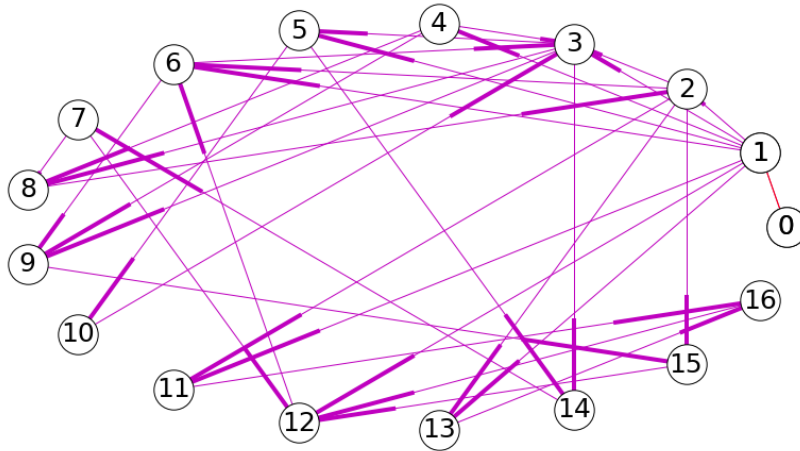


Figure 9.8: Scale-free with $K = 17$ sites. The reservoir site (site 0) feeds the first invaded site (site 1) with occurrences of the new variant X (red link). Then, site 1 (here of rank 2) influences $\text{int}(K/2) = 8$ different sites, randomly chosen, site 2 (rank 3) influences 5 sites, etc., and sites 8 to 16 influence only 1 site each. The new variant X can easily spread from site 1 to the sites which site 1 influences exclusively (e.g. site 4), but may fail to propagate to sites influenced by too many other (e.g. site 3).

Probability distribution of the links

All links of these networks are associated with a random weight γ_{ij} . We retained two ways to draw these links: by drawing them with a uniform distribution on $[0,1]$, or by drawing them with an exponential distribution of parameter α :

$$P(\gamma) = \frac{\alpha}{1 - e^{-\alpha}} e^{-\alpha\gamma}. \quad (9.7)$$

In the simulations we choose $\alpha = 2$ so that the mean value of the link, which is roughly equal to $1/\alpha$, would be close to the uniform case. The exponential distribution increases the probability for a weight to be below threshold. In that sense, it ‘turns off’ some of the links of the network, which is therefore more hollow than it would be for a uniform distribution.

Propagation failure

It appears that propagating the new variant over the network is extremely difficult, unless we posit a strong directionality in the structure, as is the case in the grid network (Fig. 9.9). If there is no such directionality, then a typical site is more likely to be fed by other nodes than the first invaded one, so that the effective frequency of the entrenched, established variant, is almost always greater than the one of the novelty. In short, if the network is too dense, then the links coming from converted sites weight too little in the expression of the effective frequency. If the network is not dense enough, then the threshold cut-off can easily sever some links and compromise the spreading over some parts of the cluster. Furthermore, as soon as we add disorder in one of the strongly directed network (e.g. long-range random links or backward links in the grid network), then the range of the propagation drops immediately.

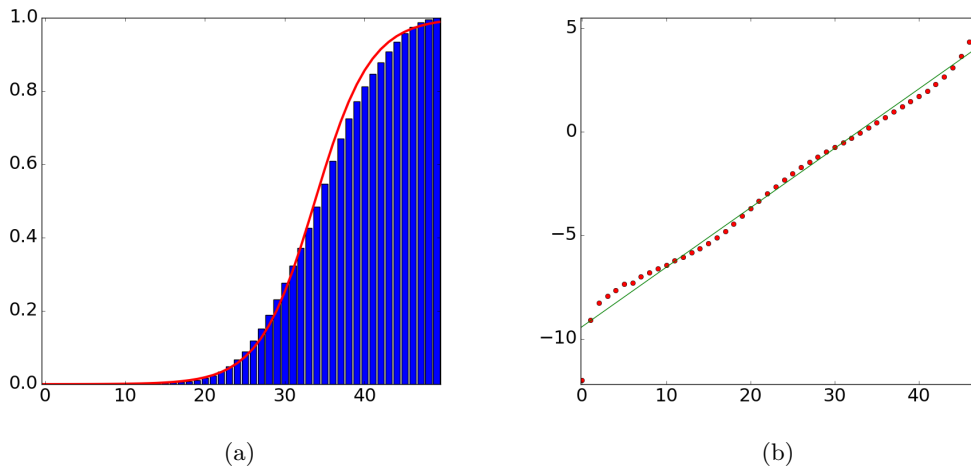


Figure 9.9: Network 3: Spreading of the new variant on a regular grid network with directed links from the upper-left corner to the bottom-right. (a) Rescaled frequency profile of the overall replacement (blue histogram) and its numerical fit by a sigmoid (red line). $K_W = 34$. (b) Logit transform of the overall frequency (red points), and its linear fit (green line). $K = 37$ ($L = 6$), $M = 5000$, $\delta = 0.0$, $\beta = \beta^*$.

The scale-free network presents the same problem. We managed to observe a propagation only when we considered that the entry point was at the top of the hierarchy (which, in practice, has little chance to happen). In this case, the new variant spreads downwards the less influencing site and shows a linear progression (Fig. 9.10). The new variant dominates, at the end of this specific run, 27 out of 29 sites. However, if the entry point is second in the hierarchy, then the propagation does not happen and the new variant barely ventures beyond the entry point.

In the other networks (random and radiant), the new variant seldom propagates beyond the entry point, and when it does it occupies but a few sites, never more than a tenth of the total cluster in our simulations.

Mechanism

We can wonder whether the speaker mechanism is the least favorable to a propagation. Indeed, the hearer mechanism is associated with a higher value of x_c , which means that the new variant can occupy a bigger fraction of memory even if it does not dominate the site. Thus it can spread more easily in the network. Even if it remains in minority everywhere, it can gradually build support, so as to take over on new sites eventually. The combined mechanism has also a higher x_c (less than the hearer mechanism though), but above all, it is associated with a much lower value for the threshold. Yet, in any case, the mechanism is variant-neutral, and though even if one of these mechanisms is stronger, it can favor the old variant as well.

Besides its relevance to the propagation process, the comparison of the three different mechanisms is interesting in the sense that, so far, we failed to provide any empirical criterion to distinguish them from another: any outcome of either of these

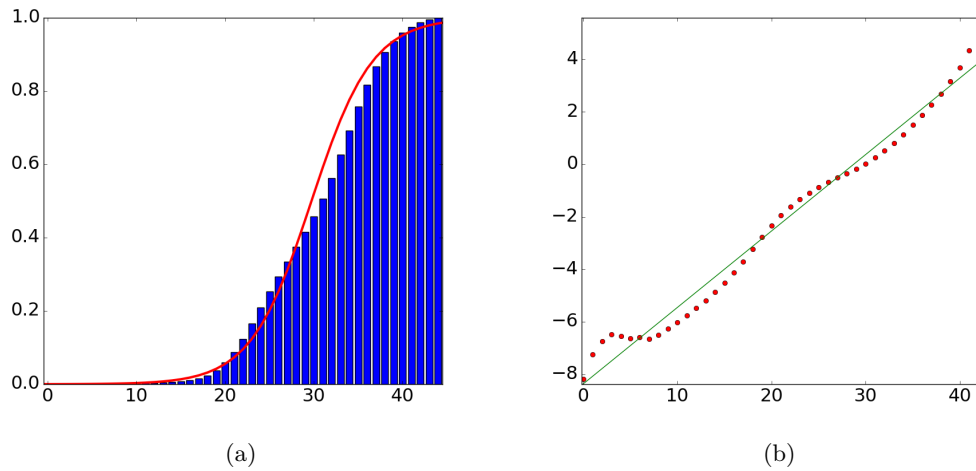


Figure 9.10: Network 4: Spreading of the new variant on a scale-free network, starting from the most influencing site. (a) Rescaled frequency profile of the overall replacement (blue histogram) its numerical fit by a sigmoid (red line). $K_W = 27$. (b) Logit transform of the overall frequency (red points), and its linear fit (green line). $K = 30$, $M = 5000$, $\delta = 0.0$, $\beta = \beta^*$.

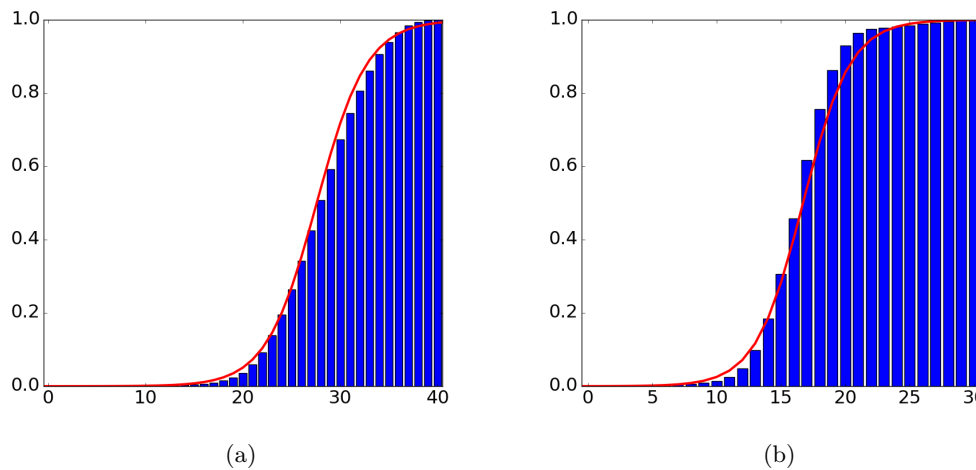


Figure 9.11: Network 3: Spreading of the new variant on a regular grid network with directed links from the upper-left corner to the bottom-right. The process relies now on the alternative mechanisms, the hearer one and the combined one, instead of the speaker one. The network on which the two runs have been performed is exactly the same as that of Figure 9.9. (a) Hearer mechanism. Rescaled frequency profile of the overall replacement. $K_W = 34$. (b) Combined mechanism. Rescaled frequency profile of the overall replacement (blue histogram) its numerical fit by a sigmoid (red line). A much better fit would be obtained if a higher criterion were set in the detection of the growth pattern (e.g. $r^2 > 0.995$ instead of $r^2 > 0.98$). $K_W = 36$. For both: $K = 37$, $M = 5000$, $\delta = 0.0$, $\beta = \beta^*$.

Sample	1	2	3	4	5	6	7	8
Speaker mechanism	34.99	4.21	33.93	30.28	27.21	19.22	23.96	10.04
Hearer mechanism	34.98	4.23	33.93	30.40	27.22	19.48	23.88	9.82
Combined mechanism	36.00	36.0	34.13	32.40	26.86	24.70	26.33	36.0

Table 9.1: Dominion of the new variant over the whole network at the end of the process. $K = 37$, $M = 5000$, $\delta = 0.0$, $\beta = \beta^*$

Sample	1	2	3	4	5	6	7	8
Speaker mechanism	3.33	2.06	2.72	4.08	X	2.67	1.39	X
Hearer mechanism	2.05	2.54	2.13	3.07	X	2.54	1.60	X
Combined mechanism	2.06	2.50	1.76	2.77	X	2.38	2.64	X

Table 9.2: Ratio of the overall growth time over the growth time within the first invade site. In some cases, the overall growth was not sigmoidal; hence, the ratio could not be computed (X). $K = 37$, $M = 5000$, $\delta = 0.0$, $\beta = \beta^*$

three was as likely to have been produced by one of the two others. Yet this was the case for the one-site competition, and we can expect to dig out more crucial differences regarding the propagation in the whole network.

We did not investigate this fully and will only provide a handful of preliminary results. For eight different samples of the grid network with size $K = 37$ (directed from the upper-left corner to the lower-right one, without any backward link), we compared the results of the three mechanisms with the following observables: the final dominion (sum of the frequencies achieved in each site of the network, bounded between 0 and $K - 1$), reported on Table 9.1, K_W , the number of sites which have been taken over according to an S-curve (with the same bounds, necessarily an integer, and lower than the dominion by definition), the mean growth time within all these K_W sites, the mean logit slope, the variance of both the growth and the slope, and finally the ratio between the overall growth time, and the growth time within the first invaded site, reported on Table 9.2.

The quantity K_W and the final dominion are close, and it might seem superfluous to consider both. Yet, it may be that one mechanism achieves a cohesive S-curve over all its dominion while another fails at it and takes over the contexts according to a different propagation pattern. We did not find, however, any difference between the mechanisms in this regard, except perhaps that the combined mechanism strengthens the cohesion. The ratio is interesting as we aim to find a longer propagation than in the one-site case. Furthermore, we can consider how it correlates with the dominion (expectedly, the larger the dominion, the greater the overall growth compared to the one-site growth: it takes more time to grab more sites).

We do not show here the results for the growth time mean and variance, nor for the slope mean and variance. As it appears, the hearer and the combined mechanisms are faster than the speaker case, which is understandable, since several occurrences of the new variant can be produced at each time step, up to 3 in the regular grid network. If

a non-saturated site is chosen, and the new variant is produced to express it so that a new occurrence of it got recorded, then additional occurrences can be added in the two neighboring site. On the contrary, only one occurrence can be produced at a time in the speaker case. As a result, the ratio of the overall growth time on the growth time in the first invaded site is also bigger in this case. As for the mean rate (mean slope of the logit), it is usually higher in the combined case (expectedly, since the mean growth is also smaller and the two are known to be anti-correlated). The growth varies more in the combined case than in the two others. Finally we did not find any significant differences between the three mechanisms regarding the variance of the slope.

A more interesting thing to observe, regarding the dominion (i.e. the depth of the network pervasion), is that the speaker and the hearer mechanisms stand as equal (Table 9.1). The combined mechanism, however, is much more efficient. This can be understood by the fact that the combined mechanism feels its surroundings in two ways (a site receives new occurrences from its neighbors and has an effective frequency of the new variant affected by their contents). Therefore, each site is more sensible to its environment, and less so to its own contents. If the novelty wanders around, it will thus be more likely to welcome it. Later on, we shall see that we can understand why the pervasion would be the same for the hearer and the speaker cases.

As for the correlation between the dominion and the ratio, its outcome is quite unexpected. In the speaker case, for which the ratio is almost always bigger than in the two other cases, we find a positive correlation of 0.52. Yet, for the hearer and the combined mechanisms, we find Pearson coefficients of -0.18 and -0.38 , respectively. This could seem, at first sight, non-sensical. However, we shall see that the actual scope of the dominion, at least in the combined case, has little impact on the ratio, for the change spreads from site to site faster than it proceeds within one. Therefore, the change will propagate almost everywhere extremely rapidly. That there is no correlation is understandable; that there should be an anti-correlation, even a slight one, is however difficult to interpret.

It appears then that the mechanisms do vary regarding the propagation further within the network; especially, the combined mechanism is more efficient. However, none of these first remarks could be easily related to empirical observations. Furthermore, it could be that some effects would be more likely in one mechanism than in another (for instance, several successive partially overlapping S-curves can happen in the hearer and speaker case, and less so in the combined case), but this remains to be explored.

Bailey-Shen law of rate acceleration

The spreading of the new variant over the different semantic units of the cluster is clearly an instance of a lexical diffusion. Lexical diffusion is associated with a law, stated first by Bailey (1973) and empirically confirmed by Shen (1997), according to which the later a context is affected by the change, the faster is the rate of change in this context. We tried to empirically support this claim in chapter 5, and failed to do so. We argued then that the determination of these ‘contexts of change’ was rather tricky in cases of semantic change, contrary to cases of phonetic change, where each word containing the changing phoneme can play the role of such a context.

Anyway, we can address this law through a modeling perspective. We already have everything we need to test numerically whether the law holds or not in our model, and under which conditions. The time at which a context starts being affected by the change is defined as the onset of the growth part, as extracted through our procedure. The rate of change is simply given by the slope of the logit transform, as was done in (Shen, 1997).

For the grid network, we find that, for the hearer mechanism, the values of the Pearson coefficient lie between 0.21 and 0.74, between 0.01 and 0.91 in the speaker case, and finally between 0.06 and 0.89 for the combined case, with means respectively equal to 0.56, 0.52 and 0.50. We seldom find nice correlations as did Shen (1997) and, even in the best case (the Pearson coefficient of 0.91, in the speaker case, is associated with only six data points, so that we choose the best combined case, with a coefficient of 0.89 and 22 data points), the correlation is not visually striking (Fig. 9.12).

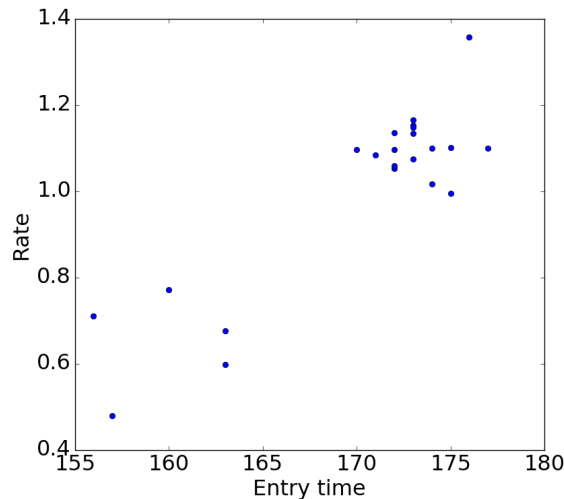


Figure 9.12: Correlation between the starting time of change and the rate, for the 5th network sample, with the combined mechanism. The associated Pearson correlation coefficient is equal to 0.89.

In our model, the origin of this correlation is not entirely clear. It makes sense to consider that later affected sites receive the input of earlier affected site, so that the change should be more and more encouraged. But in the grid network, a site is only fed by two others, except on the edges, so this effect should play little role. We can provide an attempt of an explanation, however.

Let us consider that the first invaded site is related to two sites 2 and 3, with links of respective weights γ_2 and γ_3 such that $\gamma_2 > \gamma_3 > \gamma_c$. A site can be invaded when $\gamma_k x > \gamma_c$, where x is the frequency of the new variant in the first invaded site. Clearly, site 2 is invaded first, and then site 3. Let us now consider that $\gamma_2 = 5\gamma_c$ and $\gamma_3 = 2\gamma_c$, so that these two sites can be invaded when $x = 0.2$ and $x = 0.5$, respectively. Site 2 is the first to be affected. As x continues to grow, it comes to be more and more affected. Then site 3 is affected.

Since x follows an S-curve, at the beginning it grows slowly. When x is higher, it

grows more rapidly. As a result, site 3 should receive more quickly a greater influx of the new variant than site 2 did, because $x = 1/2$ is closer to the maximum growth of the S-curve. When the frequency starts to grow in site 3, x grows faster in the first invaded site, and so it will push forward the change more efficiently in this site than it did for site 2. Hence, we have an acceleration of the rate of change for sites affected later. We would also have a deceleration of the rate for the contexts affected still later on, when it would correspond to the upper part of the sigmoid, but it would be associated with a much smaller interval for the weights ($[\gamma_c, 2\gamma_c/(1-x_c)]$, of size close to 0.2, to be compared with 0.65, the size of the interval that would be associated with an acceleration rate).

However, this picture does not hold. Indeed, we can consider a simplistic network of a chain of three sites. The first one is the reservoir, the second is the first invaded site, and the third one would play the same role as one of the two sites of the previous picture. Between the reservoir and the first invaded site, we set $\gamma_1 = 1.01\gamma_c$, and then we can investigate the effect of γ_2 on the rate of change in site 2. It turns out that the rate grows linearly with the strength of the link γ_2 (Fig. 9.13a). As the entry time indeed decreases with γ (the higher the influence of the first invaded site, the sooner the new variant enters the final site), we have, as a result, an anti-correlation between the rate of change and the entry time (Fig. 9.13b). The Pearson coefficient, in this deterministic case, is then equal to -0.59 .

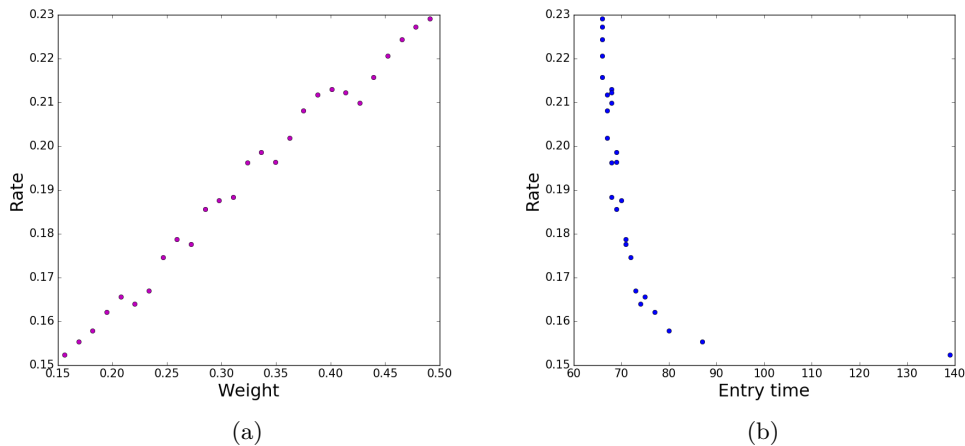


Figure 9.13: Deterministic simulation of the propagation in a chain of sites of length 3, with a link of fixed strength between the reservoir and the first invaded site, and a link of strength γ between the first invaded site and the final site. (a) Linear increase of the rate of change in the final site and the strength γ of the influence that the first invaded site exerts over it. The apparent noise is due to the fact that we recorded the evolution of the frequency every M timestep. (b) Anti-correlation of the rate of change in the final site and the time at which the new variant enters this site. The isolated point is not an artifact: it corresponds to the value of γ closest to the critical one, hence showing a latency and a much later entry time. $M = 5000$, $\delta = 0.01$, $\beta = \beta^*$.

To conclude, the Bailey-Shen law is an effect of the diffusion in the network, and

not of the particular features of the competition process. Why it holds is however unclear. Our model agrees with Bailey (1973) prediction of an accelerating rate in case of a lexical diffusion; moreover, in the model, this law is specifically an effect of this lexical diffusion. Even if we do not fully understand this feature, it is nevertheless nice to see it spontaneously emerge out of the phenomenology of the model.

Conclusion

There is a rich interplay between network structure, and the possible outcomes of a large-scale semantic expansion process, which we barely hinted at in this section. It would be extremely interesting to relate empirical features of individual changes to specific network structures. For instance, the change can be a global S-curve, or a succession of small, partially overlapping S-curves. In the latter case, it would be associated with a propagation over several small clusters. A slight semantic shift would be expected with the taking over of each successive cluster.

In this section, we have barely sketched possible lines of investigation. We believe, nonetheless, that a fruitful research perspective would be to relate, in a systematic way, the organization of the semantic network, which can be obtained through a variety of approaches (Ploux and Victorri, 1998; Gaume, 2004; Mikolov et al., 2013), and the frequency profiles, obtained from corpus data. In that sense, diachronic linguistics would be a tool to investigate the organization of knowledge at a given period, and moreover, the study of the semantic network, in synchrony, could in turn help to predict the likeliness of future language changes. This would be a vast research program, which is beyond the scope of this thesis. I believe that our model could be an interesting analytical and numerical device to explore the ties between these two extremely different but intimately related phenomena.

Concerning directionality, it is easy to implement it in the grid network. Furthermore, we can add a little bit of mixture in this network, without breaking the directionality, by drawing transverse links in the diagonals (i.e. upwards-rightwards or downwards-leftwards). This does hinder the overall propagation, but the new variant can still claim a significant dominion. Moreover, in the case of the grid network, we only considered a two dimensional grid, in which it is likely for the propagation to stop at some point, possibly early on. However, the semantic space is probably of a very high dimension, in which the intuition would be completely different. Because of the high number of dimensions, the variant will have more ease to propagate, at least in some of them. Furthermore, if we account, on top of this directional, highly dimensional hypercube, hyperplanes in which the links are not directional, then we could observe interesting configurations. Of course, adding ingredients in the model and increasing its complexity could obscure the picture, and compromise its relationship with the empirical observations of language change.

9.1.4 A theoretical approach

here, we try to investigate a bit further why propagation would be so difficult, and whether we can devise a criterion for a site to be eventually dominated or not by the new variant under conflicting inputs. Let us consider a site fed, on one side, by K_X

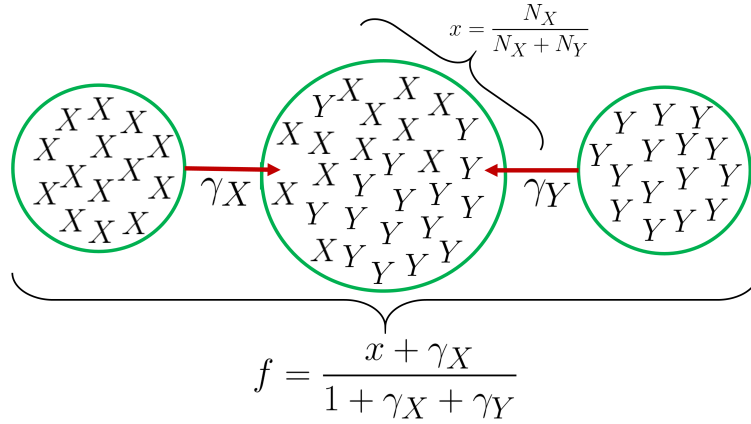


Figure 9.14: Schematic picture of the process. The site of interest (middle one) is fed by sites with X occurrences, for a total influence of γ_X , and by sites with Y occurrences only, for a total influence of γ_Y . Both contributions must be accounted for in the computation of the effective frequency f .

sites dominated by the new variant, on the other side by K_Y sites dominated by the old one (Fig. 9.14). The effective frequency will then read:

$$f = \frac{N + M \sum_{k=1}^{K_X} \gamma_k + 0 \cdot \sum_{k=1}^{K_Y} \gamma_k}{M + M \sum_{k=1}^{K_X} \gamma_k + M \sum_{k=1}^{K_Y} \gamma_k}, \quad (9.8)$$

which gives, once everything has been divided by M :

$$f = \frac{x + \gamma_X}{1 + \gamma_X + \gamma_Y}, \quad (9.9)$$

with $\gamma_X = \sum_{k=1}^{K_X} \gamma_k$ and $\gamma_Y = \sum_{k=1}^{K_Y} \gamma_k$.

To find back the expression used in the one-site competition, we can simply write:

$$f = \frac{(x - \gamma_Y) + \gamma_X + \gamma_Y}{1 + \gamma_X + \gamma_Y}. \quad (9.10)$$

The deterministic equation is still given by:

$$\dot{x} = P_0(f) - x \quad (9.11)$$

now equal to:

$$\dot{x} = P_{\gamma_X + \gamma_Y}(x - \gamma_Y) - x, \quad (9.12)$$

which can be rewritten:

$$\dot{x} = P_{\gamma_X + \gamma_Y}(x - \gamma_Y) - (x - \gamma_Y) - \gamma_Y, \quad (9.13)$$

It means that the velocity profile is now shifted both downwards and rightwards by γ_Y (and the new γ is greater than before). Note also that the completion of the change is now prevented, since if $x = 1$, we get:

$$\dot{x}|_{x=1} = P_{\gamma_X + \gamma_Y}(1 - \gamma_Y) - 1, \quad (9.14)$$

which is negative ($P(x) < 1$ for $x < 1$).

Hearer and combined cases

We can try to do the same in the hearer case. Then we have:

$$\dot{x} = P_0(x)(1 - x) + \sum_{k=1}^{K_X} \gamma_k(1 - x) - (1 - P_0(x))x - \sum_{k=1}^{K_Y} \gamma_k x, \quad (9.15)$$

which amounts to:

$$\dot{x} = P_0(x) - (1 + \gamma_X + \gamma_Y)x + \gamma_X. \quad (9.16)$$

With the transformation $\tilde{x} = (1 + \gamma_X + \gamma_Y)x - \gamma_X$, we get:

$$\frac{1}{1 + \gamma_X + \gamma_Y} \dot{\tilde{x}} = P_0\left(\frac{\tilde{x} + \gamma_X}{1 + \gamma_X + \gamma_Y}\right) - \tilde{x}. \quad (9.17)$$

A rescaling of the time suffices to get back equation (9.11) of the speaker case.

It means, therefore, that we can expect the speaker and the hearer case to have exactly the same conditions of propagation within the network. Indeed, we have observed, in the preceding discussion, that these two were always associated with the same dominion by the end of the process.

The combined case is expected to be different. We would have:

$$\dot{x} = P_0(f) - (1 + \gamma_X + \gamma_Y)x + \gamma_X. \quad (9.18)$$

But now, the change of frequency $x \rightarrow \tilde{x}$ affects the effective frequency:

$$f = \frac{\tilde{x} + \gamma_X(2 + \gamma_X + \gamma_Y)}{1 + (\gamma_X + \gamma_Y)(2 + \gamma_X + \gamma_Y)}. \quad (9.19)$$

The same trick as for the hearer case leads to:

$$\dot{\tilde{x}} = P_{\tilde{\gamma}_X + \tilde{\gamma}_Y}(\tilde{x} - \tilde{\gamma}_Y) - (\tilde{x} - \tilde{\gamma}_Y) - \tilde{\gamma}_Y. \quad (9.20)$$

with a rescaling of the time, and noting $\tilde{\gamma}_Y = \gamma_Y(2 + \gamma_x + \gamma_Y)$ as well as $\tilde{\gamma}_X = \gamma_X(2 + \gamma_x + \gamma_Y)$. Once more, the combination of the two propagation mechanisms leads to a strengthening of the links; here they are both increased by a factor $(2 + \gamma_x + \gamma_Y)$.

Pitchfork bifurcation

The set-up is quite similar to the simple competition case, in the sense that the fixed points are found at the intersection between the sigmoid and a straight curve. There will always be at least one fixed point in the $[0,1]$ interval, and no more than three.

A first case of interest would be the $\gamma_X = \gamma_Y$. Indeed, the following relation always hold:

$$\frac{f - (1 - f)}{\sqrt{f(1 - f)}} = \frac{2x - 1 + \gamma_X - \gamma_Y}{(x + \gamma_Y)(1 - x + \gamma_Y)}, \quad (9.21)$$

so that, when the two influences are equal, $x = 1/2$ is a fixed point of the dynamics. The condition of stability is given by:

$$\beta(2\gamma + 1)^2 < 4(\gamma + 1/2)^3. \quad (9.22)$$

Let us note γ_ψ the value for which we have equality between the two members, which is roughly equal to 0.31. When $\gamma > \gamma_\psi$, the fixed point $x = 1/2$ becomes stable: the population of the site is then evenly divided between the two variants, meaning that the contents of the invaded is now determined by the influx of forms, not by its own feedback (indeed it does not depend anymore on the initial condition). Below this value, two stable fixed points appear and the statu quo between the two variants point becomes unstable. This bifurcation is a pitchfork one; hence the index ψ for this new threshold value. Expectedly then, if the influx γ_X of the new variant is not greater than the influx of the old, the new variant can achieve at best a coexistence, but never takes over.

We now have a clearer picture of the equilibrium states of the system. If $\gamma_X > \gamma_Y$, the new variant can take the upper hand. If $\gamma_X > \gamma_\psi$, then there is a unique fixed point located between $1/2$ and 1 . If $\gamma_X < \gamma_\psi$ however, then there can be two fixed points, depending on the values of γ_X and γ_Y ; the saddle-node bifurcation picture of the simple invasion appears again. Therefore, above γ_ψ , $\gamma_X > \gamma_Y$ is a sufficient condition to ensure that X will become the dominant variant. Below γ_ψ , this condition becomes only necessary and γ_X must be greater enough than γ_Y for the saddle-node bifurcation to occur. This is because, below γ_Y , the site under consideration feels that it is already filled up with occurrences of the old variant and the new variant must overcome this inertia.

In the region $\gamma_Y < \gamma_\psi$, there exists a threshold value for γ_X , as was the case when we had $\gamma_Y = 0$. Below this threshold, the same saddle-node bifurcation occurs as before and a new stable fixed point appears, such that even if $\gamma_X > \gamma_Y$, the new variant remains marginal. We can compute the evolution of this critical threshold for all values of γ_Y (Fig. 9.15). In this case, as γ_X crosses the threshold, the fixed point of the system will jump from a low value to a much higher one (Fig. 9.16). This, however, is no longer true above the pitchfork bifurcation where the fixed point of the system increases continuously with γ_X .

As a consequence, there should be no S-curve above the pitchfork bifurcation threshold and the growth will be then exponential (very fast at first, and slower and slower as it goes on). In a highly connected network, where γ_X and γ_Y are expected to be big, there will be therefore no S-curve past the few first sites.

Summary

This theoretical picture clearly tells us the weakness of the new variant regarding the propagation process. Indeed, as soon as a site that the new variant tries to invade is also under the influence of a source of the old variant, then the growth becomes

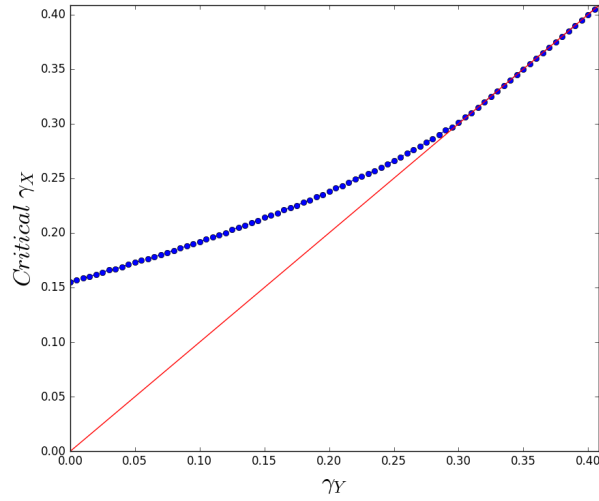


Figure 9.15: Minimal value of γ_X so that the fixed point is greater than 0.5, for any value of γ_Y . $M = 5000$, $\beta = \beta^*$. After the pitchfork bifurcation (for $\gamma_Y = \gamma_{\psi}$, there is no longer a saddle-node bifurcation and thus $\gamma_X = \gamma_Y$ is a sufficient condition for the new variant to reach majority.

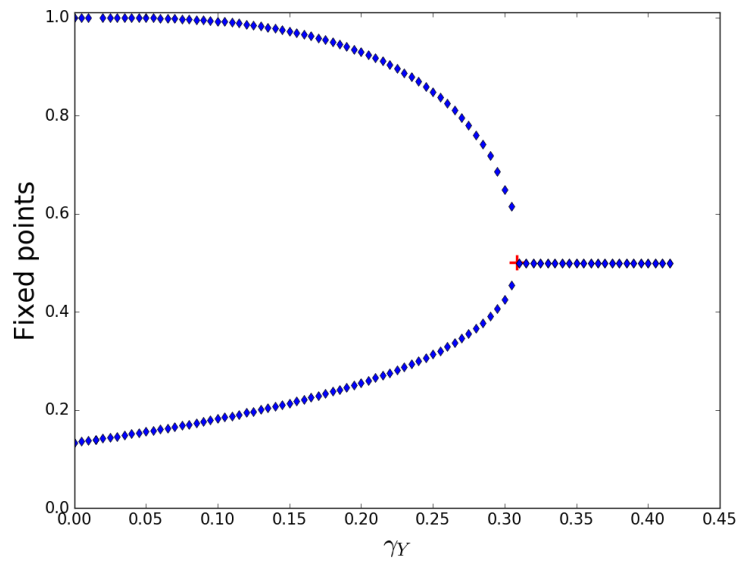


Figure 9.16: Values of the fixed point below (lower branch) and above (higher branch) the bifurcation threshold. These two merge at the pitchfork bifurcation (whose value is represented by a red cross). In this figure, the unstable branch of the pitchfork bifurcation is merged with one of the two stable branches, because the figure is associated with the saddle-node bifurcation threshold. $M = 5000$, $\beta = \beta^*$.

efficient. We can also understand why propagation is expected to be difficult, and why it would need such a specific, directional structure to proceed: competing influences over a site reduces the maximum achievable dominion, which cripples the strength of the subsequent propagation.

9.1.5 The growth distribution

We expect that the propagation within the network could solve the issues we encountered in the one-site competition, which was that the average growth time was too small compared to the latency time, and its variance too narrow. We announced in chapter 8 that these shortcomings could be overcome by considering a propagation subsequent to the first invaded site. We cannot solve this issue in this limited, perspective chapter, but at least we can check whether this claim is legitimate.

A longer growth time

In the grid network (the only one in which the propagation had good chances to reach a reasonable extent), the combined mechanism (which is the most powerful mechanism of propagation of the three available) is never associated with a high ratio between the overall growth, and the growth of the one-site competition (the time growth of the first invaded site). In Table 9.2, the maximum ratio was 2.77.

We can, however, consider a bigger network, for instance with $L = 10$ ($K = 101$). In the run we tried on, we obtained a total number of totally dominated contexts equal to 81, and a final dominion of 92.29, which is quite large. Yet, the overall growth, over the 81 fully dominated contexts, was only 2.93 times bigger than the growth within the first invaded site. This is because the change spreads faster from one site to another than it dominates a site. Indeed, the growth in the first site is equal to 15 counting windows — remember that we apply, to the surrogate data, the same procedure as we did for the corpus data, so that we record the number of occurrences produced every $K * W$ time steps — while the time between the entry in the first invaded site, and the entry in the last affected site, is of 25 counting windows — a comparable number, then.

Since the speaker mechanism presents a correlation between the extent of the dominion and the overall growth time, we might want to try, in this case, a bigger network, with the risk that, even though a wider propagation would lead to a longer overall growth time, the mechanism might not be powerful enough to achieve this propagation. On the same network of size $L = 10$ that we used for the combined mechanism (i.e. with the exact same weights), we find a number of taken contexts K_W equal to 45 over 100, and a final dominion of 57.16, which is, as expected, much lower than the previous case. However, the growth ratio is also significantly bigger since it is now equal to 9.64 (a factor of the order of magnitude we needed to match the ratio between the mean growth time and the mean latency time found in corpus data). What is, however, less satisfying, is that the overall growth curve is not fully S-shaped. We could perhaps design a modified combined mechanism (as well as a modified hearer mechanism) such that only one occurrence is recorded at a time, and not several. In that way, the growth ratio would certainly get bigger, but we might lose in efficiency. Also, the equivalence rules between the different mechanisms would be altered.

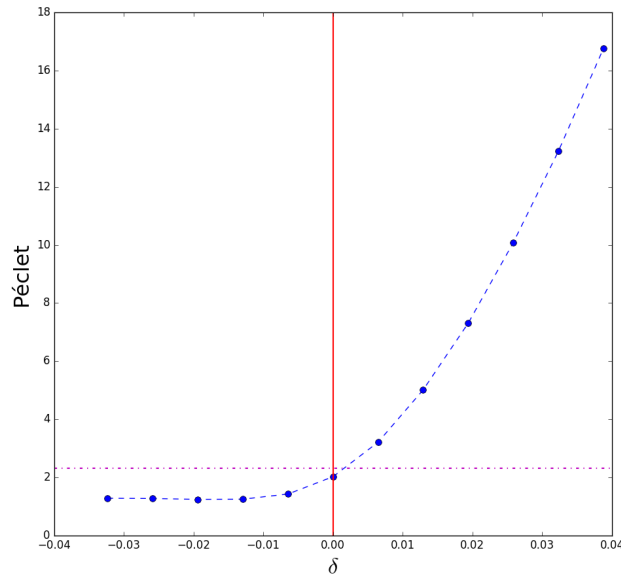


Figure 9.18: Péclet number associated with the latency distribution as a function of the distance δ to the critical threshold γ_c . The Péclet number increases very rapidly once the threshold has been crossed. The magenta dotted line indicates the value of the Péclet number found in corpus data. $M = 5000$, $\beta = \beta^*$, speaker mechanism.

Anyway, it might also be that, in the simulations, we considered that the change happened right at the threshold. As we saw in the previous subsection, latency still endures slightly away from the threshold, but diminishes rapidly. If the γ parameter is in the vicinity of the threshold value, but not exactly onto it every time, then it can be associated with shorter latencies in average. Therefore, instead of having underestimated the growth time, we might have overestimated the latency time. In such a case, that the growth time does not get much longer in the wider propagation picture should not be of much concern.

Note, however, that by moving slightly away from the threshold, we would lose the agreement with the corpus data Péclet number (Fig. 9.18), which holds only on the threshold, or slightly below. The latency phenomenon of corpus data seems thus to indicate a fine tuning to criticality, which is admittedly intriguing. And, actually, getting the right Péclet numbers is the most crucial requirement to compare properly the model outcome with empirical data. Indeed, it is those that describes the shape of the distribution. It is therefore worth investigating if a broader propagation of the new variant would allow to approach the values of the Péclet numbers found in empirical data, especially for the growing time and the slope, since the one-site set-up already matched the latency Péclet number (a good thing, as you would note, since considering a structured cluster of sites instead of a single site would change nothing to the latency part of the curve).

The Péclet numbers

We recall here that the Péclet number, defined as:

$$P_e = 2\frac{\mu^2}{\sigma^2}, \quad (9.23)$$

where μ and σ^2 are, respectively, the mean and the variance of the distribution. In the empirical data, we found a Péclet number of 13.63, while in the model, it was equal to 126.7, about ten times bigger. In short, the relative variation in growth time is too small in the numerical growth time. Similarly, for the slopes, we had found a Péclet number of 166.0 in the model, which did not match the empirical value obtained from corpus data (10.55).

We shall therefore ask a simple question: does further propagation in the network help to lower the Péclet number associated with the overall growth time? Indeed, we can expect the propagation to generate further uncertainty on the growth time, hence a higher relative variation. To answer this question, we generated 50 runs of the model, with the combined mechanism (to ensure a more efficient propagation), on a grid network of size $L = 5$ (hence $K = 26$), and a δ parameter, for the invasion of the first site, equal to 0.01 (so as to avoid to be stuck in the latency part for too long). This should be of no consequence: since the latency part only affects the first invaded site, and is not correlated with the growth time and the slope (the two quantities presently under study), the results which follow should be compatible with a tuning to criticality as well ($\delta = 0.0$). The parameters M and β have been left unchanged.

Two distinct simulations have been performed. In the first one, the directed links of the grid network are sampled anew for each run. In the second, we keep the same network over all runs, so that the variation only comes from the inherent stochasticity of the process. This would allow us to distinguish whether the additional variation that we are looking for comes from the sole process, or from the randomness of the network. Should it be the latter, it would imply that empirical data is more likely to favor a picture where the differences between the network regions on which the different changes occur explain the particular shape of the distribution.

- Random network: In this case, we find a Péclet number of 43.59 for the growth times and of 28.80 for the logit slopes. We did not manage to get satisfying distributions because of the reduced size of the sample (50), so we could not check if the distributions were still compatible with an Inverse Gaussian. These two results are still far way from their empirical counterparts (13.63 and 10.55), but at least, it shows that further propagation within the network is indeed associated with an increase of relative variation, and hence with a decrease of the Péclet numbers. An interesting question would be to assess how far the propagation must get in average to obtain the right Péclet numbers, and what is the range of values on which this dominion is expected to vary then.
- Fixed network: The results obtained in this second case are extremely surprising, for we find Péclet numbers of 286.9 and 316.5 for the growth times and the slopes, respectively. This is nevertheless not so strange if we recall that the propagation in the network is extremely deterministic: the main source of stochasticity

is latency, which plays no role in the process, and as it turned out for our reduced survey over eight different networks, the remarkable agreement between the hearer and the speaker mechanism, predicted by the deterministic picture, held remarkably well.

Therefore, an additional source of variability can arise out of the network, but because of the network itself and not because of the propagation. The variation source which can explain best the empirical data is the random and irregular character of the semantic network.

9.2 Semantic bleaching

We shall now present an extension of our model, accounting for a better representation of the semantic aspect of the process of change. It follows from the simple idea that we can, as suggested in chapter 7, associate the sites of the semantic territory with a meaning, e.g., a vectorial one. In what follows, we will try to draw the consequences of this simple proposal.

9.2.1 A conceptual picture of semantic change

There would be a lot of theoretical ways to associate a mathematical representation of meaning with each of the semantic units. In the following, we shall content ourselves with one of the simplest. To each site k is associated a S -dimensional vector $\Theta_k = \{\theta_k^1, \theta_k^2, \dots, \theta_k^S\}$, such that, $\forall s \in [1 : S]$, $0 < \theta_k^s < 1$ (in this notation, s would stand for ‘semantic trait’). As a result, all meanings are comprised within a hypercube of dimension S and of size 1.

Weights of the network links

We assumed several times in this work that the weights of the network links were reflecting the semantic distance between the different sites, so that the larger the semantic distance, the weaker the weight. It seems therefore natural to assume that the weight of the link γ_{ij} from site j to site i would be related to their respective meanings Θ_i and Θ_j . The simplest way to do so would be to define the weight as proportional to the normalized scalar product \cdot between the two vectorial meanings:

$$\gamma_{ij} = \frac{\Theta_i \cdot \Theta_j}{\|\Theta_i\| \|\Theta_j\|}, \quad (9.24)$$

where $\|\Theta_i\|$ is the norm of the semantic vector Θ_i . Note that if the two vectors do not share any semantic trait, (i.e., $\theta_i^s = 0$ or $\theta_j^s = 0$, $\forall s$), then the weight should be zero. This feature is interesting, since it implies that, for a pathway to be open between two sites with no common semantic traits, there must be a third bridging site which has semantic traits of both.

A concern with this proposal is that the links would be symmetric, that is, $\gamma_{ij} = \gamma_{ji}$ according to formula (9.24). There are two ways to deal with this asymmetry. The

first one would be to change the formula, and to propose, for instance:

$$\gamma_{ij} = \frac{1}{1 + \frac{\|\Theta_i\|}{\|\Theta_j\|}}. \quad (9.25)$$

This proposal does not pop out of nowhere. The semantic vector of site j is seen from site i as $(\Theta_i \cdot \Theta_j) / \|\Theta_i\|$. Yet this can be greater than 1 since the norm of Θ_j can be as great as \sqrt{S} . Therefore, we divide this influence by the sum of the two influences site i and site j exerts on one another. The problem is that, since this expression does not depend on the scalar product anymore, a vector associated with a very high norm will be extremely influential on all sites in the network, even if it does not share any single semantic trait with them. This non-local, infinite range influence seems not to be a realistic feature.

To get over this, we can weight how Θ_j is seen from Θ_i by the cosine of the angle α between the two angles, given by:

$$\cos \alpha = \frac{\Theta_i \cdot \Theta_j}{\|\Theta_i\| \|\Theta_j\|}, \quad (9.26)$$

so that we get:

$$\gamma_{ij} = \frac{\cos \alpha \frac{\Theta_i \cdot \Theta_j}{\|\Theta_i\|}}{\frac{\Theta_i \cdot \Theta_j}{\|\Theta_j\|} + \cos \alpha \frac{\Theta_i \cdot \Theta_j}{\|\Theta_i\|}}, \quad (9.27)$$

leading to:

$$\gamma_{ij} = \frac{1}{1 + \frac{\|\Theta_i\|}{\cos \alpha \|\Theta_j\|}}, \quad (9.28)$$

or equivalently:

$$\gamma_{ij} = \left(1 + \frac{\|\Theta_i\|^2}{\Theta_i \cdot \Theta_j} \right)^{-1}. \quad (9.29)$$

This is the formula that we used in the simulations. Naturally, it must be set equal to 0 whenever the scalar product between the two vectors is null. If the two semantic vectors have the same norm, then the weights are once more symmetric and are upper bounded by 1/2.

Directionality in the network

Note that, in this picture, directionality arises as a side effect of the difference in norms between the semantic vectors. The asymmetry then flows from the semantic unit with the highest norm to the semantic unit with the lowest one. This, however, may not be entirely satisfying. Indeed, the links should be directed from the high-features meanings to the low-features ones, which is a bit odd. As we shall see in a moment, this is not to be equated with the fact that generalist linguistic forms will take over the dominion of specialist linguistic forms, so it might not be such of a concern.

In spontaneous semantic associations tasks, for instance, people tend to relate a noun with the category it is a member of (Ferrand and Alario, 1998). Admittedly, a category member presents more specific features than the category it is a member of, so that our proposal is not aberrant from this point of view.

This directionality also echoes with Sweetser (1988) view that grammatical meanings are associated with less schematic features than the lexical ones, so this proposal should be compatible with the statistical directionality observed in grammaticalization. Changes from lexical meanings to grammatical meanings, being changes from high-features meanings to low-features ones, indeed correspond to the directionality arising out of the semantic norms in this proposal.

Additional layout

We announced two ways to deal with the asymmetry. The second one is to superpose a network layout over this first picture. That is, we could start with a complete graph, compute the weights out of the semantic vectors associated with each node, and then filter out the links according to another graph layout (e.g., a grid). The resulting network would be the superposition of the weighted complete graph, and the directed network. The adjacency matrix would therefore be given by the element-wise product of the two matrices, leading to:

$$\gamma_{ij} = \epsilon_{ij} \frac{\Theta_i \cdot \Theta_j}{\|\Theta_i\| \|\Theta_j\|}, \quad (9.30)$$

where ϵ_{ij} is either 1 and 0 according to the existence or not of a link in the directed network layout.

In the simulations, we shall consider both sources of asymmetries, since we will make use of a directed grid layout as in the first section, although we will compute the weights from equation (9.29). Indeed, the connectivity in the network must not be excessive, as we saw that influxes from the older variants reservoirs, while lesser than the influxes of the new variant, had an accumulative effect in weakening further propagation of the new variant further into the network.

9.2.2 The meaning of forms

So far, we only provided a rather arbitrary mathematical depiction of semantic units, which does not bring much to the model. But we can extend this picture to the representation of the meaning of the linguistic forms — which is the one meaning truly salient for the speakers of a language. A simple and efficient way to do so would be to consider a barycentric sum of all vectorial units, so that the semantic trait θ_X^s of the vectorial meaning Θ_X of form X is given by:

$$\theta_X^s = \frac{\sum_{k=1}^K x_k \theta_k^s}{\sum_{k=1}^K x_k}. \quad (9.31)$$

Thus, we relate the meaning of a form with its occupation over the semantic territory. If we note Θ the $S \times K$ matrix which columns are the Θ_k 's ($\Theta_{sk} = \theta_k^s$) and X the column vector of all frequencies x_k , then we would have:

$$\Theta_X = \Theta X. \quad (9.32)$$

Note that, in this view, we can speak of 'the' meaning of form X (which would be Θ_X) even if X covers several semantic units, and hence is polysemic. Also, the meaning of X , usually, does not correspond to any given semantic unit. This would explain why translations between language to another is so tricky: they may share the same semantic units, the same semantic territory, but the coverage of it by the forms will be different, and so there is never any exact match between two form meanings. Similarly, if the dominion of a linguistic form is large, it is quite pointless to ask what is its meaning. The answer would exist — Θ_X , as we pointed out — but would be blurry compared to the covered semantic units and would tell little about the actual uses of X . Another important point is that a form does not have an intrinsic meaning: this meaning stems entirely out of its uses. What we offer here is therefore a usage-based representation of meaning.

Adequacy

We can now define how much a linguistic form X is appropriate to express a given semantic unit k . In the following, we shall refer to this quantity as the 'adequacy' of form X for semantic unit k , and shall note it A_X^k . As for the weights, we have several possibilities to define the adequacy, the first one being the cosine of the angle α between the two vectors:

$$A_X^k = \frac{\Theta_k \cdot \Theta_X}{\|\Theta_k\| \|\Theta_X\|}. \quad (9.33)$$

This quantity varies between 0 and 1 and only depends on the general semantic orientations of both the form and the unit.

The second one would describe how clearly form X 'sees' the semantic unit k :

$$A_X^k = \frac{\Theta_k \cdot \Theta_X}{\|\Theta_X\|}. \quad (9.34)$$

This quantity is non-negative, and is upper-bounded by \sqrt{S} , which can be arbitrarily large. It depends on the norm of Θ_k only, not on that of Θ_X . Therefore, as long as two forms have the same orientation, they are equally fitted to express the semantic unit k , even if one has a much lower norm than the other.

Finally we can consider how the form X is perceived from semantic unit k :

$$A_X^k = \frac{\Theta_k \cdot \Theta_X}{\|\Theta_k\|}. \quad (9.35)$$

This expression does not depend on the norm of Θ_k . Now, two forms of identical orientations will not stand as equal anymore, and the one with the bigger norm will be perceived as the more adequate. If we have for instance two forms X and Y of the same orientation as Θ_k , so that Θ_Y is exactly equal to Θ_k , but Θ_X overemphasizes all

its semantic features, and hence has a higher norm than both Θ_k and Θ_Y , then X will still be seen as the most adequate. This seems not to be an issue, since it does not imply, for instance, that the most adequate forms are the most expressive, or the most vivid ones, since these pragmatic nuances could be encoded as semantic traits as well. It only means that the most adequate form is the one which goes the furthest into the semantic direction pointed out by the semantic unit.

We shall actually favor a fourth expression:

$$A_X^k = \frac{\Theta_k \cdot \Theta_X}{\|\Theta_k\|^2}, \quad (9.36)$$

which also reads:

$$A_X^k = \frac{\|\Theta_X\|}{\|\Theta_k\|} \cos \alpha. \quad (9.37)$$

The advantage of this formula is, first, that it does not scale with the dimension S of the semantic space, which is an important requirement, second, that a form is adequate only if its norm is big relatively to that of the semantic unit. This feature is in line with our aim to build an adequacy quantity which would answer the question: how much form X is fitted to express semantic unit k ? Therefore, we want the adequacy to be defined relatively to that semantic unit.

Note that, if Θ_X is exactly equal to Θ_k , the adequacy is 1, so that we have a significant mark to assess how good the adequacy is. It can actually get bigger than 1, if Θ_X magnifies all the semantic features of the semantic unit and not the others (same orientation, greater norm). This expression also means that some semantic units are more demanding than others, and will require richer semantic features so as to match their own.

Semantic bias in the dynamics

At this point it might be interesting to consider the possibility of a semantic bias, that is, to account for a favoring of the most adequate form when it comes to express a given semantic unit. There are certainly a huge number of ways to implement such a bias in the dynamics, but we shall consider only one. It could seem reminiscent to the production of T utterances in a communication event that is part of the Utterance Selection Model (chapter 6), but the main goal of the latter was to introduce a finite-size sampling of the forms, so the resemblance between this particular feature of our model and what follows is more superficial than enlightening.

We introduce a working memory of size W . Whenever semantic unit k is intended to be expressed, this working memory will be filled with occurrences until it is full. Occurrences are hired one by one in the working memory; each time, variant X is produced with probability $P(f)$ for the speaker and combined cases, and $P(x)$ for the hearer one. Otherwise an occurrence of Y is added in the working memory. Once W such occurrences has been produced, the occurrence associated with the highest adequacy with the meaning is then uttered. This little mechanism corresponds thus to a retrieval process, limited in time, aiming to find the most adequate form in a given utterance situation. If $W = 1$, then the situation is the same as before and the bias does not apply. If W is infinite, the frequency does not play a role anymore and the

most adequate form in memory will be chosen with certainty. Therefore, the actual strength of the bias depends on the size of the working memory.

The bias does not help us to introduce the new variant in the first place. Only the two mechanisms, speaker and hearer, can introduce novelty in neighboring contexts. The bias has furthermore no reason to favor the new variant, and can also act in favor of the old one, if such is the semantic set-up of the disputed site. Especially, in the one-site competition, the new variant only occupies one site and the old one another; both are therefore semantically tuned to the unit they occupy. Unless the source meaning is such that it presents exactly the same semantic features as the target meaning, only strengthened, then the semantic bias should counter the invasion of the new variant. We shall explain in the next subsection why it is more likely to act in favor of the new variant anyway, but at least, the leverage is neither granted nor intrinsic.

Furthermore, this bias is associated with non-local effects. The dynamics on one site is no longer influenced by its sole neighbors, but also becomes affected by what could happen in other, unrelated parts of the network. The linguistic forms are the ones responsible for these long range interferences, a fact which is, from a conceptual perspective, satisfying. Indeed, speakers are especially sensitive to the linguistic forms, and this can influence their choices, and their own perception of the semantic territory. A drawback from this increasingly complex picture is that the actual value of the adequacy (and consequently, of the bias) depends on various quantities, such as the solid angle and the norms, which themselves depend on the overall populational picture of the form, so that the intuition can easily get lost.

For instance, the bias might be associated with a strengthening effect. As some linguistic variant sets foot on the semantic unit, and the older one consequently steps back, the overall meaning of the new variant will shift closer to the meaning of the semantic unit, while the meaning of the old shall turn away. The adequacy of the new variant therefore increases as it takes over the site, and that of the old variant, as it retreats, diminishes. However, this bias, as we introduced it, is not sensitive to the actual difference in adequacy between the two forms. Even if a form is poorly adequate to the meaning of the semantic unit, as long as it scores better than its competitor, it should be as favored as if it were the most adequate linguistic form ever. There is therefore no actual strengthening effect, except if the invader had started semantically unfavorable, and becomes more adequate than the invadee at some point of the dynamics. This could have an effect in extremely peculiar situations only, for instance if the adequacy of the two words is very close, so that a slight change in frequency can trigger a turn of favor.

The main effect of the bias appears when the invader is benefiting from it. Thus, as soon as it appears but once in the working memory, it will be selected. As a result, the deterministic equation would be, in the speaker case:

$$\dot{x} = 1 - (1 - P_\gamma(x))^W - x, \quad (9.38)$$

and in the hearer case:

$$\dot{x} = 1 - (1 - P_0(x))^W - (1 + \gamma)x + \gamma. \quad (9.39)$$

Note that the symmetry between the two mechanisms, ensured by the change of variable $x \rightarrow (x + \gamma)/(1 + \gamma)$, still holds. Although it would seem that the bias, being

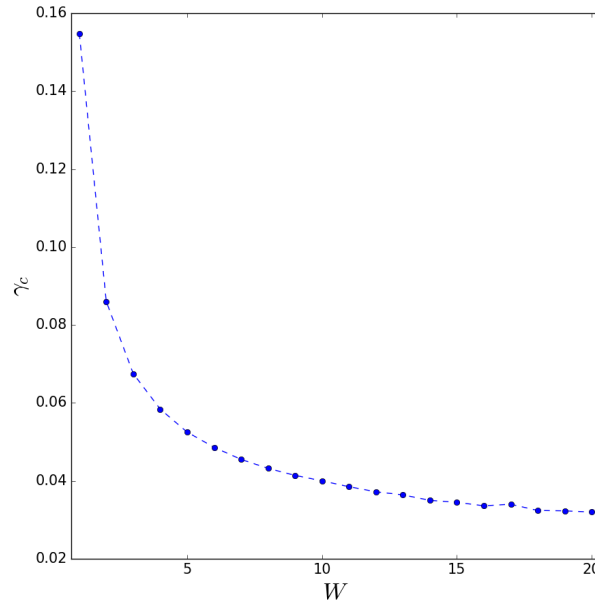


Figure 9.19: Effect of the bias (which effect increases as the size W of the working memory) on the critical threshold for the γ parameter. Speaker mechanism, $\beta = \beta^*$.

associated with the production step, could favor the speaker mechanism exclusively, this is not the case and the two mechanisms remain equal in terms of efficiency.

Let us discuss a little bit further the speaker case. At the beginning of the process, $P_\gamma(x)$ is very small, so that we can write:

$$\dot{x} = WP_\gamma(x) - x. \quad (9.40)$$

If rescale the time $t \rightarrow Wt$ and the frequency $x \rightarrow Wx$, we get:

$$\dot{x} = P_\gamma(Wx) - x, \quad (9.41)$$

and then :

$$\dot{x} = P\left(\frac{x + \gamma/W}{1 + \gamma/W + (1 - W)/W}\right) - x. \quad (9.42)$$

This, actually, is the same structure in the effective frequency as equation (9.9) for two influxes, γ_X and γ_Y , only that we have now $\gamma_X = \gamma/W$ and $\gamma_Y = (1 - W)/W$, meaning that γ_Y is negative (at least as soon as $W > 1$, which must be the case for the bias to act on the dynamics). As a result, the critical threshold γ_c is lowered (Fig. 9.19).

In the extended picture with an additional influx γ_Y of the convention variant, we would have:

$$\dot{x} = P\left(\frac{x + \gamma_X/W}{1 + \gamma_X/W + \gamma_Y/W + (1 - W)/W}\right) - x, \quad (9.43)$$

so that γ_Y must be greater than $W - 1$ to have an effect. Though this is not impossible (γ_Y , being a sum of several weights, is not upper bounded), it considerably lowers the

crippling of this influx. Therefore, the bias, though it does not change much the picture in the one-site competition, is a powerful mechanism for further propagation into the semantic network, even if W is very low.

9.2.3 Semantic bleaching and network propagation

All these previous additions to the model are especially relevant to the modeling of grammaticalization, though they apply more generally to semantic change. We can investigate, particularly, the phenomenon known as semantic bleaching.

Semantic bleaching is a common feature of grammaticalization. Albeit it is not specific to this phenomenon, it has been mostly discussed in the literature related to it. Roughly, the issue is that there are two views of grammaticalization, as we have already discussed in chapter 2: one sees the process as an impoverishment of the form (loss of semantic features, among others), while the other regards grammaticalization as an expansion (over contexts of use). On one side, we have the abstraction of a simpler, more skeletal schema (Sweetser, 1988), an increase in adaptability, a lesser dependence in the semantic context of the utterance; on the other side, the development of a new, rich, elaborated, grammatical meaning. We are not the first to argue that these two views could be combined Traugott (2015), but our attempt might be one of the first numerical modeling approach to the topic.

In our vectorial representation, the scalar product between form X and semantic unit i is given by:

$$\Theta_i \cdot \Theta_X = \frac{1}{Z} \sum_{s=1}^S \theta_i^s \sum_{k=1}^K \theta_k^s x_k, \quad (9.44)$$

where Z is a shorter notation for the sum over all sites k of the frequencies x_k of form X . This can be rewritten:

$$\Theta_i \cdot \Theta_X = \frac{1}{Z} \sum_{k=1}^K x_k \Theta_i \cdot \Theta_k. \quad (9.45)$$

Now this can be interpreted as some sort of an average over the populational presence of X which we shall write:

$$\Theta_i \cdot \Theta_X = \langle \Theta_i \cdot \Theta_k \rangle_X. \quad (9.46)$$

It is, therefore, some sort of an average correlation between all semantic vectors of units X has a foot in.

Clearly, as X spreads over the network and possibly explores different, separated dimensions of meaning, the scalar products between all the units covered will get poorer, and the average (9.46), rather low. As this scalar product directly enters the definition (9.36) of the adequacy (and contains all the dependence in X), it means that form X will cease to be perceived appropriate to express the meaning with which it is, nonetheless associated. This does not imply that other linguistic forms will be perceived better suited though, and X can remains pretty much unchallenged despite of this overall bleaching.

In this view, the semantic bleaching of form X (which translates as a global decrease in adequacy) is related to the fact that this linguistic form is associated with

heterogeneous uses, so that its meaning is no longer perceived as unified. In that sense, our model already captures the semantic bleaching phenomenon; it suffices to add a semantic representation, and then to let the meaning of form X evolve through the changes in use it undergoes as it propagates over the semantic network.

A set-up for pervasion

We could have used a simple grid network and added a semantic vector to each nodes (the main features would have been captured), but we preferred implementing all the features mentioned throughout this section, to get a flavor of how these concepts can be rooted to solid ground. Adding the biases to the picture has proved particularly problematic, since the established variant, as it occupies the network begin with, has expectedly a better overall adequacy than the new variant which is only marginal.

Worse, to tune the system at criticality, i.e. to obtain a weight equal to γ_c between the reservoir and the first invaded site (which is roughly equal to 0.075 in the combined case), it entails that the scalar product between the two must be very low. As a consequence, the adequacy between X and the first invaded site (which is initially the same as the adequacy between the reservoir and the first invaded site) is similarly poor and the bias mechanism will, in most cases, acts against it. We shall see how to deal with this technical issue.

In this simulation, we choose $M = 1000$, $\beta = \beta^*$, $\delta = 0.01$ and a grid layout of size $L = 8$, so that $K = 65$, reservoir included. Furthermore $S = 100$ and the size W of the working memory is set equal to 3. For convenience, we shall refer to the reservoir with the label 0 and to the first invaded site with the label 1.

First, we need to construct the semantic vectors of all K sites. We started with the first invaded site. We then drew 5 semantic traits out of the 100 so that, for each, θ_1^s was drawn from a uniform distribution on the $[0, 1]$ interval. Then, for each site of the network, we drew 3 semantic traits out of the 100, and whenever applicable, one semantic trait among the traits of its upper neighbor, and one among those of its left neighbor. The sites on the edges of the grid drew 2 semantic traits from their only neighbor. Then, for each site k , and for each semantic trait s (up to 5, if no trait had been drawn twice), θ_k^s was drawn out of the same uniform distribution as before.

This procedure might seem a little bizarre, but it actually ensures that the network is, on one hand, inhomogeneous enough, so that the established form is not quite adequate with any of those (and hopefully not with site 1); on the the other hand, that it still allows neighboring sites to be related with a sufficient weight so as to leave way for the semantic change to proceed. We actually managed to get a bias in favor of X in the first invaded site thanks to this set-up, though it required a relatively large network ($L = 8$) and several trials (we changed the random seed after each unsuccessful attempt).

The links of the asymmetric grid layout (it is the same directed structure as the one we used in the previous section) were then computed from the semantic vectors according to formula (9.29). Now we want the link $\gamma_{1,0}$ between the reservoir and site 1 to be equal to $\gamma = \gamma_c(1 + \delta)$. To achieve so, we set:

$$\theta_0^s = \frac{\gamma}{\max_s(\theta_1^s)} \delta_{s,s^*}, \quad (9.47)$$

with:

$$s^* = \arg \max_s (\theta_1^s). \quad (9.48)$$

Note that this definition is partially incorrect in the sense that, if $\max_s (\theta_1^s) < \gamma$, then we obtain a $\theta_0^{s^*}$ greater than 1, which is not allowed. In practice this happens with probability $\gamma^5 \approx 2.10^{-6}$, so that we should not bother too much with it.

It took then three things to implement this complex picture: 1 - the construction of the semantic units vectorial meaning; 2 - the providing of a network layout; 3 - a specific technical tweak to ensure that the first competition is tuned to criticality.

Coupling between propagation and semantic change

Including the semantic bias in the dynamics is interesting, since it shows which difficulties can arise and how we can overcome them (first by ensuring some semantic heterogeneity in the network, second by increasing the number of sites in the network, so that the established variant covers a domain of eclectic meaning and ceases to be adequate with one specific and marginal site), we prefer to illustrate the coupling between the semantic evolution and the spread associated with a simpler, non-biased dynamics, making interpretation easier. We used the same set-up, with the same procedure to generate the semantic vectors, except that the dynamics does not account for the semantic bias, and the size of the network is now $K = 50$ ($L = 7$).

In the simulation, we computed, at each time step, the adequacy of form X with all sites of the network. We then considered, among them, only those associated with a site where the form X is in majority (i.e. all sites k such that $x_k > 0.5$). They correspond, therefore, to the adequacies with sites belonging to the semantic dominion of form X . Then, we looked at the following four quantities (Fig. 9.20):

- the maximum value of these adequacies;
- the minimum value of these adequacies;
- the mean of these adequacies;
- the adequacy with the origin site (i.e. the reservoir).

The first of these quantities shows the existence of a prototypical meaning of form X . The second one quantifies how poorly adequate variant X can be with some of the sites of its own dominion (these sites would be more easily taken over by other linguistic forms in a wider picture). The mean of the adequacies measures how strong is the semantic rule of X over its dominion. It is also indicative of bleaching if this mean tends to decrease. Finally, the adequacy with the origin site leaves a trace of the semantic change of form X resulting from its propagation over the semantic network.

A last remark is in order. We could track down the evolution of these four quantities with time, but this would be irrelevant. The semantics of the forms do not evolve directly with time, but with the reshaping of their dominion (which, itself, evolves with time). Therefore, we preferred to look at the evolution of the adequacies with respect to this dominion: on Fig. 9.20, instead of plotting $A_X^k(\{x_k(t)\})$ vs t , we plotted $A_X^k(\{x_k\})$ vs $\sum_k x_k(t)$.

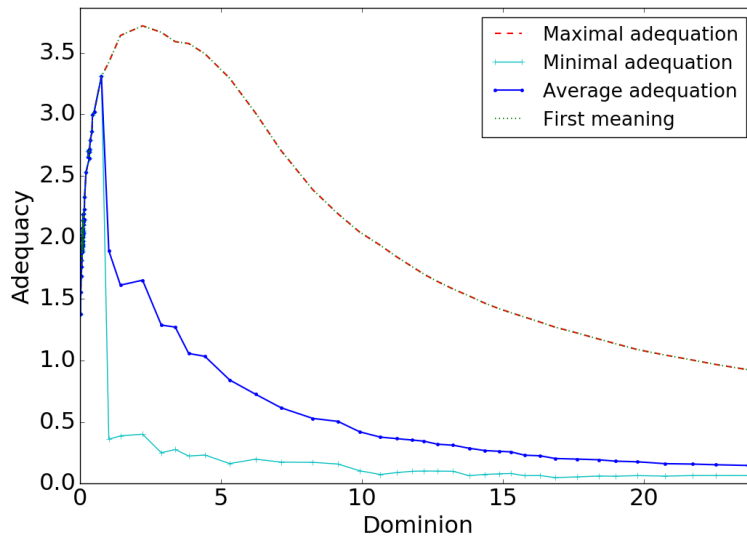


Figure 9.20: Semantic evolution of the linguistic form as it propagates over the semantic network. This semantic evolution is characterized by four quantities: the maximum and the minimum adequacies, the mean adequacy, and the adequacy with the origin site. The evolution is not tracked down in real numerical time, but in an apparent time given by the progressive extension of the semantic dominion of form X . Combined mechanism, $M = 1000$, $S = 100$, $\delta = 0.01$, $\beta = \beta^*$, $K = 50$.

We observe that, as soon as the new variant enters the cluster, its adequacy increases a lot. This is related to the fact that the norm of the projection of the semantic vector of form X in the subspace of the cluster increases suddenly as it invades the first site; so that the adequacy increases too (it scales with the norm of X). The maximal adequation remains high as long as the form does not venture too far; then it starts to decrease (after reaching a dominion of about 5). Note that, very quickly, the form comes to be associated with contexts for which it is poorly adequate. Remember that in the case of *du coup*, as soon as the grammaticalization began, the form was almost immediately associated with new, unexpected pragmatic contexts, which were not prototypical (chapter 2). As the dominion of form X extends to more and more semantically heterogeneous contexts, the average adequacy tends to the minimal one, which means that the form is mostly poorly adequate to express the semantic units it is associated with. There is, thus, a high discrepancy between what a form is used for and what it is perceived to mean. Note finally that the first site remains the one for which the form is the most adequate; this is arguably an artifact of our model, since, as we have seen, we constructed the meaning of the reservoir so as to ensure a weight at criticality. As a result, its norm was extremely small; and because the adequacy scales with the inverse of the norm of the semantic vector of a site, it is expected to be big. This is a technical issue which could be solved by adding semantic features to the the reservoir site which are not shared by anyone else in the network (that would be the ‘lexical’ features’ of the bridging context).

We illustrate this on a smaller network ($K = 25$, $L = 25$), for which 5 lexical

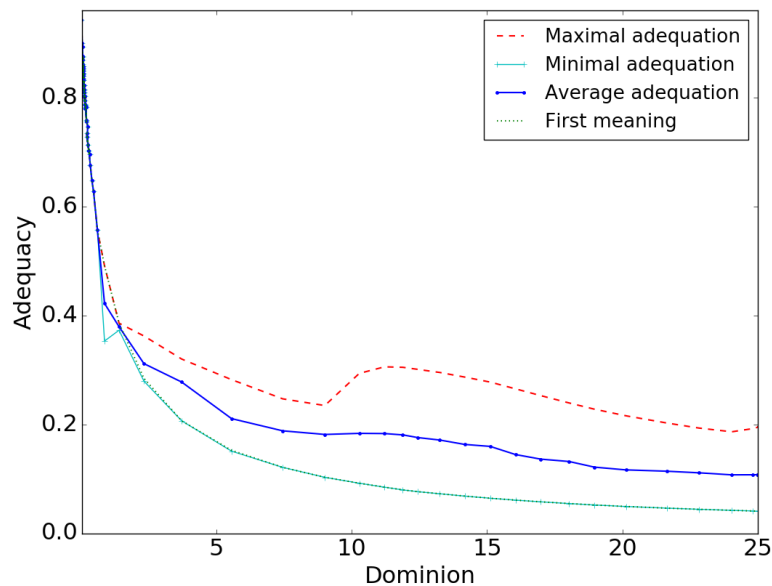


Figure 9.21: Semantic evolution of the linguistic form as it propagates over the semantic network. Lexical features has now been added for the reservoir site. Combined mechanism, $M = 1000$, $S = 105$, $\delta = 0.01$, $\beta = \beta^*$, $K = 26$.

features have been added to the reservoir (so that the norm of its semantic vector is, on average, akin to that of the grammatical sites). As the lexical features are distinct from all other features present in the network, the adequacy with the first site decreases as soon as the new variant spreads further than the first invaded site (Fig. 9.21). Interestingly, the maximal adequacy is non-monotonous: a new overall meaning orientation can emerge. More crucially, when the new variant goes beyond the first invaded site (corresponding to a dominion of slightly higher than 1), an inflexion point appears and its meaning ceases to be dominated by the original, lexical site of occupation (Fig. 9.22). This could be described as a temporally localized event of ‘grammaticalization’, even if the whole process of propagation and semantic evolution would be seen as the grammaticalization proper.

Five limitations hinder these preliminary observations. First, what we observe is extremely sensitive to the details of the semantics of the sites (much more than it is sensitive to the actual structure of the network, which determines the propagation, not the semantic evolution). Second, it might be, contrary to the set-up we presented here, that some grammatical sites present a mixture of lexical and grammatical semantic features. This could explain why the grammaticalizing form, in actual instances of grammaticalization, is still able to express efficiently its former lexical meaning. Third, we posited a situation in which the original meaning of the new variant is extremely precise and localized. Actual instances of change, conversely, hint to the fact that a semantic bleaching, on the lexical plane, usually precedes grammaticalization. This would certainly affect the start of the process. Four, we favored a semantic architecture in which the meaning is gradually regularly changing from one site to another. It

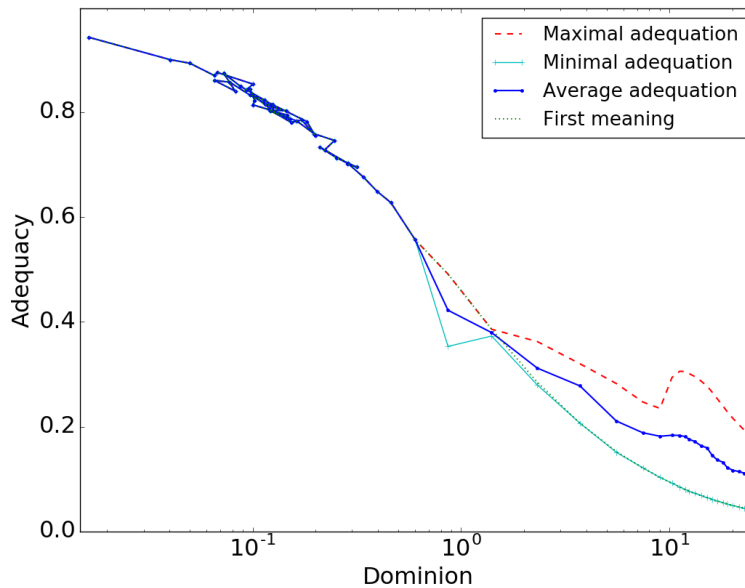


Figure 9.22: Semilog semantic evolution of the linguistic form as it propagates over the semantic network. Lexical features has now been added for the reservoir site. The fuzzy behavior of the right part of the curve corresponds to the latency step. Combined mechanism, $M = 1000$, $S = 105$, $\delta = 0.01$, $\beta = \beta^*$, $K = 26$.

would be more realistic to consider a more structured meaning organization, adding for instance small and tightly related sub-categories of meaning. Five, the process we model here involves only two forms. Yet, if variant X starts grammaticalizing, it might lose its grip over the lexical meanings it originates from, allowing for a third variant to usher it out of these sites. As a result, the adequacy of X with its newly obtained grammatical dominion will be strengthened, favoring therefore its further propagation into the grammatical plane.

From these considerations, it appears that this modeling approach cannot go much further without further hypotheses regarding the detailed organization of meaning. This joint model of linguistic replacement and semantic evolution would nevertheless allow to test, shape, and refine these hypotheses and as such help sorting out the complex semantic changes at work in the grammaticalization phenomenon.

Discussion

We can now understand a bit better a few phenomena which appears in language change. The first one is the ‘pulling’ force of change, discussed for instance by Aitchison (2013). Indeed, let us consider the fact where we have a large semantic region, entirely occupied by a single entrenched variant. On some border of this region, the new variant, X , sits in its own side of the bridge: it cannot invade the semantic domain, since the link is weak, and its meaning presents chiefly lexical features, so that it is poorly adequate to the switch context — the first grammatical context to be invaded

in the theoretical picture of Heine (2002). Yet, if the entrenched grammatical variant continues to extend further away, into other semantic domains for instance, it will progressively lose adequacy with the opposite border of its domain, precisely where the new variant is waiting. As we saw, as the semantic domain of the entrenched variant extends, its adequacy with one particular site diminishes, so that, at some point, the bias can start to act in favor of the novelty. Then, X will be able to set foot in the switch context. It will thus gain new semantic features, more adequate to the neighboring semantic units. This will allow the novelty to care with relative ease a small dominion in the semantic domain of the older variant. Therefore, the semantic expansion of one variant in one part of the semantic territory can set forth the conditions of the replacement in some other part of the network. In that sense, the semantic expansion of the old variant ‘pulls’ the new variant into the grammatical domain.

This seems to have happened in the case of *en* and *dans*. Indeed, the grammaticalization of *dans* is concomitant with an extension of the uses of *en*, which came to be used to introduce a circumstance which sets the framework of a scene, for instance:

[I]l parle du sacrifice que Silla fist à Tarente, et que **en** le faisant li apparu ou jusier d’un veel qu’il sacrefioit une couronne d’or[.]

He speaks about the sacrifice that Silla did in Taranto, and that, as he was doing so, it appeared to him, in the liver of the calf he was sacrificing, a golden crown.

DE PRESLES, Raoul, *La Cité de Dieu de Saint Augustin*, 1371-1375

Therefore, *en* was no longer at that time limited to the use of a preposition introducing a location. This, we propose, weakened its adequacy with its original meaning, and allowed *dans* to rise in its place as the standard locative preposition.

The other phenomenon is that of residue: the older variant, once evicted, is known to sometimes remain used in a few selected contexts. This we can understand: we know that the propagation of the new variant is favored by the overall loss of adequacy of the former linguistic form. Yet, as the conquest proceeds, the former variant will see its semantic dominion reduced, so that it will regain adequacy within the contexts it still populates. This can lead to the creation of a niche meaning to which the former variant becomes specialized, so that, being favored by its semantic resurge. The new variant would be then unable to dislodge it from this niche.

9.3 Summary

With these new perspectives, it appears that the further modeling of language change raises new issues, and that the propagation of the change over a cluster of related semantic units is far from guaranteed. We reached actually a conclusion similar to that of previous works (see chapter 6): the new variant can only diffuse if it benefits from a highly specific special network structure (a hierarchical, pyramidal one especially, just as in (Blythe and Croft, 2012)), or from a bias, which nature can be sociological (Blythe and Croft, 2012), inherent (Ke et al., 2008), or emerging out of a recent trend in use (Stadler et al., 2016). In the last section of this chapter, we proposed a particular bias, arising out of a vectorial representation of the meanings associated with the different semantic units. This bias evolves with the settlement of the form

over the whole network. It introduces a non-local character to the process of change, through the intermediary of the forms.

We showed, at least, that further diffusion into the network, could be the key to match the quantitative features of the empirical growth times. Though we did not manage to reach an exact agreement, we obtained preliminary results which tend to indicate that this direction of research is appropriate to solve this particular problem. Interestingly, we have seen that the increase of variation that we looked for can only occur through the intrinsic randomness, or at least disorder, of the semantic network. We have also observed that the system needed to stand right on the critical threshold for the latency to present the same statistical features as its empirical counterparts, which is reminiscent to numerous intriguing findings stemming from biological examples (Cavagna et al., 2016). However, in our case, it might be a mere coincidence.

Crucially, we also proposed a picture to relate directly and intimately meaning change, frequency change and the strength of the semantic associations. This is especially interesting, since the interplay of the three is at the core of key semantic changes, and in particular grammaticalizations. Such a unified vision, we think, had never been offered before in a modeling perspective, and it provides a promising conceptual framework to refine linguistic hypotheses, implement them, and watch them work to their outcome. In particular, the relationship between semantic bleaching and usage expansion appears as a natural consequence of the model, and does not require any more mechanism to be observed, save a proper way to represent meanings.

All of this has opened many different perspectives of research, and this network picture would benefit from further investigations. We have but pointed at several possibilities, which, I hope, speak in favor of the productivity of a modeling approach based on the cognitive organizational structure of language. As for how it would relate to sociological matters, and how the idiolectic diversity and the rich communicational aspects of language would enter and influence the picture, this is yet another story.

Conclusion

The proposed aim of this work was to investigate grammaticalization from a Statistical Physics perspective. The study of grammaticalization is filled with numerous questions: is it a specific phenomenon, or a specific outcome of a more generic phenomenon? What is its scope exactly, that is, when can we say that a linguistic form has become more ‘grammatical’? Why is it a directional process? The latter question has received considerable attention, though we did not address it in this work. On the contrary, we took directionality as granted: there is directionality in the cognitive architecture of meaning, and the directionality of grammaticalization might not lie elsewhere. A more crucial question would be: how can the cognitive features of the organization of meaning constrain and shape the change of linguistic forms? This is, specifically, the question we addressed through our model.

Physics, however, does not deal with linguistic notions on a regular basis. Physics aims to describe, explain, and classify phenomena, which are the appearance of things. Therefore, if we want to study grammaticalization from a physics perspective, we must restrain our perspective to grammaticalization as a phenomenon. The only widely attested empirical manifestation of grammaticalization is the S-curve growth in frequency of use, which is the empirical signature of virtually any language change (chapter 3), and is not specific to grammaticalization.

Most accounts of the S-curve see this empirical phenomenon as a result of a social diffusion (see chapter 6). In this view, linguistic changes do not have any empirical signature; a sharp distinction is made between the appearance of the new variant, which chiefly obeys linguistic considerations, and its societal diffusion. Only the innovation of a change is due to the language organization, while propagation is a result of the social structure, and of the complex sociological influences and differences among and between the members of a speech community (Croft, 2000).

If it were the case, the diachronic frequency profiles of linguistic forms, which constitute our main source of empirical observations on language change, would mainly reflect sociolinguistics changes and events. Yet, frequency has long been used as a tool to understand the organization of language, as the large amount of works in Corpus Linguistics can testify. Several proposed laws of change, such as the Constant Rate Hypothesis or the Bailey-Shen law of rate acceleration, also suggest that the frequency profiles, and the S-curve above all, are related to the internal, structural features of language. This does not exclude the importance of social imitation, and the two can easily be combined (Wang et al., 2004).

In this thesis, I tried to show, especially in chapters 4 and 5, that some empirical features of the S-curve and of the frequency profiles could be insightfully related to the

semantic organization of language. Once more, it does not exclude the relevance of the social diffusion in the S-curve pattern; only that the propagation is not estranged to internal, linguistic considerations. Furthermore, the long timescales of the change, and the attested fact that individuals, during a semantic replacement, can use the two variants in different, evolving proportions, hints at some sort of coherence of the process within the community of speakers: the social diffusion of the state of the change may be, in some cases, faster than the change itself.

Also, we provided, in chapter 4, empirical evidence of a latency phenomenon in frequency, before the S-curve. This phenomenon is not a propagation (actually, it can be described as the absence of a propagation), nor an innovation (the innovation already happened when latency is observed). Yet it is a part of the pattern of change. During this latency, several authors, sometimes of different generations, make use of the novelty, so that this novelty has already benefited from a social propagation. Therefore, not only does the S-curve reflect language organization in some of its features, but the social propagation can at least partially happen without being accompanied by a S-curve in frequency of use.

We have considered instead that the S-curve represents the cognitive entrenchment of a semantic change, as it is reflected in language use. Note that the semantic change does not equate to the S-curve: in fact, it is already acted at the beginning of the latency part, but not yet fully actuated into the linguistic organization. This difficulty to propagate, as we argued, is due to the fact that the language is already covering the meaning newly expressed by the expanding form. Yet, nothing entails that the linguistic means available to convey this meaning are of the same kind as the new form. A good example of this is the formulation of closed-ended questions in French: the interrogative functional meaning can be achieved 1 - through a specific word order (therefore a fully schematic construction), as in German, 2 - through the use of the interrogative marker word *Est-ce que* at the beginning of the utterance (once more a construction, [Est-ce que {Proposition} ?]), similar to the way questions are asked in Chinese (though the interrogative marker *ma* is then after the proposition), or 3 - through a specific prosody (which no longer counts as a construction), as is the case in Spanish. In most cases of grammaticalization or semantic expansion, it is therefore difficult to identify a competitor, but this does not mean that there is none.

On these grounds, we proposed a new model of semantic change, despite the already rich literature on the subject, though our own contribution could be held as a variant of the Utterance Selection Model of Baxter et al. (2006). We can ask, therefore, what does this model achieve, and why it was justified to propose a new one. The main interest of our model is that it relates the linguistic motive for the change, and its empirical unfolding: the propagation results from the innovation, and is a direct consequence of it. Therefore, it aims at drawing a direct connection between Diachronic Linguistics, and analytical or numerical modeling. However, our model, as it stands, predicts very little on the actual quantitative features of a given change. Even if we were to know the parameters on which it relies (the strength of the cognitive association between the two meanings, the size of the memory, the non-linearity parameter), we could only make predictions regarding the statistical distributions of the relevant quantities, nothing on individual changes. Furthermore, our model states that there is no correlation between the latency and the growth, so that there would be no way to infer any predictions on

the latter given the observation of the former.

In spite of this, our model has proved to achieve several successes. First, we predict, with a single mechanism, both the latency and the S-growth parts of the empirical pattern of semantic change. Second, we predict correctly the statistical features of the latency, including the shape of the distribution and the different scalings between its parameters. Third, we also successfully captured the empirical correlations between the different quantities characterizing the pattern (latency time, growth time, rate of change). It is a true possibility that other, existing models could also successfully pass this confrontation with empirical data, yet as this evidence has been put forward with the present work, they did not have the opportunity to do so until now. It is to be hoped that further empirical evidence will be gathered in the future, so that different models and mechanisms can be proposed, corroborated, or ruled out. Our work at least shows that a modeling work is able to grasp some of the very features of language change, which was certainly not granted.

Our model also offers a conceptual ground to discuss and understand better the different mechanisms involved in semantic change, and especially grammaticalization. Of particular interest to us is the view of Heine (2002) and Diewald (2002), according to which grammaticalization results from the crossing of a bridge between two semantic or constructional contexts. This view, actually, would also hold for a semantic change, though the discrepancy between the two sides of the meaning may not be as remarkable as would be the case for a grammaticalization process. In our model, we also have two contexts (which are cognitive sub-domains of memory specifically dedicated to the storage of the occurrences associated with the meanings that they respectively define), and a link between the two. Interestingly, our model shows that the strength of the link (in a sense, the semantic distance between the two sites) is actually a control parameter of a saddle-node bifurcation, which can either close or open the pathway to the semantic expansion. As grammaticalizations involve greater meaning leap than other semantic changes, they are expected to show more regularly the phenomenology associated with the vicinity of the critical bifurcation, which are the latency, and a clearer, more beautiful S-curve.

Also, we can wonder why the control parameter would increase, so as to allow the change to proceed. It may be that language changes retroactively influence the semantic organization; for instance, by inviting new analogical reasonings. We also considered the possibility for the linguistic forms to have an evolving semantic adequacy with the semantic dominion they cover — and entirely determined by this dominion — so that a form which expands on one front will lose adequacy on the other, opening the gateway for new forms to come in and replace it. We can, finally, imagine that the structure of the semantic territory does not conform to our intuitions. We might be tempted to think of it as a planar map, like an actual territory, but more probably, the semantic space is overly complex, and highly dimensional. It means that a linguistic form can get somehow lost in this space of meaning, and never find the path to grammaticalization. Indeed, if there is only one link, between only two meanings, which allows to pass from a lexical domain to a more grammatical domain of the semantic space, then it might be incredibly difficult to reach it. To locate a specific link in a grid is easy enough; to nail it within a highly dimensional manifold is, however, extremely difficult. The sparsity of the links and the complexity of the

semantic territory would jointly account for the relative rarity of grammaticalizations.

Despite these intriguing perspectives of research on semantic change and semantic structure, our model is not without shortcomings. First of all, we have been unable to reproduce the statistical features of the growth time distribution. Our attempts to overcome this failure through the extension of the model to a whole network instead of a two-sites competition have proved mitigated at best. To propagate the new variant in the semantic network, we found that specific factors should be met, be it a bias in favor of the novelty (here purely resulting from the semantic network as a greater semantic adequacy), which, in our model is more of a semantic weakening of the entrenched variant than of a semantic strengthening of the new (the novelty is pulled off rather than it pushes on), or a highly specific structure. These two kind of factors had already been explored in the literature, especially by Blythe and Croft (2012). In this regard, our model is as good as any other.

Another limitation of our model lies in its specific, semantic orientation. Indeed, it focuses almost exclusively on meaning, on the semantic territory and its structure, and linguistic forms are mere transitory inhabitants of this lasting landscape. However, linguistic forms are probably more salient to speakers than forms are. Formal analogies between forms, for instance, which are attested in language change (the productivity of a construction, e.g. the [un {N} de] construction that we investigated in chapter 4, is for instance likely to be an effect of formal analogy). We could not account for this in our modeling framework, unless they would appear as new links between existing sites. But then, the evolution of the links would be sensitive to the formal contents of the sites, which would be a completely different picture to the one we have explored in this work. Furthermore, as meaning is never readily available, and cannot be retrieved with certainty from past occurrences, it seems to part away the model from the empirical evidence. This is certainly an issue, although it would seem that we need anyhow a stable underlying semantic territory to represent semantic change.

Another matter of concern is that the S-curve is an idealization which is well-grounded and legitimate, but only accounts for part of the individual changes. Almost all frequency profiles that we presented in this work do not remain stable after the growth. Stability periods are an oddity in language change, especially if we put aside the latency periods, which are, after all, lurking changes waiting their time to leap forward. Our model, as all others in the literature, basically shows a transition between two stationary states, while language would always seem to be in a dynamical state. Furthermore, this restless dynamics of the frequency does not prevent some constructions or linguistic items to be extremely stable over time. English presents many grammatical items which shows a remarkable stability in their use: words such as *for*, *and*, *which*, *that*, have almost gone through the whole history of English language completely unchanged. French also has its lot of inexpugnable old-timers: *que*, *et*, *par*, can be traced back to Latin. This resilience would be as remarkable as the other grammatical changes, especially if there are cross-linguistic regularities in this domain. Our model of competition is therefore far to account for the language change in its whole complex and extremely diverse phenomenology, and little distinguishes itself from the rest of the literature on this matter as well.

However, different proposals which were seemingly unconnected have been unified in a single framework in our model. For instance, the bias and the structure, in (Blythe

and Croft, 2012), are two different, unrelated features of the model; in (Stadler et al., 2016), the bias results from the evolution, but requires new mechanisms altogether, such as the tracking down of two different frequencies by the speakers, and the computation of the momentum of the forms. Conversely, in our model, all quantities relevant to the diachronic evolution result from the synchronic organization of meaning and from the different settlements of the linguistic forms in this territory. The semantic bias of the linguistic form, for instance, is directly given by its perceived meaning (admittedly, we also make use of an additional mechanism with the introduction of a working memory; but it describes how the bias is accounted for in language use, and not define the bias itself). This would hopefully allow an easier comparison with empirical data in the near future, yet a lot of work must be done in that sense for this ultimate goal to be achieved.

It could be interesting at this point to provide a short outline of future research which could possibly help to make further progress in this direction. First of all, additional empirical facts are sorely needed. Indeed, there already exists a satisfying diversity of models, and all could be extended with a large array of additional features and mechanisms. Yet, they would be of no help to understand language change further, if the only thing that can be compared with is the appearance of an S-curve. If this is the only point of comparison, then almost all models stand as equal, and the favoring of one would be but a matter of taste. The Bailey-Shen law of rate acceleration is interesting, but might be more salient in phonetic changes than in semantic change. In phonetic change, the nature of the change is phonetically defined, and the diffusion is lexical, so that we can identify, even a priori, the different contexts in which the change is expected to diffuse. In semantic change, since the meaning stems from the collocates and the contexts of use, the diffusion itself retroactively affects the details of the semantic change. Yet, it would be a rewarding line of investigation, and any findings in that sense could be directly compared with the numerical outcome of a model.

Another line of empirical investigation would be to deal with frequency. Indeed, not all semantic changes are associated with the same rise of frequency; very few changes, for instance, reach the astonishing level of *dans*, which has jumped from zero frequency to become the twelfth most frequent word of French Wikipedia (according to the Wiktionnaire), in the relatively brief lapse of seven centuries. Comparatively, some semantic changes give birth to marginal forms or constructions, despite displaying a clear S-shaped frequency rise (e.g., *sur l'instant*, 'right now', the final frequency of which is equal to 2 occurrences per million). It would be relevant to investigate further these differences in frequency, and especially, to understand if these jumps of frequency make sense absolutely, or relatively. What we imply by this rather obscure formula is that the frequency height of the cliff between the start and the end of the S-curve could be in direct relation with the actual scope of the semantic expansion. This, however, assumes that all parts of language are equally frequent. Yet we can imagine the semantic territory to be layered in sub-regions associated with different ranges of frequencies; for instance, the regions of modals would be associated with a high frequency, that of encyclopedic items (names for animals, plants, technologies, natural phenomena) with a low frequency. Semantic expansions of comparable scopes in a layer with high frequency and in another layer of low frequency will be then associated

with very different frequency jumps. In that sense, the jump of frequency will make sense relatively to the associated frequency of a given layer.

It is also crucial to investigate further the semantic organization of language. We tried to offer a few perspectives along that line in chapter 5, and we reviewed some existing works in chapter 7. Though general results have been obtained (regarding the small-world structure in particular), these empirical tools have been seldom applied to make sense of individual, particular cases, with interesting exceptions such as (Ploux and Victorri, 1998) and subsequent works. Yet, this did not lead to a quantitative characterization which could be specific to each case and lead to sort out statistical regularities from the diversity of changes.

An easier way to deal quantitatively with semantic change would be to investigate how the changes of frequency are related with Zipf's law, according to which the elements of language are ranked with respect to their frequency, so that their frequency decreases as an inverse function of their rank. Indeed, Zipf's law is now to hold over any language, despite the dynamical and versatile nature of the frequencies of linguistic items. It means that the law constrains the possibilities of frequency change, and this can be investigated empirically. One could consider, for instance, how much further in rank a linguistic form gets by the end of change, and how this is related to the quantitative features of the S-curve.

Our model could be improved by relating its different parameters with attested psychological features. The size of the memory sites, the non-linearity parameter, the strength of the semantic associations between the sites, the depth of the working memory, can be all related to cognitive and psychological observations, at least presumably. By getting an idea of the range on which these parameters can realistically take their value, we would either strengthen the agreement with empirical data, or on the contrary, reject the relevance of the proposed mechanisms, in the line of what has been done for the dialectal formation of New Zealand English (Baxter et al., 2009). Regarding the physical understanding of the model, it remains to understand how exactly the semantic bias can modify the picture, and how semantic bleaching could drive the overall dynamics of the change. In the simpler version without semantic bias, one needs to understand which networks structures are compatible with a reasonable propagation of the new variant, especially in higher dimensional grid networks. Interestingly, such theoretical concerns pertain to percolation theory. The study of a multi-layer network, in which the forms of a layer are the contexts of the other (e.g., schematic constructions are contexts for their paradigm members), would be a next step further into the linguistic complexity.

By the end of this work, the time has come to go back to our initial question: what, then, have we understood about grammaticalization? In which extent could it really be a specific phenomenon? We did not find any empirical features in frequency profiles which could be held as specific of grammaticalization. True, grammaticalization often shows a latency part; but this could be the case for other semantic expansions as well (the frequency profile of the lexical *bureau*, for instance, presents a remarkable latency part). We argued that the latency phenomenon could be specific of semantic change; not of grammaticalization in particular. We could have expected grammaticalization to be associated with greater frequency leaps, but this is not the case either. One particular feature that we did not investigate is the homogeneity of uses. Indeed,

grammatical constructions are arguably topic-neutral, while lexical items depend on the overall contents of the language situation in which they appear. For instance, we do not expect to find the same lexical words in *La pyrotechnie, ou L'art du feu*, by Jacques VINCENT (1556) and in *Institution de la religion chrestienne*, by Jean CALVIN. For instance, *mercure* ('mercury') appears 145 times in the former and only 4 in the latter, while *péché* ('sin') is used 422 times by Calvin and none by Vincent. Conversely, we expect to find roughly the same grammatical constructions in both. The corpus data we gathered would allow to quantify such a phenomenon. Yet, we can instantly state that some lexical words do show the same unspecific character (e.g. *chercher, finir, trouver, comprendre, penser, temps, fois*, etc.), while some grammatical constructions are topic specific (e.g. spatial or temporal markers such as *tout autour, de loin en loin, à l'heure dite*, etc.). Still, we would expect some statistical correlation between a homogeneous repartition of the occurrences, and the grammatical level of the form.

Though we did not completely succeed in this work, we tried to point out that the specificity of the grammaticalization was the specificity of the link between the source meaning and the target meaning of the corresponding semantic change. These links may be extremely specific in the meaning space, and why exactly they stand apart from the others remains to be investigated. It could be that the grammaticalization relates different planes or layers, otherwise kept separated, of the semantic organization. Only by investigating the topology of the semantic structure could we achieve further insights on these. However, stating that grammaticalization is specific because it involves a specific link in the structure of the meaning amounts to say that language is not a specific phenomenon, but a specific result of a more generic phenomenon, which would be in this case the phenomenon of semantic expansion. Grammaticalization could however be phenomenologically specific if the links of the network are not static, but obeys a particular dynamics of formation. We have indeed put forward the idea, in chapter 2, that specific mechanisms could be involved in the creation of these semantic links. Our model, alas, did not address this particular question. In a sense then, we achieved very little in the understanding of grammaticalization, apart from outlining a general picture of semantic expansion to which one can refer, so as to narrow down the source of phenomenal specificity in the grammaticalization process.

Let us end with a little epistemological touch, by reporting the famous quote of Alfred Korzybski: "A map *is not* the territory it represents, but, if correct, it has a *similar structure* to the territory, which accounts for its usefulness. [...] If we reflect upon languages, we find that at best they must be considered *only as maps*." (Korzybski, 1994, p.58, emphasis original). We had this specific quote in mind when we referred to the space of meaning as a semantic *territory*. Indeed, a language is not just a map. It has a structure, a logic, a dynamics of its own. If it were but a map of reality, it would have no creative power; and language, in all cultures, has been the medium of tales and myths and legends. If language were but a social convention, a convenient and imperfect rendering of things, there would be no reason to deplore a language death, no reason to warn about their mass extinction that we observe nowadays. And there would be no reason either to study it from a Physics perspective. Why bother study the map, when Physics already studies what the map aims to describe? I think that this idea may have led to the particular modeling focus we observed, by which language change appeared chiefly, if not right down exclusively, as a social phenomenon. In this work,

we might have an excessive tendency to naturalize language, so as to make it a proper object of a physical study, but I, personally, truly regard language as something we live by, just as our metaphors. Clearly language could not exist without the speakers, but yet no speaker can pretend to master perfectly a language, and no speaker could invent it, in its whole complexity. Thus languages, like the heavenly meteors of yore, provides us with a rich, intriguing phenomenology, which is their own, and which we can study, observe, just as we would with fallen stars, Saint Elmo's fires, and auroras. I hope that this work would have shown, at least, how concrete and productive this seeming fantasy can eventually get.

Appendix A

List of studied semantic expansions

Table A.1: List of studied forms. The table displays, in this order: the most standard form of the researched construction; the length of the latency part, in decades; the length of the growth part, in decades; the slope of the logit transform of the growth part; the r^2 value of the linear fit of this logit transform; the average squared error; the number of occurrences associated with the form in our corpus. BUG indicates that a bug of the Frantext database has prevented from retrieving the data. NO indicates that no suitable S-curve pattern has been found in the whole frequency profile. When a form appears multiple times in the table (e.g. *à l'instar de*), it means that several non-overlapping S-curves patterns have been extracted from the frequency profile.

Table A.1

Form	Lat.	Growth	Slope	r^2	Error	# occ.
à base de	9	8	1.02	0.983	0.043	607
à bien des égards (i)	0	8	0.79	0.983	0.049	147
à bien des égards (ii)	2	6	1.57	0.984	0.041	147
à bord de	NO	NO	NO	NO	NO	1728
acabit	0	7	1.20	0.992	0.031	148
à cause de	NO	NO	NO	NO	NO	24840
à cause que	2	6	0.61	0.997	0.105	2516
à ce moment	14	13	0.46	0.990	0.080	8861
à ce propos (i)	2	7	2.22	0.988	0.011	1711
à ce propos (ii)	13	7	0.81	0.991	0.048	1711
à ce sujet	10	7	1.95	0.988	0.037	4001
à cet égard	0	6	0.85	0.993	0.087	4974
à cet instant	2	13	0.55	0.993	0.034	1198

Table A.1

Form	Lat.	Growth	Slope	r ²	Error	# occ.
à condition de	5	9	0.79	0.991	0.035	1151
à condition que (i)	11	6	1.19	0.997	0.031	1653
à condition que (ii)	4	6	0.83	0.981	0.068	1653
à contre-courant	3	11	0.75	0.995	0.021	171
à côté de	23	13	0.52	0.990	0.034	18065
à coup sûr	7	7	1.70	0.996	0.012	2546
à court terme	13	7	2.19	0.997	0.019	751
à couvert	NO	NO	NO	NO	NO	1144
actuellement	10	11	0.46	0.989	0.048	6618
à découvert	1	7	1.33	0.981	0.035	930
à défaut de	NO	NO	NO	NO	NO	1725
afin de	4	6	0.81	0.995	0.073	21833
afin que	BUG	BUG	BUG	BUG	BUG	19850
à fond de	8	6	0.91	0.987	0.059	486
à fond de train	BUG	BUG	BUG	BUG	BUG	180
à force	NO	NO	NO	NO	NO	294
à force de	NO	NO	NO	NO	NO	8178
à grand renfort de	NO	NO	NO	NO	NO	230
ainsi donc	NO	NO	NO	NO	NO	1247
à la base	NO	NO	NO	NO	NO	574
à l'accoutumée	0	8	0.99	0.988	0.027	196
à l'aide de	NO	NO	NO	NO	NO	5247
à la limite	7	11	0.73	0.983	0.076	603
à la lisière de	6	12	0.56	0.985	0.036	527
à la longue	9	7	0.57	0.987	0.115	1245
à la lumière de	0	7	1.10	0.987	0.031	1141
à la mesure de	24	9	1.14	0.988	0.056	820
à la place	22	22	0.31	0.983	0.029	5638
à la rigueur	9	8	1.14	0.983	0.049	1717
à l'écart	NO	NO	NO	NO	NO	2517
à l'écart de	6	12	0.58	0.992	0.030	854
à l'égard de	7	11	1.19	0.992	0.038	13396
à l'encontre de	9	17	0.45	0.986	0.026	1272
à l'envi	0	9	0.88	0.991	0.025	817
à l'exception de	1	8	0.81	0.995	0.080	1883

Table A.1

Form	Lat.	Growth	Slope	r^2	Error	# occ.
à l'heure actuelle	0	11	0.95	0.981	0.045	858
à l'heure dite	NO	NO	NO	NO	NO	234
à l'heure où	2	9	0.79	0.982	0.043	1779
à l'improviste	NO	NO	NO	NO	NO	1024
à l'instant	0	6	1.02	0.993	0.047	1550
à l'instar de (i)	4	10	0.61	0.981	0.044	663
à l'instar de (ii)	1	7	0.99	0.988	0.035	663
à l'insu	0	22	0.36	0.982	0.045	2776
à l'inverse	8	10	1.06	0.988	0.021	764
à l'occasion de	6	8	1.52	0.983	0.042	2032
à l'orée de	5	6	1.00	0.996	0.047	311
alors que (i)	3	7	1.01	0.983	0.044	28016
alors que (ii)	4	13	0.50	0.983	0.031	28016
à mesure de	4	8	0.71	0.990	0.087	774
à mesure que	12	7	1.74	0.993	0.017	10183
à moins que	2	8	0.90	0.981	0.059	5924
à mon avis	NO	NO	NO	NO	NO	1989
à nouveau	7	13	0.59	0.987	0.031	6039
à outrance	1	6	0.53	0.996	0.132	552
à part	0	28	0.27	0.986	0.037	12506
à part entière	0	8	1.33	0.983	0.034	180
à partir de	0	9	0.76	0.986	0.038	10996
à peine (i)	0	6	1.73	0.994	0.023	40230
à peine (ii)	0	6	1.70	0.986	0.026	40230
à peu de chose près	0	7	0.94	0.987	0.039	320
à plus d'un titre	1	7	1.02	0.995	0.031	59
à plusieurs reprises	9	7	1.22	0.994	0.023	3873
après ce	NO	NO	NO	NO	NO	101
après que	6	6	2.34	0.997	0.022	8487
après quoi	10	16	0.53	0.982	0.048	3468
après tout	NO	NO	NO	NO	NO	7741
a priori	3	9	1.21	0.985	0.046	1565
à propos	NO	NO	NO	NO	NO	1255
à propos de	0	12	0.50	0.980	0.056	9414
à proprement parler	NO	NO	NO	NO	NO	1204

Table A.1

Form	Lat.	Growth	Slope	r ²	Error	# occ.
à rebours (i)	2	6	1.09	0.994	0.070	640
à rebours (ii)	2	6	0.85	0.997	0.068	640
à qui mieux mieux	NO	NO	NO	NO	NO	247
à sa guise	0	6	1.40	0.992	0.028	1079
à son terme	1	11	0.75	0.991	0.032	359
à tel point que (i)	0	7	0.75	0.996	0.054	555
à tel point que (ii)	0	6	1.69	0.991	0.028	555
à terme	12	6	0.56	0.997	0.117	470
à titre de	5	13	0.63	0.990	0.026	1481
à tous égards	0	6	1.91	0.998	0.013	556
à tout à l'heure	0	10	0.93	0.983	0.037	280
à tout instant	NO	NO	NO	NO	NO	903
à tout moment	5	6	1.49	0.988	0.032	2262
à tout prendre	NO	NO	NO	NO	NO	480
au bord de	NO	NO	NO	NO	NO	11850
au bout de	NO	NO	NO	NO	NO	23173
au bout du compte	NO	NO	NO	NO	NO	469
au contraire	5	8	1.13	0.990	0.027	29571
au contraire de	1	8	1.14	0.989	0.028	1429
aucunefois	BUG	BUG	BUG	BUG	BUG	1248
au demeurant	0	12	0.68	0.983	0.039	1344
au dépourvu	NO	NO	NO	NO	NO	402
au détriment de	5	6	1.03	0.981	0.050	798
au dernier moment	NO	NO	NO	NO	NO	1370
au final	NO	NO	NO	NO	NO	38
au fur et à mesure	6	12	0.72	0.987	0.027	1908
au jour d'aujourd'hui	NO	NO	NO	NO	NO	87
au même moment	2	6	0.75	0.992	0.081	1437
au moment où	6	19	0.49	0.984	0.024	12729
à un moment donné	1	12	0.48	0.980	0.043	659
au passage	0	7	1.43	0.990	0.034	1754
au pire	0	6	1.63	0.994	0.026	401
au reste	0	7	1.39	0.987	0.036	4375
au sujet de	4	9	0.87	0.986	0.035	4945
au terme de	3	6	0.66	0.991	0.102	1492

Table A.1

Form	Lat.	Growth	Slope	r^2	Error	# occ.
aux troussees	NO	NO	NO	NO	NO	419
avant tout	27	10	0.91	0.986	0.029	5342
avec force	NO	NO	NO	NO	NO	324
bah	9	9	1.31	0.992	0.021	2681
bien entendu	33	18	0.44	0.984	0.042	4476
bien sûr	5	9	0.74	0.994	0.029	7997
bref	12	7	1.07	0.993	0.033	5536
brusquement	11	8	1.69	0.993	0.029	1783
carrément (i)	1	8	1.22	0.982	0.044	1207
carrément (ii)	1	7	1.55	0.982	0.042	1207
ce faisant (i)	0	6	1.88	0.992	0.027	781
ce faisant (ii)	19	8	0.64	0.994	0.059	781
ce par quoi	0	10	0.61	0.984	0.040	163
c'est alors que	6	10	0.85	0.982	0.026	3223
c'est pour le coup que	0	6	0.77	0.990	0.079	64
c'est pourquoi (i)	0	13	0.56	0.984	0.053	10994
c'est pourquoi (ii)	0	8	0.95	0.986	0.040	10994
chemin faisant	NO	NO	NO	NO	NO	641
complètement	NO	NO	NO	NO	NO	11560
compte tenu de	0	8	1.26	0.985	0.038	928
concernant	9	10	1.10	0.984	0.047	3477
considérant que	NO	NO	NO	NO	NO	191
contre mon attente	NO	NO	NO	NO	NO	102
contre toute attente	0	6	0.84	0.991	0.080	167
d'abord et avant tout (i)	0	6	0.80	0.982	0.073	62
d'abord et avant tout (ii)	1	6	1.24	0.982	0.039	62
dans ce cas	NO	NO	NO	NO	NO	4289
dans la mesure de	NO	NO	NO	NO	NO	480
dans la mesure du possible	0	11	0.62	0.988	0.050	188
dans la mesure où	NO	NO	NO	NO	NO	2753
dans le cadre de	11	6	1.27	0.988	0.069	1145
dans le même temps (i)	0	9	1.02	0.986	0.029	1217
dans le même temps (ii)	3	7	0.90	0.983	0.046	1217
dans l'ensemble (i)	0	8	0.99	0.986	0.045	1809

Table A.1

Form	Lat.	Growth	Slope	r ²	Error	# occ.
dans l'ensemble (ii)	10	7	0.82	0.981	0.068	1809
dans l'immédiat	10	9	1.10	0.984	0.033	329
dans quelque temps (i)	0	6	0.81	0.987	0.081	234
dans quelque temps (ii)	0	6	0.76	0.991	0.077	234
dans son ensemble	1	8	0.67	0.990	0.063	835
dans un autre temps	NO	NO	NO	NO	NO	143
dans un cas comme dans l'autre	1	8	1.04	0.983	0.042	111
dans une large mesure	5	8	0.66	0.994	0.062	381
dans un instant	2	10	0.87	0.981	0.035	661
dans un moment	0	15	0.47	0.984	0.037	1473
dans un premier temps	NO	NO	NO	NO	NO	229
dans tous les cas	NO	NO	NO	NO	NO	1609
d'autant plus	0	9	0.52	0.987	0.061	11584
d'autant plus que	NO	NO	NO	NO	NO	3339
d'autre part	24	12	0.64	0.982	0.045	11012
décidément	2	13	0.50	0.984	0.048	4795
de ce fait	2	8	0.66	0.993	0.105	628
de façon que	NO	NO	NO	NO	NO	1473
de fait	0	8	0.91	0.989	0.034	5018
d'année en année	NO	NO	NO	NO	NO	505
de ce côté	NO	NO	NO	NO	NO	3665
de jour en jour	NO	NO	NO	NO	NO	2217
de la part de	NO	NO	NO	NO	NO	16400
de la sorte	11	8	0.77	0.982	0.046	3752
de l'aveu de	0	17	0.34	0.986	0.054	196
de l'avis de	0	6	1.17	0.987	0.039	146
de loin	16	10	0.73	0.994	0.031	1262
de loin en loin	0	17	0.41	0.984	0.042	1348
de long en large	3	9	0.76	0.983	0.054	734
de main en main	NO	NO	NO	NO	NO	464
d'emblée	3	10	0.72	0.986	0.034	1451
de mèche	0	6	0.72	0.987	0.087	98
de mieux en mieux	0	6	1.32	0.996	0.028	445
de moins en moins	6	21	0.28	0.980	0.038	1536
de mon côté	0	14	0.71	0.981	0.043	8788

Table A.1

Form	Lat.	Growth	Slope	r^2	Error	# occ.
de mon fait	NO	NO	NO	NO	NO	467
de nulle part	24	12	0.36	0.993	0.063	289
de pair	12	7	1.04	0.983	0.050	578
de place en place	13	7	0.86	0.988	0.058	376
de point en point	0	6	0.75	0.988	0.079	247
de part en part (i)	0	7	1.49	0.995	0.019	498
de part en part (ii)	7	6	1.42	0.989	0.027	498
de part et d'autre	NO	NO	NO	NO	NO	2505
de plus en plus	0	6	1.08	0.999	0.038	18226
de près ou de loin	NO	NO	NO	NO	NO	221
de proche en proche	3	9	1.00	0.987	0.024	702
de quelque part	NO	NO	NO	NO	NO	166
des fois	5	10	0.89	0.982	0.032	1423
des fois que	2	6	0.90	0.988	0.063	182
dès l'instant	1	8	0.96	0.988	0.030	769
dès lors que	21	7	0.59	0.983	0.126	994
de sorte que	6	6	0.85	0.997	0.063	11320
de surcroît	38	9	0.91	0.988	0.032	720
de temps à autre	17	14	0.74	0.983	0.028	3547
de temps en temps	2	13	0.45	0.981	0.047	8916
de toute façon	35	7	0.94	0.996	0.066	3595
de toute manière	NO	NO	NO	NO	NO	727
de toutes façons (i)	4	6	2.04	0.989	0.039	715
de toutes façons (ii)	1	6	1.56	0.986	0.038	715
de toutes parts	0	8	1.38	0.992	0.035	4792
d'heure en heure	NO	NO	NO	NO	NO	573
d'ici là	16	9	0.60	0.985	0.096	904
dorénavant	NO	NO	NO	NO	NO	256
d'outre en outre	NO	NO	NO	NO	NO	47
du fait de	24	8	0.73	0.986	0.055	1423
du même coup	14	17	0.47	0.991	0.026	1502
du moment que	3	6	1.64	0.999	0.015	1765
d'une manière ou d'une autre	NO	NO	NO	NO	NO	320
d'une part (i)	0	8	1.13	0.985	0.038	5671
d'une part (ii)	3	7	0.80	0.982	0.061	5671

Table A.1

Form	Lat.	Growth	Slope	r ²	Error	# occ.
d'une voix claire	6	10	0.68	0.993	0.030	13511
du pareil au même	NO	NO	NO	NO	NO	92
du point de vue de	7	8	0.80	0.995	0.039	899
du reste	1	6	1.82	0.994	0.027	5510
en attendant	NO	NO	NO	NO	NO	3351
en attendant de	0	6	1.41	0.983	0.038	510
en attendant que	NO	NO	NO	NO	NO	2270
en bordure de	0	11	0.83	0.983	0.030	434
en bref	1	6	1.06	0.996	0.041	339
en ce moment (i)	5	8	1.09	0.985	0.038	12751
en ce moment (ii)	2	9	0.79	0.983	0.038	12751
en ce que	0	7	1.55	0.990	0.036	3971
en ce qui concerne	34	8	0.70	0.981	0.057	3950
en ce qui me concerne	NO	NO	NO	NO	NO	682
en considération de	NO	NO	NO	NO	NO	409
en cours de	0	14	0.54	0.984	0.074	1110
en cours de route	0	8	0.91	0.986	0.036	301
en d'autres termes (i)	0	6	0.63	0.980	0.102	1228
en d'autres termes (ii)	0	9	0.66	0.987	0.058	1228
en définitive	NO	NO	NO	NO	NO	1538
en dépit de	NO	NO	NO	NO	NO	4016
en face de	18	17	0.39	0.982	0.058	10956
en façon que	NO	NO	NO	NO	NO	48
en fait	19	11	0.57	0.985	0.047	8871
en fin de compte	27	13	0.42	0.980	0.054	1417
en gros	18	7	1.06	0.984	0.044	320
en guise de	7	10	0.93	0.982	0.035	1598
en instance de	NO	NO	NO	NO	NO	77
en l'occurrence	0	11	0.58	0.993	0.038	525
en long et en large	NO	NO	NO	NO	NO	108
en même temps	3	10	0.88	0.996	0.023	18370
en même temps que	NO	NO	NO	NO	NO	8241
en mesure de	NO	NO	NO	NO	NO	1470
en particulier (i)	2	6	1.26	0.993	0.050	8949
en particulier (ii)	17	15	0.39	0.984	0.057	8949

Table A.1

Form	Lat.	Growth	Slope	r ²	Error	# occ.
en particulier (iii)	4	7	0.80	0.987	0.072	8949
en partie	NO	NO	NO	NO	NO	5645
en passe de	NO	NO	NO	NO	NO	46
en plein	7	6	0.51	0.997	0.135	183
en plein qqch	31	10	0.86	0.982	0.035	15939
en quelque sorte	NO	NO	NO	NO	NO	6422
en sorte que	3	6	0.97	0.982	0.056	4786
en suspens	BUG	BUG	BUG	BUG	BUG	961
en tant que tel	8	6	0.91	0.997	0.065	314
entre autres	NO	NO	NO	NO	NO	4402
en vérité (i)	0	6	0.93	0.988	0.060	8194
en vérité (ii)	1	6	0.75	0.981	0.080	8194
en voie de (i)	2	6	0.57	0.996	0.200	1027
en voie de (ii)	21	22	0.31	0.980	0.052	1027
en vue de	5	6	1.31	0.995	0.026	3625
époque	7	14	0.75	0.993	0.037	32290
essentiellement	NO	NO	NO	NO	NO	5471
étant donné que	2	13	0.62	0.984	0.036	341
et après	NO	NO	NO	NO	NO	7562
excepté	5	7	0.66	0.990	0.076	5042
faute de (i)	5	7	1.76	0.990	0.024	6725
faute de (ii)	4	11	0.58	0.983	0.054	6725
faute de quoi	NO	NO	NO	NO	NO	262
force est de	NO	NO	NO	NO	NO	84
fors	BUG	BUG	BUG	BUG	BUG	4451
graduellement	3	9	1.42	0.983	0.061	827
hormis	5	6	1.01	0.987	0.070	1464
il me semble	NO	NO	NO	NO	NO	1822
il s'agit de	3	11	0.50	0.983	0.044	11558
il y a moyen	4	6	1.34	0.994	0.088	1295
j'ai l'impression	0	9	0.57	0.983	0.052	74
ja soit ce que	NO	NO	NO	NO	NO	268
je pense	5	6	1.48	0.989	0.029	4033
je suppose	0	8	1.00	0.995	0.025	1110
j'imagine	11	9	0.39	0.989	0.107	824

Table A.1

Form	Lat.	Growth	Slope	r ²	Error	# occ.
jusque là	0	6	0.81	0.980	0.075	6908
juste un	14	8	0.83	0.984	0.063	1366
l'autre jour	NO	NO	NO	NO	NO	4438
lendemain	NO	NO	NO	NO	NO	28780
le temps de	20	11	0.54	0.987	0.035	1195
liberté	2	9	0.87	0.990	0.031	46705
l'un dans l'autre	NO	NO	NO	NO	NO	69
l'un après l'autre	NO	NO	NO	NO	NO	2010
m'est avis	NO	NO	NO	NO	NO	797
nettement	0	7	0.68	0.982	0.070	6109
nommément	1	6	1.20	0.981	0.044	453
non pas tant	3	6	1.25	1.000	0.029	855
non seulement	NO	NO	NO	NO	NO	22599
non pas seulement	6	6	0.88	0.988	0.062	1605
notamment	10	8	0.55	0.991	0.091	7508
nulle part	5	12	0.57	0.980	0.045	5006
or donc	4	6	2.50	0.988	0.028	237
ouille	1	7	1.36	0.997	0.015	106
outré mesure	3	7	0.86	0.997	0.054	664
par à-coups	0	11	0.55	0.985	0.042	212
par ailleurs	27	11	0.93	0.983	0.035	2676
par avance	2	12	0.64	0.980	0.048	1265
par ce fait	0	9	1.05	0.990	0.028	101
par conséquent	6	6	1.30	0.998	0.026	12234
par contre	18	12	0.72	0.989	0.044	3014
par degrés	NO	NO	NO	NO	NO	1447
par dessus tout (i)	3	6	0.91	0.988	0.075	1433
par dessus tout (ii)	2	10	0.73	0.991	0.028	1433
pareil à	14	10	0.68	0.983	0.041	6787
par excellence	NO	NO	NO	NO	NO	1749
par faute de	4	7	0.91	0.983	0.107	353
parfois	12	17	0.56	0.983	0.030	39445
par hasard	5	8	1.20	0.994	0.023	7071
par instants	0	9	0.73	0.993	0.041	1357
par mégarde	NO	NO	NO	NO	NO	578

Table A.1

Form	Lat.	Growth	Slope	r^2	Error	# occ.
parmi d'autres	9	12	0.86	0.991	0.030	620
par moments	0	12	0.88	0.984	0.031	2774
par rapport à (i)	0	9	0.91	0.985	0.043	5290
par rapport à (ii)	3	8	0.74	0.993	0.044	5290
par surcroît	19	12	0.59	0.982	0.034	498
particulièrement	14	18	0.51	0.982	0.038	12784
par voie de	NO	NO	NO	NO	NO	976
par voie de conséquence	0	9	0.56	0.983	0.058	130
petit à petit	3	10	0.77	0.985	0.030	1547
peu à peu (i)	10	6	1.83	0.997	0.014	16450
peu à peu (ii)	1	9	0.96	0.981	0.049	16450
peu s'en faut	0	9	0.92	0.982	0.030	221
pour ainsi dire	1	13	0.77	0.982	0.038	7704
pour autant	0	13	0.79	0.994	0.015	457
pour finir	23	15	0.63	0.982	0.039	838
pour le coup	14	6	0.72	0.986	0.123	464
pour l'essentiel	0	9	0.97	0.985	0.030	284
pour le moment	1	21	0.44	0.981	0.034	2986
pour l'heure	NO	NO	NO	NO	NO	546
pour l'instant	11	14	0.63	0.988	0.027	1859
pour ma part	5	6	1.12	0.997	0.097	2744
pour peu que	0	9	0.98	0.990	0.026	2479
pour surcroît de	NO	NO	NO	NO	NO	90
pourtant que (i)	0	6	1.05	0.989	0.054	4220
pourtant que (ii)	0	6	1.70	0.989	0.024	4220
pour tout dire	2	7	1.14	0.986	0.040	655
pour un temps	NO	NO	NO	NO	NO	1333
présentement	11	6	0.88	0.981	0.070	2683
probablement (i)	2	8	1.22	0.981	0.054	8497
probablement (ii)	2	10	0.70	0.981	0.033	8497
proprement	4	6	1.05	0.989	0.044	9817
principalement	17	7	1.31	0.993	0.029	6695
progressivement	NO	NO	NO	NO	NO	2235
quand même	3	14	0.57	0.993	0.019	12171
quant à	4	11	0.61	0.989	0.036	20878

Table A.1

Form	Lat.	Growth	Slope	r ²	Error	# occ.
quant à cela	NO	NO	NO	NO	NO	91
quant à moi	4	6	0.71	0.989	0.093	4875
que dalle	NO	NO	NO	NO	NO	163
quelquefois	13	7	1.44	0.982	0.033	34408
quelque part	NO	NO	NO	NO	NO	6454
relatif à	12	10	0.57	0.980	0.081	2850
relativement à	15	7	0.54	0.992	0.119	1469
rien de plus	NO	NO	NO	NO	NO	1537
sans ambages	NO	NO	NO	NO	NO	130
sans commune mesure	2	9	1.12	0.986	0.032	112
sans crier gare	0	11	0.69	0.989	0.026	211
sans détour	9	12	0.47	0.988	0.037	467
sans façon	NO	NO	NO	NO	NO	650
sans tenir compte de	5	8	0.85	0.983	0.037	143
sauf	4	8	1.03	0.983	0.038	11138
sauf si	3	12	0.74	0.994	0.027	247
sauf que	0	13	0.58	0.985	0.026	910
selon moi	1	13	0.39	0.982	0.066	1055
si besoin est (i)	5	6	2.30	0.992	0.035	106
si besoin est (ii)	3	6	2.27	0.995	0.021	106
si bien que	4	9	0.59	0.981	0.048	4831
si ça se trouve	0	7	1.21	0.986	0.029	144
s'il en est	NO	NO	NO	NO	NO	88
si possible	22	10	0.88	0.983	0.033	760
soit dit en passant	NO	NO	NO	NO	NO	276
soudain	28	12	0.63	0.989	0.031	3498
soudainement	0	6	2.38	0.980	0.044	94
sous peu	0	6	1.83	0.993	0.028	291
sous prétexte de	NO	NO	NO	NO	NO	2341
sous prétexte que	6	6	0.69	0.997	0.112	1364
sous réserve que	NO	NO	NO	NO	NO	89
souventes fois	NO	NO	NO	NO	NO	530
spécialement	NO	NO	NO	NO	NO	3764
sur ce thème	NO	NO	NO	NO	NO	130
sur le champ	NO	NO	NO	NO	NO	5152

Table A.1

Form	Lat.	Growth	Slope	r ²	Error	# occ.
sur le moment	8	16	0.38	0.992	0.033	715
sur le sujet de	NO	NO	NO	NO	NO	292
sur le point de	2	8	0.92	0.984	0.031	3321
sur l'heure	NO	NO	NO	NO	NO	720
sur l'instant	9	14	0.53	0.987	0.044	162
un de ces jours	5	6	1.31	0.996	0.026	983
une sorte de	1	6	1.82	0.985	0.034	31306
une sorte de	2	8	1.29	0.991	0.032	31306
tandis que	BUG	BUG	BUG	BUG	BUG	39303
tant et plus	NO	NO	NO	NO	NO	155
tel quel	4	8	1.07	0.983	0.036	985
tour à tour	11	22	0.37	0.982	0.043	4480
tout à coup	3	9	1.11	0.983	0.046	20468
tout à fait	25	12	0.46	0.981	0.047	25611
tout à l'heure (i)	7	9	0.99	0.985	0.040	12853
tout à l'heure (ii)	3	9	0.88	0.995	0.022	12853
tout au long de	13	12	0.68	0.980	0.035	1363
tout au plus	11	10	0.65	0.990	0.060	2954
tout bien considéré	2	6	1.16	0.994	0.038	152
tout bien réfléchi	NO	NO	NO	NO	NO	17
tout compte fait	7	6	0.86	0.995	0.068	390
tout court	0	6	2.49	0.985	0.046	1149
tout de même	34	13	0.74	0.991	0.017	13315
tout du long	NO	NO	NO	NO	NO	302
toutefois	NO	NO	NO	NO	NO	20576
tout juste	17	13	0.45	0.982	0.038	2055
tout juste de	4	6	0.60	0.982	0.107	197
tout plein de	NO	NO	NO	NO	NO	1014
tout sauf	12	9	0.46	0.995	0.075	158
tout spécialement	4	7	0.96	0.997	0.031	164
tout un chacun	NO	NO	NO	NO	NO	260
très très	3	14	0.51	0.991	0.081	356
une espèce de	8	6	1.87	0.981	0.046	12365
un lendemain	8	8	0.56	0.993	0.072	505
un petit peu	8	14	0.60	0.983	0.048	692

Table A.1

Form	Lat.	Growth	Slope	r²	Error	# occ.
un surcroît de	NO	NO	NO	NO	NO	454
un tas de	23	9	1.30	0.990	0.023	4352
venir de	21	15	0.38	0.984	0.056	35884
vis à vis de	2	12	0.49	0.982	0.052	3384
voilà	3	13	0.75	0.984	0.033	90090
vu que	0	10	0.93	0.981	0.027	1230
zut	0	10	0.99	0.987	0.037	525

Bibliography

- Abrams, D. M. and S. H. Strogatz. Modelling the dynamics of language death. *Nature*, 424(6951):900–900, 2003.
- Aijmer, K. I think—an English modal particle. *Modality in Germanic languages: Historical and comparative perspectives*, 99(1):2009, 1997.
- Aitchison, J. *Language change: Progress or decay?* Cambridge University Press, fourth edition, 2013.
- Altmann, G., H. von Buttlar, W. Rott, and U. Strauss. A law of change in language. In Brainerd, B., editor, *Historical linguistics*, pages 104–115. Brockmeyer, 1983.
- Bailey, C.-J. N. Trying to talk in the new paradigm. *Research on Language & Social Interaction*, 4(2):312–338, 1971.
- Bailey, C.-J. N. *Variation and Linguistic Theory*. Center for Applied Linguistics, 1973.
- Baronchelli, A., V. Loreto, and L. Steels. In-depth analysis of the Naming Game dynamics: the homogeneous mixing case. *International Journal of Modern Physics C*, 19(05):785–812, 2008.
- Bascompte, J., P. Jordano, C. J. Melián, and J. M. Olesen. The nested assembly of plant–animal mutualistic networks. *Proceedings of the National Academy of Sciences*, 100(16):9383–9387, 2003.
- Baxter, G. J., R. A. Blythe, W. Croft, and A. J. McKane. Utterance selection model of language change. *Physical Review E*, 73(4):046118, 2006.
- Baxter, G. J., R. A. Blythe, W. Croft, and A. J. McKane. Modeling language change: an evaluation of Trudgill’s theory of the emergence of New Zealand English. *Language Variation and Change*, 21(2):257–296, 2009.
- Best, K.-H., E. Beöthy, and G. Altmann. Ein methodischer Beitrag zum Piotrowski-Gesetz. *Glottometrika*, 12:115–124, 1990.
- Bickerton, D. *Dynamics of a creole system*. Cambridge University Press Cambridge, 1975.
- Bisang, W. Grammaticalization and Language Contact Constructions and Positions. In Ramat, A. G. and P. J. Hopper, editors, *The Limits of Grammaticalization*, pages 13–58. John Benjamins Publishing, 1998.

- Blakemore, D. *Semantic Constraints on Relevance*. Wiley-Blackwell, 1987.
- Blakemore, D. *Relevance and linguistic meaning: The semantics and pragmatics of discourse markers*. Cambridge University Press, 2002.
- Bleys, J., M. Loetzsch, M. Spranger, and L. Steels. The grounded colour naming game. In *Proceedings of the 18th IEEE International Symposium on Robot and Human Interactive Communication (Ro-man 2009)*, 2009.
- Blythe, R. A. and W. Croft. S-curves and the mechanisms of propagation in language change. *Language*, 88(2):269–304, 2012.
- Blythe, R. A. and A. J. McKane. Stochastic models of evolution in genetics, ecology and linguistics. *Journal of Statistical Mechanics: Theory and Experiment*, 2007(July 2007), 2007.
- Boussidan, A. and S. Ploux. Using topic salience and connotational drifts to detect candidates to semantic change. In *Proceedings of the Ninth International Conference on Computational Semantics*, pages 315–319. Association for Computational Linguistics, 2011.
- Brauer, F., C. Castillo-Chavez, and C. Castillo-Chavez. *Mathematical models in population biology and epidemiology*, volume 40. Springer, 2001.
- Brugman, C. M. *The story of over: Polysemy, semantics, and the structure of the lexicon*. Taylor & Francis, 1988.
- Bybee, J. Sequentiality as the basis of constituent structure. In Givón, T. and B. F. Malle, editors, *The Evolution of Language Out of Pre-language*, pages 109–134. John Benjamins, 2002.
- Bybee, J. From usage to grammar: The mind’s response to repetition. *Language*, 82(4):711–733, 2006.
- Bybee, J. *Language, usage and cognition*. Cambridge University Press, 2010.
- Bybee, J. and S. A. Thompson. Three frequency effects in syntax. In *Proceedings of the Twenty-Third Annual Meeting of the Berkeley Linguistics Society: General Session and Parasession on Pragmatics and Grammatical Structure*, pages 378–388, 1997.
- Castellano, C., S. Fortunato, and V. Loreto. Statistical physics of social dynamics. *Reviews of modern physics*, 81(2):591, 2009.
- Cavagna, A., D. Conti, C. Creato, L. Del Castello, I. Giardina, T. S. Grigera, S. Melillo, L. Parisi, and M. Viale. Dynamic scaling in natural swarms. *arXiv preprint arXiv:1611.08201*, 2016.
- Chafe, W. *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. University of Chicago Press, 1994.
- Chafe, W. Putting grammaticalization in its place. In Wischer, I. and G. Diewald, editors, *New Reflections on Grammaticalization*. John Benjamins, 2002.

- Chambers, J. K. Patterns of Variation including Change. In Chambers, J. K., P. Trudgill, and N. Schilling, editors, *The handbook of language variation and change*, pages 349–372. Oxford: Blackwell, 2002.
- Chambers, J. K. Dialect acquisition. *Language*, pages 673–705, 1992.
- Chambers, J. K. The Canada-US border as a vanishing isogloss: the evidence of chesterfield. *Journal of English Linguistics*, 23(1-2):155–166, 1995.
- Chambers, J. K. and P. Trudgill. *Dialectology*. Cambridge University Press, 1980.
- Chambers, J. Geolinguistic patterns in a vast speech community. *Linguistica Atlantica*, 27:27–36, 2007.
- Chapin, F. S. *Cultural change*. Century, 1928.
- Chen, M. The Time Dimension: Contribution toward a Theory of Sound Change. *Foundations of Language*, 8(4):457–498, 1972.
- Corisco, A. M. H. French influence on English prepositions: A study of Ancrone Wisse. *Studia Anglica Posnaniensia*, 32:33–45, 1997.
- Coulmont, B., V. Supervie, and R. Breban. The diffusion dynamics of choice: From durable goods markets to fashion first names. *Complexity*, 21(S1):362–369, 2016.
- Croft, W. *Explaining language change: An evolutionary approach*. Pearson Education, 2000.
- Croft, W. *Radical construction grammar: Syntactic theory in typological perspective*. Oxford University, 2001.
- Croft, W. and D. A. Cruse. *Cognitive linguistics*. Cambridge University Press, 2004.
- da Silva, S. G. and J. J. Tehrani. Comparative phylogenetic analyses uncover the ancient roots of Indo-European folktales. *Royal Society open science*, 3(1), 2016.
- Dalton-Puffer, C. *The French Influence on Middle English Morphology: A Corpus-Based Study on Derivation*. Topics in English Linguistics [TiEL]. De Gruyter, 1996.
- Darwin, C. *The descent of man, and selection in relation to sex.*, volume 1. John Murray, 1871.
- Davies, M. TIME Magazine Corpus: 100 million words, 1920s-2000, 2007. Available online at <http://corpus.byu.edu/time/>.
- Davies, M. The Corpus of Contemporary American English (COCA): 520 million words, 1990-present, 2008. Available online at <http://corpus.byu.edu/coha/>.
- Davies, M. The Corpus of Historical American English (COHA): 400 million words, 1810-2009, 2010. Available online at <http://corpus.byu.edu/coha/>.
- de Saussure, F. *Cours de linguistique générale (ed. Tullio de Mauro)*. Payot, 1995.

- Degand, L. and J. Evers-Vermeul. Grammaticalization or pragmaticalization of discourse markers?: More than a terminological issue. *Journal of Historical Pragmatics*, 16(1):59–85, 2015.
- Delaurenti, B. *La puissance des mots. «Virtus verborum». Débats doctrinaux sur le pouvoir des incantations au Moyen Age*. Éditions du Cerf, 2007.
- Denison, D. Log(ist)ic and simplistic S-curves. In Hickey, R., editor, *Motives for Language Change*, pages 54–70. Cambridge University Press Cambridge, 2003.
- Desagulier, G. Visualizing distances in a set of near synonyms: rather, quite, fairly, and pretty. In Glynn, D. and J. Robinson, editors, *Polysemy and Synonymy : Corpus Methods and Applications*. John Benjamins, 2012.
- Desagulier, G. Le statut de la fréquence dans les grammaires de constructions: simple comme bonjour? *Langages*, 1:99–128, 2015.
- Diessel, H. Frequency effects in language acquisition, language use, and diachronic change. *New ideas in psychology*, 25(2):108–127, 2007.
- Diessel, H. and M. Hilpert. Frequency Effects in Grammar, 2016. URL <http://dx.doi.org/10.1093/acrefore>.
- Diewald, G. A model for relevant types of contexts in grammaticalization. In Wischer, I. and G. Diewald, editors, *New Reflections on Grammaticalization*, pages 103–120. John Benjamins Publishing Company, 2002.
- Diewald, G. Context types in grammaticalization as constructions. *Constructions*, 1(9), 2006.
- Dixon, R. Adjective classes in typological perspective. In Dixon, R. and A. Y. Aikhenvald, editors, *Adjective Classes: A Cross-Linguistic Typology*, pages 1–49. Oxford University Press, 2004.
- Dubossarsky, H., Y. Tsvetkov, C. Dyer, and E. Grossman. A bottom up approach to category mapping and meaning change. In *NetWordS*, pages 66–70, 2015.
- Dubossarsky, H., D. Weinshall, and E. Grossman. Verbs change more than nouns: a bottom-up computational approach to semantic change. *Lingue e linguaggio*, 15(1):7–28, 2016.
- Evert, S. Corpora and collocations. In *Corpus linguistics. An international handbook*, pages 1212–1248. Mouton de Gruyter, 2008.
- Fagard, B. and B. Combettes. De en à dans, un simple remplacement? Une étude diachronique. *Langue française*, 178(2):93–115, 2013.
- Fagyal, Z., S. Swarup, A. M. Escobar, L. Gasser, and K. Lakkaraju. Centers and peripheries: Network roles in language change. *Lingua*, 120(8):2061–2079, 2010.
- Farrar, F. W. Philology & Darwinism. *Nature*, 1(21):527–529, 1870.

- Ferrand, L. and F.-X. Alario. Normes d'associations verbales pour 366 noms d'objets concrets. *L'Année psychologique*, 98(4):659–709, 1998.
- Ferrer i Cancho, R. and R. V. Solé. The small world of human language. *Proceedings of the Royal Society of London B: Biological Sciences*, 268(1482):2261–2265, 2001.
- Fischer, O. On the role played by iconicity in grammaticalisation processes. In Nänny, M. and O. Fischer, editors, *Form Miming meaning: Iconicity in Language and Literature*, pages 345–374. John Benjamins, 1999.
- Fischer, O. On analogy as the motivation for grammaticalization. *Studies in Language. International Journal sponsored by the Foundation "Foundations of Language"*, 32(2):336–382, 2008.
- Fischer, O. Grammaticalization as analogically driven change. *The Oxford handbook of grammaticalization*, pages 31–42, 2011.
- Frantext. Base textuelle FRANTEXT. ATILF - CNRS & Université de Lorraine. Website: <http://www.frantext.fr>.
- Gaume, B. Balades aléatoires dans les petits mondes lexicaux. *I3 information interaction intelligence*, 4(2):39–96, 2004.
- Gavranović, V. M. Compound Nouns in Old and Middle English. *Journal for Languages & Literatures of the Faculty of Philosophy in Novi Sad/Zbornik za Jezike i Knjizevnosti Filozofskog Fakulteta u Novom Sadu*, 5, 2015.
- Geurts, B. Explaining grammaticalization (the standard way). *Linguistics*, 38(4):781–788, 2000.
- Ghanbarnejad, F., M. Gerlach, J. M. Miotto, and E. G. Altmann. Extracting information from S-curves of language change. *Journal of The Royal Society Interface*, 11(101):10–44, 2014.
- Givón, T. *On understanding grammar*. New York: Academic Press, 1979.
- Givón, T. *The Diachrony of Grammar*. John Benjamins Publishing Company, 2015.
- Goldberg, A. E. *Constructions: A construction grammar approach to argument structure*. University of Chicago Press, 1995.
- Goldberg, A. E. *Constructions at work: The nature of generalization in language*. Oxford University Press, 2006.
- Gong, T., L. Shuai, and M. Zhang. Modelling language evolution: Examples and predictions. *Physics of life reviews*, 11(2):280–302, 2014.
- Gracia-Lázaro, C., L. Hernández, J. Borge-Holthoefer, and Y. Moreno. The joint influence of competition and mutualism on the biodiversity of mutualistic ecosystems. *arXiv preprint arXiv:1703.06122*, 2017.

- Gries, S. 50-something years of work on collocations. *International Journal of Corpus Linguistics*, 18(1):137–166, 2013.
- Gries, S. and M. Hilpert. The identification of stages in diachronic data: variability-based neighbour clustering. *Corpora*, 3(1):59–81, 2008.
- Grieve-Smith, A. B. *The spread of change in French negation*. PhD thesis, The University of New Mexico, 2009.
- Hamilton, W. L., J. Leskovec, and D. Jurafsky. Diachronic word embeddings reveal statistical laws of semantic change. *arXiv preprint arXiv:1605.09096*, 2016.
- Hansen, M.-B. M. The semantic status of discourse markers. *Lingua*, 104(3-4):235–260, 1998.
- Harris, Z. *Mathematical structures of language*. John Wiley, 1968.
- Haspelmath, M. Why is grammaticalization irreversible? *Linguistics*, 37(6):1043–1068, 1999.
- Haspelmath, M. Why can't we talk to each other? *Lingua*, 110(4):235–255, 2000a.
- Haspelmath, M. The relevance of extravagance: a reply to Bart Geurts. *Linguistics*, 38(4):789–798, 2000b.
- Haspelmath, M. The European linguistic area: Standard Average European. In Haspelmath, M., editor, *Language typology and language universals. (Handbücher zur Sprach- und Kommunikationswissenschaft)*, pages 1492–1510. de Gruyter, 2001.
- Heine, B. *Cognitive foundations of grammar*. Oxford University Press, 1997.
- Heine, B. On the role of context in grammaticalization. In Wischer, I. and G. Diewald, editors, *New Reflections on Grammaticalization*, pages 83–102. John Benjamins Publishing Company, 2002.
- Heine, B. On discourse markers: Grammaticalization, pragmaticalization, or something else? *Linguistics*, 51(6):1205–1247, 2013.
- Heine, B. and T. Kuteva. *World lexicon of grammaticalization*. Cambridge University Press, 2002.
- Heine, B., T. Kuteva, and G. Kaltenböck. Discourse Grammar, the dual process model, and brain lateralization: some correlations. *Language and Cognition*, 6(01):146–180, 2014.
- Heine, B., H. Narrog, and H. Long. Constructional change vs. grammaticalization. *Studies in Language. International Journal sponsored by the Foundation "Foundations of Language"*, 40(1):137–175, 2016.
- Hilpert, M. Distinctive collexeme analysis and diachrony. *Corpus Linguistics and Linguistic Theory*, 2(2):243–256, 2006.

- Hilpert, M. *Germanic future constructions: A usage-based approach to language change*. John Benjamins Publishing, 2008.
- Hilpert, M. *Construction grammar and its application to English*. Edinburgh University Press, 2014.
- Hilpert, M. and D. C. Saavedra. The unidirectionality of semantic changes in grammaticalization: an experimental approach to the asymmetric priming hypothesis 1. *English Language & Linguistics*, pages 1–24, 2016.
- Himmelmann, N. P. Lexicalization and grammaticization: Opposite or orthogonal. In Bisang, W., N. Himmelmann, and B. Wiemer, editors, *What makes grammaticalization? : A look from its Fringes and its Components*, pages 21–42. Walter de Gruyter, 2004.
- Hoffmann, S. Are low-frequency complex prepositions grammaticalized? In Lindquist, H. and C. Mair, editors, *Corpus approaches to grammaticalization in English*, pages 171–210. John Benjamins Publishing Company, 2004.
- Hopfield, J. J. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8):2554–2558, 1982.
- Hopper, P. J. and E. C. Traugott. *Grammaticalization*. Cambridge University Press, 1993.
- Hudson, R. A. *Word grammar*. Blackwell, 1984.
- Hunston, S. *Pattern grammar*. Wiley Online Library, 2000.
- Hutchison, K. A. Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psychonomic Bulletin & Review*, 10(4):785–813, 2003.
- Idiatov, D. Antigrammaticalization, antimorphologization and the case of Tura. In Seoane, E. and M. J. López-Couso, editors, *Theoretical and Empirical Issues in Grammaticalization*, pages 151–170. John Benjamins, 2008.
- Israel, M. The way constructions grow. *Conceptual structure, discourse and language*, pages 217–230, 1996.
- Jäger, G. and A. Rosenbach. Priming and unidirectional language change. *Theoretical linguistics*, 34(2):85–113, 2008.
- Ji, H., S. Ploux, and E. Wehrli. Lexical knowledge representation with contextonyms. In *9th MT summit Machine Translation*, pages 194–201, 2003.
- Kallel, A. The loss of negative concord in Standard English: Internal factors. *Language Variation and Change*, 19(1):27–49, 2007.
- Ke, J., T. Gong, and W. S.-Y. Wang. Language change and social networks. *Communications in Computational Physics*, 3(4):935–949, 2008.

- Keller, R. Invisible-hand theory and language evolution. *Lingua*, 77(2):113–127, 1989.
- Keller, R. *On language change: The invisible hand in language*. Routledge, 1994.
- Kendall, D. G. Stochastic processes and population growth. *Journal of the Royal Statistical Society. Series B (Methodological)*, 11(2):230–282, 1949.
- Kirby, S. and J. R. Hurford. The emergence of linguistic structure: An overview of the iterated learning model. In Cangelosi, A. and D. Parisi, editors, *Simulating the evolution of language*, pages 121–147. Springer, 2002.
- Kirillov, A. B., C. Myre, and D. Woodward. Bistability, switches and working memory in a two-neuron inhibitory-feedback model. *Biological cybernetics*, 68(5):441–449, 1993.
- Koplenig, A. Using the parameters of the Zipf–Mandelbrot law to measure diachronic lexical, syntactical and stylistic changes—a large-scale corpus analysis. *Corpus Linguistics and Linguistic Theory*, ahead of print, 2015.
- Korzybski, A. *Science and sanity: An introduction to non-Aristotelian systems and General Semantics*. Institute of General Semantics, fifth edition, 1994.
- Kövecses, Z. *Metaphor in culture: Universality and variation*. Cambridge University Press, 2005.
- Kroch, A. S. Function and grammar in the history of English: Periphrastic do. In Fasold, R. W. and D. Schiffrin, editors, *Language Change and Variation*, pages 133–172. John Benjamins, 1989a.
- Kroch, A. S. Reflexes of grammar in patterns of language change. *Language variation and change*, 1(3):199–244, 1989b.
- Kroch, A. S. Modeling language change and language acquisition. 2006.
- Labov, W. *Principles of linguistic change. Vol. 1: Internal features*. Oxford: Blackwell, 1994.
- Lakoff, G. and M. Johnson. *Metaphors we live by*. Chicago University, 1980.
- Langacker, R. W. *Foundations of cognitive grammar: Theoretical prerequisites*. Stanford university press, 1987.
- LaPolla, R. J. Grammaticalization as the Fossilization of Constraints on Interpretation: Towards a Single Theory of Communication, Cognition, and the Development of Language. *unpublished*, 1997.
- Lass, R. *Historical linguistics and language change*. Cambridge University Press, 1997.
- Lehecka, T. Collocation and colligation. In *Handbook of Pragmatics Online*. Benjamins, 2015.

- Lehfeldt, R. The Normal Law of Progress. *Journal of the Royal Statistical Society*, 79 (3):329–332, 1916.
- Lehmann, C. *Thoughts on Grammaticalization*. LINCOM studies in theoretical linguistics. LINCOM Europa, 1995.
- Lenzner, T., L. Kaczmirek, and A. Lenzner. Cognitive burden of survey questions and response times: A psycholinguistic experiment. *Applied Cognitive Psychology*, 24(7): 1003–1020, 2010.
- Lévi-Strauss, C. *La pensée sauvage*. Plon, 1962.
- Lichtenberk, F. On the gradualness of grammaticalization. In Traugott, E. C. and B. Heine, editors, *Approaches to grammaticalization - Volume I*, pages 37–80. John Benjamins, 1991.
- Loreto, V., A. Baronchelli, A. Mukherjee, A. Puglisi, and F. Tria. Statistical physics of language dynamics. *Journal of Statistical Mechanics: Theory and Experiment*, 2011(04), 2011.
- Majid, A., F. Jordan, and M. Dunn. Semantic systems in closely related languages. *Language Sciences*, 49:1–18, 2015.
- Marchello-Nizia, C. Les grammaticalisations ont-elles une cause? Le cas de beaucoup, moult et tres en moyen français. *L'information grammaticale*, 87(1):3–9, 2000.
- Marchello-Nizia, C. *Grammaticalisation et changement linguistique*. De Boeck-Duculot, 2006.
- Marchello-Nizia, C. Grammaticalisation et pragmatization des connecteurs de concession en français: cependant, toutefois, pourtant. *Revue roumaine de linguistique*, 54(1-2):7–20, 2009.
- Marchello-Nizia, C. L'oral représenté: un accès construit à une face cachée des langues 'mortes'. In Guillot, C., B. Combettes, A. Lavrentiev, E. Oppermann-Marsaux, and S. Prévost, editors, *Le changement en français. Etudes de linguistique diachronique*, pages 247–264. Peter Lang, 2012.
- Marthelot, P. *Karl Bühler: Du contexte à la situation, la signification*. Armand Colin, 2012.
- Masterman, M. The Nature of a Paradigm. In Lakatos, I. and A. Musgrave, editors, *Criticism and the Growth of Knowledge*. Cambridge University Press, 1970.
- Meillet, A. *L'État actuel des études de linguistique générale, leçon d'ouverture du cours de grammaire comparée au Collège de France, lue le 13 février 1906*. Durand, 1906.
- Meillet, A. L'Évolution des formes grammaticales. *Scientia (Rivista di scienza)*, XII (24, 6), 1912.
- Melis, L. and P. Desmet. La grammaticalisation: réflexions sur la spécificité de la notion. *Travaux de linguistique*, 36:13–26, 1998.

- Meyer, D. E. and R. W. Schvaneveldt. Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of experimental psychology*, 90(2):227, 1971.
- Meyer, D. E., R. W. Schvaneveldt, and M. G. Ruddy. Activation of lexical memory. In *Meeting of the Psychonomic Society, St. Louis, 1972*.
- Michaelis, L. A. Sign-Based Construction Grammar. In Hoffmann, T. and G. Trousdale, editors, *The Oxford handbook of construction grammar*. Oxford University Press, 2013.
- Michard, Q. and J.-P. Bouchaud. Theory of collective opinion shifts: from smooth trends to abrupt swings. *The European Physical Journal B-Condensed Matter and Complex Systems*, 47(1):151–159, 2005.
- Mikolov, T., K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.
- Minkova, D. and R. Stockwell. *English words: History and structure*. Cambridge University Press, 2009.
- Morange, M. *Une histoire de la biologie*. Éditions du Seuil, 2016.
- Motter, A. E., A. P. De Moura, Y.-C. Lai, and P. Dasgupta. Topology of the conceptual network of language. *Physical Review E*, 65(6):65–102, 2002.
- Müller, F. M. *Lectures on the science of language: Delivered at the Royal Institution of Great Britain in April, May, & June 1861*. Longmans, Green, 1866.
- Müller, M. The Science of Language. *Nature*, 1(6):256–259, 1870.
- Neely, J. H. Semantic priming and retrieval from lexical memory: Evidence for facilitatory and inhibitory processes. *Memory & Cognition*, 4(5):648–654, 1976.
- Nerlove, S. and A. K. Romney. Sibling Terminology and Cross-Sex Behavior. *American Anthropologist*, 69(2):179–187, 1967.
- Nettle, D. Using social impact theory to simulate language change. *Lingua*, 108(2):95–117, 1999.
- Nevalainen, T. and H. Raumolin-Brunberg. *Historical sociolinguistics: language change in Tudor and Stuart England*. Pearson, 2003.
- Newman, M. E. and J. Park. Why social networks are different from other types of networks. *Physical Review E*, 68(3):036122, 2003.
- Nicolle, S. A relevance theory perspective on grammaticalization. *Cognitive Linguistics (includes Cognitive Linguistic Bibliography)*, 9(1):1–36, 1998.
- Nielsen, A. and J. Bascompte. Ecological networks, nestedness and sampling effort. *Journal of Ecology*, 95(5):1134–1141, 2007.

- Niyogi, P. *The computational nature of language learning and evolution*. MIT press, 2006.
- Niyogi, P. and R. Berwick. The Logical Problem of Language Change. Technical report, Massachusetts Institute of Technology Artificial Intelligence Laboratory, 1995.
- Niyogi, P. and R. Berwick. A dynamical systems model for language change. *Complex Systems*, 11(3):161–204, 1997.
- Noël, D. Diachronic construction grammar and grammaticalization theory. *Functions of language*, 14(2):177–202, 2007.
- Nowak, A., J. Szamrej, and B. Latané. From private attitude to public opinion: A dynamic theory of social impact. *Psychological Review*, 97(3):362, 1990.
- Nowak, M. A., J. B. Plotkin, and D. C. Krakauer. The evolutionary language game. *Journal of theoretical biology*, 200(2):147–162, 1999.
- Ogura, M. and W. S.-Y. Wang. Snowball Effect in Lexical Diffusion. In Britton, D., editor, *English historical linguistics 1994: papers from the 8th International Conference on English Historical Linguistics (8. ICEHL, Edinburgh, 19-23 September 1994)*, pages 119–141, 1996.
- Osgood, C. E. and T. A. Sebeok. *Psycholinguistics: A Survey Theory and Research Problems*. Indiana University Press, 1965 edition, 1954.
- Patriarca, M. and T. Leppänen. Modeling language competition. *Physica A: Statistical Mechanics and its Applications*, 338(1):296–299, 2004.
- Pearl, R. and L. J. Reed. On the rate of growth of the population of the United States since 1790 and its mathematical representation. *Proceedings of the National Academy of Sciences*, 6(6):275–288, 1920.
- Pechenick, E. A., C. M. Danforth, and P. S. Dodds. Characterizing the Google Books corpus: Strong limits to inferences of socio-cultural and linguistic evolution. *PloS one*, 10(10), 2015.
- Pemberton, H. E. The curve of culture diffusion rate. *American Sociological Review*, 1(4):547–556, 1936.
- Pintzuk, S. Variation and change in Old English clause structure. *Language variation and change*, 7(2):229–260, 1995.
- Ploux, S. Modèles et sémantique lexicale. In Kayser, D. and C. Garbay, editors, *Informatique et sciences cognitives : influences ou confluence ?*, pages 1–18. Maison des sciences de l’homme, 2011.
- Ploux, S. and B. Victorri. Construction d’espaces sémantiques à l’aide de dictionnaires de synonymes. *Traitement automatique des langues*, 39:161–182, 1998.

- Ploux, S., A. Boussidan, and H. Ji. The semantic atlas: an interactive model of lexical representation. In *Proceedings of the seventh conference of International Language Resources and Evaluation*, pages 1–5, 2010.
- Preston, D. R. Variationist perspectives on second language acquisition. In Bayley, R. and D. R. Preston, editors, *Second language acquisition and linguistic variation*, pages 1–45. John Benjamins Amsterdam, 1996.
- Prévost, S. La grammaticalisation: unidirectionnalité et statut. *Français moderne*, 71 (2):144–166, 2003.
- Prévost, S. Grammaticalisation, lexicalisation et dégrammaticalisation: des relations complexes. *Cahiers de praxématique*, 46:121–140, 2006.
- Reali, F. and T. L. Griffiths. Words as alleles: connecting language evolution with Bayesian learners to models of genetic drift. *Proceedings of the Royal Society of London B: Biological Sciences*, 277(1680):429–436, 2010.
- Redner, S. *A guide to first-passage processes*. Cambridge University Press, 2001.
- Reid, J. R. A French Word-Frequency Distribution Curve. *Language*, 20(4):231–237, 1944.
- Rey, A. et al. *Dictionnaire historique de la langue française*. Le Robert, 2010.
- Rickford, J. R. Implicational scales. In Chambers, J. K., P. Trudgill, and N. Schilling-Estes, editors, *The handbook of language variation and change*, pages 142–167. Blackwell Publishing Ltd, 2002.
- Rossari, C. Analyse contrastive, grammaticalisation et sémantique des connecteurs. In *Marqueurs de hiérarchie et grammaticalisation.*, pages 115–126. Rijksuniversiteit van Gent, 1998.
- Rossari, C. and J. Jayez. Du coup et les connecteurs de conséquence dans une perspective dynamique. *Linguisticae investigationes*, 23(2):303–326, 2000.
- Ruppin, E. and Y. Yeshurun. Recall and recognition in an attractor neural network model of memory retrieval. *Connection Science*, 3(4):381–400, 1991.
- Santorini, B. The rate of phrase structure change in the history of Yiddish. *Language Variation and Change*, 5(3):257–283, 1993.
- Sapir, E. *Language: An introduction to the study of speech*. Harcourt, Brace & World, Inc., 1949.
- Shen, Z. Exploring the dynamic aspect of sound change. *Journal of Chinese Linguistics Monograph Series*, 11:i–189, 1997.
- Sinclair, J. The lexical item. In Weigand, E., editor, *Contrastive Lexical Semantics*, pages 1–24. John Benjamins, 1998.

- Solé, R. V., B. Corominas-Murtra, and J. Fortuny. Diversity, competition, extinction: the ecophysics of language change. *Journal of The Royal Society Interface*, page rsif20100110, 2010.
- Sperber, D. and D. Wilson. *Relevance: Communication and Cognition*. Blackwell, 1986.
- Stadler, K., R. A. Blythe, K. Smith, and S. Kirby. Momentum in Language Change. *Language Dynamics and Change*, 6(2):171–198, 2016.
- Stauffer, D. and C. Schulze. Microscopic and macroscopic simulation of competition between languages. *Physics of Life Reviews*, 2(2):89–116, 2005.
- Steels, L. A self-organizing spatial vocabulary. *Artificial life*, 2(3):319–332, 1995.
- Steels, L. Modeling the cultural evolution of language. *Physics of Life Reviews*, 8(4): 339–356, 2011.
- Steels, L. *The talking heads experiment: origins of words and meanings*. Language Science Press, 2015.
- Strauss, U. and G. Altmann. Change in language, 2007. http://lql.uni-trier.de/index.php/Change_in_language.
- Sweetser, E. E. Grammaticalization and semantic bleaching. In *Annual Meeting of the Berkeley Linguistics Society*, volume 14, pages 389–405, 1988.
- Talmy, L. *Toward a cognitive semantics*, volume 1. MIT press, 2000.
- Thompson, S. A. A discourse approach to the cross-linguistic category ‘adjective’. In Corrigan, R., F. R. Eckman, and M. P. Noonan, editors, *Linguistic categorization*, pages 245–266. John Benjamins Publishing, 1989.
- Tolkien, J. *The Lost Road and Other Writings*. Grafton, 1992.
- Tolkien, J. *On Fairy-stories*. HarperCollinsPublishers, 2014.
- Traugott, E. C. Toward a coherent account of grammatical constructionalization. In Barðdal, J., S. Gildea, L. Sommerer, and E. Smirnova, editors, *Diachronic construction grammar. Amsterdam: John Benjamins*, pages 51–80. John Benjamins Publishing Company, 2015.
- Traugott, E. C. Literacy and language change: The special case of speech act verbs. *Interchange*, 18(1-2):32–47, 1987.
- Traugott, E. C. Pragmatic strengthening and grammaticalization. In *Annual Meeting of the Berkeley Linguistics Society*, volume 14, pages 406–416, 1988.
- Traugott, E. C. On the rise of epistemic meanings in English: An example of subjectification in semantic change. *Language*, pages 31–55, 1989.

- Traugott, E. C. Constructions in grammaticalization. In *The handbook of historical linguistics*, pages 624–647. Blackwell Publishing, 2003.
- Traugott, E. C. and E. König. The semantics-pragmatics of grammaticalization revisited. In *Approaches to grammaticalization - Volume I*, pages 189–218. John Benjamins Publishing, 1991.
- Traugott, E. C. and G. Trousdale. *Constructionalization and constructional changes*. Oxford University Press, 2013.
- Tria, F., V. Loreto, V. D. P. Servedio, and S. H. Strogatz. The dynamics of correlated novelties. *Scientific reports*, 4:5890, 2014.
- Trousdale, G. On the relationship between grammaticalization and constructionalization. *Folia Linguistica*, 48(2):557–578, 2014.
- Trudgill, P. *New-dialect formation: The inevitability of colonial Englishes*. Oxford University Press, 2004.
- Van Kampen, N. G. *Stochastic processes in physics and chemistry*. Elsevier, 1992.
- Victorri, B. Homo Narrans: Le rôle de la narration dans l'émergence du langage. *Langages*, pages 112–125, 2002.
- Victorri, B. Continu et discret en sémantique lexicale. *Cahiers de praxématique*, 42: 75–94, 2004.
- Vogt, P. and B. De Boer. Language evolution: Computer models for empirical data. *Adaptive behavior*, 18(1):5–11, 2010.
- Waltereit, R. Grammaticalization and discourse. In Heine, B. and H. Narrog, editors, *Handbook of Grammaticalization*, pages 413–423. Oxford University Press, 2011.
- Wang, W. S.-Y. Competing changes as a cause of residue. *Language*, 45(1):9–25, 1969.
- Wang, W. S.-Y. and C.-C. Cheng. Implementation of phonological change: The Shaung-feng Chinese case. In *The lexicon in phonological change*, pages 148–158. Walter de Gruyter, 1977.
- Wang, W. S.-Y. and J. W. Minett. The invasion of language: emergence, change and death. *Trends in ecology & evolution*, 20(5):263–269, 2005.
- Wang, W. S.-Y., J. Ke, and J. W. Minett. Computational studies of language evolution. *Computational linguistics and beyond*, pages 65–106, 2004.
- Wang, X.-J. Probabilistic decision making by slow reverberation in cortical circuits. *Neuron*, 36(5):955–968, 2002.
- Weinreich, U., W. Labov, and M. I. Herzog. *Empirical foundations for a theory of language change*. University of Texas Press, 1968.

- Wierzbicka, A. Semantic Complexity: Conceptual Primitives and the Principle of Substitutability. *Theoretical linguistics*, 17(1-3):75–98, 1991.
- Wierzbicka, A. *Semantics: Primes and universals: Primes and universals*. Oxford University Press, UK, 1996.
- Wierzbicka, A. *Understanding cultures through their key words: English, Russian, Polish, German, and Japanese*. Oxford University Press, 1997.
- Williams, J. M. *Origins of the English Language: A Social and Linguistic History*. The Free Press, 1975.
- Yanovich, I. Genetic drift explains Sapir’s “drift” in semantic change. In *The Evolution of Language: Proceedings of the 11th International Conference (EVOLANG11)*, pages 321–329, 2016.

Résumé

Cette thèse se propose d'étudier la grammaticalisation, processus d'évolution linguistique par lequel les éléments fonctionnels de la langue se trouvent remplacés au cours du temps par des mots ou des constructions de contenu, c'est-à-dire servant à désigner des entités plus concrètes. La grammaticalisation est donc un cas particulier de remplacement sémantique. Or, la langue faisant l'objet d'un consensus social bien établi, il semble que le changement sémantique s'effectue à contre-courant de la bonne efficacité de la communication ; pourtant, il est attesté dans toutes les langues, toutes les époques et, comme le montre la grammaticalisation, toutes les catégories linguistiques. Dans cette thèse, nous étudions d'abord le phénomène de grammaticalisation d'un point de vue empirique, en analysant les fréquences d'usage de plusieurs centaines de constructions du langage connaissant une ou plusieurs grammaticalisations au cours de l'histoire de la langue française. Ces profils de fréquence sont extraits de la base de données de Frantext, qui permet de couvrir une période de sept siècles. L'augmentation de fréquence en courbe en S concomitante du remplacement sémantique, attestée dans la littérature, est confirmée, mais aussi complétée par l'observation d'une période de latence, une stagnation de la fréquence d'usage de la construction alors même que celle-ci manifeste déjà son nouveau sens. Les distributions statistiques des observables décrivant ces deux phénomènes sont obtenues et quantifiées. Un modèle de marche aléatoire est ensuite proposé afin de reproduire ces deux aspects du changement. La latence s'y trouve expliquée comme un phénomène critique, au voisinage d'une bifurcation nœud-col. Une extension de ce modèle articulant l'organisation du réseau sémantique et les formes possibles de l'évolution est ensuite discutée.

Mots Clés

Changement linguistique ; étude sur corpus ; grammaticalisation ; marche aléatoire ; criticalité ; réseaux.

Abstract

This work aims to study grammaticalization, the process by which the functional items of a language come to be replaced with time by content words or constructions, usually providing a more substantial meaning. Grammaticalization is therefore a particular type of semantic replacement. However, language emerges as a social consensus, so that it would seem that semantic change is at odds with the proper working of communication. Despite of this, the phenomenon is attested in all languages, at all times, and pervades all linguistic categories, as the very existence of grammaticalization shows. Why it would be so is somehow puzzling. In this thesis, we shall argue that the components on which lies the efficiency of linguistic communication are precisely those responsible for these semantic changes. To investigate this matter, we provide an empirical study of frequency profiles of a few hundreds of linguistic constructions undergoing one or several grammaticalizations throughout the French language history. These frequencies of use are extracted from the textual database Frantext, which covers a period of seven centuries. The S-shaped frequency rise co-occurring with semantic change, well attested in the existing literature, is confirmed. We moreover complement it by a latency part during which the frequency does not rise yet, though the construction is already used with its new meaning. The statistical distribution of the different observables related to these two phenomenal features are extracted. A random walk model is then proposed to account for this two-sided frequency pattern. The latency period appears as a critical phenomenon in the vicinity of a saddle-node bifurcation, and quantitatively matches its empirical counterpart. Finally, an extension of the model is sketched, in which the relationship between the structure of the semantic network and the outcome of the evolution could be discussed.

Keywords

Language change; corpus-based studies; grammaticalization; random walk; criticality; networks.