

Habilitation à diriger des recherches. Pour une romanistique moderne : la microtypologie.

Benjamin Fagard

▶ **To cite this version:**

Benjamin Fagard. Habilitation à diriger des recherches. Pour une romanistique moderne : la microtypologie. : Mémoire de synthèse et Volume de publications.. Linguistique. Université François-Rabelais de Tours, 2015. tel-01241740

HAL Id: tel-01241740
https://shs.hal.science/tel-01241740

Submitted on 15 Dec 2015

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.

UNIVERSITÉ FRANÇOIS-RABELAIS

DE TOURS

Année Universitaire : 2015-2016

HABILITATION A DIRIGER DES RECHERCHES

Discipline : Sciences du langage

présentée et soutenue publiquement

par :
Benjamin Fagard

Le 9 décembre 2015

POUR UNE ROMANISTIQUE MODERNE : LA MICROTYPOLOGIE

RAPPORTEURS

Mme Christiane	MARCHELLO-NIZIA	Professeur Emérite	ENS Lyon
M. Daniel	PETIT	Professeur des Universités	ENS Paris
M. Mario	SQUARTINI	Professeur	Université de Turin, Italie

JURY

M. Gabriel	BERGOUNIOUX	Professeur des Universités	Université d'Orléans
M. Ekkehard	KÖNIG	Professeur Emérite	Université de Berlin, Allemagne
Mme Béatrice	LAMIROY	Professeur	Université de Louvain, Belgique
Mme Christiane	MARCHELLO-NIZIA	Professeur Emérite	ENS Lyon
M. Daniel	PETIT	Professeur des Universités	ENS Paris
Mme Paola	PIETRANDREA	Maître de Conférences, HDR	Université de Tours

Remerciements

Je tiens à remercier ici, en premier lieu, Paola Pietrandrea, pour son enthousiasme, son écoute et ses conseils.

Je voudrais également remercier les rapporteurs, Christiane Marchello-Nizia, Daniel Petit et Mario Squartini, ainsi que les autres membres du jury, Béatrice Lamiroy, Gabriel Bergougnoux et Ekkehard König, d'avoir accepté cette tâche assez ingrate.

Je remercie aussi les quelques collègues, amis et proches qui ont bien voulu relire des passages (et même l'intégralité pour certains, qui sauront se reconnaître : merci !) de ma synthèse : Walter de Mulder, Quentin Feltgen, Frédéric Landragin, Marie, Remi et Jacotte.

Pour leurs conseils réguliers, depuis le début de mon parcours scientifique, je tiens à remercier tout particulièrement Fabio Zinelli, Christiane Marchello-Nizia, Bernard Combettes, Michel Charolles, Bernard Victorri, Liesbeth Degand. Un pensée, également, pour Peter Koch, avec qui je préparais depuis 2010 un projet de recherche qui devait m'amener à passer un à deux ans à Tübingen pour travailler, avec lui, sur l'expression du mouvement dans les langues romanes – et le sarde en particulier.

Mes remerciements aussi à ceux, nombreux, collègues, amis et proches, qui m'ont écouté lors de mes complaintes rituelles, ou conseillé sur quelque point précis de mes recherches : Olivier Bertrand, Julie Glikman, Dejan Stosic, Karolina Krawczak, Sophie Prévost, Frédérique Becquet, Claudio Iacobini, Laure Sarda, Alexandru Mardale, Jean-Michel Fortis, Catherine Bolly, Kim-Teck Ly, Laurent Barberon, Roberta Mastrofini, Jean-Philippe Busson, Céline Guillot, Mathilde Fagard, Thierry Poibeau, Catherine Fuchs, Massimo Cerruti, Denis Vigier.

Pour leur soutien et leurs nombreuses contributions en conseils, blagues et plaques de chocolat, je remercie enfin les collègues, mais aussi les doctorants, post-doctorants et stagiaires du Lattice. Et je ne voudrais pas oublier les collègues et équipes qui m'ont accueilli ici et là : feu le département de linguistique de Rome 3, les collègues de Turin, ceux du CECC à Lisbonne, du romanisches Seminar de Tübingen, de Louvain-la-Neuve, de Lund. Pour nos discussions sur d'autres directions qu'aurait pu prendre cette habilitation, je tiens également à remercier Alain Lemaréchal.

Pour finir, je me dois de remercier Adèle, Robin, Jules et Emile d'avoir attendu patiemment (parfois) le temps que je finisse une page d'un article, une diapositive d'une présentation ou un chapitre de mon mémoire de synthèse… et Sylvie Bordin, sans qui rien ne se fait ici-bas !

Note liminaire

J'ai glosé les exemples, sauf pour le français et les langues romanes (roumain mis à part) ; je propose également une traduction, sauf pour le français classique, moderne et contemporain. Pour les exemples français, je précise la période, à titre indicatif (ancien français, moyen français, français préclassique et classique, prémoderne et moderne, contemporain) ; pour les langues romanes, je distingue uniquement les périodes médiévale et moderne.

Pour faciliter la consultation de mes publications, je les ai classées en six tomes avec numérotation continue (fichiers PDF joints), et j'ai indiqué entre crochets, dans ce mémoire, les renvois à mes publications incluses dans ces six tomes, avec la ou les page(s) concernées, comme suit : [I : 1] *renvoi à la première page du premier tome*.

Principales abréviations utilisées dans ce mémoire :

Gloses

1	première personne
2	deuxième personne
3	troisième personne
ACC	accusatif
DAT	datif
DEI	adverbe ou préverbe déictique
DET	déterminant défini
F	féminin
GEN	génitif
IMP	impératif
IND	déterminant indéfini
LOC	locatif
M	masculin
N	nom
NOM	nominatif
NT	neutre
PART	participe
PL	pluriel
PRS	présent (par défaut, à l'indicatif)
PST	passé (par défaut, à l'indicatif)
REL	pronom relatif
SG	singulier
V	verbe

Autres abréviations

BFM	Base du français médiéval
CFPP	Corpus du français parlé parisien
CdP	Corpus do Português
DMF	base du *Dictionnaire du moyen français*
Fm	données récoltées à l'aide de l'outil d'élicitation sur le mouvement fictif (Blomberg 2014)
fr	français
gm	allemand
MDO	marquage différentiel de l'objet
NCA	Nouveau Corpus d'Amsterdam
pl	polonais
pm	piémontais
sw	suédois
th	thai
Traj	données récoltées à l'aide de l'outil d'élicitation sur la trajectoire (Ishibashi et al. 2006)

Sommaire

0 Prologue 5

I Prolégomènes méthodologiques 7

1 Questions de méthode : le corpus comme outil de base 9
1.1 Avantages des corpus 10
1.2 Conditions et limites du recours aux corpus 15

2 Eclairages 23
2.1 L'approche diachronique 23
2.2 Approche contrastive et linguistique romane 26
2.3 L'approche typologique 31
2.4 La linguistique cognitive 35

II Questionnements (et quelques débuts de réponse) 39

3 Le localisme 41
3.1 Panorama des études actuelles et antérieures – Le localisme : introduction 41
3.2 Etudes sur le localisme 44
3.3 Résultat : un localisme revisité 62

4 L'espace dans la langue 65
4.1 Les relations spatiales statiques 67
4.2 L'expression du déplacement 72
4.3 Le mouvement fictif 85
4.4 L'asymétrie source/but 93

5 La grammaticalisation 99
5.1 Panorama des études actuelles et antérieures – Définition et état de l'art 101
5.2 Etudes menées sur la grammaticalisation 105
5.3 Grammaticalisation et syntaxe 114
5.4 Sources de la grammaticalisation et fin anticipée du processus 115
5.5 Conclusions sur la vitesse des changements 120

III Suites 123

6 Projet de recherche 125
6.1 De la synthèse au programme 125
6.2 Une nouvelle linguistique 126
6.3 Un exemple 129
6.4 Interdisciplinarité 130
6.5 Applications 131

Bibliographie 135
Corpus et bases de données 156

Table des matières 135

If you know what you're going to find, you're just not doing research.
(Peter Lawrence, 2011 : 26)

Il ne faut pas que la description du langage humain nous fasse oublier l'homme, qui en est à la fois le principe et la fin, puisque tout, dans le langage, procède de lui et s'adresse à lui.
(Michel Bréal, 1866 : 67)

Der Mensch ist nur Mensch durch Sprache
« l'homme n'est homme que par le langage »
(Wilhelm von Humboldt, 1903-1936 [1820], t. IV : 15)[1]

0 Prologue

Le mémoire de synthèse de l'habilitation à diriger des recherches est un exercice quelque peu étrange, mais qui oblige à prendre du recul et permet ainsi de remettre l'ensemble des travaux effectués dans une même perspective, de tester leur cohérence. Celle-ci est parfois difficile à trouver et, pour qualifier celle de mes propres productions scientifiques et des recherches que j'ai encadrées ou dirigées, j'hésiterais (par déformation professionnelle, cela va sans dire, et pour être fidèle à mes thématiques de recherche favorites), entre RESSEMBLANCE DE FAMILLE et STRUCTURE PROTOTYPIQUE. Ma fascination pour la variation en langue(s), ses limites et ses origines m'a en effet amené à conduire un grand nombre d'études ponctuelles dont on pourrait dire que le prototype est une série d'études sur le **localisme**, sous la forme d'**études diachroniques sur corpus dans plusieurs langues romanes** cherchant à répondre à la question suivante : **dans quelle mesure l'évolution linguistique est-elle guidée par des contraintes liées aux modes de perception propres à l'être humain ?** Plus précisément, puisque nous parvenons à peu près correctement à rendre compte des changements linguistiques *a posteriori*, pourquoi est-il si difficile de les prévoir ? À partir de ces questions de fond se déclinent un certain nombre d'études de cas, dans plusieurs directions, en fonction des résultats, des corpus disponibles, des collaborations, des projets : grammaticalisation de prépositions, d'adverbes, de verbes, d'autres classes de mots, évolution sémantique.

Dans ce mémoire, je pars de considérations méthodologiques (§1) et théoriques (§2), en essayant de replacer chaque pan de mes travaux dans le contexte des études actuelles (et parfois plus anciennes). Je suis à grands traits mon cheminement d'un intérêt pour l'histoire de la langue et donc la VARIATION DIACHRONIQUE d'abord vers la romanistique – qui combine histoire et comparaison en synchronie – puis vers la typologie – qui s'intéresse d'abord à la

[1] Cité (et traduit) par Trabant (2008 : 240, 2014 : 584).

comparaison synchronique, et est donc focalisée sur la VARIATION DIATOPIQUE (au sens très large de potentiels linguistiques). Le fil rouge de mes recherches est l'étude de l'espace (perçu) et de sa place dans la langue. Je détaille mes recherches dans les sections suivantes (§3 à §5). Je montre en fin de compte (§6) que ces recherches préfigurent le projet dont je présente l'ébauche dans le dernier chapitre, et que je conçois comme une façon de compenser certains biais actuels de la linguistique, combinant l'étude de deux types de variation (diachronique et diatopique) : la *microtypologie*. Ce projet étant avant tout une proposition de méthode, la boucle est bouclée.

I Prolégomènes méthodologiques

> Armchair linguistics does not have a good name in some linguistics circles. A caricature of the armchair linguist is something like this. He sits in a deep soft comfortable armchair, with his eyes closed and his hands clasped behind his head. Once in a while he opens his eyes, sits up abruptly shouting, 'Wow, what a neat fact!', grabs his pencil, and writes something down. Then he paces around for a few hours in the excitement of having come still closer to knowing what language is really like. (There isn't anybody exactly like this, but there are some approximations.)
>
> Corpus linguistics does not have a good name in some linguistics circles. A caricature of the corpus linguist is something like this. He has all of the primary facts that he needs, in the form of a corpus of approximately one zillion running words, and he sees his job as that of deriving secondary facts from his primary facts. At the moment he is busy determining the relative frequencies of the eleven parts of speech as the first word of a sentence versus as the second word of a sentence. (There isn't anybody exactly like this, but there are some approximations.)
>
> These two don't speak to each other very often, but when they do, the corpus linguist says to the armchair linguist, 'Why should I think that what you tell me is true?', and the armchair linguist says to the corpus linguist, 'Why should I think that what you tell me is interesting?' (Fillmore, 1992 : 35)

1 Questions de méthode : le corpus comme outil de base

Je commence par la méthode parce que c'est pour moi un point fondamental, et un des points communs entre mes différents travaux (publications, présentations, projets), dont la quasi-totalité repose sur la linguistique de corpus. La rigueur de la méthode conditionne à mes yeux la validité des résultats, et donc leur intérêt même. Parallèlement à mes questionnements théoriques, je mène en effet depuis le début de ma carrière une réflexion sur la méthodologie, et chacune de mes études constitue à la fois une recherche théorique et une occasion de questionner mon approche méthodologique, et me permet à l'occasion de tester une nouvelle technique – ce qui explique également mon intérêt pour des corpus construits selon différentes méthodologies : romans, théâtre, presse, textes techniques, corpus oraux, élicitation, enquêtes de terrain, questionnaires. Mes principales exigences de ce point de vue portent sur les points suivants :

(a) La référence constante à des données solides, ce qui impose selon moi le recours aux corpus (préexistants ou bien créés pour l'occasion), dans la grande majorité des cas.
(b) Un maximum de précautions dans le passage des résultats spécifiques d'une étude à des conclusions générales, ce qui a pour corollaire :
    a. Une attention particulière à la constitution et à l'utilisation des corpus ;

b. L'utilisation attentive et prudente des statistiques[2] ;
c. La répétition d'études parfois très proches pour vérifier que les conclusions sont les mêmes ;
d. La réplication de tests semblables sur plusieurs langues, plusieurs périodes, avant de se poser la question d'une possible généralisation des résultats.

(c) L'utilisation raisonnée des données linguistiques – autrement dit, j'essaie de travailler autant que possible sur les langues que je connais ;
(d) Le recours à une théorie globale, ouverte, évolutive et non-aprioriste, c'est-à-dire à des concepts cohérents mais compréhensibles (permettant le dialogue avec d'autres chercheurs), et qui peuvent, le cas échéant, être modifiés pour s'adapter aux données, voire abandonnés – plutôt que l'inverse.

Ces points sont repris et développés dans les sous-sections qui suivent (§1.1 à 1.3).

1.1 Avantages des corpus

1.1.1 Le corpus, outil indispensable pour le diachronicien

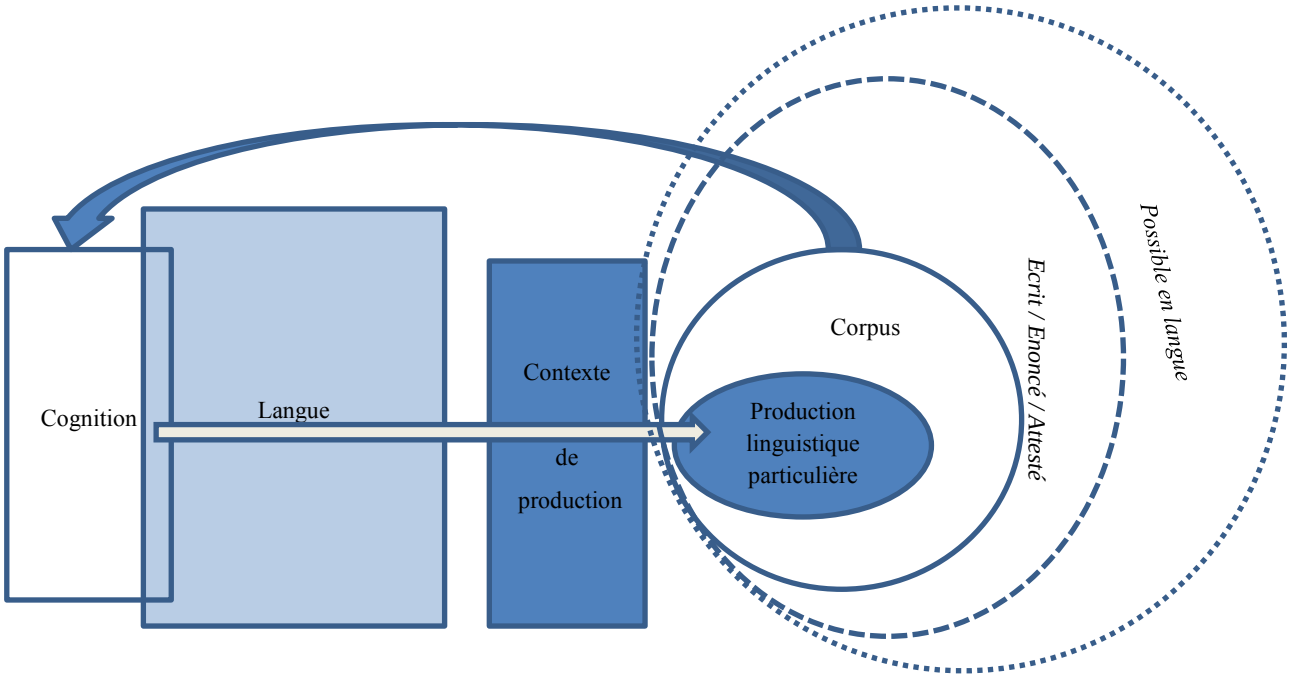
Pour un diachronicien de formation, l'utilisation de corpus est presque une évidence. En effet, un problème fondamental de l'approche diachronique est qu'il est par définition impossible d'avoir des locuteurs à consulter, à moins de travailler en diachronie très courte, sur la langue moderne, avec une étude longitudinale ou une diachronie artificielle avec des locuteurs d'âges différents. Pour les linguistes travaillant sur des états de langue disparus, « le corpus est indispensable, puisqu'il conditionne l'existence même de l'objet à décrire » (Prévost 2005 : 147). Si la consultation des locuteurs pose elle aussi certains problèmes, si l'introspection elle-même ne permet pas toujours de décrire correctement le phénomène étudié (cf. Bresnan 2007), l'impossibilité même du recours au locuteur est indéniablement un handicap pour le linguiste. On ne peut pas apprendre l'ancien français (ou le moyen français, le français classique…) comme on apprend l'italien, l'allemand ou le tchèque. Pour apprendre une langue moderne, et acquérir un niveau de maîtrise un tant soit peu proche des locuteurs natifs, on peut aller sur place, s'immerger, prendre un bain linguistique total, s'exposer aux mille variations qui font la richesse d'une langue, observer ses mutations en fonction du lieu, du locuteur, du niveau de langue, de la modalité (écrite ou orale), etc. Pour la langue ancienne, on ne dispose que de témoignages extrêmement limités : des textes littéraires, des documents, des lettres. Il serait vain de croire que l'on peut, à partir de cela, avoir une idée complète du système linguistique sous-jacent.

Une conséquence directe de cet état de fait est que, comme le note Marchello-Nizia (1995 : 22), les médiévistes ne possèdent pas de compétence de production. Lorsqu'un chercheur se permet de recréer de l'ancien français, c'est toujours avec la conscience du danger que cela

[2] Un point sur lequel il me reste une grande marge de progression, qui s'explique en partie par une certaine circonspection à l'endroit des statistiques. On pensera à ce propos à l'adage suivant, dont la paternité est couramment attribuée (à tort) à Mark Twain : « There are three kinds of lies: lies, damned lies, and statistics ».

représente : ainsi, Banniard (1993 : 50) parle bien d' « oser » traduire le latin de la vie de saint Riquier en très ancien français[3]. Habert, Nazarenko & Salem (1997 : 132-133) notent pour leur part que « [l]'érudit contemporain ne saurait affirmer : cet énoncé n'est pas acceptable » (Fagard 2008b [VI : 936]). Sans aller aussi loin, on admettra que le jugement moderne ne peut être tranché et que, devant une construction inattendue ou un hapax, le linguiste ou philologue ne peut faire appel à un quelconque « sentiment du locuteur », comme pour les langues modernes.

Le développement des corpus électroniques, grâce notamment aux progrès des plates-formes d'annotation (voir par exemple Heiden 2010, Heiden et al. 2010, 2012), constitue donc une chance pour les linguistes de compenser en partie ce manque. En effet, ils permettent d'avoir à notre disposition des outils de mesure objectifs, et peuvent ainsi apporter un complément à la subjectivité du chercheur, et remédier à nos propres limitations quand il s'agit de nous prononcer – y compris sur notre langue ou même sur nos propres pratiques langagières. Le fait qu'il existe désormais pléthore de corpus, diachroniques et synchroniques, écrits et oraux, formels et informels, etc. permet de plus de limiter le défaut majeur du travail sur corpus, à savoir que – comme le graphique (1) l'illustre très schématiquement – la langue est un système abstrait dont le corpus est un produit : on ne peut pas réduire la langue au corpus. Cependant, la multiplicité des corpus permet d'avoir une idée plus nette du système linguistique, dans la mesure où chaque corpus instancie un type de contrainte du contexte sur la langue ; en analysant la variabilité entre différents corpus, on peut mesurer l'impact de cette contrainte, et donc remonter (théoriquement du moins) un peu plus en amont, vers le système linguistique ou cognitif, c'est-à-dire vers ce qui, à terme, m'intéresse : la cognition, et le rapport entre la cognition, la perception du monde et le monde lui-même.



Graphique 1 : De la langue au corpus, du corpus à la langue

[3] Je dois à ce titre plaider coupable – j'ai moi-même essayé, avec le concours de Claude Buridant, que je remercie ici à nouveau, de traduire quelques phrases en ancien français, à la demande d'un écrivain qui voulait « mettre en situation » un personnage d'enquêteur médiéval.

1.1.2 Compenser les limites des grammaires

Un avantage certain des corpus est qu'ils permettent de compenser en partie les limites des ouvrages de référence. Les grammaires et dictionnaires, qui constituent un point de départ précieux pour le linguiste, ont le défaut d'être nécessairement figés, historiquement et théoriquement. Ainsi, des nouveaux concepts comme celui de l'intersubjectivation (*intersubjectification*, Traugott 2003, 2007b, 2010), ou des nouvelles classes de mot comme les marques discursives, ne sont pas pris en compte. L'utilisation de grammaires pour des études contrastives est également problématique, puisqu'elles ne sont pas nécessairement compatibles entre elles.

Cela pose aussi problème pour l'étude de constructions rares, dont les ouvrages de référence ne rendent pas toujours compte. Enfin, tout ce qui est à la marge, notamment entre lexique et grammaire, risque de ne pas être bien documenté, ni par les dictionnaires, ni par les grammaires ; c'est le cas, entre autres, des prépositions, ou encore de divers types de locutions.

Dans tous ces cas, une recherche sur un corpus suffisamment complet permet de compenser les failles éventuelles des données lexicologiques existantes, avec l'avantage supplémentaire d'une information quantifiée.

1.1.3 Analyse quantitative et subjectivité

Sans aller aussi loin que certains (cf. « *langage* as 'language in use' cannot be studied without statistics », Herdan 1964 : 5), on reconnaîtra que l'analyse quantitative est la première vertu des corpus (Fagard 2008b [VI : 933]), qui permettent d'objectiver les observations, de les rendre moins intuitives :

> corpus research may provide a clearer, less intuitive insight. (Oakey 2002 : 111)

Ainsi, pour un phénomène linguistique complexe comme l'alternance, en anglais, entre datif et construction prépositionnelle (*to give **someone** a book / to give a book **to someone***), Bresnan a bien montré les limites du jugement de grammaticalité et l'apport du corpus :

> The experiment points to ways of establishing sounder empirical foundations for syntactic and semantic theory and suggests why **the older ways of doing syntax – by generalizing from linguistic intuitions** about decontextualized constructions **and ignoring research on actual usage**, especially quantitative corpus work – **produce unreliable and inconsistent findings**. (Bresnan 2007 : 91 ; je souligne)

Un élément clef de cette approche est l'analyse quantitative (Mair 1995 : 260). Cette dernière doit respecter des règles précises : il ne s'agit pas uniquement de donner des relevés chiffrés, mais de montrer dans quelle mesure les données observées sont significatives (Biber et al. 1998 : 268). La question est bien de savoir si tel ou tel effet est ou non significatif, par rapport à quoi et avec quelle représentativité ; se pose donc le problème de savoir comment on calcule cela. J'ai ainsi cherché, pendant mes années au CNRS, à affiner ma connaissance de ces règles d'application des statistiques. Dans de nombreuses disciplines scientifiques, les

résultats sont invariablement donnés avec une indication de leur degré de fiabilité statistique. C'est le cas dans certaines sous-disciplines de la linguistique, mais ce n'est pas (encore ?) très répandu en linguistique diachronique, où le recours aux corpus est à la fois plus ou moins obligatoire et toujours partiel : tandis que les spécialistes de langues modernes peuvent espérer avoir un corpus représentatif, et qu'il peut donc leur paraître important de voir à quel point les résultats qu'ils obtiennent sont solides, les spécialistes de langues anciennes sont souvent contraints d'admettre comme un présupposé que leur corpus n'est pas représentatif. On pourrait donc penser qu'il devient superflu et même vain de chercher à vérifier leur degré de fiabilité. Cependant, on peut trouver deux objections à ce désistement de la vérification : d'une part, les corpus ont largement progressé, et deviennent maintenant nettement plus conséquents et donc plus proches de ce que l'on peut attendre d'un corpus représentatif ; d'autre part, même si le corpus n'est pas réellement représentatif d'un état de langue en général, cela ne doit pas nous empêcher de vérifier la valeur des résultats obtenus.

Afin d'avancer en ce sens, je me suis donc formé peu à peu à l'utilisation de méthodes statistiques. Je l'ai fait tout d'abord de manière très empirique, dans ma thèse et mes premiers articles. J'ai ensuite pu progresser grâce à mes collaborations avec des collègues plus habitués que moi à ces méthodes, notamment en collaborant avec Liesbeth Degand (UCL), dans le cadre du projet GRAMIS[4]. J'ai aussi cherché à affiner mes connaissances en suivant plusieurs formations : Initiation aux statistiques (2011), Notions fondamentales en statistiques (2011), Introduction au logiciel R pour les statistiques (2012, 2015). L'utilisation de tests statistiques est ainsi mise en œuvre dans les articles que j'ai écrits en collaboration avec Liesbeth Degand, ainsi que d'autres plus récents (Stosic et al. 2015 [V : 631 sqq.], Fagard et al. sous presse [V : 613 sqq.]). On peut noter que c'est là un domaine en pleine expansion, et qui reçoit désormais l'attention des utilisateurs de corpus. C'est le cas dans le domaine de la recherche (voir Blumenthal 2008, Gries & Hilpert 2008, 2010, Hilpert & Gries 2009, Prévost 2011a,b, et tout récemment le numéro 197 de *Langages*, Loiseau 2015), mais aussi dans celui de l'enseignement (voir le nombre croissant de manuels de linguistique de corpus orientés vers le TAL, la quantification, les statistiques, mais aussi les formations de statistiques proposées dans le cadre du Labex EFL, ou encore les séminaires de Guillaume Desagulier à Paris 8, par exemple). Malgré cela, le recours aux statistiques ne semble pas encore être la norme dans la linguistique diachronique.

1.1.4 Phénomènes rares et structure du langage

Précisément en lien avec l'utilisation des statistiques, un atout particulièrement frappant de l'étude sur corpus est que cette approche permet de compenser au moins en partie les limites de l'introspection, ce qui signifie que cette dernière, de toute façon discutable pour les périodes anciennes, n'est plus la seule méthode disponible pour vérifier le degré d'acceptabilité d'une construction. L'étude sur corpus permet ainsi de dépasser la question

[4] Il s'agit d'un projet de recherche financé par le gouvernement fédéral de Belgique, regroupant des chercheurs de plusieurs universités en Flandres, Wallonie et Allemagne pour travailler sur la question des rapports entre grammaticalisation et (inter-)subjectivation (Gramis, IAP P6/44, 2007-2011).

assez problématique de l'acceptabilité. Problématique, en premier lieu, parce qu'il est souvent difficile d'obtenir un consensus auprès des locuteurs, comme j'ai pu m'en rendre compte lors de mes cours d'introduction à la linguistique à Paris 7, pendant lesquels je soumettais régulièrement aux étudiants des phrases-tests, par exemple sur le problème du subjonctif avec *après que* ou encore l'acceptabilité de *malgré que* : je n'ai jamais réussi à obtenir des réponses homogènes, à part sur des questions évidentes (et encore !). De plus, l'expérience de Bresnan (2007) mentionnée plus haut confirme qu'il ne s'agit pas là d'un biais lié au manque d'expérience des étudiants, puisqu'elle a interrogé des linguistes chevronnés, spécialistes de la question. Par ailleurs, l'utilisation de tests d'acceptabilité dans la littérature me semble souvent peu probante, du moins lorsqu'elle est formulée en ces termes : bien souvent, il est assez facile de trouver un contexte où la phrase réputée inacceptable devient tout à fait naturelle, sans pour autant nécessairement remettre en question l'intérêt ou la solidité des résultats. Cela tient au fait que la question la plus intéressante n'est pas celle de la possibilité ou non d'une construction donnée. Le plus intéressant, me semble-t-il, est plutôt de savoir si une construction ou une séquence est naturelle, usuelle, en d'autres termes d'évaluer ses probabilités d'apparition :

> Modern electronic corpora now make it possible to explore the nature and use of linguistic phenomena in a much wider variety of texts. **Such descriptions** go beyond exploring what is grammatically and semantically possible, and **add a distributional dimension which characterizes linguistic features in terms of probability of occurrence**. (Kennedy 2002 : 73 ; je souligne)

L'attention se porte ainsi sur l'usage, plutôt que sur ce que le système linguistique permet ou non :

> Rather than looking at what is **theoretically possible** in a language, we study the actual language used in naturally occurring contexts. (Biber & al., ibid. : 1 ; je souligne)

On s'attardera cependant sur la citation de Biber ci-dessus, et notamment sur le terme *naturally* : c'est bien cela qui pose problème, à savoir *qu'est-ce que le langage sous sa forme naturelle ?* La question n'appelle pas nécessairement de réponse, mais elle mérite d'être posée, tant le langage est un animal culturel : son habitat (pour filer la métaphore), ou plutôt son expression même, mais aussi sa condition d'existence, est la culture. Quoi qu'il en soit, je reste convaincu que le recours aux corpus, en combinaison avec l'utilisation des statistiques, permet de mieux rendre compte de la structure même du langage.

Les limites de l'introspection sont particulièrement frappantes en ce qui concerne les constructions rares ou très rares. Le corpus permet non seulement de séparer ce qui est usuel de ce qui est exceptionnel (Mair 1995 : 260), mais encore de travailler, en comptant sur des corpus diversifiés et incluant des registres non standard, sur des constructions généralement considérées comme hors du système linguistique. C'est ce qui m'a permis par exemple de travailler sur le marquage différentiel de l'objet (MDO) en français, une construction rejetée par les grammaires (Fagard & Mardale 2014 [I : 9]) : en effet, par nécessité (de place, de temps), les ouvrages de référence ne recensent pas systématiquement les phénomènes rares ou très rares. Ainsi, concernant le MDO, on pourrait simplement accepter ce que nous dit la

tradition grammaticale, à savoir qu'il n'existe pas en français, et que les exemples qu'on peut trouver ne sont que l'effet du contact linguistique, comme le note Claire Blanche-Benveniste :

> Une tournure comme *parce que tu nous aimes à nous*, avec la préposition *à* devant le complément [direct], est facilement mise sur le compte d'un régionalisme, à expliquer par l'influence d'une langue voisine ou d'un dialecte. (Blanche-Benveniste 1997 : 37, section « Les fautes »)

Il est vrai que le statut de ce phénomène en français ne peut se comparer ni aux langues romanes où il est obligatoire, grammatical, comme l'espagnol et le roumain, ni à celles où il est fréquent sans être obligatoire, comme l'italien méridional ou le portugais. Cependant, nous montrons dans notre étude que le MDO en français mérite davantage d'attention, et que le recours aux corpus permet d'en savoir un peu plus sur le statut qu'il faut lui accorder, malgré son extrême rareté. D'une part parce que sa présence n'est pas limitée aux zones de contact avec des langues à MDO – le contact linguistique n'est donc pas la seule force en présence. D'autre part parce que ce que l'on peut voir du MDO en français correspond aux contextes d'émergence du phénomène en espagnol médiéval et en roumain classique :

> comme dans toutes les langues où le MDO a été étudié, deux types de facteurs entrent en jeu, des facteurs inhérents (à l'objet) et des facteurs globaux. (…) Concernant la hiérarchie des facteurs internes, le MDO en français touche les positions les plus hautes, comme en espagnol et en roumain : pronoms personnels et GN lexicaux référent à des humains, avec un déterminant possessif ou démonstratif. (Fagard & Mardale 2014 : 165 [I : 23])

1.2 Conditions et limites du recours aux corpus

Mon enthousiasme concernant l'intérêt des corpus ne doit pas cacher les difficultés que présente cette méthode ; j'expose ci-dessous les principales conditions – et limites – du recours aux corpus.

1.2.1 Problèmes liés à l'étude sur corpus

Les remarques faites dans la section précédente (§1.1) sont valables pour des études menées sur un « corpus de référence » idéal. Cependant, on rencontre plusieurs types de problèmes lorsque l'on utilise cette méthode – et en premier lieu le fait qu'aucun corpus n'est idéal. D'abord parce qu'il y a, comme je le suggère dans le graphique (1), un fossé entre ce qui est linguistiquement possible et ce qui est énoncé ; et un autre entre ce qui est énoncé et tout corpus, aussi grand soit-il. Le défi que représente la description d'une langue à partir d'un corpus de quelques millions de mots est comparable – mais d'un ordre de grandeur bien plus élevé encore me semble-t-il – à celui du sondeur qui, à partir de moins de mille intentions de vote, tente de prédire le résultat d'une élection à laquelle participent des dizaines de millions d'électeurs.

Même un « corpus de référence », qui se doit de reproduire la variation propre à la langue (Biber et al. 1998 : 247-248, Biber 1990, Prévost 2005), ne peut donc rendre compte de la masse de données produite en permanence par les locuteurs d'une langue, et bien moins encore des possibilités offertes par la langue. La limite la plus criante de tout corpus est celle du genre textuel : un corpus dit « de référence » doit présenter un certain équilibre entre différents genres, mais les corpus disponibles sont toujours de fait limités à quelques genres.

Un facteur aggravant est que la modalité écrite entraîne nécessairement une distance supplémentaire par rapport à la parole « spontanée », tandis que les corpus oraux sont, pour des raisons techniques, généralement plus limités et plus difficiles à consulter. Pour les périodes anciennes, ils sont même évidemment absents. L'étude de l'oral représenté (Marchello-Nizia 2012), que nous évoquerons plus bas, permet de compenser ce manque en partie – mais en partie seulement.

Un problème supplémentaire est celui de la comparabilité : comment faire pour avoir des corpus comparables pour différents états de langue ou différentes langues ? Le mieux est que les différents corpus soient équilibrés et présentent les mêmes types de textes (au sens très large d'ensemble d'énoncés). Mais les genres textuels et discursifs ne sont pas les mêmes pour différentes périodes et différentes langues. Pour ne prendre qu'un exemple, la définition du roman a bien changé à travers les siècles, de « traduction en langue romane d'un texte latin » à « récit en prose ». Mon expérience dans le projet Presto[5], dans le cadre duquel j'ai supervisé la constitution d'un corpus pour la partie médiévale ($9^{ème}$-$15^{ème}$), m'a permis de vivre ce dilemme de l'intérieur : étant donné les choix faits pour la période $16^{ème}$-$20^{ème}$, avec donc la sélection d'un certain nombre de genres textuels et de critères de sélection, comment faire pour obtenir un corpus comparable pour la période ancienne ? Plusieurs problèmes concrets se posent en effet : il n'y a pas d'équivalent au genre « journalistique », semble-t-il – par ailleurs, ce dernier connaît une telle progression entre le $18^{ème}$ et le $20^{ème}$, et peut-être même entre le $20^{ème}$ et le $21^{ème}$, qu'il est difficile de le considérer comme un genre stable. Pour d'autres genres, comme le roman, même en admettant qu'il y a une certaine continuité, le peu de textes conservés (qu'explique en partie la destruction d'innombrables manuscrits, notamment au cours des dernières guerres) rend impossible la constitution d'un corpus comparable quantitativement. Enfin, il faudrait mentionner également le problème de l'étiquetage pour un corpus diachronique : le choix entre des étiquettes continues – identiques de l'ancien français au français moderne – et des étiquettes distinctes pour chaque période est complexe. Doit-on mettre sur le même plan *cil*, *cist* et *celui-ci*, *celui-là*, *ce(tte)* ? Quelle étiquette choisir, qu'il s'agisse de morpho-syntaxe ou de lemmatisation ? Pour les articles indéfini et partitif, à partir de quand faut-il les étiqueter comme tels ? Pour les mots lexicaux qui ont changé nettement de sens, faut-il adopter des lemmes distincts ? Autant de questions d'autant plus douloureuses qu'il faut les résoudre en amont, avant d'avoir les retours des utilisateurs[6]. Et ces questions, déjà complexes pour une langue donnée, ne font qu'empirer

[5] Un projet financé par l'ANR et la DFG, piloté par Denis Vigier et Peter Blumenthal, qui rassemble des chercheurs de laboratoires à Paris, Lyon et Cologne, et vise à produire une description de l'évolution des prépositions en français, à partir d'un corpus diachronique annoté et de ressources informatiques dédiées.

[6] Du moins dans l'état actuel des financements : dans le cadre qui s'est progressivement imposé au monde de la recherche, on est contraint à envisager un projet comme ayant une date de début et de fin déterminées à l'avance, qui ne permettent guère des aller-retour de ce type.

lorsqu'il s'agit de comparer plusieurs langues, qui plus est en diachronie. La liste des corpus que j'utilise régulièrement, et qui figure après la bibliographie, montre bien l'étendue du problème : on peut constater une grande variabilité à tous points de vue, notamment concernant leur taille (de moins de 100 000 mots à plusieurs milliards !), leur niveau de sophistication (texte brut, encodage simple des métadonnées, lemmatisation, étiquetage morpho-syntaxique, schémas syntaxiques, étiquetage sémantique…), ou encore les langues, périodes, modalités et genres concernés.

Un autre problème de la recherche sur corpus – qui n'est pas sans lien avec le précédent – est celui du temps de travail : le chercheur est souvent amené à limiter son analyse à un nombre réduit d'occurrences de la construction étudiée, ne serait-ce que parce que, lors d'une étude qualitative répétée sur plusieurs centaines d'occurrences similaires, l'analyse elle-même fait évoluer l'opinion du chercheur : insensiblement, et même lorsque l'on pose des critères stricts d'analyse, on peut finir par changer d'avis, et par coder comme « métaphorique » ce que l'on codait au départ comme « abstrait », ou inversement. Il est extrêmement difficile de coder de manière uniforme une très grande quantité de données, comme on peut s'en rendre compte à chaque recodage. Cette problématique est d'ailleurs bien (re)connue, notamment dans le domaine du TAL : comme le note Fort (2012 : 50),

> Il est aujourd'hui difficile, voire impossible, de publier un article de recherche sur un corpus annoté sans fournir de mesure précise de sa qualité (en l'occurrence sa fiabilité), si possible sous forme d'une métrique unique et bien connue, ladite métrique devant en outre permettre de comparer un corpus annoté à un autre.

Et pourtant il n'est guère aisé de trouver des indications métriques fiables (Artstein & Poesio 2008), à plus forte raison pour tout ce qui relève du sens ou de la fonction : « Applying functional labels to linguistic units is a highly subjective activity » (Oakey 2002 : 115). Si l'on retient comme critère de qualité scientifique la reproductibilité des études, le codage doit être reproductible lui aussi, ce qui est complexe, voire utopique dès qu'il s'agit d'aspects fonctionnels, sémantiques ou pragmatiques.

Enfin, même si l'on parvient à limiter les dommages liés à ces problèmes, il reste une limite qui affecte presque nécessairement la linguistique de corpus :

> The exclusive focus on corpora of standard variants of languages, often produced by highly literate speakers (or their offspring), provides the linguist only with limited external validity. This is a trivial claim that has been made by sociolinguists for at least 50 years and by dialectologists for more than 200 years. (Berthele 2013 : 57)

On peut tenter de contourner cette difficulté en créant ses propres corpus, comme le font les typologues. C'est ce que j'ai fait pour le piémontais (en collaboration avec Massimo Cerruti ; voir §5). Comme le souligne Marianne Mithun, cela peut également présenter des avantages, notamment celui de la maîtrise des données : « Each use (…) can be examined in its full syntactic and discourse context » (Mithun 2012 : 2167) ; mais cela requiert énormément de temps et d'énergie.

1.2.2 Précautions d'emploi (1) : Constitution du corpus

Si toutes les difficultés mentionnées ci-dessous ne suffisent pas selon moi à remettre en cause l'utilité des corpus, c'est principalement parce qu'une utilisation raisonnée des corpus permet de les contourner en partie.

Une première précaution à prendre est de combiner la recherche sur corpus à une approche plus traditionnelle à partir des outils philologiques (grammaires, dictionnaires). On utilisera alors le corpus comme complément (Biber & al. 1998 : 9-10). Un autre élément important est que la constitution du corpus doit se faire de manière réfléchie et ne doit pas reposer sur la seule disponibilité des données (Lebel 2003) : on peut ainsi tirer pleinement profit du caractère dynamique du corpus, qui « peut (et doit) s'adapter à chaque fois à l'objet d'étude » (Fagard 2008b [VI : 936]). La constitution du corpus ne peut en effet être décidée indépendamment de la question posée, du phénomène observé. Ce qui pouvait sembler utopique il y a quelques années est désormais faisable, grâce à la multiplication des corpus électroniques à la fois riches et facilement disponibles : il devient possible de sélectionner de manière réfléchie un corpus pertinent *pour une question de recherche spécifique* au sein d'une série de bases de données. Par ailleurs, il reste possible de construire son propre corpus, comme je l'ai fait pour l'étude de la trajectoire (§4.2) et du mouvement fictif (§4.3). Cela devient même plus facile, notamment grâce à des structures comme les consortiums IRCOM « corpus oraux et multimodaux » et « Corpus Ecrits », du TGIR Huma-Num.

Concernant la constitution du corpus, on peut se demander en particulier si certains lieux sont plus propices à l'observation du changement linguistique, qu'il s'agisse de types de textes (avec par exemple la correspondance plus proche de la langue orale, plus relâchée, le théâtre plus artificiel, plus codé, et le roman intermédiaire ?) ou d'autres paramètres (voir notamment la notion d'oral représenté déjà mentionnée plus haut, cf. Marchello-Nizia 2012). J'ai abordé cette problématique notamment dans Fagard & Degand (2010 [III : 313 sqq.]) puis dans Degand & Fagard (2012 [III : 269 sqq.]), en comparant les emplois de *parce que*, *car* et *alors* à l'écrit et à l'oral : l'oral n'est pas assimilable à un état futur de la langue, mais il me semble en tout état de cause intéressant d'inclure, dans les études diachroniques, une comparaison entre écrit et oral de la langue moderne. Cela nous offre donc une alternative à la vaine recherche du corpus de référence, c'est-à-dire le choix d'un corpus si possible proche de l'oral, de la langue réelle, et aussi peu que possible touché par les règles littéraires ou autres influences de la tradition discursive propre (pour les textes techniques, lexique figé, tournures figées, etc. sans lien fort avec le système linguistique global).

Il reste un point fondamental : l'analyse sur corpus doit à mon sens respecter le principe de reproductibilité des études (cf. Oakey 2002 : 115), comme d'ailleurs les méthodes expérimentales de manière générale (Pashler & Wagenmakers 2012, Makel & Plucker 2014, Prosser 2010, Smith & Smith 2014). On notera à ce sujet que la permanente évolution des bases de données, qui tendent à croître régulièrement (ainsi de Frantext ou Corde), peut également avoir des contre-indications : en effet, il devient difficile de répliquer exactement une étude, puisque le corpus peut avoir changé entre deux utilisations de la base (cependant, certaines bases préservent la possibilité d'accéder aux anciennes versions). Autre aspect fondamental, le caractère circulaire de son utilisation : le médiéviste est contraint d' « étudier

des textes avec une connaissance de la langue exclusivement fondée sur ces mêmes textes » (Prévost 2005 : 149).

1.2.3 Précautions d'emploi (2) : Utilisation des corpus

Une fois le corpus constitué, il reste à voir comment l'utiliser. Il faut ici suivre un certain nombre de règles pour s'assurer de la qualité des résultats. La première difficulté à cet égard est le **choix du nombre d'occurrences** à analyser. Ce dernier doit être suffisant pour prétendre à une certaine représentativité et suffisamment faible pour rester gérable. La méthode qui me semble préférable de ce point de vue est de combiner des analyses quantitative (par exemple sur les propriétés distributionnelles : position, contextes typiques, etc.) et qualitative :

> On pourra à cette fin associer une étude semi-automatisée sur l'ensemble des occurrences, et une étude de détail sur une partie de ces dernières, sélectionnées aléatoirement, chronologiquement ou sur un autre critère (pour autant qu'il soit justifié, et surtout explicité ; voir Biber & al. 1998 : 4). (Fagard 2008b [VI : 938])

Cette démarche n'est pas révolutionnaire (voir Jick 1979), mais elle permet de compenser en partie les défauts des approches purement quantitatives ou qualitatives indépendamment du domaine (Ramsay 2007), les deux se complétant mutuellement (Kelle 2006, Rheinländer 2011 : 119). Elle est connue sous le nom d'approche mixte (*mixed-methods* approach) et recoupe en partie celle de triangulation (Netanda 2012, Mertens & Hesse-Biber 2012), dont l'acception est cependant variable selon son domaine d'application (Rheinländer 2011 : 112). L'idée générale est que ces deux approches, qualitative et quantitative, ne sont pas incompatibles (ce point a longtemps fait débat, et n'est toujours pas consensuel, cf. Teddlie & Tashakkori 2003, Denzin 2012), et que chacune est le pendant de l'autre. Ainsi, l'analyse **qualitative** permet de compenser en partie le bruit et le silence : le bruit, inévitable, consiste à avoir dans les résultats autre chose que la structure recherchée, tandis que le silence, qu'il convient au contraire d'éviter autant que possible, résulte de l'oubli de certaines variantes, par exemple graphiques, de cette même structure. Les deux, bruit et silence, sont également difficiles à repérer de manière automatique, tandis qu'une analyse qualitative permet d'exclure totalement le bruit et de repérer des variantes non prises en compte au départ, surtout si le contexte donné avec le pivot des requêtes est suffisamment important. C'est là entre autres qu'intervient la complémentarité des deux approches, sur corpus et traditionnelle.

L'analyse **quantitative**, quant à elle, permet comme nous l'avons expliqué de contrebalancer la subjectivité du chercheur. Les tests statistiques développés pour les sciences dites dures et largement utilisés dans la littérature scientifique de certains domaines comme la psychologie ou même la psycholinguistique ne semblent pas toujours adaptés au traitement des données linguistiques (cf. Kilgariff 2005). Il y a un effort continu de la communauté pour produire des méthodes adaptées aux données linguistiques (voir notamment les travaux de Blumenthal, Hilpert, Gries : Blumenthal 2008, Gries & Hilpert 2008, 2010, Hilpert & Gries 2009). Il me semble cependant que, à la différence d'autres disciplines, il n'y a pas encore de consensus sur les mesures qu'il faut appliquer aux données pour les considérer valables. De plus, on

notera qu'avoir des quantités énormes de données ne suffit pas à produire une recherche de qualité ; il convient avant tout de savoir quels types de réponses on peut attendre des tests statistiques, et d'éviter d'appliquer des tests aveuglément (Hilpert 2013 : 42).

Inversement, on voit parfois des utilisations purement statistiques de données linguistiques, dont l'intérêt scientifique n'est pas toujours évident – voir par exemple l'article de Lupyan & Dale (2010), qui prennent les données du WALS comme base pour identifier statistiquement des langues « exotériques », ayant moins de mille locuteurs, caractérisés par un groupe social homogène et une localisation géographique réduite, ayant moins de marques casuelles et une probabilité plus grande de ne pas avoir de marques de personnes sur les adpositions. Ce type d'étude semble avoir le vent en poupe, et on lira avec intérêt le billet de Haspelmath concernant la fin de la linguistique au Max Planck de Leipzig, notamment le passage suivant :

> Of course, Russell Gray is not originally and not only a linguist, and his methods and publication venues are not the ones of traditional linguists. But this is precisely why the Max Planck Society decided to make him the director of this new department: **They feel that historical linguistics holds a lot of potential, and that by applying new quantitative methods to the available data, significant progress can be made in understanding human history at a world-wide scale**. This is not how most linguists see things. The great majority regard the relation between language and cognition as much more interesting than (large-scale) historical linguistics, and this is understandable because most linguists just work on a single language. But outside of our field, the big historical questions and the possible answers that linguists can give seem to have a much greater attraction (note that very few linguists were involved in deciding on this new department). (http://dlc.hypotheses.org/710, 18/10/2014 ; je souligne)

Une fois que tous ces points sont réglés, il reste un point qui a souvent été oublié dans l'exploitation des résultats de la recherche sur corpus en diachronie : le problème à la fois théorique et épistémologique de la contribution au renouvellement linguistique en fonction de l'âge, c'est-à-dire le mode d'apparition et de propagation des innovations (progressive ou abrupte ?). C'est ce que j'ai appelé la question de « l'âge des innovateurs » (cf. Bybee & Slobin 1982, Hurford 1991, Nettle 1999). Cette question, essentielle pour l'étude de l'évolution linguistique, pourrait avoir une influence fondamentale sur la manière de travailler sur corpus en diachronie. En effet, si les locuteurs ne changent pas de façon de parler – et donc, si les écrivains ne changent pas de façon d'écrire – alors la prise en compte de la date de publication peut entraîner des erreurs d'appréciation : la présence massive d'auteurs âgés dans une tranche du corpus pourrait ainsi masquer une évolution. En sociolinguistique, la réflexion sur ce point remonte au moins au début du siècle dernier (Gauchat 1905, Hermann 1929), et a été théorisée vers le milieu du siècle (notion de « temps apparent », Hockett 1950 puis Labov 1963, 1994). Elle est centrale dans l'approche de quelques collègues (Ayres-Bennett 2004, Tristram & Ayres-Bennett 2012, Hansen & Malderez 2004), mais à peu près absente de la réflexion méthodologique sur l'utilisation des corpus.

J'ai abordé ce problème en collaboration avec Bernard Combettes (Université de Lorraine & ATILF), et nous avons pu montrer que, selon les phénomènes, la prise en compte de l'âge des écrivains avait ou non une importance, peut-être en lien avec la difficulté d'acquisition variable des phénomènes linguistiques (Kerswill 1996). La réponse que nous avons pu

apporter pour l'instant à cette problématique est qu'il y a à la fois des phénomènes de mode (ce qui implique que les locuteurs participent alors au changement linguistique indépendamment de leur âge) et des phénomènes de génération (où l'âge des locuteurs doit être pris en compte). D'un point de vue méthodologique, cela signifie qu'il faudrait toujours en théorie deux calculs concurrents lors de l'étude d'un phénomène en diachronie : le premier à partir des dates de publication, le second à partir des dates de naissance des auteurs (Fagard & Combettes 2013 [VI : 819 sqq.]). En pratique, cependant, cette méthode est difficile à mettre en œuvre, entre autres parce que nous ne disposons pas toujours des dates de naissance des auteurs – en particulier pour la période médiévale.

2 Eclairages

Chaque positionnement théorique, chaque (sous-)discipline des sciences du langage éclaire à sa manière le fonctionnement du langage et, quelle que soit la question linguistique que l'on souhaite poser, il me semble que l'on peut trouver des réponses au moins satisfaisantes dans l'un ou l'autre cadre théorique. Cela ne signifie pas que notre tâche est accomplie et qu'il n'y a plus rien à faire. Cela implique plutôt que l'on dispose désormais d'outils et d'informations suffisants pour obtenir des réponses, à condition de faire appel à l'ensemble des ressources disponibles et de ne pas se cantonner à son propre champ scientifique. C'est pourquoi je préconise de **mettre en commun les résultats de plusieurs domaines**, et donc de **combiner plusieurs approches théoriques et méthodologiques : la linguistique de corpus, la linguistique diachronique, la linguistique contrastive, la linguistique expérimentale et la linguistique typologique.** Je décris dans cette section ce que je retiens de ces différentes approches du fait linguistique.

S'il est pour moi fondamental d'étudier chaque phénomène non seulement dans une langue à un moment donné mais aussi (a) en diachronie, (b) dans des langues proches et (c) dans d'autres langues, à divers degrés de détail, avec autant que possible (d) un contrôle expérimental, c'est parce que chacun des pans de cette approche éclaire les autres, et que ce ne sont que les quatre réunis qui, de mon point de vue, permettent une vision suffisamment globale du phénomène pour espérer en avoir (e) une approche cognitive. Il s'agit pour moi de différentes façons de faire de la linguistique : chaque approche implique un certain bagage théorique, un certain nombre d'habitudes de travail, de présupposés, de méthodes qui leur sont en partie propres. Un même problème peut ainsi être traité de cinq manières différentes (si l'on s'en tient à la courte liste ci-dessus) selon l'approche adoptée, ou encore combiner deux ou plus de ces approches afin d'obtenir des résultats on l'espère plus solides[7]. Dans cette section, j'illustre une à une chacune de ces méthodes afin de montrer ce que j'entends respectivement par *diachronie*, *linguistique contrastive*, *linguistique romane*, *typologie* et *linguistique cognitive*, en théorie mais également en renvoyant à ma propre pratique scientifique, et en reprenant à l'occasion certains de mes travaux.

2.1 L'approche diachronique

La diachronie est pour moi doublement essentielle. Je ne conçois pas d'explication d'un état de langue donné – d'une synchronie donnée – autrement que (aussi) comme une explication diachronique. Cela n'invalide pas l'étude synchronique en soi : la diachronie ne peut pas non plus se passer de la synchronie ; pour reprendre Mithun (1993 : 521), « neither can be understood without reference to the other ». Mais il me semble qu'en refusant d'intégrer les

[7] Les problèmes spécifiques liés à la triangulation sont en effet non négligeables, et la plus-value n'est pas automatique (voir par exemple Mertens & Hesse-Biber 2012 (eds) et les articles du numéro).

aspects diachroniques, on risque bien souvent de rendre la synchronie inintelligible. La linguistique synchronique garde toute sa légitimité, et doit pouvoir expliquer le système tel qu'il fonctionne pour un locuteur donné ou un ensemble de locuteurs, mais bien souvent des choses qui semblent incongrues, anomales en synchronie prennent tout leur sens en diachronie. C'est parce que le locuteur doit faire sens d'un système qui est (nécessairement ?) imparfait qu'il contribue lui-même à introduire de nouveaux déséquilibres.

De plus, mon but étant de décrire les conditions et les contraintes pesant sur l'évolution linguistique, l'objet de ma recherche est par essence diachronique. Je décris dans cette section ma vision de l'approche diachronique et de ce qu'elle suppose, en illustrant donc à la fois ma pratique réelle de l'approche diachronique, et ma vision idéale de ce que devrait être l'approche diachronique.

On sait que la linguistique diachronique, après l'âge d'or du 19ème siècle, a connu une certaine disgrâce au 20ème siècle – en lien avec l'accent mis par Saussure sur la synchronie et peut-être plus encore, si l'on en croit Peyraube, avec la réception de Saussure aux Etats-Unis au début du 20ème siècle :

> Les études de grammaire diachronique ont été incontestablement la parente pauvre de la recherche linguistique au cours du siècle dernier. On rend habituellement Saussure responsable de cette situation : sa fameuse dichotomie synchronie/diachronie aurait constitué une véritable coupure épistémologique par rapport à la période prédédente (dernier quart du XIXe siècle), dominée par les néogrammairiens. Rien n'est moins sûr. (…) Bloomfield et les structuralistes américains ont davantage pris la dichotomie comme un dogme, et cette position a été perpétuée par les différents courants issus ensuite des grammaires formelles. (Peyraube 2004 : 135, premières phrases de l'article)

Le générativisme a achevé ce mouvement, dans la deuxième moitié du siècle, avec un intérêt quasiment exclusif pour la synchronie, comme en témoigne admirablement l'anecdote suivante :

> Vers 1962, le regretté Paul Imbs, l'un des maîtres de la linguistique française, déclarait à peu près ceci à un petit groupe d'universitaires qui l'entouraient : « La linguistique historique ou diachronique a fait son temps ; elle a été incapable d'expliquer, par exemple, une phrase aussi simple que 's'il faisait beau demain ; j'irais à la campagne'. Aujourd'hui nous avons heureusement des méthodes synchroniques et des théories pour faire la lumière sur ce point et d'autres ». (Lanly 1999 : 199)

Cependant, on constate depuis une trentaine d'années un renouveau certain de la diachronie, qui est sans doute lié en partie aux travaux sur la grammaticalisation (en particulier depuis la fin des années 80) et a été facilité par la constitution de bases de données électroniques qui ont à peu près révolutionné la méthode des diachroniciens[8]. Le lien avec l'essor de la typologie noté par Peyraube dans la suite de son article est intéressant, et recoupe mes préoccupations :

> Dans les années 1970, toutefois, à la suite des travaux de Greenberg (1966[a]) sur les universaux de langues, de nombreux travaux sur la typologie, mais aussi sur la diachronie – tant il est vrai que les deux domaines sont vite apparus liés – ont vu le jour. Depuis une

[8] On notera cependant que les études sur corpus (y compris des corpus étendus) sont bien antérieures à la mise en place des bases de données informatisées. Ce qui a changé, c'est la possibilité de croiser les requêtes de manière instantanée et presque infinie.

> quinzaine d'années, à l'instar de la linguistique typologique dont un des programmes essentiels est la recherche d'universaux dans les langues, la grammaire historique a considérablement élargi sa perspective en se voulant en partie tributaire de la linguistique cognitive dans son ensemble. (Peyraube 2004 : 135, suite de la citation précédente)

Il y a bien un renouveau, d'une approche plutôt philologique vers une approche cognitive[9], noté dans la dernière phrase de la citation ci-dessus et porté par des chercheurs comme Koch (Koch 1997b, Blank & Koch 1999) ; ce renouveau est tel qu'aujourd'hui – à la différence d'il y a un demi-siècle – le diachronicien n'est plus nécessairement considéré comme un scientifique dépassé ou un vestige archaïsant d'une époque révolue, et que les collaborations avec des spécialistes de la langue moderne se multiplient.

Une bonne pratique de la linguistique diachronique suppose de mon point de vue le respect de quelques règles élémentaires :

(a) Elle n'implique pas nécessairement une période longue ;
(b) Elle ne se résume pas à l'étude d'états de langue passés ;
(c) Elle implique la comparaison entre différents états de langue ;
(d) Elle ne peut partir uniquement de découpages chronologiques effectués a priori.

Ainsi, il n'y a pas d'empan minimal pour la diachronie – on peut étudier l'évolution d'un mot ou d'une construction à une échelle de temps très réduite, par exemple à l'aide des nouvelles technologies : la diffusion d'un néologisme dans un réseau social (twitter, sms, etc.) peut donner des indications sur les modalités de la diffusion des innovations en général. Mais il s'agit nécessairement de l'étude d'une évolution à travers la comparaison de différentes synchronies. Cela pose le problème de la définition d'une synchronie – un problème tout aussi complexe que la définition d'une langue ou d'un dialecte, et qui rejoint la délicate question de la périodisation (Combettes & Marchello-Nizia 2008). Ainsi, pour deux dialectes, il est complexe de savoir où mettre la frontière, et la décision finale est de toute façon en partie arbitraire – liée à un poids plus grand attribué à tel ou tel critère. Il en est de même pour deux périodes : quand commence le français classique ? Quand finit le moyen français ? Faut-il parler, entre les deux, de français préclassique (Combettes & Marchello-Nizia 2010) ? Ou bien faut-il encore, comme le suggère Hilpert (2013 : 32 sq.), partir des données, avec des outils statistiques qui permettent de trouver la meilleure périodisation pour un phénomène spécifique ? Les deux approches ne sont en fait pas du tout incompatibles, comme nous allons le voir.

Dans mes recherches, je combine les représentations sur la base des périodisations « traditionnelles » (par exemple, pour le français : ancien, moyen, préclassique, classique…) et d'un découpement arbitraire en tranches chronologiques allant de la décennie au siècle, selon le degré de granularité souhaité. Ce sont les résultats obtenus qui permettent de décider quel découpage chronologique adopter : comme pour les dialectes, pour un phénomène linguistique donné, on ne peut savoir a priori où va se trouver l'isoglosse (dans ce cas, isoglosse chronologique et non spatial), c'est-à-dire qu'on ne peut pas être certain qu'il se

[9] Cela ne signifie pas que la linguistique cognitive constitue nécessairement un progrès sur l'approche philologique. Je pense même, au contraire, que les linguistes ont tout à gagner à une collaboration étroite avec les philologues.

trouve là où d'autres linguistes ont, sur la base d'un certain nombre de phénomènes, décidé qu'il était le plus juste de mettre la frontière. La périodisation établie par d'autres permet d'apprécier l'évolution du phénomène étudié – et inversement, le découpage obtenu à partir des données peut apporter des informations sur l'intérêt et l'adéquation des périodisations préexistantes.

2.2 Approche contrastive et linguistique romane

> si intristì un pò, voleva dirgli di parlare in dialetto, di non preoccuparsi che lo capiva, ma poi desistette. Non voleva sentirsi come un antropologo che va in mezzo ai selvaggi per vedere quanto sono selvaggi e loro, per fargli piacere, si mettono proprio a fare i selvaggi. (Gianni Biondillo, *Per cosa si uccide* [Parma : Ugo Guanda editore], 2004 : 84)

L'intérêt de la linguistique contrastive est qu'il est (relativement !) facile de construire une théorie du rapport entre sens et système, par exemple, sur la base d'une langue… mais que, dès que l'on compare les langues, on est bien obligé de prendre en compte la complexité des rapports entre langue et monde, entre système et fonction, etc. En étant plus optimiste, on peut supposer avec Mithun (2012 : 2181) – en étendant quelque peu son propos – que ces différences « reflect a deeper characteristic common to speakers of all languages ».

Mon intérêt pour la linguistique est naturellement contrastif, et ma formation est également contrastive et comparative, avec l'étude dès le collège et le lycée de quatre langues (allemand, espagnol, latin, grec) outre mes deux langues premières (pour ne pas dire maternelles : français et anglais), puis en cycle universitaire la découverte de l'indo-européen, du sanscrit, de l'occitan, de la linguistique comparée. L'émerveillement de l'italien pendant la thèse, et l'apprentissage depuis, à des degrés très divers, d'une longue série de langues (notamment portugais, roumain, polonais, serbe, thaï, hongrois, russe, néerlandais, khmer, malais, langue des signes française), n'ont fait que renforcer cet attachement à la diversité linguistique. Enfin, mon intérêt pour la linguistique contrastive – au sens de l'étude d'une question à partir de l'observation d'au moins deux langues – est lié à celui pour la variation. La linguistique contrastive rencontre encore un grand succès, comme en témoigne l'apparition de séries éditoriales dédiées (Marzo et al. 2012, Aijmer & Altenberg 2013), de conférences (par exemple la série de conférences Dia : Gand 2010, Copenhague 2012), ou encore de journaux comme *Languages in Contrast* (premier numéro en 1998).

Dans mes propres recherches, la comparaison interlangues a le plus souvent – mais pas exclusivement – pris la forme d'études romanes. Ainsi, dans ma thèse, j'ai étudié les systèmes prépositionnels des langues romanes pour tenter d'évaluer leur degré de ressemblance, et de voir quels mécanismes permettent de rendre compte des différences et des ressemblances entre les différents systèmes. J'ai ensuite travaillé sur les systèmes de locutions prépositionnelles dans les langues romanes (Fagard & Mardale soumis), sur les conjonctions de subordination romanes (Fagard 2011a [III : 293 sqq.], Fagard 2009c [III : 343 sqq.], Fagard

2008b [VI : 929 sqq.], Degand & Fagard 2012 [III : 269 sqq.], Evers-Vermeul et al. 2011 [VI : 895 sqq.], Fagard & Degand 2008 [III : 389 sqq.], Fagard et al. soumis), sur certains marqueurs discursifs romans (Fagard 2010b [IV : 587 sqq.], 2012b). Dans le cadre de mes recherches sur l'expression de l'espace dans les langues, et sur les marqueurs discursifs, j'ai également mené des études contrastives notamment entre français, italien, allemand, suédois, néerlandais et polonais (communication, Mortier & Fagard 2008), français, allemand et polonais (communication, Fagard, Kopecka & Sarda 2009), allemand et polonais (communication, Kopecka & Fagard 2014), français, piémontais, polonais, allemand, suédois et thaï (communications, Fagard et al. 2013a,b), polonais et serbe (communication, Kopecka, Fagard & Stosic 2015).

Cela explique l'importance de l'approche contrastive dans les recherches que j'ai dirigées, co-dirigées ou suivies officiellement : thèse de Daniel Rojas-Plata sur l'espagnol et le français médiévaux, thèse d'Emilia Tuuri sur le finnois, stage de Camille Corcia sur le hongrois, thèse de Johan Blomberg sur le français, le suédois et le thaï. Cela explique également l'orientation résolument contrastive de mes projets de recherche (notamment le projet « Espace et grammaticalisation », un projet innovant financé par l'Université Paris III Sorbonne Nouvelle, 2011-2012).

La comparaison n'est pas une fin en soi : elle doit surtout permettre de mieux évaluer la spécificité de chaque système linguistique. Elle n'exclut pas une étude plus fine et plus complète du phénomène choisi dans une des langues : au contraire, les deux approches sont complémentaires. L'approche contrastive est nécessaire parce qu'elle permet de vérifier que l'on n'a pas laissé de côté une partie du problème.

2.2.1 Approche onomasiologique, approche sémasiologique

Pour des langues apparentées, et de manière assez aiguë pour les langues romanes en raison des données dont on dispose sur le latin, se pose la question du choix entre deux approches, (i) *onomasiologique* (étudier les signes utilisés dans différentes langues pour une même fonction) et (ii) *sémasiologique* (étudier les fonctions remplies par un même signe dans différentes langues). La richesse des langues romanes est justement qu'elles offrent la possibilité de combiner ces deux approches. On peut étudier l'expression d'une même réalité dans différentes langues romanes, comme on pourrait le faire pour tout autre groupe de langues, apparentées ou non. Mais on peut aussi partir d'un étymon latin et comparer son évolution dans différentes langues romanes ; les travaux de Blank (1997) et Koch (2003, 2006), notamment, ont suffisamment montré l'intérêt de ces deux approches pour que j'évite de m'étendre sur ce sujet.

En revanche, dans des études contrastives sur des langues non apparentées, seule la première méthode est directement accessible ; et lorsque les langues sont apparentées mais que la langue mère n'est pas suffisamment documentée, on en est réduit à reconstruire l'étymon commun, et la part de conjecture que cela implique limite d'autant l'intérêt des résultats.

2.2.2 Le codage en contrastif

On peut retenir un *caveat* qui s'applique à la linguistique contrastive en général (Coseriu 1970) : le but de la linguistique contrastive ne peut être simplement de comparer un état dans une langue cible à ce que l'on connaît déjà dans une langue source. En d'autres termes, il ne s'agit pas de plaquer une analyse toute faite (et qui pourrait être adaptée pour la langue source) sur une langue cible (pour laquelle cette analyse peut être tout à fait inadaptée). Au contraire, si l'analyse a déjà été faite sur la langue source, l'analyse de la langue cible doit servir à questionner cette analyse précédente et, bien souvent, elle amène à la réviser au moins en partie. Une des vertus de la linguistique contrastive est en effet, à mes yeux, de forcer le chercheur à sortir de son carcan théorique habituel.

D'un point de vue méthodologique, on notera également que les difficultés de codage notées au § 1.2 sont démultipliées dès lors qu'il s'agit d'appliquer un même codage (ou des codages équivalents, ou même simplement suffisamment proches) à des systèmes linguistiques différents : soit avec une maîtrise linguistique nécessairement inégale, soit en collaboration, ce qui implique une phase préalable, où les chercheurs doivent trouver des repères scientifiques – théoriques, méthodologiques et terminologiques – communs. Par exemple, lorsque l'on étudie l'expression du mouvement dans différentes langues d'Europe, la question du statut des préfixes verbaux est assez épineuse. On range généralement dans la même catégorie les particules verbales de l'anglais (*walk* ***in***, *walk* ***out***), les préverbes de l'allemand (***(hin)ein****laufen*, ***(hin)aus****laufen*) et du polonais (***w****chodzić*, ***wy****chodzić*), tandis que les préverbes du français ne sont pas pris en compte car trop peu productifs, et trop lexicalisés (Kopecka 2006) ; effectivement, on traduira difficilement *he walked in* par #*il a marché dedans*, ou encore **il a emmarché*, et par ailleurs on ne peut pas analyser *il est entré* comme comportant un préfixe verbal correspondant à l'anglais *in*, à l'allemand *ein* ou au polonais *w-*. Il s'agit en fait d'un continuum qui va d'adverbes rattachés sémantiquement (plutôt que syntaxiquement) au verbe, comme ce qu'on reconstruit pour l'indo-européen, à des préverbes inséparables et dont le sens n'est parfois même plus identifiable. Mais en pratique, pour l'analyse statistique des résultats, il faut bien trancher, et cela pose problème.

Pour illustrer un point plus sémantique, si l'on essaie de voir la part des verbes de manière dans l'expression du mouvement, on peut se poser la question de l'analyse à adopter pour le suédois *gå* « marcher, aller » : tandis que l'anglais et le français marquent nettement l'opposition entre *aller* ou *go* et *marcher* ou *walk*, ce n'est pas le cas pour le suédois. Le verbe *gå* n'est pas tout à fait désémantisé, comme le montre le contraste existant avec *åka* (comme entre l'allemand *gehen* et *fahren*), ou encore l'exemple (1), des données trajectoire (§4.2) :

(1) (suédois)

en	kvinna	**kommer**	**gående**	**ur**	en	grotta
'IND	femme	venir.3SG.PRS	aller.PART.PRS	hors.de	IND	grotte

på	en	skogsstig
sur	ind	forêt.sentier'

« une femme **vient en marchant** hors d'une grotte sur un chemin forestier »
(Traj23_sw_06)

Mais il l'est partiellement, comme le montrent les constructions du type (2) :

(2) (suédois)

en	kvinna	som	**går**	ifrån	ett	träd	**springande**
‘IND	femme	REL	aller.3SG.PRS	depuis	IND.NT	arbre	courir.PART.PRS’

« une femme qui **part** d’un arbre **en courant** »
(Traj33_sw_04)

Des deux interprétations possibles (pour la description du déplacement humain) du verbe *gå*, le premier exemple semble exclure la possibilité d’une lecture déictique, qui serait incompatible avec le verbe *komma* ‘venir’, et le second la possibilité d’une lecture comme verbe de manière de mouvement, qui serait incompatible avec le verbe *springa* ‘courir’. Dans d’autres exemples du même corpus, il est plus difficile de trancher, en l’absence d’indices aussi nets.

Ce problème de verbes de mouvement proches étymologiquement mais dont seul certains ont conservé le poids sémantique (que l’on peut supposer) d’origine se retrouve ailleurs. Ainsi, en polonais, les verbes formés sur *iść* ‘aller (en marchant)’ s’opposent typiquement aux verbes formés sur *jeździć* ‘aller (en voiture, en train…)’, une opposition que l’on retrouve en russe et en tchèque, mais pas en serbe où les verbes de la série *iti* ont été neutralisés sémantiquement, et signifient ‘aller’, ce qui explique des constructions comme *iti autom « aller en voiture »*, impensables en polonais (**idę samochodem* « je vais (à pied) / je marche en voiture »).

2.2.3 La question de la maîtrise de la langue

La linguistique contrastive suppose un certain niveau de maîtrise des langues prises en compte. Il me semble qu’il n’y a pas un lien direct entre le niveau de maîtrise d’une langue et la qualité des analyses menées ; cependant, il me paraît évident que la finesse de l’analyse linguistique est conditionnée dans une large mesure par la connaissance de la langue elle-même. De ce point de vue, la romanistique présente à la fois un grand attrait et quelques risques. L’attrait est la relative facilité avec laquelle on peut passer d’une langue à l’autre, une fois que l’on a acquis les bases de quelques langues romanes (cf. la méthode panromane et le projet Eurom[10]). Le risque est celui d’une fausse aisance, les faux-amis étant légion, que ce soit sur le plan lexical ou sur celui des constructions : on pensera par exemple au passé catalan, formé sur le même mode que le futur français (*vaig fer*, littéralement ‘je vais faire’, signifie ainsi « j’ai fait », « je fis »).

2.2.4 La question de la granularité

Cette question se pose de manière plus aiguë encore pour les dialectes et langues régionales : non moins intéressants que les standards, ils sont malheureusement moins accessibles, d’une part parce qu’ils sont moins bien décrits, d’autre part parce qu’ils sont moins bien documentés (cf. Berthele, op.cit.), enfin parce que les locuteurs eux-mêmes sont moins faciles à consulter. On peut noter tout de même que, dans le domaine des langues romanes, on dispose de

[10] Voir le site : http://www.eurom5.com/.

données écrites, de corpus oraux et écrits pour de nombreux dialectes, dont j'ai pu bénéficier par exemple pour une étude sur l'évolution des prépositions du type « hors de » et « à l'exception de » (communication, Fagard 2008a). Reste le problème du repérage des ressources, de leur accessibilité et de leur compatibilité ; pour que ce type d'approche soit réellement pertinent, il faut à mon avis recourir à des collaborations, comme celle que j'ai entamée il y a quelques années avec Massimo Cerruti (Università di Torino) sur le piémontais, qui se poursuit avec plusieurs articles en préparation, ou encore celles que j'ai lancées dans le cadre du projet « Espace et grammaticalisation ». Ce dernier, dont le but était de poursuivre la réflexion sur le localisme entamée dans le cadre de ma thèse, reposait entièrement sur la collaboration entre linguistes d'horizons divers, aussi bien concernant leur langue de travail (langues romanes, germaniques et slaves) que leur formation (linguistique de corpus, linguistique cognitive, linguistique contrastive). Malgré la durée limitée du projet (un an), les résultats scientifiques ne sont pas négligeables : outre l'organisation d'un atelier dans le cadre de la conférence Aflico IV en 2011, ce projet a renforcé mon intérêt pour les marqueurs discursifs et les déictiques, deux sujets sur lesquels je travaille en ce moment dans le cadre de divers projets : projet du laboratoire – opération « marqueurs discursifs », projet du Labex TransferS sur les topiques du discours, projet TUL (2015-2016) sur la deixis. Tous ces projets sont fondés sur une approche contrastive et typologique ou, plus précisément, sur une approche « microtypologique » (cf. section 6).

2.2.5 Les corpus en linguistique contrastive

L'étude sur corpus en linguistique contrastive est particulièrement complexe à mettre en œuvre. Plusieurs possibilités s'offrent au chercheur, notamment l'utilisation de corpus **parallèles** ou **comparables**, ou encore de corpus construits à partir de données élicitées.

Parasol (http://www.slavist.de/)
Sprik (http://www.hf.uio.no/ilos/english/services/omc/)
Europarl (Koehn, 2005, http://www.statmt.org/europarl/)

Quelques corpus parallèles

Un des problèmes majeurs est celui de la compatibilité des corpus, en particulier l'importation des données issues de requêtes spécifiques : celle-ci est parfois entravée par les concepteurs eux-mêmes, ce qui s'explique probablement par les relations parfois problématiques avec les éditeurs. Ainsi, les gestionnaires de la BFM ont dû fermer la base pendant un certain temps, puis renoncer à une partie des textes qui la composaient, pour des questions de droits, suite aux pressions de plusieurs éditeurs, notamment Droz. Des questions de droits qui paraissent absurdes en regard de la date des textes et de l'utilisation non commerciale qui en est faite.

Une autre solution peut être l'utilisation d'internet, notamment en version améliorée – je pense ici à la série de corpus Tenten (disponibles à travers *The Sketch Engine*, http://www.sketchengine.co.uk/), qui compensent par la masse des données les limites de leur composition – on pourrait en effet les appeler des corpus sales : il s'agit de pages web récoltées par des logiciels dédiés, qui se limitent principalement à les nettoyer en ôtant divers éléments parasites, à les référencer et à ôter les doublons. Ce type de corpus peut se révéler

intéressant pour étudier des phénomènes rares : par exemple, étudiant l'existence d'une forme de marquage différentiel de l'objet en français, nous avons constaté que les corpus traditionnels, oraux comme écrits, ne permettent pas d'en rendre compte, tandis que le corpus Tenten en contient un certain nombre d'exemples (Fagard & Mardale 2014 [I : 10-16]). Or il est clair que pour des phénomènes en phase d'apparition, ou marginaux, on a besoin de corpus de ce type. Dans le cas du marquage différentiel de l'objet, sans le corpus Tenten nous aurions été limités aux occurrences récupérées à la volée, pour reprendre l'expression de Corminboeuf (2008). Il est possible de procéder ainsi, et les occurrences comme la suivante sont précieuses car elles fournissent la preuve de l'existence du phénomène :

(3) (français contemporain)
Voyez l'hypocrisie : on nous empêche **à nous** de produire des maïs OGM, par contre les commerciaux sont autorisés à vendre le maïs OGM
(locuteur du Lot et Garonne, *France Info*, 22 février 2012)

Mais elles ne suffisent guère pour établir les limites et les conditions d'apparition de ce dernier. Cela justifie donc, de mon point de vue, l'utilisation de ce type de corpus.

2.3 L'approche typologique

Pour étudier la variation linguistique, l'approche typologique pourrait sembler plus indiquée encore que la linguistique diachronique ou contrastive. Elle a en effet elle aussi porté ses fruits dans les dernières décennies, avec la mise au jour de nombreuses régularités dans les langues du monde, des universaux statistiques (Dryer 1998), la question des universaux conditionnels (à commencer par ceux de Greenberg 1966a,b), etc. La production de descriptions (dictionnaires, grammaires et corpus) de langues inconnues ou peu connues, qui a commencé il y a quelques siècles déjà (Colombat 2007), a elle aussi explosé dans les trente dernières années, et contribue à la possibilité d'une approche typologique englobante que j'appellerais volontiers typologie de surface, prenant en compte une quantité de langues parfois impressionnante, de 50 à plusieurs centaines (Svorou 1994, Haspelmath 1997, Corbett 2000, etc.). La production d'atlas typologiques comme le WALS (Haspelmath et al. 2005, Dryer & Haspelmath 2011) ou l'atlas comparatifs des langues créoles (Michaelis et al. 2013) va également dans ce sens.

2.3.1 Apports et limites de la typologie

Dans la mesure où le rôle de la typologie est d'apporter une ouverture sur la diversité linguistique, comme l'indique l'introduction du WALS, elle le remplit parfaitement :

> [we wish to provide] an overview both of the typological diversity of the world's languages — including reference to phenomena that might be judged highly unusual by the standards of the world's best known languages — and of the ways in which that

> diversity is often patterned geographically, with particular areas often being characterized by particular feature values or particular combinations of feature values (whence the term "areal typology"). (Comrie et al. 2013)

C'est également un des buts des ouvrages généraux de typologie :

> one of the aims of the handbook, as stated in the preface, is to help experts on individual languages 'place their language within the space of what is possible and common in the languages of the world'. (Haspelmath et al. 2001 : vi, cité par Siewierska 2004 : 286)

Il me semble, cependant, que la typologie ne peut valoir que si ses résultats sont confrontés à ceux de la linguistique générale. Ainsi, les généralisations proposées dans le WALS posent un certain nombre de problèmes, notamment celui des catégories appliquées de manière trop uniforme, une question qui revient souvent dans les écrits typologiques, comme en attestent les citations suivantes :

> Cette typologie [a] un défaut majeur : elle implique l'identité des fonctions dans les langues différentes. (Coseriu 1997 : 223)

> La plupart des auteurs font usage sans états d'âme des notions grammaticales traditionnelles telles que sujet, objet direct, objet indirect (notion usuelle en grammaire anglaise), génitif, adjectif, pronom, etc., sans se soucier de savoir quel est leur statut et si leur emploi est légitime dans la comparaison des langues. Ils ne font en cela que suivre l'exemple de Greenberg, qui a donné le ton dès son article initiateur. (Lazard 2008 : 17)

> two potential dangers in typology, namely 'premature statistics' and 'not comparing like with like'. (Corbett 2005 : 25)

> Categories are not cross-linguistically comparable. (…) The emerging categories [dans l'acquisition du langage] are not identical from language to language (Raible 2001 : 5)
> Descriptive formal categories cannot be equated across languages because the criteria for category assignment are different from language to language. (Haspelmath 2010 : 663)

Ce problème est le même que celui que nous avons évoqué pour la sémantique ; dans un cas comme dans l'autre, je suis assez sceptique à l'égard du remède proposé dans Hartmann et al. (2014), qui consiste finalement à sauter l'étape intermédiaire des catégories inter-langues pour passer directement de la microanalyse de chaque langue (des *microrôles*) à l'identification de regroupements (*clusters*) sur la base d'une analyse statistique comparée. En prenant le parti de ne pas comparer les langues, et en se contentant d'une analyse spécifique pour chaque langue, pour laisser l'analyse statistique faire émerger les points communs et les différences entre les langues, les auteurs évitent peut-être le danger des catégories faussement partagées, mais tombent à mon avis dans un autre piège, en abandonnant finalement le regard de linguiste pour celui du statisticien. Ce danger se trouve de plus renforcé par le fait que, comme dans nombre d'études typologiques, ce qui est étudié, ce n'est pas la langue dans sa variation, mais les constructions typiques de chaque langue telles qu'on peut les trouver dans les grammaires.

Un autre point me semble particulièrement problématique : faut-il vraiment mettre sur le même plan les analyses de langues bien connues et documentées et celles qui ne portent que sur une description limitée, comme c'est le cas par exemple dans le WALS, mais aussi dans bien d'autres survols typologiques ? Au nom de l'exigence statistique de couverture maximale

et de différenciation maximale[11], ne risque-t-on pas en visant la scientificité de perdre en qualité des analyses linguistiques ? Plus précisément : entre une collection de données tirées d'ouvrages de référence pour une centaine de langues, et une analyse de données linguistiques (sur corpus, ou bien élicitées) pour quelques langues – ce que Colette Grinevald appelle la *working typology* (Grinevald 2011) –, peut-on être certain que c'est la première méthode qui est la plus scientifique, et celle qui nous en apprend le plus sur le fonctionnement du langage ? Si la prise en compte de nombreuses langues me semble intéressante, elle nécessite à mon sens un certain nombre de garde-fous. Les données des grammaires sont à utiliser avec précautions – en particulier pour une langue que l'on ne maîtrise pas. De plus, la confrontation de traditions linguistiques différentes peut entraîner des incompréhensions, avec des équivalences partielles entre termes identiques, ou bien des incompatibilités terminologiques. J'ai pu en prendre conscience en préparant l'index du *Dictionnaire des langues* (Bonvini et al. 2011) : la mission confiée par Emilio Bonvini était de rendre compte des correspondances terminologiques entre les différentes traditions ; dans un ouvrage rédigé par une centaine d'auteurs, auxquels de surcroît les éditeurs avaient pris le parti de laisser une grande liberté, il était inévitable que certains appellent *sibilante* ce que d'autres appellent *sifflante*, que les uns se réfèrent à la *base* plutôt qu'au *radical*, sans parler de cas plus complexes comme *admiratif~exclamatif*, *allongement~métathèse*, *article~déterminant*, *phrase~période~proposition* : pour chaque paire ou série de termes, la question est bien sûr celle de leur (non-)équivalence ou (in)compatibilité éventuelles.

2.3.2 Conditions d'une typologie scientifique

Etant donné ces réserves, je suis somme toute assez d'accord avec Lazard concernant aussi bien les qualités que les défauts de la typologie moderne :

> La typologie, telle qu'elle s'est développée depuis quarante-cinq ans, a accumulé beaucoup de données, d'hypothèses et de conclusions plus ou moins affirmées. Ces résultats forment une masse considérable. On doit cependant se demander s'ils sont tels que l'on puisse les qualifier de scientifiques. Je pense que non. (Lazard 2008 : 16)

Mais je pense qu'il est possible de faire de la typologie de manière scientifique, à certaines conditions. Ces conditions portent d'une part sur la pratique de la typologie elle-même, d'autre part sur l'utilisation des résultats issus de cette typologie :

(i) **une maîtrise minimale des langues incluses dans le corpus :** on peut admettre qu'un linguiste qui a l'habitude des grammaires finit par pouvoir se faire une idée acceptable d'un système linguistique assez vite, mais le recours exclusif à quelques grammaires ou à quelques informateurs me semble problématique ;

(ii) **une documentation suffisante pour ces langues :** parce qu'une grammaire ne suffit généralement pas à venir à bout de la description d'une langue, que la

[11] On pourra citer par exemple van der Auwera (1998 : 1-2) : « The evidence must be cross-linguistic, ideally arrived at on the basis of a justified sample of the languages of the world or of the region studied. » ; il note plus loin (ibid : 10) « This method yields 'variety samples', i.e., samples that are geared towards maximal variety. It attemps to avoid a genetic bias (…) ».

consultation de plusieurs grammaires permet de se faire une idée plus précise du degré de variation d'une langue, que l'existence de corpus permet, comme on l'a vu, de compenser les lacunes de ces dernières ;

(iii) **l'utilisation de données diversifiées** – ce qui va dans le même sens que le point (ii) : ne partir que d'une série de grammaires, ou d'un questionnaire[12], limite forcément la portée des résultats, et il faudrait idéalement combiner les résultats de plusieurs méthodes (grammaires, questionnaires, élicitation, corpus) ;

(iv) **la prise en compte des différences interlinguistiques :** cela peut paraître une évidence, mais la comparaison des langues ne doit pas se faire sur l'hypothèse que tout se vaut (c'est-à-dire que toutes les langues n'ont pas nécessairement la catégorie « adjectif », par exemple, et que cette catégorie n'a pas les mêmes caractéristiques d'une langue à l'autre[13]) ;

(v) **une exploitation prudente des résultats :** en particulier au cas où les conditions (i) et (ii) ne seraient pas remplies, des précautions extrêmes sur la possibilité d'extrapolation des résultats ; en tout état de cause, une certaine prudence dans les généralisations.

2.3.3 La question de la diversité

Cela n'invalide en rien les pratiques typologiques qui amènent leurs auteurs à produire des grands atlas portant sur un ou plusieurs points du système linguistique : ces résultats demeurent un excellent point de départ pour mener des études plus détaillées. Cela signifie-t-il que nous devons renoncer à inclure des langues peu connues ou peu décrites ? Non, et j'irais même plus loin : je suis tout à fait d'accord avec les analyses comme celles d'Evans & Levinson (2009) : il faut absolument élargir le socle des langues bien décrites. Mais ce n'est qu'une fois qu'une langue est bien décrite, et que l'on dispose pour l'analyser d'outils philologiques de référence, que l'on peut à mes yeux légitimement l'inclure dans des études typologiques. Cela ne rend que plus urgente l'entreprise de documentation des langues en voie de disparition, entreprise qui doit rester secondaire à l'aide aux individus et au respect de l'Autre (Kapuściński 2009) : les personnes qui abandonnent leur langue sont souvent dans des conditions défavorisées. Et surtout, il faut garder à l'esprit le point (v) et donc rester prudent lorsque l'on inclut des données non contrôlées. Il y a d'autres possibilités comme la collaboration avec des spécialistes de chaque langue, comme cela a été le cas dans le cadre du projet Trajectoire (financé par la fédération TUL du CNRS, voir §4.2), mais cela n'est pas compatible par exemple avec les sélections « statistiques aléatoires » de langues qui sont de

[12] Cf. le travail de Ricca (1993) sur la deixis, fort intéressant au demeurant.

[13] Pour donner un exemple concret, le fait que les cas en polonais et en hongrois soit mis sur le même plan dans les cartes du WALS me semble problématique, et ne me paraît acceptable que comme première hypothèse de travail assez grossière. En hongrois, les cas sont des morphèmes identifiables, et l'accusatif par exemple prend la forme d'une dentale sourde ajoutée à la base, avec quelques accidents (comme pour l'accusatif de *pain* par exemple : le nominatif *kenyér* donne *kenyeret* à l'accusatif). En polonais, le cas est une reconstruction du linguiste sur la base d'alternances qui varient d'un groupe de mots à l'autre. On peut voir aisément qu'il est impossible de le déduire sur la base de la *forme* : le nominatif *kobieta* « femme » donne *kobietę* à l'accusatif (substantif féminin), le nominatif *chłopiec* « garçon » donne *chłopca* (substantif masculin non animé), et qu'il faut donc partir des *fonctions* pour identifier le cas.

règle en typologie (avec différentes techniques d'échantillonnage, cf. van der Auwera (1998) et Croft 2003 : 19 sqq.).

2.4 La linguistique cognitive

Etonnamment, le gros des publications de linguistique cognitive porte sur une seule langue, d'après Catherine Fuchs :

> most works on language as a cognitive activity are based on a **single** language (…) treated as some arbitrary representative of language in general. (Fuchs 1999 : 4)

Or, comme elle le montre bien, il y a un lien naturel entre linguistique cognitive et approches contrastives, souligné aussi par Geeraerts & Kristiansen (2014) :

> We will argue that Cognitive Linguistics embodies a far-reaching paradigm shift in linguistics, and that the interest in interlinguistic and intralinguistic language variation constitutes the cornerstone of that paradigm shift. (Geeraerts & Kristiansen 2014 : 203)

Il me semble naturel de commencer par étudier le langage comme un système indépendant du monde et de l'homme, un peu comme un mathématicien cherche à établir des lois qui ne valent que comme abstraction indépendante. C'est là une partie fondamentale du travail du linguiste, qu'il étudie la syntaxe, la morphologie, la sémantique, etc., et que ce soit en synchronie ou en diachronie. Mais il est également important de prendre en compte les paramètres cognitifs qui entrent en jeu, en suivant l'hypothèse de Leonard Talmy (voir Bergen et al. 2010 pour une analyse critique) :

> In terms of major cognitive systems, the language system and the culture system (…) were the last to evolve. In forming, they may have copied, or developed connections to, mechanisms of cognitive structuring already present for other major cognitive systems, ones long in place, such as those of visual perception, motor control, and reasoning/inferencing. In that case, the language system would have incorporated some of those extant structuring mechanisms. (Talmy 2003 : 37-38, vol. I)

Comme le note Catherine Fuchs, l'approche cognitiviste revient à rejeter la vision modulariste du langage (ce qui est d'ailleurs confirmé par l'imagerie cérébrale, cf. Stowe et al. 2004 : 1032-1035), et à le penser plutôt

> as a property which emerges from the general mechanisms of cognition and has numerous homologous relationships with other cognitive activities including perception. (Fuchs 1999 : 12)

Il nous revient d'analyser l'importance des relations entre cognition et langage : **dans quelle mesure le fonctionnement du cerveau humain détermine-t-il la forme et la structure du langage ?** Inversement, dans quelle mesure la langue détermine-t-elle les conditions de la pensée, comme le suggère la citation ci-dessous ?

> Further, research has shown that culture and linguistic diversity have profound effects on differences in core cognitive domains like spatial cognition (Haun, Rapold, Call, Janzen,

> & Levinson, 2006; Levinson, 2003), perception (Winawer et al., 2007), abstract representations (Boroditsky & Gaby, 2010), numerical knowledge (Frank, Everett, Fedorenko, & Gibson, 2008), and perspective taking (Wu & Keysar, 2007). (Liszkowski et al. 2012: 699-700)

Ces questions renvoient au vieux débat sur la part respective de nature et culture, inné et acquis, cerveau et contexte (milieu, environnement), et en fin de compte au débat Sapir-Whorf sur la relativité linguistique, une problématique rendue célèbre par Sapir et Whorf, et qui est à l'origine d'un débat sans cesse renouvelé, repris entre autres par Berlin & Kay (1969) (voir Fortis 2010). Cette problématique est en fait bien plus ancienne, et présente notamment dans les écrits de Wilhelm von Humboldt :

> Ihre Verschiedenheit [= der Sprachen] ist nicht eine von Schällen und Zeichen, sondern eine Verschiedenheit der Weltansichten selbst.
> Leur diversité [= des langues] ne réside pas (que) dans les sons et les signes, mais relève de regards différents sur le monde (von Humboldt 1820, cité par Trampe 2013 : 52 ; je traduis)

Ce débat est loin d'être dépassé, comme le montre la polémique récente (évoquée plus haut) sur l'importance de la description de langues inconnues, lancée par Nicholas Evans et Stephen Levinson. Ces derniers montrent, à partir d'un état de l'art sur la linguistique descriptive, contrastive et typologique, l'importance du travail de description de nouvelles langues pour l'avancée scientifique en linguistique. Pour eux, l'urgence n'est pas à la recherche des universaux mais au défrichage des quelques milliers de langues encore non décrites :

> [Our] article summarizes decades of cross-linguistic work by typologists and descriptive linguists, showing just how few and unprofound the universal characteristics of language are, once we honestly confront the diversity offered to us by the world's 6-8000 languages. (Evans & Levinson 2009)

Cette charge virulente me semble exagérée bien que je partage en partie l'analyse des auteurs (cf. la section précédente) : de mon point de vue, l'absence d'universaux parfaits n'invalide pas l'intérêt de la 'typologie de surface', à la condition (v) notée ci-dessus (§ 2.3). Je pense en effet qu'à partir des variations que l'on peut observer en comparant les langues (et en étudiant les langues non décrites, mais aussi les langues déjà décrites, pour lesquelles les données sont d'autant plus fiables que l'on dispose déjà d'un corpus de travail – textes, dictionnaires, grammaires, etc.), ce que Croft appelle « patterns of variation » (2000 : 86), on peut remonter effectivement à la nature du langage : le but est de (mieux) comprendre la nature du langage pour (mieux) comprendre le fonctionnement du cerveau humain en général. C'est là le but de la linguistique cognitive, du moins telle que la conçoit Talmy, qui parle d'une « trajectory toward unification with the psychological » (2000 : 3) au sens où, à long terme, la linguistique cognitive doit non seulement expliquer les structures cognitives propres au langage, mais permettre de remonter plus loin vers les bases de la cognition humaine. C'est exactement ainsi que je conçois mon approche de la langue : **je suis cognitiviste en ce sens que je me situe dans une trajectoire qui veut tendre vers la compréhension non seulement du fait linguistique, mais des structures cognitives sous-jacentes**. Cependant, ma pratique linguistique se démarque au moins dans une certaine mesure des approches de certains linguistes qui se réclament de la linguistique cognitive, et dont l'appareil théorique ne me

semble pas offrir toutes les garanties du principe de vérifiabilité. On peut ainsi se poser la question pour les schémas d'image (*image schemas*, Langacker 1987, 1991, 1997, Johnson 1987, Tyler & Evans 2003, Evans 2010), définis comme suit par Johnson (1987 : xiv) :

> An image schema is a recurring dynamic pattern of our perceptual interactions and motor programs that gives coherence and structure to our experience.

Que représentent-ils au juste ? Malgré leur intérêt pour la compréhension du fonctionnement cognitif (voir notamment les indices positifs apportés par plusieurs séries d'expériences sur leur validité cognitive, cf. Rohrer 2005), on peut noter avec Hampe (2005ba : 3) que le débat concernant le statut des schémas d'image a été intense, à tel point que

> [n]either the original characterization, nor the entire subsequent research to date can thus be said to have provided the Cognitive-Linguistics community with a set of clear-cut criteria to set image-schematic representations apart from other basic or schematic concepts.

Le livre de Hampe (2005a) contient d'ailleurs plusieurs contributions très intéressantes du concept de schéma d'image, rappelant entre autres qu'il est important de ne pas laisser de côté la dimension de variation socio-culturelle (Zlatev 2005). Quoi qu'il en soit, le flou qui entoure ce concept depuis son introduction par Johnson & Lakoff le rend plus difficile à manier (Hampe 2005b). La question mérite donc d'être creusée davantage.

II Questionnements (et quelques débuts de réponse)

Cette deuxième partie est consacrée à la description des trois principaux « champs d'investigation » qui forment le cœur de mes recherches : le *localisme* (§3), la *grammaticalisation* (§4) et l'*espace* (§5). Pour chacun de ces champs, je situe mes recherches, puis je présente mes principaux résultats, sans chercher l'exhaustivité, en tirant parti des études publiées ainsi que, ponctuellement, de quelques études non publiées. Les trois domaines étant reliés, certaines études peuvent apparaître plusieurs fois. Comme on le verra, les résultats rapportés ici reposent pour une bonne part sur la combinaison de deux approches ou plus parmi celles qui sont décrites dans les sections précédentes. Tout n'est pas 'diachro-contrastivo-typolo-cognitif' ; cela n'aurait guère de sens. Mais je considère que la combinaison des approches permet souvent à chacune de contrôler les résultats de l'autre, et aide en fin de compte le chercheur à prendre du recul sur ses propres résultats.

Com as palavras todo cuidado é pouco,
mudam de opinião como as pessoas.

« Avec les mots, toute précaution est vaine,
ils changent d'opinion comme les personnes »

(José Saramago)

3 Le localisme

Dans cette section, je présente brièvement le localisme (3.1.1), puis j'illustre cette thématique avec un choix de résultats issus de mes recherches (3.1.2).

3.1 Panorama des études actuelles et antérieures – Le localisme : introduction

3.1.1 Définition préliminaire du localisme

Le localisme est l'idée que la conceptualisation des relations spatiales ou des déplacements (motion events) est sous-jacente à la structure grammaticale[14] :

> In the 1970s, some linguists went a considerable distance along this road under the banner of 'localism', treating grammatical categories such as tense and aspect as spatially grounded, as well as, for example, causatives, modals, transitivity, instrumental adverbs, possessive and existential constructions (cf. Lyons 1977: 718-724; Anderson 1971). (Chilton 2010 : 4)

Il semble bien qu'il s'agisse d'un phénomène cognitif :

> Spatial expressions are linguistically more basic, according to the localists, in that they serve as structural templates, as it were, for other expressions; and **the reason why this should be so, it is plausibly suggested** by **psychologists, is that spatial organization is of central importance in human cognition**. (Lyons 1977 : 718 ; je souligne)

L'universalité de ce phénomène, implicite dans la citation de Lyons, est parfois explicitée :

> Parmi les métaphores les plus structurelles, Lakoff insiste sur l'universalité de la conception du temps comme un double déplacement spatial. (Victorri 2004a : 87)
>
> **Across languages**, spatial terms come over time to acquire abstract meanings (Sweetser 1990), which results synchronically in words that are polysemous, with both spatial and abstract meanings. It has been suggested that this relation between abstract and concrete goes beyond mere language, such that understanding of abstract concepts is grounded in experiences with their concrete spatial counterparts, such as states being understood as containers (Lakoff 1993). (Bergen et al. 2010 : 80 ; je souligne)

[14] Sans exclure le lexique ; mais je m'intéresse ici principalement à la grammaire.

Il y a un lien entre l'explication cognitive et l'hypothèse universaliste (cf. la section 2.4 sur l'approche cognitive). Plus remarquable encore, cette centralité du sens spatial est non seulement considérée universelle, mais souvent **posée comme une évidence.**

> The intuition that prepositions specify 'locative' relationships cannot just be dismissed. (Langacker 2010)

Cette universalité supposée n'est pas nouvelle, et date même des origines du localisme. Je rappelle ci-dessous les origines de cette théorie.

3.1.2 Bref historique des théories localistes[15]

L'idée que le sens procède du concret à l'abstrait, ou encore (ce qui ne revient pas tout à fait au même) de l'espace à d'autres domaines, n'est pas neuve, loin de là. Depuis Hjelmslev (1935-7), on la fait remonter à Planude (13ème siècle), mais cette hypothèse a été réfutée (Chanet 1985 ; voir aussi Murru 1979), et l'origine réelle des théories modernes du localisme est à chercher beaucoup plus tardivement. En fait, d'après Fortis (2013), le localisme moderne prend ses origines dans la philologie du 19ème siècle avec notamment un livre sur les cas écrit par Wüllner (1827), qui considère que la signification fondamentale des cas est spatiale. Dès 1833, Bopp écrit notamment que « dans le principe, les cas n'exprimèrent que des relations dans l'espace ; mais on les fit servir ensuite à marquer aussi les relations de temps et de cause » (Bopp 1865 [1833] : 275, traduction de Fortis, 2013).

Dans le courant du 19ème siècle, les théories localistes évoluent, notamment avec l'avènement du psychologisme, car le localisme « a en effet une justification cognitive séduisante » : Wundt « sort du domaine indo-européen et s'intéresse aux marques de relation en général (adpositions, particules, position et composition en font partie) » (Fortis 2013). Il est présent pendant tout le 20ème siècle, dans les travaux de Hjelmslev sur les cas (1935-7), Anderson (1971), Lyons (1977), Groussier (1997), O'Keefe (1996) et d'autres.

3.1.3 Etat de l'art – le localisme aujourd'hui

On peut trouver des arguments non linguistiques à l'appui du localisme : ainsi de l'argument philosophique selon lequel la pensée abstraite est fondée sur (et dans) la sensation, la perception, ou encore est vide si elle n'a pas un contenu expérientiel. Du côté de la neurologie, on a pu noter qu'il y a un système neurophysiologique dédié au traitement des relations spatiales (Landau & Jackendoff 1993, Lakoff & Johnson 1999 : 55), et les (linguistes) cognitivistes considèrent volontiers qu'il y a une base neurologique au localisme (Talmy 2003 : 37-38, I, 2.4), une hypothèse à mettre en lien avec la non-spécialisation des aires du cerveau :

[15] Cet historique ne vise qu'à donner une idée globale de l'importance épistémologique du localisme, et de brosser à grands traits sa place dans le paysage des sciences du langage aux 19ème-21ème siècles. Pour une étude plus aboutie, on se rapportera avec profit à Fortis (2013).

> « brain areas formerly thought to be purely sensorimotoric are turning out to have important roles in the so-called 'higher' cognitive processes ». (Rohrer 2005 : 165 ; voir également Mazoyer 2004)

Ces questions sont souvent reprises – sans lien avec les structure neurologiques –, particulièrement chez un certain nombre de linguistes étudiant les relations entre espace et langage (Svorou 1994, Haspelmath 1997, etc.[16]) ; elles sont parfois présentes uniquement en arrière-fond et sans même mention du terme de localisme, en particulier chez les cognitivistes :

> Although the term 'localism' is no longer used, many approaches in cognitive linguistics are consistent with this idea. (Chilton 2010 : 4)

L'approche cognitive du localisme (Sweetser 1990, Lakoff 1993, Langacker 2010, Evans 2010, Chilton 2010) se double d'autres approches, avec des travaux plus ou moins localistes en typologie (Svorou 1994, Haspelmath 1997), en psycholinguistique (notamment sur l'acquisition, Johnston & Slobin 1979, Slobin 1985, Morgenstern & Sekali 1997, McDonough et al. 2003), ou encore en diachronie (Groussier 1997).

Le localisme est également au cœur d'études plus ou moins critiques (par ex. Pottier 2001 : 18, Tenbrink 2007, Vandeloise 1986, Victorri 2010). Ces études ont permis de progresser d'une vision trop géométrique de l'espace à une vision plus réaliste (à mon sens) de la spatialité comme fonctionnelle (cf. Herskovits 1986, 1988, Vandeloise 1986, 1987, 1999, Victorri 2010), soulignant l'importance de la perception, cf. Parret (2009) : « la spatialisation est organisée à partir du *corps*, de ses *vécus* et de ses *esthésies* ». Cette approche est liée à la notion d'*embodiment* (cf. Zlatev 2008), que l'on peut sans doute faire remonter plus loin (Zlatev empruntant largement, par exemple, à Merleau-Ponty, et se réclamant explicitant de lui) ; on pensera par exemple à Michel Bréal :

> On doit commencer à voir à quel point de vue l'homme a agencé son langage. La parole n'a pas été faite pour la description, pour le récit, pour les considérations désintéressées. Exprimer un désir, intimer un ordre, marquer une prise de possession sur les personnages ou sur les choses – ces emplois du langage ont été les premiers. (Bréal 1924 [1897] : 243, cité par Nerlich 2014 : 618)

La version moderne de cette position n'est finalement pas si éloignée de celle de Bréal :

> People's subjective, felt experiences of their bodies in action provides part of the fundamental grounding for language and thought. Cognition is what occurs when the body engages the physical, cultural world and must be studied in terms of the dynamical interactions between people and the environment. Human language and thought emerge from recurring patterns of embodied activity that constrain ongoing intelligent behavior. We must not assume cognition to be purely internal, symbolic, computational, and disembodied, but seek out the gross and detailed ways that language and thought are inextricably shaped by embodied action. (Gibbs 2003 : 2)

Elle a pris son essor en psychologie puis dans d'autres domaines scientifiques (philosophie, sociologie, robotique, psychologie, sciences cognitives, neurosciences), surtout dans les vingt dernières années (voir entre autres Thelen et al. 2001, Thompson et al. 2013, Welsh 2014,

[16] Sans exclure Guillaume (1973 : 234), dans une optique théorique il est vrai très différente.

Pitts-Taylor 2015). L'*embodiment* a ainsi reçu depuis des confirmations de différents domaines (Iverson & Thelen 1999, Gibbs & Wilson 2002, Xue et al. 2015), y compris en linguistique (Bergen et al. 2007), mais toutes les recherches expérimentales ne vont pas nécessairement dans ce sens. Ainsi, Morgenstern & Sekali (1997) montrent que les enfants français utilisent des prépositions abstraites avant les prépositions spatiales, mais elles reconnaissent que les relations spatiales sont peut-être trop évidentes pour que les enfants ressentent le besoin de les expliciter. Du côté des arguments à l'appui, Groussier (1997) parle de « stérilité lexicogénétique » des relateurs non-spatiaux, à laquelle on peut opposer la « versatilité sémantique » des relateurs spatiaux, et montre qu'il y a pour les prépositions, en anglais, une asymétrie des évolutions sémantiques – ce qu'elle appelle la « primauté du spatial ». Cette asymétrie est conforme à l'hypothèse localiste, avec des glissements du spatial au non spatial et aucun glissement dans le sens inverse, confirmant ainsi en grande partie les chaînes de grammaticalisation proposées dans (Heine et al. 1991) :

OBJECT (BODY PART) > SPACE > TIME > QUALITY
PROCESS (MOTION VERBS) > SPACE

La question que je pose à cet égard est la suivante : **s'agit-il là de tendances interlinguistiques** (cf. Haspelmath 1997)**, ou réellement d'une règle valable universellement ?** Evidemment, il est difficile d'apporter des réponses « universelles », mais il suffit en revanche d'un contre-exemple valable pour démontrer qu'il ne s'agit pas d'un phénomène universel ; on peut alors supposer qu'il s'agit plutôt d'une tendance. Je passe en revue dans la section suivante (§3.2) un certain nombre d'exemples décrits en détail dans mes publications, afin d'illustrer ce qui constitue et ne constitue pas, pour moi, un argument ou un contre-argument.

3.2 Etudes sur le localisme

La « primauté de l'espace » de Groussier et les chaînes de grammaticalisation notées ci-dessus constituent une réponse possible à la question des contraintes sémantiques sur l'évolution (*quelles sont les contraintes sémantiques portant sur l'évolution linguistique, dans le cadre de la grammaticalisation ?*, cf. §5). Un des buts de mes recherches est de chercher à vérifier le bien-fondé de ces réponses : ce n'est pas de prouver que l'espace est partout, mais de mesurer son importance, nécessairement limitée. Ce qui est intéressant, c'est de trouver ces limites, précisément.

Par ailleurs, l'idée n'est pas de trouver des réponses définitives, mais au moins des témoins (ou arguments) valables empiriquement, c'est-à-dire **des évolutions dont on puisse démontrer l'existence par des attestations dans des corpus diachroniques, par opposition à celles dont on ne peut que supposer l'existence par reconstruction rétrospective**. En effet, si l'on pousse le raisonnement localiste jusqu'à l'absurde, on obtient quelque chose de peu satisfaisant et, à mon sens, peu scientifique :

(a) on remarque que de très nombreux termes (prépositions, adverbes, etc.) ont des sens spatiaux et non spatiaux, dans toutes les langues étudiées à ce jour dans cette optique ;
(b) on part du principe que l'évolution naturelle est l'évolution ESPACE > AUTRES DOMAINES SEMANTIQUES ;
(c) on considère donc que, pour tous ces termes polysémiques ayant un sens spatial et un sens non spatial (ou plusieurs), le sens spatial est premier ;
(d) on a ainsi un vivier virtuellement infini de preuves du localisme !

Or il est certain que la reconstruction est souvent prise comme argument suffisant :

> It goes without saying that there are languages whose *have*-constructions are synchronically opaque. Nevertheless, in a number of these languages there is evidence in the form of pattern transparency (…) that allows us to **reconstruct** the schema involved. (Heine 1997b : 91 ; je souligne)

Je me suis donc efforcé, depuis ma thèse, de tester l'hypothèse localiste sans a priori, sur des éléments variés d'un point de vue formel et sémantique, dans plusieurs langues et à plusieurs périodes – prépositions simples et complexes, conjonctions et locutions conjonctives, marqueurs discursifs, adverbes, principalement dans les langues romanes (médiévales et modernes) mais aussi, ponctuellement, en anglais, allemand, néerlandais, polonais, serbe et russe. Dans le cadre de ces études sur corpus, j'ai surtout cherché à distinguer, aussi nettement que possible :

(i) les *preuves* (des éléments pour lesquels on trouve des attestations spatiales antérieures aux emplois non spatiaux)
(ii) les *illustrations potentielles* (des éléments pour lesquels on ne trouve pas ce type d'attestation, mais qui restent compatibles avec l'hypothèse localiste)
(iii) les *contre-exemples* (des éléments pour lesquels on trouve des attestations non spatiales antérieures aux emplois spatiaux)

J'illustre dans le reste de la section les résultats obtenus, sans viser l'exhaustivité, mais plutôt en fournissant quelques exemples représentatifs.

3.2.1 Du spatial au comitatif

Une première illustration des résultats que j'ai obtenus dans ce domaine est une série d'exemples d'évolution du spatial au comitatif – une chaîne sémantique assez répandue. Elle est rapportée par Svorou (1994 : 145 sqq.) pour des prépositions ayant comme sens de départ « derrière », et je l'ai notée (Fagard 2010a) pour des prépositions gallo-romanes ayant pour sens principal « avec » mais toutes issues du latin *apud* « près de » (4) :

(4)
ancien français *o*, *atout*, *otout*, *avuecques*, français moderne *avec*
catalan médiéval *amb*
occitan médiéval *amé*

Je l'ai également observée pour un certain nombre de prépositions dérivées de *coste* et *lez*, mais de manière inégale. L'extension est déjà présente au 12ème siècle pour les prépositions

lez et *delez*, avec une fréquence certes assez faible (une quinzaine d'occurrences dans notre corpus, soit 3 % des occurrences environ) et comme l'illustre l'exemple suivant :

(5) (ancien français)
Quant ge ai **delez moi** ma fille,
tot le mont ne pris une bille ;
c'est mes deduiz, c'est mes deporz,
c'est mes solaz et mes conforz,
c'est mes avoirs et mes tresors,
je n'ain tant rien come son cors.
« Quand j'ai **à mes côtés / avec moi** ma fille, je ne donnerais pas une bille pour le monde entier : c'est mon plaisir, mon bonheur, ma joie, mon confort, mon avoir, mon trésor, je n'aime rien autant qu'elle »
(BFM, *Erec et Enide*, v. 541, 1170)

Elle est bien plus tardive pour *coste* et ses dérivés, et on n'en trouve trace que dans un exemple du 14ème siècle :

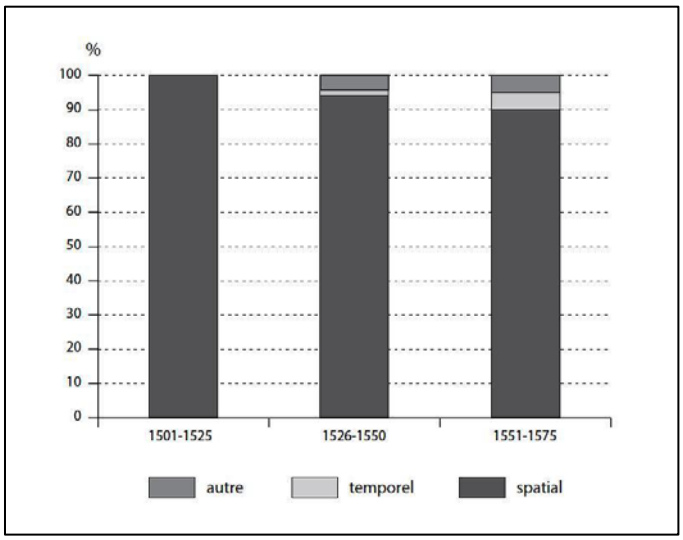
(6) (moyen français)
Et en la fin de son sermon dit ainsi que il avoit leue la Bible et les livres qui vont **encoste la Bible**, ne onques n'avoit veu, ne ou livre des creans ne ou livre des mescreans, que nul royaume ne nulle seigneurie feust onques perdue ne changee de seigneurie en autre ne de roy en autre fors que par defaut de droit.
« Et il dit ainsi, à la fin de son sermon, qu'il avait lu la Bible et les livres qui vont **avec la Bible** et que jamais il n'avait vu, ni dans le livre des croyants ni dans celui des mécréants, qu'aucun royaume ou seigneurie soit perdu ou change de seigneur ou de roi autrement que par injustice »
(BFM, Jean de Joinville, *la Vie de Saint Louis*, p. 328, 1305)

Cependant, dans le cas des prépositions *coste* et *lez* et de leurs dérivés, il faut bien noter que les deux lectures sont compatibles : dans ces exemples, une interprétation purement spatiale reste possible. Il s'agit donc de glissements sémantiques sans suite, c'est-à-dire qu'on ne passe pas au troisième stade de l'inférence contextuelle selon Heine (2002), où le sens de départ est exclu. Les cas des prépositions dérivées sur *coste* et *lez* ne constituent pas des arguments forts pour l'hypothèse localiste, puisque le sens spatial reste prépondérant, avec des effets de sens qui ne semblent pas systématiques.

3.2.2 Du spatial au temporel

Une seconde illustration de mes résultats est une évolution bien connue et largement décrite (pour ne citer que quelques références, voir Haspelmath 1997, Boroditsky 2001, Boroditsky & Ramscar 2002) est l'utilisation fréquente des prépositions spatiales avec un sens temporel ou bien, pour le formuler différemment, le fait que les prépositions ou adverbes d'origine spatiale prennent fréquemment un sens temporel. Mais bien souvent il est difficile de montrer clairement qu'il y a eu un glissement de sens, parce que – à la différence de certaines évolutions sémantiques – il ne semble pas y avoir eu de phase monosémique, ou du moins cette phase n'est pas attestée. C'est pourquoi, parmi les prépositions (*vers*, *envers*, *devers* et autres, cf. Fagard 2002 [I : 137 sqq.], 2010) et adverbes (*au fond*, Fagard & Sarda 2014 [IV : 446 sqq.] ; adverbes déictiques dans les langues slaves : communication, Fagard, Sikora & Stosic 2011) pour lesquels j'ai observé un glissement du spatial au temporel, je choisis *dans*

pour illustrer ce phénomène : son émergence tardive, au 16ème siècle, permet de montrer clairement qu'il y a effectivement une évolution, comme on le voit dans le graphique (2) ci-dessous.



Graphique 2 : Sémantisme de *dans* au 16ème siècle (Frantext, 585 occurrences, Fagard & Sarda 2009 [I : 98])

Le premier exemple (et le seul pour la première tranche chronologique, dans le graphique 2 ci-dessus) que l'on trouve dans Frantext est clairement spatial :

(7) (moyen français)
Et, pour s'en aller, le roy luy donna douze cens escus, qui luy furent delivrés par les mains de monseigneur de Comminges, qui l'avoit fait **entrer dans la chambre** du roy.
(DMF, Jean Le Clerc, *Interpolations et variantes de la Chronique scandaleuse*, p. 156, 1502)

Mais il ne faut attendre qu'une trentaine d'années – dans notre corpus – pour voir la première attestation du sens temporel :

(8) (français préclassique)
Je vous annonce pour vray que **dans quinze jours** tel exemple par la justice du vray Amour, lequel tousjours vous avez desprisé, se fera en vous (…)
(DMF, Jeanne Flore, *Contes amoureux*, p. 192, 1537)

Cela montre la vitesse du changement sémantique[17] – on pourrait dire aussi la vitesse de l'extension des contextes d'emploi – et par là même pourquoi il est si difficile d'en trouver des preuves matérielles. Mais on peut donc bien accepter l'idée que, des deux sens illustrés par les exemples modernes suivants, c'est bien le premier qui est étymologique :

(9) (français moderne)
On entend Dora parler **dans l'antichambre**.

(10) (français moderne)
Annenkov Pavel a-t-il confirmé le renseignement ? Kaliayev le grand-duc ira cette semaine au théâtre. **Dans un moment**, Pavel connaîtra le jour exact et remettra un message au portier. (Il se tourne vers Dora et rit.) Nous avons de la chance, Dora.
(exemples 9 et 10 : Frantext, Albert Camus, *Les Justes*, p. 314 et 315, 1950)

[17] Ou plutôt : cela suggère que ce changement a pu être très rapide (cf. §4.5).

3.2.3 Du spatial à la possession

Une autre évolution bien connue est l'utilisation de schémas locatifs pour l'expression de la possession (Heine 1997a,b, Koch 2012a). J'ai ainsi pu observer à plusieurs reprises un glissement de la proximité spatiale à la possession, comme l'illustre l'emploi de l'italien *appo* (issu du latin *apud*) dans l'exemple suivant :

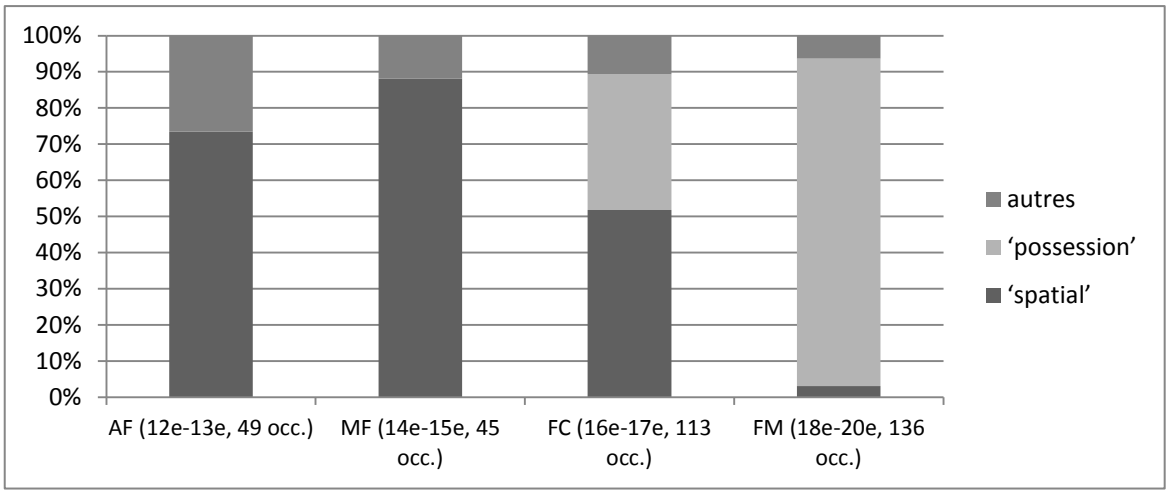
(11) (italien médiéval)
E che altro cotidianamente pericola e uccide le cittadi… tanto quanto lo nuovo raunamento d'avere **appo alcuno** ?
« Qu'est-ce qui met en danger et tue les villes autant que d'ammasser les richesses **aux mains de quelques privilégiés** ? »
(OVI, Dante Alighieri, *Convivio*, p. 336, 1307 ; le texte est glosé ainsi : « l'adunamento di ricchezze nelle mani di uno o di pochi privati »).

On le retrouve, en français, avec la préposition *pardevers* : dans l'exemple suivant, par exemple, ce n'est pas un sens purement spatial (glosable par 'j'ai encore près de moi') mais plutôt un sens de possession (glosable par 'j'ai encore en ma possession').

(12) (français classique)
J'ai encores **par devers** moi deux livres d'Epigrammes françois, deux de latins, que je vous promets à la premiere commodité ; et puis des Polemicques en diverses langues, oeuvres de sa jeunesse (…)
(Frantext, *Les tragiques*, p. 11-12, 1630)

(13) (français classique)
Quand mesmes nous n'aurions faict que visiter et examiner ensemble les raretez de son coffre d'Aegypte, que nous eussions faict revenir de la purification de Tollon, et ce que j'en avois retenu **par devers** moy, il y eust bien eu de bonnes heures de passe temps, qui ne luy eusse pas esté, je m'asseure, si dezagreable.
(Frantext, *Lettres aux frères Dupuy*, p. 209-210, 1631)

Il s'agit là encore du schéma locatif repéré par Heine, qui apparaît après les emplois spatiaux (et un certain nombre d'emplois relationnels, du type *être débiteur par-devers X*, qui semblent calqués sur la préposition *devers* et, à terme, disparaissent), comme l'illustre le graphique (3) ci-dessous.



Graphique 3 : Evolution sémantique de la préposition *pardevers*.

On peut encore noter un glissement ultérieur au 16ème siècle, avec une notion de possession qui s'étend à des éléments abstraits :

(14) (français prémoderne)
Il ne tiendroit qu'à moi de penser autrement ; car j'ai, sans vanité, quelques aventures **par-devers** moi dont un autre se feroit un honneur infini.
(Frantext, Diderot, *Lettres à Sophie Volland*, t. 2, p. 162, 1774)

(15) (français prémoderne)
L'idée de destruction étant plus composée, ne frappe pas de même ; l'image de la mort touche plus tard et plus foiblement, parce que nul n'a **par devers** soi l'experience de mourir ; il faut avoir vu des cadavres pour sentir les angoisses des agonisans.
(Frantext, Jean-Jacques Rousseau, *Emile ou De l'éducation*, p. 511, 1762)

3.2.4 Du spatial à l'exclusion

Ces résultats, obtenus principalement à partir d'études sur le français en diachronie, ont été testés également à travers la comparaison avec d'autres langues. Ainsi, à partir d'une étude diachronique sur plus d'une trentaine de langues (29 langues et dialectes romans, plus quelques langues germaniques et slaves), j'ai pu vérifier que des prépositions signifiant 'à l'exception de' résultent bien souvent de l'évolution sémantique d'une construction signifiant 'en dehors de', mais que l'évolution inverse n'est pas attestée dans ces langues (communication, Fagard 2008a).

Ma recherche a consisté à repérer toutes les constructions prépositionnelles signifiant (i) 'à l'exception de' et celles signifiant (ii) 'en dehors de', puis à voir dans quelle mesure les deux séries sont identiques, proches ou compatibles, selon les langues : les prépositions ayant principalement un sens abstrait peuvent-elles être employées avec un sens spatial, ou inversement ? Un des résultats obtenus est le suivant : sur 64 formes repérées, la moitié ont un sens spatial d'origine et un sens abstrait secondaire, comme les prépositions *behalve* et *povrh* (16-17).

(16) (néerlandais)

Ik	ben	nergens	schuldig	aan,	**behalve**	aan	te veel	vertrouwen
'je	être.PRS.1SG	nulle.part	coupable	à	**sauf**	à	trop	confiance'

« je ne suis coupable de rien, sauf d'un excès de confiance »
(*De Standaard*, 28 mars 2014, requête Webcorp)

(17) (serbe)

Povrh	liderstva	u	tehnologijama	i	proizvodima,	Cisco	
'excepté	excellence.GEN	en	technologie.LOC.PL	et	produit.LOC.PL,	Cisco.NOM	
je	prepoznatljiv	i	kao	inovator	u	poslovanju	(…)
être.3SG	bien.connu.NOM	aussi	comme	innovateur.NOM	en	savoir-faire.LOC'	

« Outre son excellence en technologies et produits, Cisco est bien connue comme innovateur en savoir-faire »
(internet)

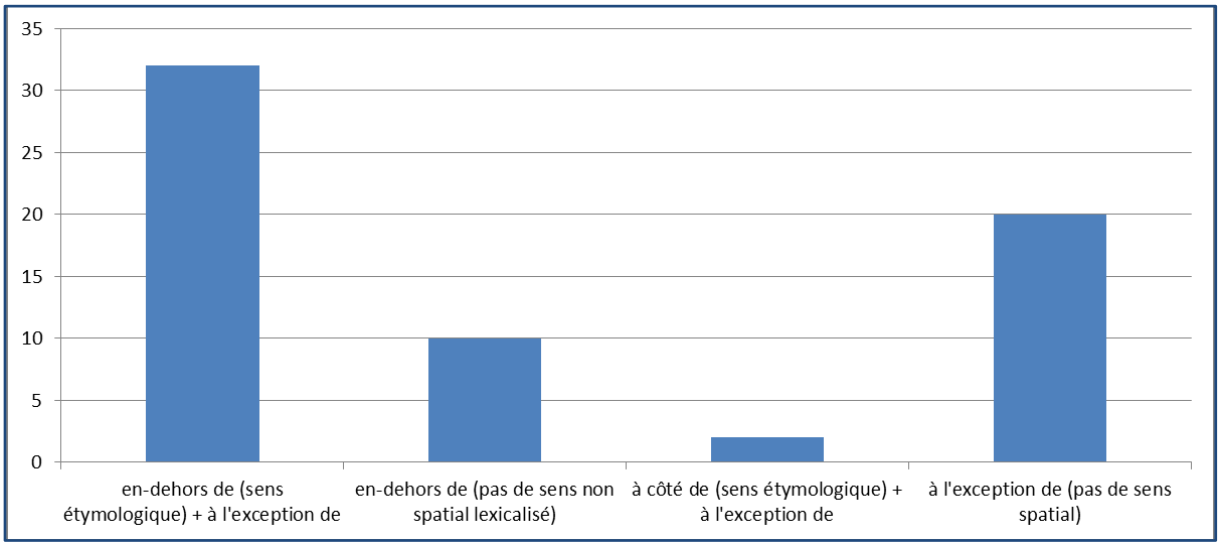
Des vingt prépositions qui ont un sens abstrait d'origine, comme *salvo* en (18), aucune n'a de sens spatial secondaire (voir le graphique 4).

(18) (espagnol médiéval)

Et mando que ninguna moger non se liege a mio lecho, **salvo** mis parientes aquellos que y quisieren venir.

« et je commande qu'aucune femme ne s'allonge dans mon lit, **mis à part** mes proches parents, ceux qui voudraient y venir »

(Corde, Anonyme, documents notariaux, 1294)



Graphique 4 : Prépositions signifiant 'en dehors de' et 'à l'exception de' dans quelques langues d'Europe (communication, Fagard 2008a)

3.2.5 Du spatial au comparatif

Heine (1997b : 114-115) note l'existence de plusieurs schémas de base, selon les langues, pour l'expression de la comparaison. L'un est le schéma locatif (*the location schema*), qu'il ramène à la formule « *X* is *Y* at *Z* » où

> the notion 'at' stands for a variety of static locative functions such as 'at', 'on', 'above', 'in', 'by', and the like. The formula can be described as conveying roughly the meaning '*X* has property *Y*, and if *Z* is placed in the same location as *X*, *X* has more of *Y* than *Z* does'. (ibid : 114)

Il note l'existence de ce schéma en naga, en hongrois, en kawaiisu, etc. ; d'après Heine & Kuteva (2002), on trouve un schéma proche en twi, baka, kisi, turku, et créole ndjurka :

> PASS (to pass (by), to pass through) > COMPARATIVE (Heine & Kuteva 2002 : 229-230)

Ce schéma se rapproche de deux autres à base spatiale, *source* et *but* (Heine 1997b : 115-116), le schéma *source* étant illustré d'après lui par l'italien (*sono più grande **di** lui*) et le schéma *but* par l'anglais (*inferior/superior **to** Z*).

Ces schémas peuvent également être observés pour des prépositions dans les langues romanes, mais il faut distinguer deux cas de figure assez différents : d'une part des prépositions d'origine latine dont seule une très faible proportion des emplois peut s'interpréter comme ayant un sens comparatif, et pour lesquelles on ne peut que reconstruire

l'évolution sémantique ; d'autre part des créations romanes de prépositions complexes qui ont un sens presque exclusivement comparatif.

3.2.5.1 Les évolutions reconstruites

Pour un certain nombre de prépositions des langues romanes, on observe ainsi des emplois comparatifs qui illustrent probablement un des schémas à base spatiale de Heine, mais pour lesquelles on ne peut être sûr qu'il s'agisse réellement d'une évolution romane – le sens comparatif peut en effet être hérité directement du latin, classique ou tardif. Le fait que les mêmes formes (aux préfixes ou suffixes et à la graphie/phonétique près) acquièrent ce nouveau sens dans plusieurs langues incite en tout cas à la prudence[18].

Ainsi, on peut observer ces emplois dans le cas des prépositions issues du latin *versus*, comme je l'ai montré pour le français (Fagard 2002 [I : 137 sqq.]) et les langues romanes (Fagard 2010a). Les exemples ci-dessous illustrent ces emplois comparatifs pour des prépositions issues du latin :

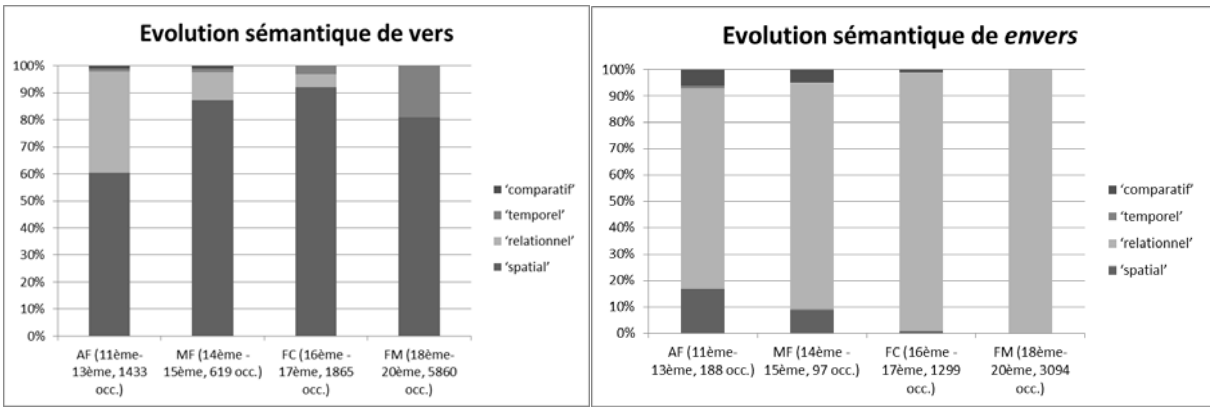
(19) (ancien français)
Theüter ot Povrepeleux,
qui mout fu bien chevalerex ;
n'i ot cheval tant amast chier,
vers lui sont li autre eschacier
« Theüter avait Povrepeleux, qui était très courageux ; il l'aimait plus qu'aucun autre cheval : les autres ne sont qu'échassiers **comparés à** lui »
(BFM, *Le roman de Thèbes*, 1150-1155, v. 6298)

(20) (italien médiéval)
è sillogismo che la m'ha conchiusa
acutamente sì, che **'nverso** d'ella
ogne dimostrazion mi pare ottusa.
« [l'inspiration du saint-esprit] est un raisonnement qui m'a infusé [cette joie] si bien que, **en comparaison**, toute démonstration me paraît obscure »
(OVI, Dante Alighieri, *Paradiso*, XXIV, v. 91-96, 1321)

(21) (occitan médiéval)
Li port amor tan fin e natural,
Que tuit son fals **ves** mi li plus leial
« je lui voue un amour si fin et naturel, que **comparés à** moi tous les autres sont faux »
(COM, Bernart de Ventadorn, *Can par la flors*, 12ème siècle)

La rareté de ces emplois comparatifs, dès la période médiévale, et leur quasi-disparition au plus tard à l'époque classique (voir les graphiques 5a-b ci-dessous) constituent des arguments supplémentaires pour n'y voir que des survivances d'un glissement à époque latine.

[18] Il faut noter cependant que certaines extensions ont pu justement se propager d'une langue à l'autre, comme en témoigne la remarque de l'*Enciclopedia Dantesca* sur l'exemple de Dante donné ci-dessous, attribué à une possible influence provençale.



Graphiques 5a & 5b : Evolution sémantique des prépositions *vers* et *devers* (Fagard 2010a)

Les seuls emplois comparatifs qui restent pour ces prépositions sont ressentis comme littéraires ou formels, qu'il s'agisse de *envers* en français moderne, de *verso* en italien moderne ou de *envers* en catalan moderne (22), dont je n'ai trouvé aucun emploi équivalent dans des textes modernes.

(22) (catalan moderne, litt.)
Què és la formiga **envers** l'elefant ?
« Qu'est-ce que la fourmi, comparée à l'éléphant ? »
(exemple de dictionnaire)

Pour ces prépositions, on ne peut donc exclure une évolution sémantique conforme à l'hypothèse localiste et aux schémas proposés par Heine, mais les recherches sur corpus ne permettent pas de le démontrer. On peut faire la même analyse pour toute une série de prépositions romanes issues des prépositions latines *contra*, *ante*, *circa*, *apud*, et dont certains emplois sont comparatifs :

(23) (occitan médiéval)
Paucs en conosc
C'**ab** vos vaillon un gan
« [ami Bertrand, toi que tant de gens connaissent], j'en connais peu qui vaillent un gant **à côté de** [comparés à] toi »
(COM, Giraut de Borneil, *No.s pot sofrir ma lenga qu'ilh no dia*, 12ème siècle)

(24) (occitan médiéval)
Sembleran tut leugier
Contra la mort del jove rei engles
« [toutes les douleurs de ce pauvre siècle réunies] sembleraient toutes légères **en comparaison de** la mort du jeune roi anglais »
(Bertrand de Born, *Elégie sur la mort du prince Henri*, Bayle 1879 : 216)

(25) (espagnol médiéval)
ca estas nuestras pinturas
cerca dellas son como lumbre d'estrellas
antel sol en sus alturas
« [là-bas les couleurs sont si bigarrées] que nos peintures comparées aux leurs sont comme la lumière des étoiles comparée au soleil quand il est au zénith »
(Corde, Íñigo López de Mendoza, Marqués de Santillana, *Bias contra Fortuna*, 15ème siècle)

Il faut tout de même noter qu'un argument pourrait être avancé pour soutenir l'hypothèse d'une évolution sémantique (de l'espace à la comparaison) pendant la période romane : le fait que, en latin classique, l'expression de la comparaison se fait plutôt à l'aide de la préposition *prae* et de quelques prépositions déjà très polysémiques, *ad* et *juxta* (toutes trois compatibles avec le schéma locatif, puisque leurs sens de base sont respectivement « devant », « à côté de, vers » et « à côté de »).

3.2.5.2 Les évolutions romanes

Une confirmation plus probante pourrait venir de prépositions plus tardives. On trouve, pour quelques prépositions simples ou complexes grammaticalisées à l'époque romane, des glissements tout à fait comparables à ceux que je viens d'illustrer pour *versus*, *contra*, *ab*, etc. et leurs dérivés romans. Ces glissements se comprennent très facilement à la lecture de l'exemple suivant, qui illustre bien la possibilité d'une double interprétation et donc les conditions mêmes de l'évolution sémantique, puisque les deux sens (spatial et comparatif) sont encore accessibles :

(26) (ancien français)
Mes nus soleuz ne l'anlumine,
qu'il est d'une couleur si fine,
si clers et si replandissanz
que li soleuz esclarcissanz
en l'autre eve les cristauz doubles
lez lui seroit occurs et troubles.
« Le soleil ne peut illuminer ce cristal, car il est d'une couleur si fine, si claire et si éclatante que le soleil, illuminant dans l'autre fontaine les cristaux doubles, serait [**à côté de lui** >] en comparaison obscur et trouble »
(BFM, Jean de Meun, *Le roman de la rose*, 1269, v. 16494)

Il s'agit ici d'une extension contextuelle, sans qu'il y ait création d'un sens nouveau ; ce sens n'émergera d'ailleurs pas dans l'évolution subséquente de la préposition *lez*. Pour d'autres prépositions, cependant, l'émergence du sens comparatif est plus nette. C'est le cas du portugais *face a*, *à vista de*, de l'italien *in confronto a*, dont je donne quelques exemples ci-dessous :

(27) (portugais)
Não era nada **à vista do** Louredo, nosso vaqueiro; pode acreditar, que é a verdade.
« Ce n'était rien **à côté de** Louredo, notre vacher ; vous pouvez me croire, c'est la vérité »
(CdP, José de Alencar, *O Sertanejo*, 1891)

(28) (portugais)
Défice do Estado agravou-se 7,7 por cento em Setembro **face a**o mês anterior
« Le déficit de l'état s'est agravé de 7,7% **par rapport au** mois précédent »
(internet, presse, 21ème siècle)

(29) (roumain)

în	**faţa**	unei	asemenea	frumuseţi,	ea	păleşte
'**en**	**face**.DET.F	IND.F.SG.GEN	tel	beauté.F.PL,	elle	pâlir.3SG.PRS'

« **comparé à** [litt. 'en la face de'] une telle beauté, elle pâlit »
(exemple construit)

On remarquera que toutes ces constructions sont formées sur l'idée de la mise côte à côte ou face à face des deux éléments à comparer – compatible avec les schémas locatifs de Heine. Si l'on excepte les locutions formées sur diverses bases nominales et qui sont dédiées à l'expression de la comparaison (fr. *en comparaison de*, it. *a paragone di*, port. *em comparação de*, etc.), les prépositions passées en revue dans cette section sont donc toutes au moins compatibles avec le schéma locatif.

3.2.6 Du spatial au discursif

J'ai étudié dans cette optique quatre types d'évolution aboutissant à la formation d'un marqueur discursif, et partant :

(a) de l'adverbial (préposition + groupe nominal, type *au fond*)
(b) de l'adverbe (type *parallèlement*)
(c) de l'adverbe déictique (type *ici*, *là*)
(d) de formes verbales (type *regarde*, *voici*)

Pour chaque type, j'ai trouvé soit des confirmations très nettes de l'hypothèse localiste, soit des éléments compatibles avec cette hypothèse, soit, enfin, des éléments qui m'ont amené à la réviser légèrement.

3.2.6.1 Arguments forts

Ainsi, l'étude sur corpus, en diachronie, des adverbiaux *au fond* et *d'un côté* (Fagard & Sarda 2014 [IV : 435 sqq.]) a bien montré que les emplois spatiaux sont nettement antérieurs aux emplois discursifs (la situation est moins nette, cependant, pour *d'un côté*, voir ci-dessous). Les premiers emplois repérés dans les corpus pour *au fond* ne présentent aucune ambiguïté, tandis que les emplois abstraits n'apparaissent que légèrement plus tard, et les emplois discursifs nettement plus tard. On peut faire exactement la même observation concernant *parallèlement* (ibid. et Sarda & Charolles 2009), bien que ses emplois discursifs ne soient pas du même type que ceux de *au fond* en français moderne. Les premiers emplois sont spatiaux (30a-b), les emplois métaphoriques ou abstraits (31a-b) et discursifs (32a-b) apparaissant plus tard :

(30a) (ancien français)
celui trova qui se tapist desoz **au fonz**.
« il en trouva un qui s'était tapis dessous, **tout au fond** »
(BFM, *Enéas*, p. 176, 1150)

(30b) (français prémoderne)
Ce sont des couches de nuages minces, qui marchent **parallèlement** les unes aux autres
(Frantext, Charles Bonnet, *Considérations sur les corps organisés*, p. 161, 1762)

(31a) (ancien français)
il ne puet estre a nul fuer
Qu'il n'ait emprainte **ou fons** dou cuer
Et seelee en son corage
Cele semblance, cele ymage
Ou estudie adez et pense

« Il ne peut être en aucune manière qu'il n'ait, gravé **au fond** du cœur et celée en sa volonté, cette image qui fait sans cesse l'objet de son étude et de sa pensée »
(BFM, Gautier de Coinci, *Miracles de Nostre Dame*, p. 313, 13ème siècle)

(31b) (français prémoderne)
Ces paroles étaient, pour ainsi dire, le point de jonction de deux scènes qui s'étaient jusque-là développées **parallèlement** dans le même moment, chacune sur son théâtre particulier, l'une, celle qu'on vient de lire, dans le Trou aux Rats, l'autre, qu'on va lire, sur l'échelle du pilori.
(Frantext, Victor Hugo, *Notre-Dame de Paris*, *1482*, p. 296-297, 1831)

(32a) (français prémoderne)
Il y a dans notre position des choses embarrassantes, mais il faut les supporter et nous en tirer ensemble. **Au fond,** j'ai eu tort de la laisser exposée à une situation dont elle-même m'avait averti. Si elle se refuse à mes prières, c'est autre chose.
(Frantext, Benjamin Constant, *Journaux intimes*, p. 420-421, 1816)

(32b) (français moderne)
Cette substitution est devenue chez Pirandello une critique de la notion de réalité. Et, **parallèlement,** chez les dramaturges psychologues, plus particulièrement chez les français, un thème théâtral nouveau, (…) le "que suis-je" a pris la place du "comment agir ?" (…).
(Frantext, Collectif, *Arts et littérature dans la société contemporaine*, t. 2, p. 3008, 1936)

L'apparition des emplois discursifs se fait, pour les deux adverbiaux, selon un processus complexe de grammaticalisation qui sera décrit plus en détails dans la partie § 4.

3.2.6.2 Illustrations

Pour *d'un côté*, on peut également observer des emplois discursifs, où toute interprétation spatiale est exclue :

(33) (français contemporain)
en même temps moi **d'un côté** j'suis jeune
(interview d'Ali (nom d'emprunt), 23 ans, banlieue parisienne, *France Info*, le 6 décembre 2012)

Cependant, la situation est moins claire dans la mesure où, dès les premières attestations, on trouve des emplois spatiaux et non spatiaux (Fagard & Sarda 2014 [IV : 457-461]). S'il est raisonnable de penser que les emplois spatiaux sont premiers, étant donné le sémantisme des éléments composant l'adverbial (*de*, *un*, *côté*), on ne peut donc pas l'avancer comme argument pour l'hypothèse localiste, puisque l'étude sur corpus n'a pas établi l'antériorité des emplois spatiaux.

Il en est de même pour les adverbes déictiques dans les langues romanes et slaves, étudiés pour mon projet « Espace et grammaticalisation » (communication, Fagard, Sikora & Stosic 2011). Dans les deux groupes de langues, nous avons trouvé des phénomènes compatibles avec l'hypothèse localiste, sans pouvoir démontrer que les emplois spatiaux sont premiers, du moins sur la base d'études sur corpus. Ainsi, on trouve des emplois évaluatifs et discursifs pour le serbe *tu* et *ovde* « ici » (médial pour *tu*, proximal pour *ovde*) et pour le polonais *tu* « ici » et *tam* « là, là-bas », illustrés ci-dessous pour le polonais *tam* (34-35) ; ces emplois se trouvent ailleurs, comme en espagnol *allá* (36) (Fortis & Fagard 2011 [VI : 734]) :

(34) (polonais)

Mów **tam** sobie!
'parler.IMP.2SG [**là**] soi.DAT !'
« vas-y, cause toujours ! »
(exemple construit)[19]

(35) (polonais)
Kupił sobie **tam** nowy samochód
'acheter.3SG.M.PST soi.DAT [**là**] nouveau.ACC.NT voiture.ACC'
« il s'est acheté une nouvelle voiture (et je m'en fiche bien) »
(exemple construit)

(36) (espagnol)
Pon el libro **allá** arriba
« met le livre **par là**, en hauteur »
(Carbonero Cano 1979 : 89)

Nous avons trouvé confirmation sur corpus de l'existence de ces emplois (spatiaux et discursifs, mais aussi temporels) pour la plupart des adverbes déictiques en serbe, polonais, espagnol, portugais, italien et français, mais nous n'avons pas trouvé d'éléments attestant de l'antériorité des emplois spatiaux, parce que ces constructions sont polysémiques dès les premiers textes. Cela nous renvoie à la question de la vitesse des changements, à laquelle je reviens dans la section §4.5.

3.2.6.3 Reformulations

L'étude de conjonctions et de marqueurs discursifs figurait déjà dans mon programme de recherches présenté pour le concours CNRS (2007), précisément parce qu'il me semblait que, a priori, l'analyse de ces parties du discours était susceptible de fournir des contre-exemples à l'hypothèse localiste. Il est vrai que les origines de ces constructions sont rarement spatiales – ou moins souvent, du moins, que pour les prépositions. Mais certaines études, plutôt que d'apporter de réels contre-exemples, m'ont amené à reformuler ce que j'entends par localisme. Le cas des marqueurs discursifs formés sur un impératif (*regarde*[20], *écoute*, etc. : Fagard 2010b [IV : 587 sqq.], 2012b) est particulièrement intéressant, puisqu'on peut observer un glissement qui n'est pas exactement du type [ESPACE > AUTRES DOMAINES]. De la même manière que les études sur corpus d'un certain nombre de prépositions m'ont amené à suivre les propositions de Herskovits (1982) et Vandeloise (1986) concernant l'analyse fonctionnelle, l'étude des marqueurs discursifs m'a amené à suivre une hypothèse plus large que le localisme simple : ce n'est pas simplement **l'espace** qui est utilisé pour construire notre vision du monde, mais **l'ensemble du système de perception**, pour lequel l'espace constitue un facteur fondamental. En d'autres termes, pour rendre compte de la structure linguistique, il faut se rattacher à ce que l'on connaît des processus cognitifs en général (Langacker 1987 :

[19] Un exemple dont la validité est facilement vérifiée sur corpus, ou même sur internet, où on trouve des constructions comme *więc mów tam sobie co chcesz ale mnie i tak nie przekonasz* (qu'on pourrait traduire par « alors raconte [là] toujours ce que tu veux mais moi même comme ça tu ne pourras pas me convaincre », http://www.malopolaki.darmowefora.pl/index.php?topic=332.30).

[20] Dont l'équivalent en français de Louisiane pourrait bien, à la différence du français de métropole, avoir atteint le statut de marqueur discursif qui échappe à *regarde*, comme le suggère l'exemple suivant : *Ga ! tu vas perdre ton balan* (traduit « watch out! you're going to lose your balance » par le *Dictionary of Louisiana French*). Cependant, l'absence d'occurrences dans le corpus cadien invite à la prudence : comme en français métropolitain, c'est plutôt *écoute* qui a des fonctions de marqueur discursif.

12-13). Je suis de ce point de vue en plein accord avec ce que Bernard Victorri appelle les « grammaires cognitives » :

> [dans le courant des *grammaires cognitives*, c'est-à-dire de la linguistique cognitive au sens de Lakoff, Langacker, Talmy, Fauconnier, Sweetser, Goldberg, Turner] la perception visuelle et l'expérience sensori-motrice jouent un rôle central dans la compréhension de la structure sémantique du langage. (Victorri 2004 : 74)

Les emplois concrets – que l'on pourrait également appeler 'situés' – des impératifs de verbes de perception sont en effet largement antérieurs à leurs emplois discursifs ; on trouve d'abord des emplois intégrés et proches du sens d'origine (37-38) avant de trouver des emplois plus ambigus (39-40), puis enfin des emplois clairement discursifs (41-42) :

(37) (espagnol médiéval)
¿Non bastan tus propios males e particulares penas, que con las glorias ajenas sientas tormentos mortales? **mira** que todos yguales en este mundo venimos, e assí mesmo morimos, mas bivimos desiguales.
« Est-ce que tes propres douleurs et peines ne te suffisent pas, pour que tu ressentes des tourments mortels à voir la gloire des autres ? [**regarde**] que nous arrivons tous égaux dans ce monde, et que nous en partons de même, mais nous vivons inégaux »
(Corde, *Cancionero castellano*, p. 283, 15ème siècle)

(38) (ancien français)
regarde chu ki est presens et chu ki est a uenir.
« [**regarde**] ce qui est présent et ce qui est à venir »
(NCA, *Gloses wallonnes*, p. 201, 1290)

(39) (moyen français)
car **regarde** a ton afaire si te retien de mal a faire dont tu mes t'ame en grant dolor
« [**regarde**] donc à ton affaire et retiens-toi de faire du mal et soumettre ainsi ton âme à des douleurs atroces »
(NCA, *Verite*, l. 1181, 1450)

(40) (espagnol médiéval)
E llamaron al portero del castillo e dixeron: – Cata aquí presos que traemos al rey Artur; e liévagelos e **mira** que no pierdas ninguno dellos, ca te dezimos, cierto, que tu señor nunca tan grand plazer uvo ni vio como éste.
« Ils appelèrent le garde à la porte du château et lui dirent : – vois ici les prisonniers que nous apportons au roi Arthur ; toi, amène-les-lui, et [**regarde**] à n'en perdre aucun, car nous te disons, avec certitude, que ton seigneur n'a jamais eu un si grand plaisir que celui-ci »
(Corde, *El baladro del sabio Merlín con sus profecías*, Fol. 60 R°, 15ème siècle)

(41) (italien médiéval)
Se tu voli biasimare o riprendere altrui, **guarda** che tu non sia magagnato di quello vizio medesimo
« Si tu veux blâmer ou reprendre autrui, [**regarde**] à ne pas être dévorer par ce vice »
(OVI, *Tesoro di Brunetto Latini volgarizzato*, p. 253, 1300)

(42) (espagnol médiéval)
No, mujer, no me altero… pero a lo menos que oiga el que yo la propongo, que le conozca y le trate, y después… **mira,** Bernardo a la hora esta debe haber llegado ya de Barcelona; habrá consagrado los primeros instantes a sus parientes; pero de un momento a otro le tendremos aquí, y es preciso recibirle como a quien viene a ser mi yerno

> « Non, femme, je ne suis pas en colère… mais au moins entends ce que j'ai à te proposer, apprends à le connaître et traite-le bien, et ensuite… [**regarde**], Bernardo à cette heure doit être arrivé déjà de Barcelone ; il aura consacré ses premiers instants à ses parents ; mais d'un moment à l'autre nous l'aurons ici, et il faut le recevoir comme le mérite celui qui deviendra mon gendre »
> (Corde, Mariano José de Larra, *No más mostrador*, p. 18, 1831)

Ce type de résultat m'a amené à une interprétation plus proche de l'*embodiment* que du localisme : ce qui compte n'est pas tant l'espace que l'espace vécu, ce qui renvoie à la notion de perception notée plus haut (comme l'a montré, entre autres, Vandeloise).

3.2.7 Après l'espace

Je présente dans cette section des évolutions ayant un autre point de départ que l'espace, mais qui ne remettent pas en question l'hypothèse localiste puisqu'elles montrent simplement que, à partir d'un sens second, l'évolution peut continuer. Le cas le plus fréquent dans les éléments que j'ai étudiés est constitué par les évolutions partant d'un sens temporel. Etudiant l'évolution sémantique de prépositions, conjonctions et marqueurs discursifs en français et dans les langues romanes, j'ai trouvé trois types d'extensions partant de sens temporels : (a) des évolutions internes au domaine temporel – l'évolution de *puis*, du sens 'après' au sens 'depuis' ; (b) des évolutions du temporel à d'autres domaines abstraits – l'extension sémantique de *puis* temporel au logique ; (c) enfin, des extensions du temporel au spatial – l'extension sémantique de *puis* et *depuis*, du temporel vers le spatial : un contre-exemple seulement en apparence (cf. Fagard 2007 [I : 116-117]). J'ai pu montrer que, dans ce type d'évolution, les mêmes mécanismes de grammaticalisation (et, dans certains cas, de subjectivation, au sens de l'anglais *subjectification*, Traugott 2003, 2007b, 2010) sont à l'œuvre que pour les constructions ayant une origine spatiale. Ainsi, l'affaiblissement sémantique, qui est en fait plutôt une perte en référencialité, se fait de la même façon. Nous verrons ces derniers exemples dans la section suivante.

(a) évolutions internes au domaine temporel

La préposition *puis* a presque exclusivement le sens 'après' en ancien français. Elle institue alors une distinction entre deux états, un « avant » et un « après ». Cependant, il se développe progressivement une distinction entre des emplois avec ce sens d'origine et un sens que l'on peut gloser par 'depuis'. C'est pour cette raison que j'ai désigné dans ma thèse l'extension sémantique de 'après' (*puis demain* « après demain ») à 'depuis' (*depuis deux ans*) comme un phénomène de subjectivation. Cette évolution sémantique est accompagnée d'une évolution morphologique, le nouveau sens étant marqué peu après son apparition par une forme spécifique, préfixée en *de-* : la préposition *depuis*, qui apparaît au 12ème siècle et supplante assez rapidement la forme *puis* dans les emplois prépositionnels.

(b) évolutions du temporel à d'autres domaines abstraits

Les extensions du temporel au causal sont bien connues également, et j'ai effectivement pu observer un certain nombre d'occurrences isolées montrant une tendance à ce type d'inférence pour des prépositions dont le sens prototypique est temporel. Ainsi, pour *puis*, depuis *et*

despuis, on trouve quelques occurrences où au sens temporel s'ajoute par inférence un sens logique, qui peut être causal ou concessif :

(43) (moyen français)
En la dicte conté nagueres avoit ung jeune filz orphenin, qui bien riche et puissant demoura **puis** le trespas de son pere et sa mere.
« Il y avait il y a peu, dans la comté, un jeune orphelin, qui demeura riche et puissant après la mort de ses parents. »
(BFM, *Les cent nouvelles nouvelles*, p. 131, 1515)

(44) (français préclassique)
Elle, se mocquant, secoua la teste, me disant : « je voudrois avoir ce qu'il s'en faut ». **Despuis ceste fortune** la damoiselle s'est recongneue, et n'a plus esté si nice.
« (…) Depuis cette aventure, la demoiselle (…) n'a plus été aussi naïve. »
(Frantext, *Le moyen de parvenir*, p. 250-251, 1610)

La rareté de ces exemples semble indiquer qu'il ne s'agit là que d'inférences contextuelles, qui ne se sont pas figées en un sens lexicalisé de la préposition.

On trouve ce type d'extension de manière beaucoup plus nette pour les conjonctions (Heine & Kuteva 2002 : 275-276, Traugott & König 1991 : 194, Fagard 2008b [VI : 929 sqq.], Fagard 2009b [III : 343 sqq.]), selon plusieurs chaînes de grammaticalisation possibles (cf. §5) :

[PENDANT] > [CONCESSIF] (anglais *while*, allemand *während*, baka, bulgare, créole des Seychelles)
[APRES] > [CAUSE] (it. *dal momento che*, fr. *puisque*, angl. *since*),
[APRES] > [ADVERSATIF] (vaï, lingala),
[APRES] > [CAUSAL] (vieux haut allemand *weil*, latin, français *quand*, finnois, estonien, roumain)
[APRES] > [CONDITIONNEL] (hittite, tagalog, indonésien, karok) (Heine & Kuteva 2002)

Evolutions possibles (et attestées) sur la base de conjonctions temporelles

Si l'évolution sémantique suit des parcours similaires, l'aboutissement sémantique de ces évolutions n'est jamais spatial. Une question demeure, celle de la fréquence relative des deux types d'évolution : d'après la théorie localiste, même redéfinie selon mes propres critères énoncés ci-dessus, on pourrait s'attendre à ce que la grammaticalisation s'opère plus fréquemment sur des termes spatiaux que sur des termes non spatiaux. Mais comment tester cette hypothèse ? Cela me semble une gageure. Même un inventaire complet des grammaticalisations dans une langue donnée (inventaire qui reste à faire, et me paraît déjà une entreprise hasardeuse) ne saurait résoudre définitivement le problème qui se pose au chercheur : comment pondérer ? La solution que j'ai adoptée jusqu'ici est de faire des études séparées pour différentes classes de mots. À partir d'inventaires des prépositions de 7 langues romanes (catalan, espagnol, français, italien, occitan, portugais, roumain), j'ai pu montrer que l'on peut reconstruire une origine spatiale pour une grande majorité des formes grammaticalisées (Fagard 2010a). Cependant, une étude similaire (bien que moins étendue) sur les conjonctions de cause dans les langues romanes a donné – sans surprise – un résultat inverse (Fagard 2009c [III : 343 sqq.]), confirmé par une étude sur les conjonctions en général (Fagard 2011a [III : 293 sqq.]). Il semble donc qu'il y a, selon les classes de mots, une proportion plus ou moins importante de phénomènes de grammaticalisation à partir de termes spatiaux. Les études menées dans le cadre du projet « Espace et grammaticalisation » laissent d'ailleurs penser que les marqueurs discursifs se situent entre ces deux pôles : on trouve un

nombre important de marqueurs formés sur des termes spatiaux (par exemple *au fond*, *parallèlement*, Fagard & Sarda 2014 [IV : 435 sqq.]), mais il ne s'agit probablement pas de la majorité comme pour les prépositions.

3.2.8 Des contre-exemples ?

Ces questions m'ont amené à étudier l'origine et l'évolution sémantique de termes ou constructions semblant contredire l'hypothèse localiste. Ainsi, on trouve de nombreuses prépositions principalement temporelles qui ont des emplois spatiaux, ou du moins sont employées dans des contextes spatiaux. Les exemples ci-dessous illustrent bien cette possibilité :

(45) (ancien français)
Les Turs nous vindrent hardoiant jusques en nostre ost ; pour ce commanda le roy que l'en cloussit nostre ost de fossés par devers Damiete **depuis le fleuve de Damiete jusques au flum de Rexi**.
« Les turcs vinrent nous narguer jusque dans notre campement ; c'est pourquoi le roi ordonna que l'on entour le camp de fossés, du côté de Damiette, depuis le fleuve de Damiette jusqu'à celui de Rexi. »
(BFM, Jean de Joinville, *La vie de saint Louis*, p. 197, 1305)

(46) (français contemporain)
Pendant des années, j'ai vu cette chambre et ces rideaux comme je les voyais **depuis** le lit où j'étais couchée.
(Frantext, Annie Ernaux, *L'événement*, p. 85, 2000)

Cependant, on peut donc se demander, pour ce type d'emploi, (i) quel est le détail des étapes qui ont permis leur émergence, en particulier quel est le rôle des contextes de déplacement, et des emplois de mouvement fictif ; (ii) s'il s'agit réellement de sens lexicalisés, ou simplement d'inférences contextuelles ; (iii) si d'autres phénomènes entrent en jeu, comme celui de répartition des emplois pour des prépositions proches formellement, ou l'apport sémantique d'un préfixe spatial.

On peut noter que les emplois spatiaux du moyen français se rapprochent du mouvement fictif (exemple 45), et qu'il y a toujours un mouvement sous-entendu (dans l'exemple en question, le mouvement des ouvriers, d'un fleuve à l'autre), à la différence de leur emploi en français moderne (dans l'exemple 46, 'il' ne bouge pas). En effet, dans le cas de la trajectoire avec mouvement réel, on peut continuer à avoir une lecture temporelle :

(47) (français contemporain)
Elle vient du Trocadéro, mais a dû marcher **depuis** la place Clichy parce qu'un soldat allemand s'est jeté sous le métro.
(Frantext, Anne-Marie Garat, *Dans la main du diable*, p. 235, 2006)

pourra ainsi être lu comme « elle a marché depuis le moment où elle était place Clichy ». Il y a donc extension sémantique avec conservation des deux lectures, même si la lecture spatiale prévaut. Or l'extension spatiale avec mouvement fictif, la plus éloignée du sens temporel d'origine, est également la plus tardive. On a ici une parfaite illustration d'une extension métonymique, avec des extensions radiales progressives à partir du sens de base : [DEPUIS

(TEMPOREL)] > [DEPUIS (SPATIAL AVEC MOUVEMENT)] > [DEPUIS (MOUVEMENT FICTIF)]. Cependant, l'évolution de *depuis* du temporel au spatial (à laquelle se réfère Haspelmath 1997 : 142) est en fait clairement liée à l'ajout du préfixe *de-*, lui-même spatial. Il ne s'agit donc pas d'une évolution spontanée ou naturelle du temps à l'espace ; donc pas d'un contre-exemple au localisme.

Dans les autres cas que j'ai étudiés, comme *envers*, *vers*, *devers* (Fagard 2002 [I : 137 sqq.]), *avant*, *devant* (Fagard 2012d [I : 29 sqq.]), etc., pour lesquels on pourrait penser qu'il y a un passage du temporel au spatial, le processus à l'œuvre n'est pas réellement une évolution mais un rééquilibrage, c'est-à-dire une perte des emplois temporels au profit des emplois spatiaux.

Enfin, un dernier mécanisme est celui de l'apparition d'emplois spatiaux de prépositions temporelles, mais à la marge, dans des contextes très spécifiques, comme

(48) (français contemporain)
Suivre la direction Audrieu (3 km), rentrer dans le village, passer devant l'église ; le Château se situe à peu près à 100 mètres **après** l'église sur la gauche.
(internet, http://www.chateaudaudrieu.com/fr/situation-et-plan-acces.php)

Un argument pour considérer qu'il ne s'agit pas d'une réelle lexicalisation est que pour *depuis* par exemple, ces emplois sont très rares, et le restent pendant toute la période où la préposition est attestée. De plus, ces emplois sont très contraints contextuellement : pour *après* en (48), il faut un contexte de description d'itinéraire pour que l'exemple soit vraiment naturel. En conclusion, les exemples d'extension du temporel au spatial sont assez complexes à démêler ; en réalité, il ne s'agit pas de simples extensions sémantiques, mais plutôt de cas d'apport sémantique du préfixe.

Enfin, les contre-exemples sont peut-être les cas les plus intéressants, car ils demandent plus de travail de compréhension. Une analyse plus fine montre néanmoins, pour tous les cas que j'ai étudiés, que ce sont des contre-exemples apparents, qui s'expliquent par l'interaction entre différents types de contraintes portant sur l'évolution sémantique : il est clair, en effet, que la « primauté du spatial » n'est qu'une des contraintes en jeu (Fagard 2010a : 409 sqq.).

Il s'agit ici d'un résultat que j'ai documenté pour une longue série de mots de l'espace, à l'aide d'études diachroniques sur corpus. J'ai ainsi pu montrer que les emplois initiaux ne sont jamais purement géométriques, et que la part concrète diminue constamment, en diachronie, jusqu'à ce qu'il devienne, selon le terme étudié, marqué ou impossible de l'utiliser pour décrire une situation spatiale. C'est devenu impossible par exemple pour *envers* ; marqué pour *dans* et plus encore pour *en*. Pour bien des prépositions, il reste des emplois spatiaux mais ils sont limités à certains contextes, et on peut en conclure qu'ils décrivent plutôt des rapports fonctionnels, comme l'a bien montré Michel Aurnague pour la préposition *à* (Aurnague 2012).

3.3 Résultat : un localisme revisité

Ces études m'ont ainsi permis d'affiner les conclusions de ma thèse, où j'affirmais prudemment qu'un certain nombre de chaînes sémantiques étudiées « respecte tout à fait le principe de la primauté du spatial » (Fagard 2010a : 396). Je reformulerais comme suit :

(i) **il ne s'agit pas vraiment, ou pas seulement, d'espace** – mais de perception du monde au sens large : espace-temps, modes de perception (surtout visuelle et auditive)… l'hypothèse de la primauté du spatial ne peut être préservée qu'au prix d'une reformulation ou d'une précision terminologique : il ne s'agit pas réellement d'espace, ou pas d'espace réel, objectif. **Ce qui est en jeu est l'espace perçu, subjectif, vécu**. C'est aussi la question étudiée par Camille Colin dans sa thèse (que je co-dirige avec Sophie Prévost), qui porte spécifiquement sur les relations entre espace et temps. D'après elle,

> le temps ne pouvant être vécu indépendamment de sa relation à l'espace (avant tout parce que le sujet est doublement ancré dans le temps et dans l'espace), et inversement, l'espace ne pouvant être vécu indépendamment de la situation temporelle d'un observateur, les deux dimensions ne peuvent être dissociées. (Camille Colin, résumé de communication)

(ii) **La configuration spatiale est toujours première en diachronie** (pour les adverbes et prépositions ayant des emplois spatiaux). Cependant, elle peut ne pas être seule, même à l'origine ; et elle tend souvent à donner lieu à des interprétations non spatiales, dès ses premiers emplois, comme on peut le voir pour les emplois de *derrière* ou encore *a chief de* en français médiéval. Les premiers emplois spatiaux que l'on observe peuvent même être métaphoriques, comme dans le cas de *au-dessus de*, ou fonctionnels, comme dans le cas de *en lieu de* (Fagard & De Mulder 2007 [II : 244-246]).

(iii) **Un adverbe ou une préposition peuvent gagner en spatialité, mais uniquement dans des circonstances exceptionnelles.** Par ailleurs, cela ne donne généralement pas lieu à une lexicalisation, et enfin, il ne s'agit pas réellement d'une évolution sémantique, mais plutôt d'un rééquilibrage (on peut se demander si cela est également vrai pour les verbes, cf. Aurnague 2011 sur les verbes de mouvement). Par exemple, une préposition ayant des emplois spatiaux et temporels peut perdre ses emplois temporels, comme *devant* (Fagard & De Mulder 2010 [I : 53 sqq.]) ; une préposition ayant des emplois spatiaux et notionnels (c'est-à-dire indiquant non des relations spatiales mais des relations abstraites) peut perdre ses emplois notionnels, comme *vers* (cf. Fagard 2009a [I : 67 sqq.]) ; une préposition ayant des emplois uniquement temporels peut, par composition, aboutir à une préposition ayant des emplois spatiaux, comme *depuis* (cf. Fagard 2010a).

(iv) **En synchronie, on ne peut pas considérer le sens spatial comme central**, même s'il n'est pas exclu qu'il ait un statut à part, lié à la manière dont nous conceptualisons le monde, dont nous appréhendons la réalité.

Par ailleurs, le statut privilégié de l'espace par rapport au temps dans la cognition est sujet à caution (cf. Colin, en préparation), même si un certain nombre d'éléments vont dans ce sens, comme l'ordre d'acquisition des repères spatiaux et temporels par l'enfant (cf. Morgenstern et al. 2010).

L’espace dont il est question ici ne saurait donc se résumer à sa conception géométrique ; c’est un espace vécu, perçu (correspondant à une vision fonctionnelle, voir la sous-section précédente). À cette précision près, il faut bien constater qu’une grande quantité de chaînes sémantiques que j’ai pu observer respecte parfaitement le principe de la primauté du spatial : SPATIAL > TEMPOREL, SPATIAL > COMITATIF, SPATIAL > POSSESSION, SPATIAL > COMPARATIF, SPATIAL > RELATIONNEL (Fagard 2002 [I : 137 sqq.], 2004 [I : 121 sqq.], 2006a,b [VI : 979 sqq.], 2010a). Les évolutions observables ne sont pas toujours complètes, dans la mesure où le nouveau sens peut n’apparaître qu’en tant qu’inférence possible, sans que le sens d’origine soit pour autant exclu.

4 L'espace dans la langue

Il semble évident que, pour avancer dans le déchiffrage du rôle de l'espace dans l'évolution des langues et en particulier la grammaticalisation, il est important de comprendre comment chaque langue permet de conceptualiser l'espace. La première question est donc de savoir comment les locuteurs de différentes langues décrivent l'espace. Mes travaux sur le localisme m'ont ainsi progressivement amené à m'intéresser à l'espace non seulement comme source de grammaticalisation mais pour lui-même, avec une attention particulière portée à la question suivante : puisque les mots spatiaux d'une langue n'ont pas réellement ou du moins pas uniquement un sens spatial, qu'utilisent-elles pour décrire l'espace ? Les recherches issues de ces interrogations s'inscrivent dans un large courant de recherche de la linguistique d'inspiration cognitive des trente dernières années, avec quelques détours notables par la psycholinguistique et la typologie. Les études sur l'espace dans la langue sont légion, et il serait vain de tenter d'en rendre compte exhaustivement – citons juste en passant Aurnague (2004), Borillo (1998), Fillmore (1997), Heine (1989, 1997a, 1997b), Hickmann & Robert (2006), Landau (2009, 2012), Langacker (1987, 1991a, 1991b), Levinson (2003), Slobin (1996a, 2004), Svorou (1994), Talmy (2000), Vandeloise (1986), Zelinsky-Wibbelt (1993).

Une des questions les plus fascinantes, de ce point de vue, est la réflexion déjà mentionnée sur la relativité linguistique. En fait, la question est double : (a) quelles contraintes les systèmes linguistiques posent-ils au locuteur lors de la description du réel ? (b) ces contraintes ont-elles des répercussions cognitives ? On peut également les formuler différemment : y a-t-il des invariants dans la manière dont l'être humain perçoit (spatialement) le monde qui l'entoure ? Cette perception est-elle simplement codée ou plutôt informée (transformée ?) par la langue, conformément à l'hypothèse de Slobin (« thinking for speaking », Slobin 1996b) ? La langue est-elle un reflet de notre conception de l'espace (et du monde) ou bien une grille de lecture, qui nous amène à mieux percevoir certaines choses, nous empêche d'en voir d'autres ? Toutes ces questions ont été au centre des débats de ces dernières années, en particulier en lien avec les travaux de Bowerman, Slobin, Talmy, Evans et bien d'autres.

> In recent years, a major focus of analysis in the domain of space has been the relationship between language and cognition (Landau and Jakendoff, 1993; Levinson, 2003; Majid et al., 2004). (Ibarretxe-Antuñano 2010 : 251)

Mes propres recherches ont porté plus particulièrement sur les points suivants :

I. Les relations spatiales statiques
II. Le déplacement
III. Le mouvement fictif
IV. L'asymétrie source / but

La question posée par mes recherches porte principalement sur les outils présents dans les langues pour décrire chaque type de relation spatiale. J'ai entamé cette réflexion pendant ma thèse et je l'ai poursuivie dans le cadre de ces trois projets de recherche, Trajectoire, ET et EoSS. Les résultats présentés dans cette section sont issus principalement de mon travail dans le cadre de trois projets de recherche. J'ai collaboré aux deux premiers, Trajectoire, un projet de la fédération Typologie et Universaux du Langage du CNRS (2006-2008 & 2010-2012), et EoSS (Evolution of semantic systems), un projet du Max Planck de Nimègue (2011-2014).

Le projet Trajectoire, dirigé par Colette Grinevald, Jean-Michel Fortis, Anetta Kopecka et Alice Vittrant, avait pour objectif d' « Etudier l'expression de la Trajectoire en décrivant les stratégies employées par un échantillon des langues du monde et en élaborant une typologie des modes d'expression de la trajectoire fondée sur ces descriptions » (http://www.ddl.ish-lyon.cnrs.fr/trajectoire/). Il a réuni jusqu'à vingt collaborateurs travaillant sur des langues d'Europe, d'Afrique, d'Océanie, d'Asie et des Amériques, avec pour philosophie la production d'une méthodologie commune, qui soit adaptée à la diversité linguistique représentée dans le projet. Le résultat le plus probant du projet, outre une série de publications dont l'une est parue en 2011 (numéro spécial des *Cahiers de Faits de langue*) et les autres en préparation, est la mise au point d'un outil d'élicitation que je décris plus bas.

Le projet EoSS (Evolution of Semantic Systems), dirigé par Michael Dunn, Fiona Jordan et Asifa Majid (MPI, Nimègue), vise à étudier la variation sémantique dans l'espace et dans le temps, à partir d'élicitations dans une cinquantaine de langues sur quatre catégories : les récipients, la couleur, les parties du corps et les relations spatiales (http://www.mpi.nl/departments/other-research/research-consortia/eoss). À la différence du projet précédent, la réflexion théorique et méthodologique est pilotée entièrement par les directeurs du projet, qui ont confié la récolte des données pour chaque langue à un chercheur ou une équipe de chercheurs spécialiste(s). Pour ce projet, j'ai piloté la récolte des données en occitan, effectuée sur le terrain par Xavier Bach (doctorant à l'Université d'Oxford).

J'ai également obtenu, en 2011, un financement du Labex TransferS pour mon projet ET (« De l'espace au temps : cadres de référence et interférences culturelles », 2011-2016), visant à réutiliser les données récoltées dans le cadre du projet Trajectoire, en collaboration avec Jordan Zlatev (Université de Lund, Suède). Le but du projet ET est de « s'interroger sur la part d'arbitraire dans l'évolution linguistique, en comparant les invariants linguistiques et le rôle joué par la culture dans un type spécifique d'évolution : le passage de l'espace au temps ». Il s'agit de partir des données récoltées à l'aide d'outils d'élicitation pour dresser la liste des termes utilisés dans différentes langues pour décrire les relations spatiales, puis de voir dans quelle mesure ces mêmes termes sont utilisés pour décrire les relations temporelles : je considère en effet que les paradigmes lexicaux et grammaticaux ne suffisent pas à avoir une bonne idée du fonctionnement de la langue, et qu'il faut justement partir de l'usage pour comparer les langues. Cela permettrait donc de compléter avec profit des études de typologie de surface, qui s'intéressent avant tout aux outils grammaticaux et aux inventaires lexicaux.

C'est également ce type d'interrogation qui m'a amené à étudier des langues dont le traitement de l'espace est particulier – ou, du moins, différent de ce que l'on trouve le plus

souvent dans les langues indo-européennes. C'est notoirement le cas de la langue des signes, comme en témoigne l'extrait suivant :

> An examination of signed language shows that its structural representation of space systematically differs from that in spoken language in the direction of what appear to be the structural characteristics of scene parsing in visual perception. Such differences include the following: Signed language can mark finer spatial distinctions with its inventory of more structural elements, more categories, and more elements per category. It represents many more of these distinctions in any particular expression. It also represents these distinctions independently in the expression, not bundled together into pre-packaged schemas. And its spatial representations are largely iconic with visible spatial characteristics. (Talmy 2010 : 319)

On trouve également des modes d'expression et de description de l'espace assez atypiques dans d'autres langues, qui ont à leur disposition des moyens morphologiques particuliers pour exprimer l'espace, ce qui m'a amené à les étudier à divers degrés de détail : des systèmes casuels à la fois étendus et omniprésents, pour le basque, le finnois et le hongrois ; un système de directionnels complexe et lui aussi omniprésent, pour les langues maya ; des séries verbales, pour le thaï. L'étude de ces langues, entamée pour certaines il y a déjà plusieurs années, n'a pas (du moins pas encore) systématiquement donné lieu à publication, notamment par respect de la maxime (i) (cf. § 2.3).

4.1 Les relations spatiales statiques

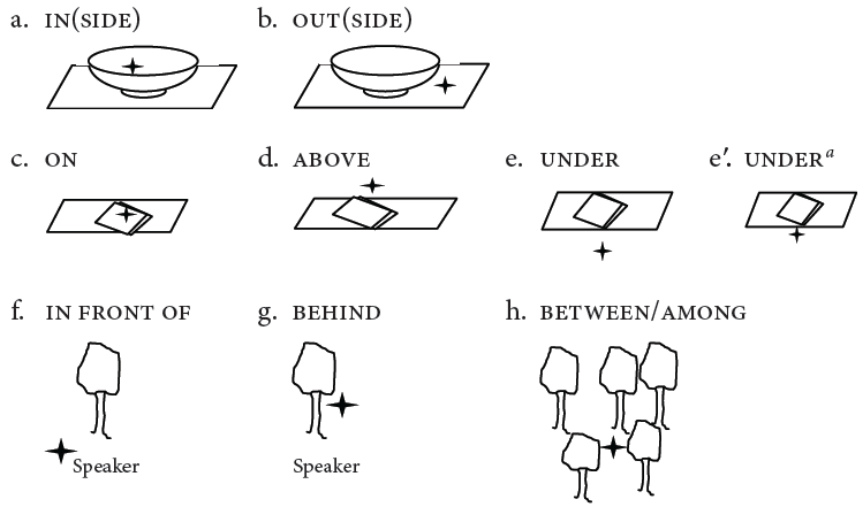
Mes recherches m'ont permis de montrer que certaines relations spatiales sont codées linguistiquement de manière assez stable, avec une faible variation intralinguistique, d'autres non, avec une plus grande variabilité. Il y a une corrélation entre le degré de complexité morphologique et le nombre de contextes d'emploi : les prépositions simples sont plus polysémiques que les prépositions complexes.

Mes premières interrogations sur l'espace dans la langue sont liées à mes recherches sur les prépositions, et partent du constat (suivant entre autres Vandeloise 1986) que le sens des prépositions spatiales n'est pas si spatial que cela :

> languages do not capture metric relationships in a simple way, a point first elaborated by Talmy (1983). (Landau et al. 2010 : 56)

Je ne reviens pas outre mesure sur ce point, déjà traité dans la section § 3, mais je développe un point connexe et également intéressant, qui est celui des contextes d'utilisation de chaque préposition dans un système prépositionnel, d'un point de vue contrastif.

J'ai ainsi étudié la correspondance entre prépositions et relations spatiales pour un article en collaboration avec Alexandru Mardale (Fagard & Mardale 2012 [VI : 857 sqq.]). Nous avons cherché à savoir, sur la base d'une étude sur corpus, quelles prépositions étaient couramment utilisées par différentes langues romanes pour une série de relations spatiales basiques (graphique 6 ci-dessous), dans une démarche inspirée par Levinson et al. (2003) :



Graphique 6 : Quelques relations spatiales (Fagard & Mardale 2012 [VI : 875])[21]

Pour cette étude, la difficulté principale était de distinguer les emplois purement spatiaux (49) des emplois non-spatiaux (50), avec la question épineuse du statut des emplois métaphoriques ou figés (51), que nous avons choisi d'écarter du corpus dans un premier temps :

Clairement spatial

(49) (portugais)

Enquanto conversavam, captavam detalhes um do outro; **sob a** mesa pequena, os joelhos se esbarravam ocasionalmente.

« Tandis qu'ils discutaient, ils percevaient des détails l'un de l'autre ; **sous** la petite table, leurs genoux se touchaient de temps en temps »

(CdP, Luiz Alfredo Garcia-Roza, *O Silêncio da Chuva*, 1996)

Clairement non spatial

(50) (portugais)

Sempre pensei que teria de tomar esta opção, **sob** o risco de cair em situações desagradáveis.

« J'ai toujours pensé que je finirais par devoir choisir cette option, au risque (litt. **sous** le risque) de tomber dans des situations désagréables. »

(CdP, *Jornal de Notícias*, 20ème siècle)

Métaphorique

(51) (portugais)

Houve a candidatura de Otelo… a última bandeira **sob** a qual se acolheu a ultima [*sic*] unidade.

« Il y a eu la candidature d'Otello – le dernier drapeau **sous** lequel nous avons formé un front uni »

(CdP, Internet, 20ème siècle)

Les premiers résultats de cette étude montrent une variabilité inégale entre les langues romanes selon le type de relation spatiale envisagée. Ainsi, pour la relation spatiale étiquetée ON, on trouve principalement une ou deux prépositions fréquentes pour chaque langue (donc peu de variation intralinguistique), avec une variabilité interlinguistique relativement grande, puisque les prépositions sont issues de trois prépositions latines différentes, *in*, *super* et *per* (tableau 1) :

[21] Les termes anglais constituent des étiquettes. Ils ne renvoient pas aux prépositions correspondantes en anglais, mais à une relation spatiale (et fonctionnelle) idéalisée. Le choix de l'anglais s'explique par la langue de rédaction de l'article.

	Fréquence relative (par million de mots)	
	> 5000	> 1000
espagnol	*en*	*sobre*
français		*sur*
italien		*su*[22]
portugais	*em*	*sobre*
roumain	*pe*	

Tableau 1 : Prépositions romanes dénotant la configuration spatiale (c) (ON) (Fagard & Mardale 2012 [VI : 884])

En revanche, pour la relation spatiale étiquetée BETWEEN, on trouve une variabilité intralinguistique importante, et une variabilité interlinguistique limitée : la préposition la plus fréquente est issu du latin *inter*, pour toutes les langues retenues, et seules deux autres racines sont représentées avec une fréquence notable, l'une correspondant au latin *medium* « centre », l'autre au latin *sinus* « creux » :

	Fréquence relative (par million de mots)			
	> 500	> 100	< 50	< 10
espagnol	*entre*	*dentro de*		*por entre*, *por dentro (de)*
français	*entre*		*parmi*, *au sein de*	
italien	*tra*, *fra*		*in mezzo a*	*nel mezzo di*
portugais	*entre*	*no meio de*	*dentre*	
roumain		*între*	*printre*	*în sânul*, *din sânul*

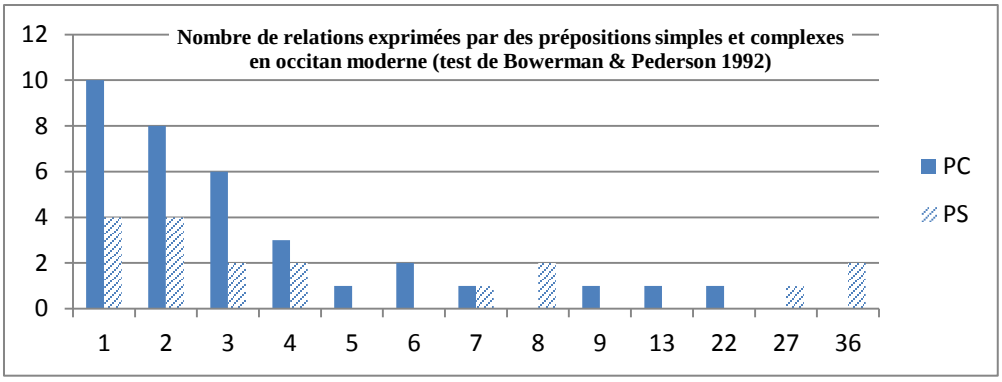
Tableau 2 : Prépositions romanes dénotant la configuration spatiale (h) (BETWEEN) (Fagard & Mardale 2012 [VI : 885])

Ces deux tableaux (1-2) illustrent bien un autre résultat important, à savoir que **certaines relations spatiales semblent être codées linguistiquement de manière assez stable (c'est-à-dire avec peu de variation intralinguistique), tandis que pour d'autres on observe une grande variabilité**. Cette hypothèse semble confirmée par les résultats préliminaires d'une autre étude, menée dans le cadre du projet EoSS, et dont une tâche est centrée sur l'expression d'un certain nombre de relations spatiales statiques en occitan moderne.[23]

Ces résultats permettent d'annoncer les remarques formulées plus loin sur la grammaticalisation (§ 5) : il semble y avoir une corrélation entre le degré de complexité morphologique et le nombre de contextes d'emploi, avec les prépositions simples couvrant en moyenne un nombre d'emplois plus large que les prépositions complexes. Le graphique (7), ci-dessous, illustre bien ce point : dans les réponses au test de Bowerman & Pederson (1992), la plupart des prépositions complexes (24) ne sont utilisées que pour décrire un tout petit nombre relations spatiales (3 au maximum) ; inversement, plusieurs prépositions simples peuvent être utilisées dans un grand nombre de contextes différents (jusqu'à 36 pour *a*).

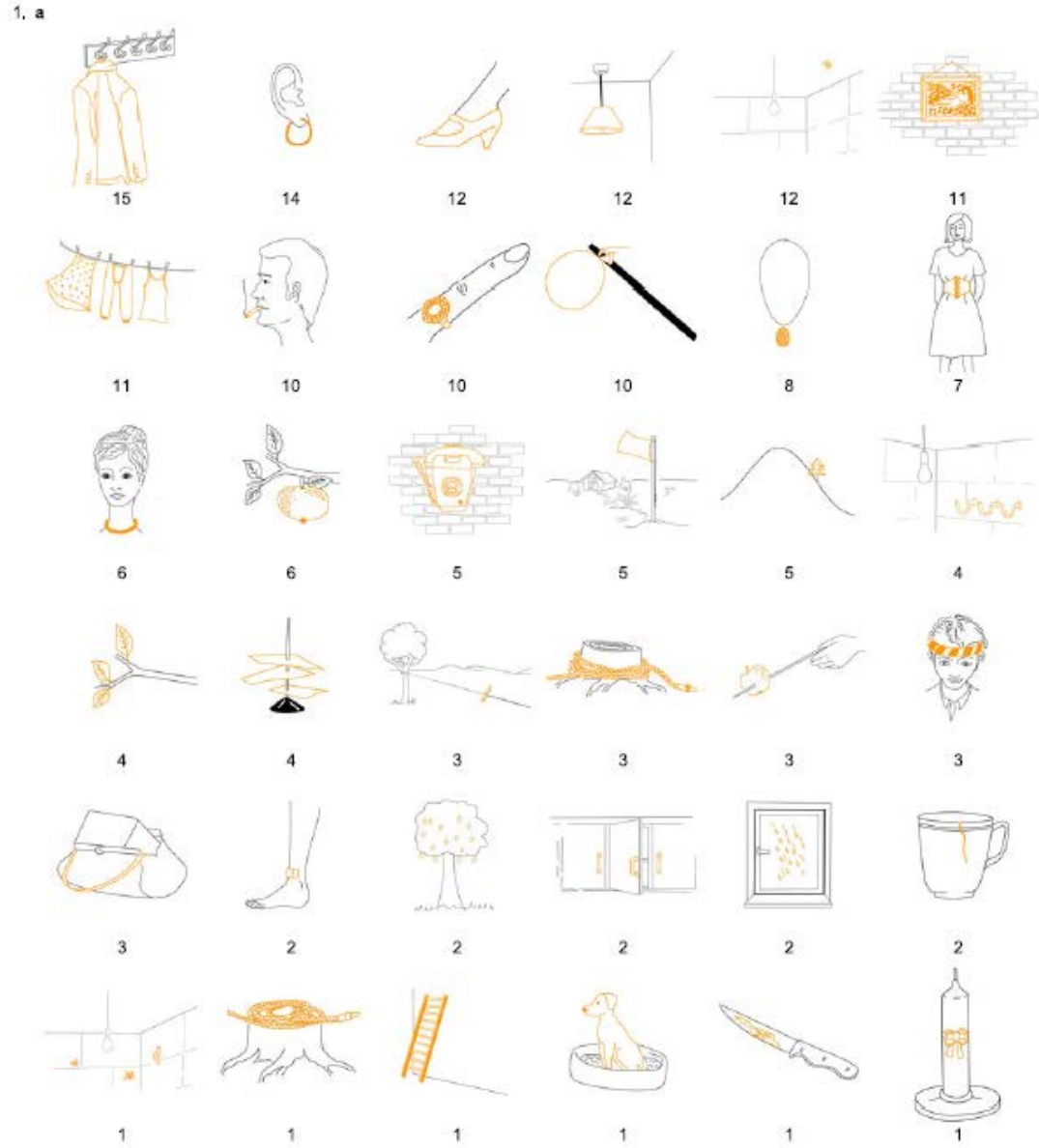
[22] Le but de l'étude étant d'identifier les prépositions utilisées *par défaut*, nous avons exclu les constructions (simples ou complexes) trop peu fréquentes. C'est le cas par exemple de l'italien *sopra*, 20 fois moins fréquent environ que *su*.

[23] Les expérimentations ont été menées sur le terrain par Xavier Bach, avec l'enregistrement de 4 tâches auprès de 20 locuteurs. Nous avons ensuite transcrit et codé ensemble les enregistrements. Il reste beaucoup de travail pour analyser l'ensemble des données récoltées : les vérifier, notamment en les comparant à des données spontanées, comme des corpus écrits et oraux, ou encore essayer de voir dans quelle mesure la variation entre les différents participants est liée à des différences dialectales, au degré de maîtrise de la langue, ou encore au contexte sociolinguistique. Cependant, même en l'état actuel, elles sont déjà précieuses.



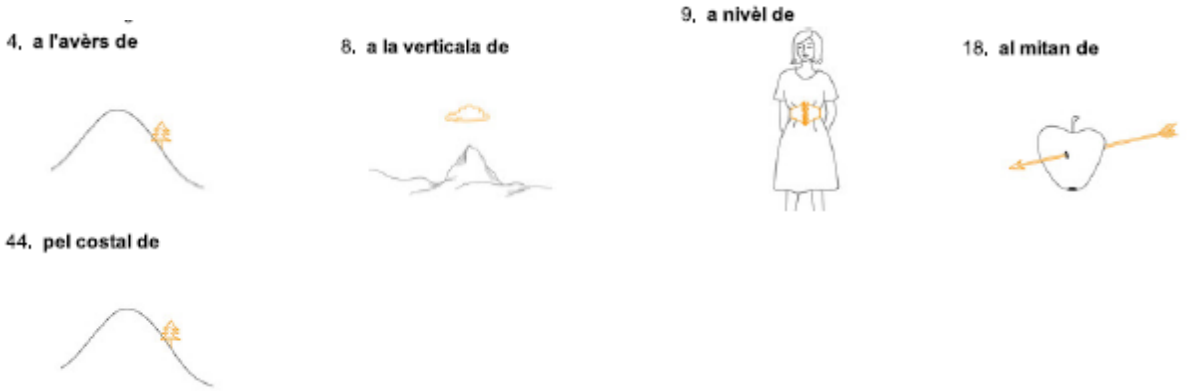
Graphique 7 : Relations sémantiques et complexité morphologique (projet EoSS ; PC = prépositions complexes, PS = prépositions simples, en abscisses, le nombre de relations sémantiques ; en ordonnées, le nombre de prépositions concernées)

La polysémie attendue de certaines prépositions simples apparaît clairement au vu des images décrites par la préposition *a*, par exemple :



Graphique 8 : Relations spatiales désignées par la préposition *a* (projet EoSS)

Tandis que la tendance des prépositions complexes à présenter à la fois une grande variabilité morpho-syntaxique et une faible polysémie est bien illustrée par les exemples suivants, où l'on voit que chaque forme n'a été utilisée dans notre corpus que pour un type de relation spatiale :



Graphique 9 : Relations spatiales désignées par quelques prépositions complexes (projet EoSS)

Il faut noter, de plus, qu'un grand nombre d'emplois de *a* est commun à plusieurs locuteurs, comme en attestent les nombres souscrits (graphique 8), qui indiquent le nombre d'occurrences de la préposition pour chaque image : pour dix images différentes, au moins la moitié des locuteurs a utilisé la préposition *a* ; seules six images correspondent à un emploi apparemment idiosyncratique (avec un seul emploi de *a*). À l'inverse, on ne trouve qu'une seule occurrence de chacune des prépositions complexes illustrées ci-dessus.

Le détail des occurrences montre qu'une analyse plus fine est nécessaire. Ainsi, pour *a*, on remarque dans les quelques exemples ci-dessous que si la configuration géométrique varie beaucoup, la relation fonctionnelle est assez stable (cf. Aurnague 2012, pour le français) :

(52) (occitan)
la baga **es al** det
« la bague **est au** doigt »

(53)
la cauçura **es al** pè
« la chaussure **est au** pied »

(54)
de cirièiras **a** l'arbre
« des cerises **à** l'arbre »

(55)
una bocla d'aurelha, **a** l'aurelha de qualqu'un
« une boucle d'oreille, **à** l'oreille de quelqu'un »

(56)
la correja **es al** sac, **estacat** d'aquel biais **al** sac
« la bandoulière **est au** sac, **attachée** de cette manière **au** sac »
(exemples 52-56 : données recueillies dans le cadre du projet EoSS)

Dans presque tous les cas, il s'agit de la position attendue : la bague au doigt, la chaussure au pied, la cerise dans l'arbre, la boucle d'oreille sur l'oreille. C'est peut-être moins vrai pour la bandoulière accrochée au sac ; justement, les hésitations, dont cet énoncé constitue une bonne illustration (56), semblent indiquer que, lorsque la relation est moins typique, les locuteurs sont tentés de préciser la configuration géométrique. Il me semble que la variation tient en partie à ce choix entre description fonctionnelle et description géométrique, le sens des prépositions étant constitué des deux types de traits.

Si l'on analyse de plus près la variation interpersonnelle, on voit par exemple que, pour l'image de la veste pendue au crochet, les réponses suivent deux schémas de base, avec un verbe d'état (de position, qui tend à décrire la relation géométrique) ou une construction locative (plutôt fonctionnelle ?), et presque toujours avec la préposition *a* dans les deux cas :

i. **pendre/être pendu + PREP** : 14 occurrences – 12 *a*, 1 *sus*, 1 *per*
ii. **être/ø + PREP** : 6 occurrences – 5 *a*, 1 *sus*

La variation est donc relativement limitée : [(la|une) veste (est|est pendue|pend|ø) (à|sur|par) le porte-manteau], comme l'illustrent les exemples suivants :

(57) (occitan)
la vèsta **es penjada al** pòrta-mantèl
« la veste **est pendue au** porte-manteau »

(58)
una vèsta **al** pòrta-mantèl
« une veste **au** porte-manteau »
(exemples 57-58 : données recueillies dans le cadre du projet EoSS)

Une piste intéressante pour poursuivre ces recherches et les approfondir serait d'une part la comparaison plus systématique des données de corpus (en suivant la méthodologie de Stosic et al. 2015 [V : 631 sqq.], ou de Fagard & Mardale 2012 [VI : 857 sqq.]) et des données élicitées, d'autre part l'exploitation des données récoltées et la comparaison avec les données obtenues pour d'autres langues, à commencer par les langues romanes, qui sont bien représentées dans le projet EoSS (portugais, espagnol, galicien, français, romanche, italien, frioulan, roumain). L'équipe de coordinateurs du projet EoSS a répondu de manière enthousiaste à ma proposition d'organiser avec Anetta Kopecka (Université Lyon 2, DDL) un colloque sur les données romanes du projet, qui devrait se tenir fin 2016 si je parviens à rassembler des fonds : cette piste de travail est donc prometteuse.

4.2 L'expression du déplacement

Mes recherches sur le déplacement m'ont permis de revoir la typologie de l'expression du déplacement, et m'ont amené à considérer qu'il faut revoir radicalement la typologie de Talmy (1991, 2000 : langues à satellite vs à cadre verbal) : les langues proches génétiquement ne sont pas nécessairement proches structurellement (Fagard et al., sous presse [V : 613 sqq.]), et il faut de plus prendre en compte la différence entre structure linguistique et usage

de la langue (Fagard et al., soumis ; voir également Schøsler 2008 sur l'expression du mouvement dans les langues romanes) ; inversement, l'éloignement génétique n'implique pas toujours des différences structurelles fortes. La typologie doit être plus fine.

Ma réflexion sur le déplacement a d'abord pris la forme d'une participation à un projet de la fédération TUL, le projet Trajectoire, dont j'ai été membre dès sa fondation – et même avant, puisque j'étais présent à la journée d'étude (en 2003) où il s'est, pour la première fois, discuté du dépôt du projet. J'ai ainsi participé à l'élaboration d'un outil d'élicitation, le DVD Trajectoire, qui rassemble 76 scènes d'une dizaine de secondes environ. Il y a deux scènes d'entraînement et un tiers de distracteurs, où l'on peut voir par exemple un jeune homme qui dort, qui mange une banane, une jeune femme lui versant à boire, etc. Les deux tiers restant sont les scènes test, qui ont été calibrées pour voir comment des locuteurs de langues éloignées typologiquement décrivent la trajectoire. On peut donc y voir une personne qui sort d'une caverne, qui entre dans une caverne, qui court sur le sable, qui monte quelques marches, etc. Un certain nombre d'éléments varient :

(a) le point de vue. On voit la personne venir vers la caméra, s'éloigner de la caméra ou suivre une trajectoire perpendiculaire à la caméra.
(b) le type de mouvement. On voit la personne courir, marcher plus ou moins vite, sauter.
(c) la personne filmée. On voit diverses personnes se mouvoir : quelques enfants, quelques jeunes hommes et femmes.
(d) le nombre de personnes filmées. La plupart des scènes ne montrent qu'une personne, mais dans quelques cas il y en a de deux à cinq.
(e) le(s) point(s) saillant(s) de la trajectoire. On peut schématiser une trajectoire comme un mouvement en trois temps : (1) début, (2) milieu, (3) fin (Grinevald 2011). Nous avons rendu saillant l'un ou l'autre de ces éléments, de différentes manières (saut au-dessus d'un tronc d'arbre, d'une pierre, quelques pas vers un arbre, vers une personne couchée, etc.).
(f) le franchissement de frontière. Le point initial ou final de la trajectoire est rendu saillant, dans un certain nombre de scènes, par divers types de frontières, plus ou moins marquées : on une personne entrer dans une grotte, dans la forêt, dans la mer, ou bien en sortir.

4.2.1 Expérimentation « Trajectoire » : Récolte des données

Les différents membres du groupe ont récolté des données à l'aide de cet outil. J'ai pour ma part récolté une grande quantité de données, sur plusieurs langues, pour certaines en collaboration avec un certain nombre de collègues[24] :

[24] Les collègues en question, que je remercie encore une fois ici, sont indiqués dans le tableau par leurs initiales : Laure Sarda (LS), Anetta Kopecka (AK), Johan Blomberg (JB), Jordan Zlatev (JZ), Massimo Cerruti (MC), Elisa Omodei (EO), Erica Cosentino (EC), Snežana Todorović (ST), Dejan Stosic (DS), Alexandru Mardale (AM).

langue	participants	lieu principal	enregistrement	transcription	codage
allemand	19	Tübingen	BF & LS	BF	BF
polonais	15	Cracovie	AK & BF	AK & BF	BF
thaï	20	Lund	JB	JZ	BF, JB & JZ
suédois	20	Lund	JB	JB	BF, JB & JZ
serbe	12	Belgrade	ST	ST	BF, ST & DS
français	12	Paris, Louvain-la-Neuve, Montréal	LS & BF	LS & BF	BF
piémontais	12	Foglizzo & Turin	MC & BF	MC & BF	MC & BF
italien	14	Rome	EO & EC	BF	BF
portugais	10	Lisbonne	BF	BF	BF
roumain	10	Bucarest	AM	AM	BF

Pour ces dix langues, les données me semblent suffisantes pour bâtir des hypothèses. J'ai de plus récolté des données limitées pour quelques autres langues : hongrois (6 locuteurs), catalan (3 locuteurs), galicien (2 locuteurs), occitan (1 locuteur), slovaque (1 locuteur) et espagnol (1 locuteur). Ces données sont précieuses, malgré leur faiblesse quantitative, car elles peuvent servir de point de départ pour identifier les langues les plus intéressantes pour tester certaines hypothèses ; en revanche, il est clair qu'elles ne suffisent pas pour obtenir des résultats publiables.

4.2.2 Expérimentation « Trajectoire » : Analyse des données

Si la récolte des données demande du temps et de l'énergie, l'analyse des données pose, elle, une série de problèmes. Le plus important est que la complexité des éléments à tester était sans doute trop grande. Ainsi, les efforts du groupe pour mettre au point un système de codage valable pour toutes les langues représentées (une quinzaine à ce stade) ont abouti à la rédaction d'un manuel de codage très détaillé, et qui n'a malheureusement pas pu être mis à profit. On peut se demander s'il n'aurait pas mieux valu se contenter d'un codage plus simple, ou du moins incrémental, ou d'un nombre de scènes moins important, avec une seule question simple à la clef, par exemple : « est-ce que le point de vue a une importance ? » ou encore « est-ce que le but (point final de la trajectoire) est plus important que la source (point initial) ? ». Ceci dit, j'ai la conviction que les données ainsi récoltées sont de qualité, et permettent de formuler des hypothèses intéressantes, à poursuivre éventuellement à l'aide d'autres outils d'élicitation et à contrôler à l'aide de données de corpus. J'ai pu présenter différents résultats obtenus à l'aide de cet outil d'élicitation (Fagard, Kopecka & Sarda 2009, Fagard et al. 2013, 2014, Kopecka & Fagard 2014, Kopecka et al. 2015). Je détaille dans (Fagard et al., sous presse [V : 613 sqq.]) une partie de ces résultats : nous avons cherché, à partir des données récoltées en français, piémontais, allemand, suédois, polonais et thai, à voir s'il fallait revoir la typologie de Talmy (1991, 2000), opposant les langues 'à cadre verbal' (*verb-framed*), qui expriment typiquement la trajectoire (en gras dans l'exemple ci-dessous)

dans le verbe, aux langues ‘à satellite’ (*satellite-framed*), qui expriment typiquement la trajectoire dans une particule et peuvent par conséquence exprimer la manière du mouvement dans le verbe principal :

(59) (français)
Il est **entré** dans la maison (en courant)
(exemple construit)

(60) (anglais)
He ran **into** the house
« il est **arrivé** en courant **dans** la maison [depuis ailleurs, avec franchissement de frontière] »
(exemple construit)

Notre question était précisément de savoir si et surtout comment il convenait de revoir cette typologie. Plusieurs possibilités existent en effet. La première serait d’ajouter un « troisième type », qui serait constitué par des langues où manière et trajectoire sont sur le même plan, et peuvent être co-présentes dans l’énoncé sans que l’on puisse identifier une hiérarchie syntaxique ou informationnelle claire entre les deux. Cela semble être le cas de certaines langues où manière et trajectoire sont exprimées dans des préverbes cumulables, comme le jaminjung (Schultze-Bernd 2000, 2006, Slobin 2004), ainsi que de certaines langues à séries verbales, comme le thai (Zlatev & David 2003, Zlatev & Yangklang 2004). Ainsi, dans nos données, les locuteurs de thai ont souvent recours pour décrire le déplacement humain à une série verbale constituée de trois verbes, le premier désignant la manière, le second la trajectoire et le dernier la deixis. Par exemple, en (61), *wîŋ* ‘courir’ et *khâw* ‘entrer’ coexpriment le mouvement avec un troisième verbe, *pay* ‘aller’.

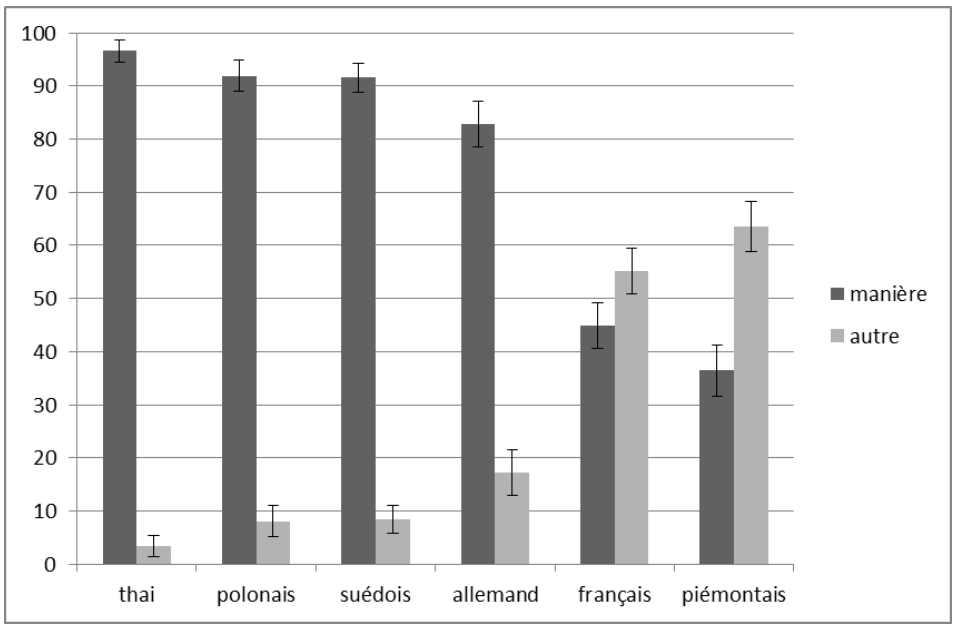
(61) (thai)

kháw	**wîŋ**	**khâw**	**pay**	nay	bâan
‘3SG	**courir**	**entrer**	**aller**	dans	maison’

« Il est **entré en courant** dans la maison [**en s’éloignant de moi**] »
(exemple construit)

S’il est possible de considérer que le verbe déictique est plus grammaticalisé (comme les préverbes déictiques *hin-* et *her-* en allemand, il est très fréquent et parfois employé sans lien avec l’organisation spatiale de la scène), l’identification du verbe principal n’est pas automatique, bien qu’elle puisse se faire, par exemple, sur des bases sémantiques (voir Thepkanjana 2008, pour qui le verbe principal de (62) serait *wîŋ*, suivant Chuwicha 1993).

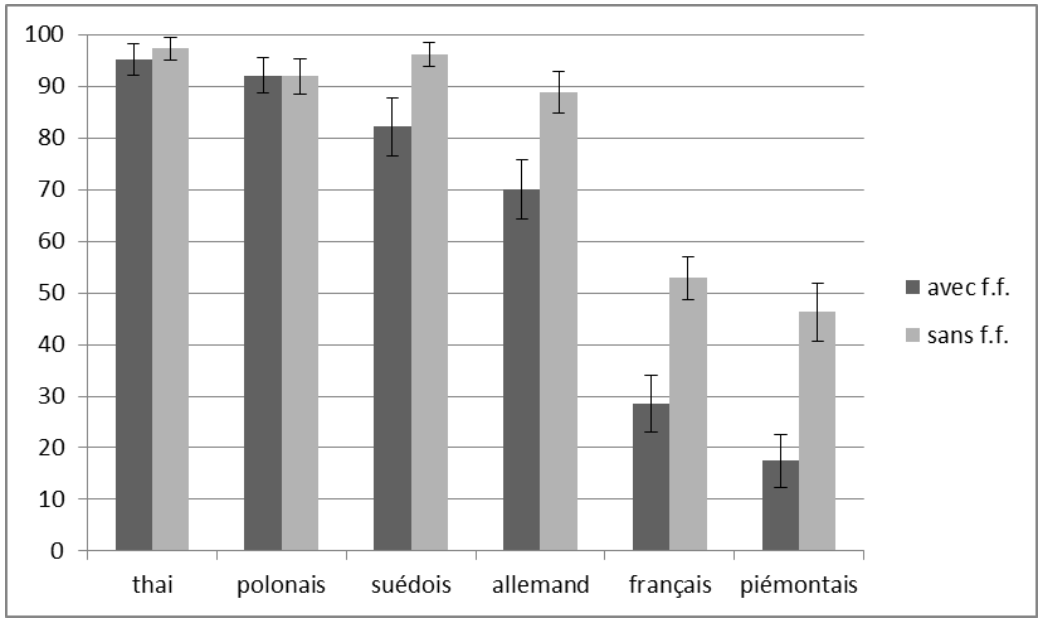
Une autre possibilité serait de substituer aux types la notion de continuum, avec des langues exprimant plus ou moins la MANIERE (Slobin 2004) ou la TRAJECTOIRE (Ibarretxe-Antuñano 2009). Nos données ne permettent, me semble-t-il, ni de soutenir, ni d’exclure cette possibilité, dans la mesure où on peut voir dans le graphique (10) ci-dessous à la fois un décrochement important entre langues à satellites et langues à cadre verbal, et une variation au sein de chaque groupe qui permettrait peut-être, en ajoutant d’autres langues, de retrouver un continuum (communication, Fagard et al., 2013).



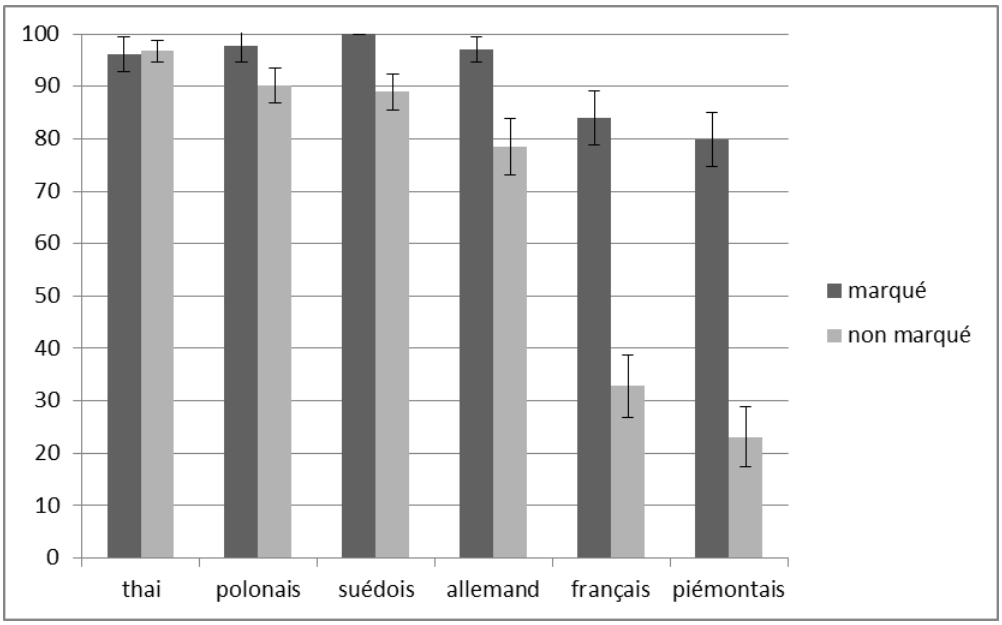
Graphique 10 : Fréquence de l'expression de la manière dans le verbe (données Trajectoire, six langues : pourcentage d'énoncés avec expression de la manière, ou sans (« autre »), avec indication de la variation individuelle sous forme de barres d'erreur)

Une autre possibilité encore serait de substituer aux types de langues les types de construction, chaque langue étant constituée d'un ensemble complexe de structures de l'un ou de l'autre type (Kopecka 2006, Beavers, Levin & Tham 2010, Croft et al. 2010). C'est cette dernière possibilité que je suis en train d'explorer avec mes collègues (Fagard et al., soumis).

Les résultats de ces recherches montrent que la différence attendue de fréquence d'expression de la manière illustrée par le graphique (10) se retrouve surtout dans les scènes où la manière est peu marquée (c'est-à-dire où la figure marche, voir graphique 11), et dans les scènes où il y a franchissement de frontière (voir graphique 12) :



Graphique 11 : Fréquence de l'expression de la manière dans le verbe (données Trajectoire, six langues : pourcentage d'énoncés avec manière dans les descriptions de vidéos avec et sans franchissement de frontière (f.f.), avec indication de la variation individuelle sous forme de barres d'erreur)



Graphique 12 : Fréquence de l'expression de la manière dans le verbe (données Trajectoire, six langues : pourcentage d'énoncés avec manière dans les vidéos avec manière de mouvement marqué en gris foncé, et même chose pour les vidéos avec manière de mouvement non marqué en gris clair, avec indication de la variation individuelle sous forme de barres d'erreur)

L'importance du franchissement de frontière a déjà été remarquée, et constitue ce qui a été appelé "boundary-crossing constraint" (cf. Aske 1989, Slobin & Hoiting 1994) : dans les langues à cadre verbal, la description verbale de scènes contenant un franchissement de frontière se fait typiquement à l'aide de verbes de trajectoire (*entrer*, *sortir*, *passer*), au détriment des verbes de manière. Il faut noter cependant que lorsque ces deux éléments sont combinés (franchissement de frontière + manière marquée), la manière peut tout de même être exprimée. Les phrases suivantes, en piémontais, illustrent bien ce phénomène : on voit qu'une scène comprenant un mouvement avec franchissement de frontière est décrite avec ou sans verbe de manière selon que (respectivement) ce mouvement est marqué ou non :

(62) (piémontais)
na fia c a **intra** nt na crota
'IND.F fille REL 3SG **entrer** dans IND.F grotte'
« une fille qui **entre** dans une grotte »
(Traj22_pm_08)

(63) (piémontais)
an cit c a **cur** andrinta a l' eva
'IND.M garçon REL 3SG **courir**.PRS.3SG dans à DET eau'
« un garçon qui **court** dans l'eau »
(Traj59_pm_09)

À l'inverse, dans les langues à satellites, la manière est typiquement exprimée même lorsqu'elle n'est pas marquée dans la scène, comme l'illustrent bien les deux exemples suivants, en polonais et suédois :

(64) (polonais)
mężczyzna wy- **szedł** z krzaków
'homme.NOM hors.de- **marcher**.3SG.M.PST hors.de buisson.GEN.PL'
« L'homme est sorti des buissons **en marchant** »
(Traj55_pl_01)

(65) (suédois)

en	kvinna	**går**	in	i	en	grotta
‘IND	femme	**marcher**.3SG.IND.PRS	dans	dans	IND	grotte’

« Une femme entre dans une grotte **en marchant** »
(Traj54_sw_13)

Bien que tout ne soit pas encore codé et analysé, les premiers résultats de ces recherches sont donc encourageants, et permettent de pointer les limites de la typologie de Talmy. Il apparaît en effet assez clairement, au vu de ces résultats, qu'il est difficile de classer les langues avec pour seul critère le lieu d'expression de la trajectoire (verbe ou satellite) : ce classement ne fonctionne qu'imparfaitement pour prédire la façon dont les langues retenues dans cette étude expriment la manière de mouvement, et très mal pour prédire l'expression de la trajectoire et de la deixis, pourtant des éléments fondamentaux dans la description du mouvement. De plus, nos résultats suggèrent que la proximité génétique n'est pas à elle seule un bon prédicteur de la proximité de la production dans l'expression de la trajectoire. Pour la deixis, la fréquence d'expression est similaire en allemand et en thaï, en français et en suédois, enfin en polonais et en piémontais ; il y a donc à la fois des similarités indépendantes de la famille linguistique (polonais / piémontais, français / suédois, allemand / thaï) et des écarts au sein des familles (français-piémontais, allemand-suédois). De même, pour l'expression de la manière, il y a des affinités assez fortes indépendamment des familles de langues, notamment entre allemand, polonais et thaï. Décrire des types de construction plutôt que des types de langue, comme le suggèrent Croft et al. (2010), pourrait être une piste intéressante, permettant de rejoindre en partie les hypothèses de Talmy, mais en les affinant.

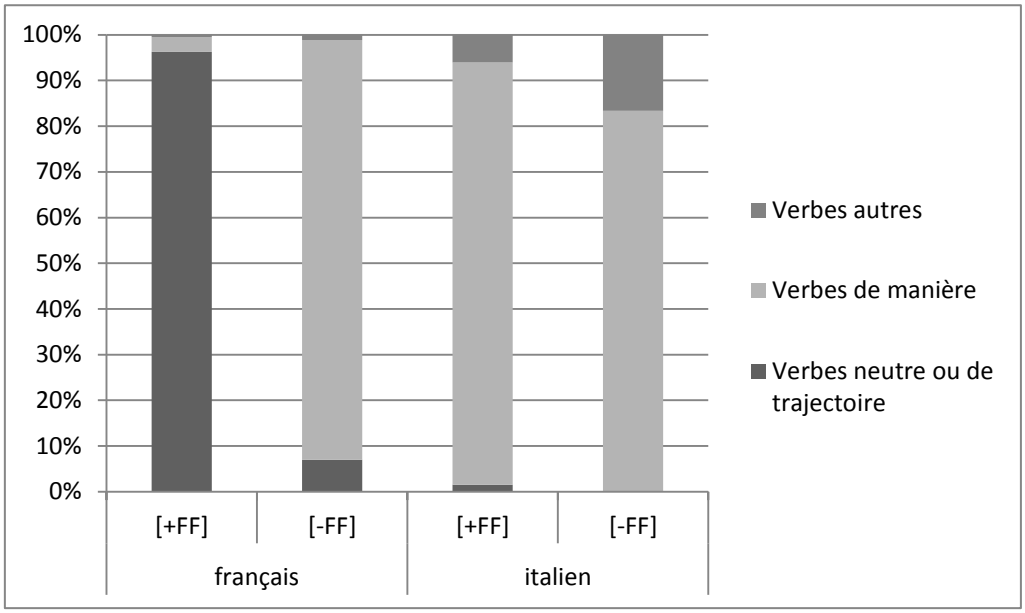
4.2.3 Etude de l'expression du mouvement : approches diachroniques

Etant donné que le type du latin dans la typologie de Talmy est l'opposé du type roman, on ne peut que s'interroger sur ce qui a causé le glissement de l'un à l'autre : pourquoi et comment passe-t-on d'une langue à satellites à une langue à cadre verbal, ou inversement ? J'ai abordé cette question dans le cadre d'une série d'études menées avec Claudio Iacobini (Iacobini & Fagard 2011 [V : 675 sqq.], soumis). Ces études, ainsi que les résultats préliminaires d'une comparaison entre différentes langues romanes pour des étapes anciennes, semblent indiquer que la différenciation n'est pas récente (voir aussi, sur ce thème, Stolova 2015). Ainsi, une étude que j'ai menée sur corpus en français (base BFM), italien (base OVI), portugais (base CdP) et espagnol (base CORDE) semble indiquer des différences assez marquées. Afin de déterminer l'importance du critère de franchissement de frontière, j'y ai cherché toutes les cooccurrences des verbes *aller*, *courir*, *entrer* avec les prépositions équivalentes à *en* et *dans*, ainsi que toutes les occurrences de constructions du type *dans la salle*, *dans le château*. J'ai ainsi codé 300 occurrences par langue environ, et les occurrences exploitables indiquent bien une différence entre les langues (tableau 3) :

Langue	Type de scène	Type de verbe			Total
		verbe neutre ou verbe de trajectoire	verbe de manière	autres verbes	
espagnol	[+FF]	49	1	12	*62*
	[-FF]	0	1	2	*3*
français	[+FF]	358	12	2	*372*
	[-FF]	6	79	1	*86*
italien	[+FF]	1	61	4	*66*
	[-FF]	0	5	1	*6*
portugais	[+FF]	85	17	33	*135*
	[-FF]	0	1	4	*5*
Total		*499*	*177*	*59*	***735***

Tableau 3 : Types de verbes de déplacement dans un corpus médiéval (espagnol, français, italien, portugais)[25]

Il s'agit de données trop limitées pour en tirer des conclusions définitives, d'autant qu'il faudrait disposer de corpus strictement comparables pour aller plus loin. Elles permettent au moins de postuler quelques hypothèses, à partir de deux faits marquants : il n'y a qu'en français qu'on trouve une opposition nette en fonction du type de scène, avec des verbes de manière surtout dans les scènes sans franchissement de frontière ; il n'y a qu'en italien qu'on trouve fréquemment des verbes de manière dans les scènes avec franchissement de frontière (graphique 13).



Graphique 13 : Types de verbes en français et en italien (données du tableau 3)

On peut donc formuler les hypothèses suivantes, à tester sur de plus larges corpus et sur d'autres constructions :

[25] Les chiffres donnés correspondent aux occurrences pour lesquelles il m'a été possible de déterminer (d'après le contexte) s'il y avait ou non franchissement de frontière (FF), ce qui explique que le total soit réduit.

1. en français médiéval, la manière de mouvement est exprimée quand il n'y a pas franchissement de frontière (cf. Aske 1989, Slobin & Hoiting 1994, Fagard et al. sous presse [V : 613 sqq.]) ; s'il y a franchissement de frontière, l'expression de la manière est marquée, et donne lieu à l'utilisation de constructions spécifiques (*courir jusque + SP* ou *verbe de mouvement + (en) corant*)
2. en italien médiéval, probablement en lien avec la présence fréquente d'adverbes de trajectoire (Iacobini & Masini 2006), la manière de mouvement est exprimée fréquemment, même quand il y a franchissement de frontière
3. en espagnol et en portugais médiévaux, les adverbes de trajectoire sont présents mais pas suffisamment fréquents pour inhiber l'expression de la trajectoire dans le verbe

Il ne s'agit là que d'hypothèses, qui demandent à être (in)validées. Elles reposent sur les observations que je décris brièvement dans ce qui suit. En français médiéval, on trouve principalement des exemples avec franchissement de frontière et sans manière : dans ce type de construction, aussi bien le verbe que les syntagmes prépositionnels expriment la trajectoire, comme dans les deux exemples ci-dessous.

(66) (moyen français)
en ceste presente sepmaine, en l'ostel du chauderonnier demourant soubz les pilliers, ès halles, ouquel il avoit esté logié, **il entra par une luquerne en une chambre** dudit hostel, et illec print et embla une courte houppelande de drap pers
« [Il reconnut et confessa, outre ce que nous avons rapporté ci-dessus, que] cette semaine, en l'hôtel du chaudronnier qui demeure sous les pilliers, aux halles, où il avait été logé, **il est entré par une lucarne dans une chambre** de cet hôtel, et y a pris et dérobé une courte houppelande de drap bleu »
(BFM, *Registre criminel du Châtelet*, t. 2, p. 217, 1389)

(67) (moyen français)
(…) quant ilz se partirent, ilz cuidierent **entrer en l'ostel** d'une poissonniere **par les fenestres** de son hostel, qui estoient ouvertes, pour la cuidier desrober
« Quand ils se séparèrent, ils pensèrent **entrer dans le logis** d'une poissonnière **par les fenêtres** de son logis, qui étaient ouvertes, pour la voler »
(BFM, *Registre criminel du Châtelet*, t. 1, p. 98, 1389)

La trajectoire peut également être exprimée par un adverbe, ce qui n'entraîne pas l'expression de la manière dans le verbe (cf. les exemples ci-dessous), de manière assez étonnante (et contraire aux hypothèses de Talmy 1985, 2000) :

(68) (moyen français)
il **entra arriere** en l'yaue.
« Il **entra** dans l'eau **en arrière** »
(BFM, Gaston Phébus, *Livre de chasse*, p. 208, ch. 45, 1387)

(69) (moyen français)
Le fel cuvert esploita tant qu'il vint en la mer et **entra ens**, et les damoiselles avecques lui
« Le vil traître fit tant qu'il arriva à la mer et y entra [litt. « **entra dedans** »], et les demoiselles avec lui »
(BFM, *Bérinus*, t. 1, p. 324, 1350)

Alors même que l'ancien français présente certaines caractéristiques des langues à satellite (Kopecka 2009, Burnett & Tremblay 2009, Troberg 2011), et que l'expression du mouvement

sous la forme [Verbe de manière + Adverbe de trajectoire] est possible, ce type de construction n'est pas très fréquent. Cependant, la manière peut être exprimée, et même dans le verbe principal. C'est le cas lorsque la trajectoire est exprimée par un syntagme introduit par *jusque* (70), en particulier lorsqu'il y a distribution des informations spatiales sur plusieurs propositions (71) :

(70) (moyen français)
(…) toutez cez chosez faittez, ainsy que je vous dis, le conte d'Urgel lez siens mis en bonne ordonnance, **coururent** de grant radeur **jusquez en l'ost** et sans espargnier et de prime fache se ferirent ez Castelains, desquelz ilz firent grant occision pour ce que lez pluiseurs estoient nudz, c'est a dire sans armeurez et pris en desarroy
« le comte d'Urgel mit les siens en bon ordre, et ils **se précipitèrent dans la mêlée** »
(BFM, *Le roman du comte d'Artois*, p. 40, 1453)

(71) (moyen français)
Et Roland ne cessa de **courir, jusques il fut dedens la tour**. Et quant les barons furent leans, ilz fermerent les portes bien seurement, sans avoir paour que on leur fist empeschement ne viollence par lors nullement.
« Et Roland ne cessa de **courir jusqu'à ce qu'il soit dans la tour**. »
(DMF, Jehan Bagnyon, *L'histoire de Charlemagne*, p. 122, 2nd livre, 2nde partie, ch. XVI, 1465)

Par ailleurs, la manière peut être exprimée par un gérondif, donc comme un élément secondaire, le verbe principal exprimant la trajectoire (au sens large) :

(72) (moyen français)
Je m'en vois a l'ostel **courant,** Pour mon fil faire avoir baptesme.
« **Je m'en vais en courant** à mon logis, pour faire baptiser mon fils »
(DMF, *Miracle de l'enfant donné au diable*, p. 14-15, 1339)

En italien médiéval, les constructions du type *andare* + *in* sont très nettement privilégiées. L'italien n'exprime donc très souvent le franchissement de frontière que dans un syntagme prépositionnel ou adverbial (73-74), plutôt que dans le verbe – et dans ce cas, c'est que l'accent est mis sur le franchissement de frontière… (75) :

(73) (italien médiéval)
Ma ddappoi che T. e madonna Isotta fuorono ala magione dela savia damigiella, e T. ismontoe da cavallo e **andoe dentro ala magione**, e vide bene ch'ella iera più bella e ppiù delettevole a vedere che ggiamai fosse veduta al mondo.
« Mais quand Tristan et madame Iseult furent chez la sage demoiselle, Tristan descendit de cheval et **alla dans la maison**, et vit bien qu'elle était la plus belle et plaisante à voir qu'on ait jamais vue au monde »
(OVI, *Tristano Riccardiano*, p. 176, 1300)

(74) (italien médiéval)
Ella si lieva dinanzi a suo barone e **corre dentro a sua camera** e prese suo picciolo fanciullo Arnantes e lo porta dinanzi a suo padre e glil mise a' piei.
« Elle se leva devant son baron et **courut** [litt. « court »] **dans sa chambre**, puis prit son petit garçon Arnantes et l'apporta devant son père, et le mit à ses pieds »
(OVI, Binduccio dello Scelto, *La storia di Troia*, p. 340, 1322)

(75) (italien médiéval)
E quando nascie, la prima sua vocie è di dolore, però ch'**escie del caldo luogo** e viene nel contradio.

« Et quand [le bébé] naît, son premier cri est de douleur, parce qu'**il sort d'un lieu chaud** et arrive dans un lieu contraire [c'est-à-dire froid]. »
(OVI, Fr. da Barberino, *Del reggimento e costumi di donna*, p. 372, 1318-20)

En espagnol médiéval, les constructions du type *entrar* + *en* sont privilégiées – la trajectoire étant donc le plus souvent exprimée simultanément dans le verbe et un syntagme adverbial (76) ou prépositionnel (77) ; les énoncés contenant un verbe de manière et un syntagme prépositionnel indiquant la trajectoire sont bien plus rares (78), et on trouve plutôt pour indiquer la manière des constructions complexes (79) :

(76) (espagnol médiéval)
Bien seas venido. **entra dentro** et yantaras.
« Sois le bienvenu. **Entre** et tu mangeras. »
(Corde, *Esopete ystoriado*, fol. 134v, 1482)

(77) (espagnol médiéval)
(…) atándolo debaxo de un portal, **entró en la casa** y vio un hombre fincado de inojos rezando
« Ayant attaché [son cheval] sous un portique, **il entra dans la maison** et vit un homme qui se tenait à genoux et priait »
(Corde, Rodríguez de Montalvo, *Garci Amadís de Gaula*, p. 1105, 1482-1492)

(78) (espagnol médiéval)
le dexo salir dende mandando le que **anduviesse dentro en casa**
« il le laissa sortir de là, ordonnant-lui de **se rendre dans la maison** »
(Corde, *Esopete ystoriado*, fol. 130v, 1482)

(79) (espagnol médiéval)
quando él lo supo, **tyróse** más que de paso **dentro de las montañas**.
« Quand il le sut, **il se lança dans les montagnes** à toute vitesse »
(Corde, *La corónica de Adramón*, p. 574, 1492)

Dans le détail, on voit qu'en espagnol médiéval, dans notre corpus, le franchissement de frontière est le plus souvent exprimé par *entrer*, *venir*, *revenir en/dans* ; en italien médiéval, plutôt par *aller*, *courir en/dans* et très peu par *entrer en/dans*. En portugais médiéval, enfin, on trouve les deux types de construction (80-82) ; les verbes de manière semblent être plus fréquents qu'en espagnol et en français, mais moins qu'en italien :

(80) (portugais médiéval)
per que os imiigos possam **entrar** a mall fazer **ao castello**.
« pour que les enemis puissent entrer dans le château pour méfaire »
(CdP, *Castelo Perigoso*, 15ème siècle)

(81) (portugais médiéval)
o viram sayr, foron se logo rriigos e dereytos aa porta e **entraram no castello** e lançarom logo fora a molher do alcayde e os que com ella estavam
« ils le virent sortir, allèrent droit à la porte et **entrèrent dans le château**, puis expulsèrent de là la femme du gouverneur et ceux qui étaient avec elle »
(CdP, *Estoria de Dom Nuno Alvares Pereyra*, 15ème siècle)

(82) (portugais médiéval)
el rey d'Arryona vyo que aquelles cavaleiros se poseron a tal perigoo por **acorrer ao castello** e que erã ja dentro

« Le roi d'Arragone vit que ces chevaliers s'exposaient à de tels riques pour **accourir au château** et qu'ils étaient déjà entrés »
(CdP, *Crónica Geral de Espanha*, 14ème siècle)

4.2.3.1 Limites et pistes pour des recherches futures

Cependant, comme je l'indiquais plus haut, il s'agit ici de pistes de réflexion plutôt que de résultats probants. En effet, il reste un certain nombre de problèmes méthodologiques et théoriques à régler avant d'étendre le corpus et de pouvoir vérifier mes hypothèses, notamment (a) le codage de la manière ; (b) le traitement des emplois métaphoriques.

Le point le plus ardu est bien celui du statut de la MANIERE, que j'illustre ci-dessous à l'aide des emplois de l'ancien français *courir*. Ce dernier est en effet souvent employé avec un sens assez éloigné de la MANIERE ; ainsi lorsqu'il peut être glosé par « se dépêcher », comme dans l'exemple ci-dessous :

(83) (moyen français)
Chiers sires, il vous fault ces bonnes gens conforter, car se vous perdés la forte ville de Calais, ce vous sera uns trop grans prejudisses et a vostre roiaulme, et averont les Englois trop biel a venir et a armer a Calais et **courir en** France et la retraire et retourner en lor pais.
« Mes seigneurs, vous devez réconforter ces bonnes gens, car si vous perdez la ville-forte de Calais, cela sera un trop grand préjudice pour vous et votre royaume, et il sera facile pour les anglais de venir armer Calais et **courir en France** puis faire demi-tour et retourner dans leur pays »
(BFM, Froissart, *Chroniques*, p. 819-820, 1400)

De même lorsqu'il désigne le mouvement d'un non-humain, qu'il s'agisse d'un animal ou, pire (dans le sens où l'on s'éloigne de la « course » prototypique d'un être humain), d'un non-animé :

(84) (moyen français)
Je ne le retins pas bien ; ung jour qui passa, si m'en print mal. Car, **en courant** [en] une valée après le lievre et mes chiens, **mon cheval** se rompit le col, et je fuz tresbien blecié. Et en memoire de ce est la pendue la peau du cheval que alors je perdy.
« Je ne l'ai pas bien retenu [le deuxième enseignement de mon père, à savoir de ne pas laisser courir son cheval] et, un jour, mal m'en a pris. Car **en courant** dans une vallée après le lièvre et mes chiens, **mon cheval** s'est cassé le cou, et j'ai été gravement blessé. C'est en souvenir de cette aventure que j'ai pendu la peau du cheval que j'y ai perdu »
(BFM, *Les cent nouvelles nouvelles*, p. 335-336, 1456)

(85) (moyen français)
Dieu voulut ainsi disposer des choses que ceste nuyct sourdit une grande tormente, et telle qu'il faillit que l'armée dudict duc de Bourgongne fuist, et **coururent les ungs des navires en Escosse, les autres en Hollande**, et en peu d'heures après se trouva le vent bon pour ledict conte, lequel passa sans peril en Angleterre.
« Dieu a voulu ainsi disposer les choses – cette nuit-là une grande tempête se déchaîna, faisant fuir l'armée du duc de Bourgnonne : les navires **coururent les uns en Ecosse, les autres en Hollande**, et en quelques heures après cela le vent fut favorable au comte, qui passa sans péril en Angleterre »
(BFM, Philippe de Commynes, *Mémoires*, t. 1, p. 199, 1489)

Dans un premier temps, j'ai considéré que les emplois métaphoriques (par exemple ci-dessous), trop éloignés de l'expression du mouvement réel, devaient être traités à part.

(86) (italien médiéval)
Una donna mi passa per la mente, **ch'a riposar si va dentro nel core**; ma trova lui di sì poco valore, che de la sua vertù non è possente; sì che si parte disdegnosamente e làsciavi uno spirito d'amore
Une femme me passe par la tête, et **va reposer dans mon cœur**, mais le trouve de si peu de valeur qu'il n'est pas digne de sa vertu, et elle repart donc dédaigneusement et y laisse un soupir d'amour
(OVI, Cino da Pistoia, *Rime*, p. 566, 1336)

Un dernier point à retenir est celui de la répartition des informations au sein de l'énoncé (Fagard et al., soumis). C'est d'ailleurs là un angle d'attaque intéressant pour revoir la typologie de Talmy, en la complétant par une analyse constructionnelle, comme annoncé plus haut. L'exemple suivant est une bonne illustration, montrant qu'une langue à cadre verbal peut exprimer la manière du mouvement mais, dans ce cas, les informations relatives à la manière et à la trajectoire sont typiquement réparties sur deux propositions :

(87) (moyen français)
Et chevauça li rois d'Engleterre et **s'en vint au siege devant Vennes**, et la se tint.
« **Le roi d'Angleterre chevaucha** et **vint au siège devant Vennes**, et y resta. »
(BFM, Froissart, *Chroniques*, p. 588, 1400)

J'ai testé cette hypothèse en comparant les données Trajectoire en français et suédois, avec des résultats préliminaires encourageants (communication, Fagard et al. 2013b) qui permettent d'aller un peu plus loin que la simple vérification de la présence ou absence des traits sémantiques MANIERE et TRAJECTOIRE selon le type de langue (tableau 4).

Traits sémantiques présents dans l'énoncé	TRAJECTOIRE & MANIERE (coprésents dans une même proposition)	TRAJECTOIRE & MANIERE (distribués sur plusieurs propositions)	TRAJECTOIRE seule	MANIERE seule	Total
français	333	75	427	45	880
suédois	749	37	81	55	922
Total	1082	112	508	100	1802

Tableau 4 : Répartition de l'information dans les données trajectoire – français et suédois (communication, Fagard et al. 2013b)

Ainsi, il apparaît que la différence principale entre les deux langues concernant la répartition des informations est que, conformément aux attentes, les propositions combinant manière et trajectoire sont bien plus fréquentes en suédois (81 % des énoncés) qu'en français (36 % des énoncés) : non seulement la manière est moins fréquente en français, mais elle est généralement exprimée à part, avec une répartition des information sur plusieurs propositions qui est deux fois plus fréquente qu'en suédois (8 % des énoncés en français, 4 % en suédois). Enfin, les énoncés exprimant uniquement la trajectoire, au détriment de la manière, sont plus fréquents en français (46 % des énoncés, contre 9 % en suédois). On notera que ces résultats sont statistiquement significatifs : la coexpression manière et trajectoire est nettement plus fréquente en suédois ($khi^2 = 353{,}43$, $p<0{,}01$), la distribution des deux plus fréquente en français ($khi^2 = 15{,}71$, $p<0{,}01$). Ils n'ont rien de révolutionnaire et tendent plutôt à confirmer ce que l'on sait, mais permettent d'avoir une idée plus précise des différences interlinguistiques ; il reste à les affiner encore et à répéter l'analyse pour les autres langues du corpus (Fagard et al., soumis).

4.3 Le mouvement fictif

« Papa, pourquoi les routes vont ? »
(Jules, 2 ans et 10 mois, 2012)

Plus encore que l'espace, le mouvement semble fondamental pour l'être humain (Berthoz 2013). De ce point de vue, le mouvement fictif est un phénomène de choix pour étudier les liens entre langage et cognition : il s'agit par exemple de l'emploi de verbes de mouvement pour décrire des situations statiques, et les explications avancées pour justifier l'existence de ce phénomène dans de nombreuses langues semblent bien indiquer qu'il est dû à l'importance du mouvement pour la cognition – plusieurs auteurs parlent, par exemple, de simulation mentale (« mental simulation », Barsalou 1999, Matlock 2004, Richardson & Matlock 2007).

Il regroupe en fait plusieurs phénomènes proches, qui correspondent peu ou prou aux différentes nomenclatures que l'on peut trouver : mouvement virtuel (« virtual motion », Talmy 1983), mouvement subjectif (« subjective motion », Langacker 1987, Matsumoto 1996), mouvement fictif (« fictive motion », Talmy 2000, Matlock 2004), mouvement abstrait (« abstract motion », Matlock 2010), descriptions de trajectoire (« path settings », Huumo 2014).

Les recherches menées dans le cadre du projet ET nous ont permis de confirmer l'existence de motivations diverses pour ce phénomène, et surtout de montrer que la variation interlinguistique s'explique en partie par l'impact des systèmes linguistiques particuliers, en partie par leur sensibilité à différentes motivations (Stosic et al. 2015 [V : 631 sqq.]).

4.3.1 Données diachroniques sur le mouvement fictif

J'ai abordé ce phénomène en premier lieu à travers des manifestations assez atypiques : les énoncés où il semble y avoir une équivalence entre mouvement d'une entité (typiquement, animée humaine) et des formes moins concrètes de mouvement, liées à la perception auditive et visuelle :

Mouvement réel
[x va depuis A jusqu'à B, vers B]

Mouvement fictif
[x voit B depuis A]
[x entend B depuis A]
[x regarde vers B]

On peut observer ce phénomène en diachronie, et j'ai pu montrer (Fagard 2010a) la fréquence des extensions sémantiques qui s'expliquent ainsi. On trouve par exemple de nombreuses illustrations de la chaîne de grammaticalisation [POINT DE DEPART] > [ORIGINE D'UN MOUVEMENT FICTIF], illustrée ci-dessous :

(88) (moyen français)
Dont regarda Aigres **contremont la mer** et perçut une nef moult acourre roidement.
« Aigre regarda alors **au loin sur la mer** et aperçut un navire, qui s'approchait très vite. »
(BFM, *Bérinus*, p. 256, 1350)

(89) (français contemporain)
Petit enfant, je reste sur le trottoir d'en face, à regarder le couloir **depuis** la rue jusqu'à un jardinet, au fond, de l'hôpital de notre village.
(Frantext, Pierre Guyotat, *Coma*, p. 23, 2006)

(90) (français contemporain)
Ceux qui les entourent, sur la piste, sont électrisés, et ceux qui observent **depuis** le bord, eh bien, ils ne peuvent pas faire autrement, ils sont emportés ; de la tête ou du menton, des hanches ou de la main, des genoux ou des épaules, à leur corps défendant, ils dansent aussi.
(Frantext, Gérald Tenenbaum, *L'Ordre des jours*, p. 56, 2008)

(91) (ancien français)
Les Turs nous vindrent hardoiant jusques en nostre ost ; pour ce commanda le roy que l'en cloussit nostre ost de fossés par devers Damiete **depuis le fleuve de Damiete jusques au flum de Rexi**.
« Les Turcs vinrent nous narguer jusque dans notre campement ; c'est pourquoi le roi ordonna que l'on entoure le camp de fossés, du côté de Damiette, **depuis le fleuve de Damiette jusqu'à celui de Rexi**. »
(BFM, Jean de Joinville, *La vie de saint Louis*, p. 197, 1305)

(92) (catalan)
ell s'ho mirava **des de**l balcó
« Il le regardait **du** balcon »
(exemple de dictionnaire)

(93) (espagnol)
el periodista está hablando en directo **desde** el estadio de fútbol
« le journaliste parle en direct **du** stade de foot »
(exemple de dictionnaire)

(94) (ancien français)
le chastel qui siet sur la mer, **dès** l'une mer jusques à l'autre
« le château qui est installé sur la mer, **d'**une mer jusqu'à l'autre »
(Frantext, Jean de Joinville, *La vie de saint Louis*, p. 280, 1305)

Il semble qu'il faut aussi inclure dans ce phénomène les extensions du type [DIRECTION] > [BUT], et donc toutes les grammaticalisations de prépositions sur la base du mot « tête » ou « tourné vers ». C'est le cas des formes romanes formées sur le latin *facies* 'aspect, figure, visage' (espagnol *hacia*[26], français *face à*, portugais *face a*, etc.) ou *caput* 'tête' (catalan *cap a*) – et d'autres grammaticalisations du même type dans d'autres langues (Heine & Kuteva 2002 : [FACE] > [FRONT] en mixtec, copala trique, quiché colonial, alamblak, ||ani, gimira, halia, vaï ; [FOREHEAD] > [FRONT] en dullay et boulou ; [HEAD] > [FRONT] en masaï, alamblak et anglais…).

[26] Il faut noter que l'étymologie de *hacia* est débattue, et parfois ramenée à l'arabe *hatta* (cf. le portugais *até*) ; il pourrait aussi y avoir eu fusion entre les deux origines.

(95) (ancien français)
Guardez amunt **devers** les porz d'Espaigne : Veeir poez, dolente est la rereguarde
« Regardez en haut, **vers** les cols d'Espagne : vous pouvez le voir, l'arrière-garde est en mauvaise posture »
(BFM, *La chanson de Roland*, v. 1103, 1100)

(96) (ancien français)
Carles guardat amunt **envers** le ciel, Veit les tuneires e les venz e les giels E les orez, les merveillus tempez, E fous e flambes i est apareillez
« Charles regarda en haut, **vers** le ciel, vit le tonnerre, le vent, le gel et les brises, les tempêtes formidables, le feu et les flammes prêts à tomber »
(BFM, *La chanson de Roland*, v. 2532, 1100)

(97) (ancien français)
Cume David s'en fud turnez é estut luinz al sumet d'un munt, é un grant val fud entre lúi é l'ost, criad **vers** ces de l'ost é vers Abner le fiz Ner é dist : « Dun ne me respunderas tu, Abner ? Dun ne me respunderas tu, Abner ? »
« [...] il cria **en direction de** la troupe [...] »
(BFM, *Les quatre livres des Rois*, p. 53, 1190)

Il faut noter cependant l'opposition intrigante entre d'un côté un glissement de sens très fréquent, illustré par les exemples ci-dessus, et d'un autre des emplois qui sont très rares pour au moins certaines prépositions, comme l'illustre le tableau (5) ci-dessous pour la préposition *depuis* :

Sens de *depuis*	en nombre d'occurrences	en pourcentage des occurrences
'spatial'	*4436*	99,15
'spatial – mouvement fictif'	*30*	0,67
autres	*8*	0,18
Total des occurrences	*4474*	100

Tableau 5 : Emplois de la préposition *depuis* dans un corpus moderne (Frantext – 20ème siècle, Fagard 2010a).

4.3.2 Mouvement fictif : données élicitées

J'ai également travaillé sur le mouvement fictif à partir de données élicitées, dans le cadre du projet ET. L'outil d'élicitation utilisé consiste en une série d'images créées par Johan Blomberg et Jordan Zlatev. Il s'agit de 38 images : 2 d'entraînement, 12 distracteurs et 24 images « cibles », qui représentent différentes configurations permettant plus ou moins facilement la projection ou la conceptualisation d'un mouvement (voir ci-dessous). Il a été conçu afin de tester une hypothèse précise : on sait que les langues permettent plus ou moins facilement d'exprimer avec des verbes de mouvement des trajectoires ne correspondant pas à un mouvement réel. Ainsi, en français, *la route monte vers le sommet* est tout à fait acceptable, mais ?*les maisons s'égrènent à l'horizon* paraît marqué. L'hypothèse à tester est que plusieurs facteurs sont en jeu, et que le poids de chaque facteur diffère d'une langue à

l'autre. Parmi les motivations possibles et connues d'après les travaux antérieurs (travaux cités plus haut de Talmy, Matlock, Langacker, Matsumoto), Blomberg & Zlatev (2014) retiennent les suivantes :

(i) la nature active (*(en)active*) de la perception,
(ii) la nature relationnelle (*correlative*) de l'intentionnalité
(iii) l'imagination (*re-enactive imagination*).

Ce caractère hybride des réalisations linguistiques liées au mouvement fictif expliquerait les variations typologiques observées (par exemple dans. Pour tester ces motivations, les auteurs ont mis en œuvre dans l'outil d'élicitation deux paramètres :

(1) la compatibilité avec une trajectoire effective :

certaines images présentent une route ou un chemin : en regardant ces images, on peut facilement imaginer qu'on se déplace le long de la route ou du chemin ;

d'autres présentent des trajectoires dans lesquelles il est moins facile de se projeter : un fil électrique, une barrière, une canalisation ;

(2) le point de vue :

la trajectoire peut être représentée au premier plan (égocentrée) ;
elle peut être représentée en arrière-plan (allocentrée).

Nous avons récolté avec ce matériel des données pour les langues suivantes :

- thaï, suédois, français, italien, allemand (12 à 20 locuteurs par langue),
- serbe (6 locuteurs), portugais (5 locuteurs), hongrois (2 locuteurs).

Les prédictions formulées sur la base des hypothèses de travail étaient les suivantes :

1 : toutes les catégories donneront lieu à la production d'énoncés contenant des expressions de mouvement fictif, à l'exclusion des distracteurs ;

2 : en fonction de la motivation principale du phénomène, on s'attend à des effets différents :

2a) si c'est la nature (en)active de la perception, le mouvement fictif devrait être plus fréquent pour les images de type [PROPICE AU DEPLACEMENT HUMAIN], indépendamment du point de vue ;

2b) si c'est le parcours mental (*mental scanning*), le mouvement fictif devrait être plus fréquent pour les images de type [DISTAL], indépendamment du trait [+/- PROPICE AU DEPLACEMENT HUMAIN] ;

2c) s'il s'agit d'un phénomène hybride, on s'attend à trouver plus de mouvement fictif dans le cas [PROXIMAL + PROPICE AU DEPLACEMENT HUMAIN].

Je résume, dans le tableau (6) ci-dessous, les hypothèses retenues en fonction du type d'image où les énoncés de mouvement fictif sont les plus fréquents :

Type d'image		trait [PROPICE AU MOUVEMENT]	
		+	-
PERSPECTIVE	proximale	*motivation hybride (perception + parcours mental)*	*autre motivation*
	distale	*perception*	*parcours mental*

Tableau 6 : Configurations et hypothèses pour l'outil d'élicitation FM, en fonction du type d'image suscitant le plus d'énoncés de mouvement fictif (Blomberg & Zlatev 2014)

D'après Blomberg & Zlatev (2014) et Blomberg (2014), les résultats obtenus à partir de l'analyse des données pour trois langues (français, suédois, thaï) tendent à confirmer globalement les prédictions, en particulier (1), (2a) et (2c). On notera en particulier les résultats suivants :

(i) **les quatre types d'images ont donné lieu à la production d'énoncés avec du mouvement fictif**, ce qui semble confirmer l'hypothèse de motivations multiples.

(ii) **on observe des différences marquées entre les langues**, ce qui confirme l'interaction entre motivations cognitives et contraintes propres à chaque système linguistique – ainsi, 53.3 % des descriptions en suédois contenaient au moins un verbe de mouvement et environ 50 % en thaï, contre seulement 35 % en français.

Les exemples (98-99) illustrent bien les oppositions entre français et suédois vues en §4.2 : tandis qu'on trouve en suédois un verbe de manière (*går* « va », exemple 98), on trouve en français, assez typiquement, un verbe de trajectoire (*un pont… pour traverser*, exemple 99)

(98) (suédois)

en	hängbro	som	**går**	mellan	två	berg
'IND	pont.suspendu	REL	aller.3SG.PRS	entre	deux	montagne'

« un pont.suspendu qui **va** entre deux montagnes »
(Fm09_sw_23)

(99) (français)
Un pont suspendu **pour traverser** un grand précipice au dessus des nuages
(Fm09_fr_12)

Cette projection des caractéristiques observées pour le déplacement n'est pas aussi nette en thai, en revanche : comme l'a noté Blomberg (2014), on ne trouve pas de verbes de manière dans les énoncés de mouvement fictif (100-101), alors que ceux-ci étaient presque constants dans l'expression du déplacement. Les séries verbales qu'on trouve dans les données FM sont plutôt du type [verbe d'action + verbe de trajectoire] (100) ou [verbe de trajectoire + verbe déictique] (101) :

(100) (thai)

saphan	**yong**	**kham**	rawang	song	napha
'pont	**relier**	**traverser**	entre	deux	falaise'

« Un pont **relie en traversant** entre deux falaises »
(Fm09_th_07)

(101) (thai)

mi	tho	prapa	**khâw**	**pai**	nai	chong	(…)
il.y.a	canalisation	eau	**entrer**	**aller**	dans	trou	

« une canalisation d'eau **entre** dans un trou [**en s'éloignant de moi**] »
(Fm19_th_14)

(iii) les images de type [PROPICE AU DEPLACEMENT HUMAIN] ont favorisé la production d'énoncés avec mouvement fictif (102-103), ce qui confirmerait l'hypothèse de l'énaction comme cause principale du phénomène

(102) (suédois)

en	väg	som	**kommer**	ut	ur	en	tunnel
'IND	route	REL	venir.3SG.PRS	dehors	hors.de	IND	tunnel'

« une route **vient** en sortant d'un tunnel [litt. vient dehors hors.d'un tunnel] »
(Fm07_sw_25)

(103) (français)
c'est une sortie de tunnel qui **débouche** sur une route qui s'en va vers la campagne
(Fm08_fr_11)

L'expression du mouvement fictif dans les images ne relevant pas de ce type, notamment en français, est parfois assez intrigante – et on peut se demander, dans l'exemple (104) suivant, s'il n'y a pas projection malgré le type d'image en question :

(104) (français)
une canalisation d'égout qui semble fuir qui **traverse** qui s'apprête à entrer dans un tunnel
(Fm19_fr_06)

(105) (suédois)

ett	staket	som	**går**	på	en	strand	ut	mot	vattnet
'IND.NT	clôture	REL	aller.3SG.PRS	sur	IND	plage	dehors	vers	eau.DET.NT'

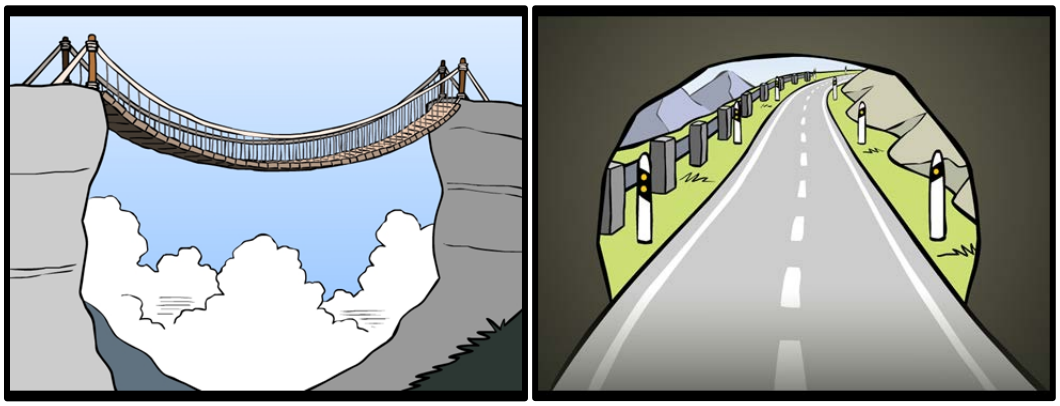
« une barrière qui **va** sur une plage dehors vers l'eau »
(Fm14_sw_25)

(106) (français)
On voit une espèce de clôture ou barrière sur la plage qui **va** de la mer jusqu'à l'autre extrémité de la plage
(Fm14_fr_10)

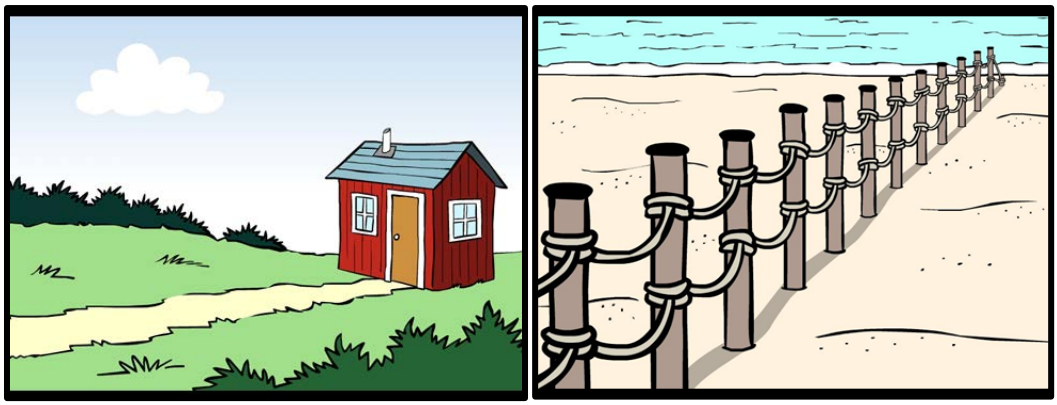
(iv) enfin, dans les trois langues, les images de type [+ PROPICE AU DEPLACEMENT HUMAIN ; + PROXIMAL] ont favorisé la production d'énoncés avec mouvement fictif, ce qui confirmerait l'hypothèse d'une motivation hybride. Cet effet est, d'après Blomberg (2014), particulièrement marqué en français, avec 48 % d'énoncés pour ce type et seulement 30 % pour les trois autres.

Cependant, dans une nouvelle étude faite à partir des données obtenues ainsi pour le français, l'allemand, l'italien et le serbe (Stosic et al. 2015 [V : 631 sqq.]), nous avons obtenu des résultats moins probants. Ainsi, en suivant le schéma expérimental d'origine, nous n'avons trouvé de différence significative pour aucune des quatre configurations. Intrigués, nous avons réévalué le matériel d'élicitation, cette fois-ci uniquement sur la base des images elles-mêmes

(et non des énoncés). Nous avons découvert un fort biais dans les images, montrant que l'opposition entre proximal et distal est en fait assez floue, avec certaines images ayant un statut intermédiaire entre proximal et distal. Plus précisément, tandis que certaines images sont clairement distales (14a) ou clairement proximales (14b), d'autres qui sont codées comme distales (14c) ou proximales (14d) sont finalement assez proches pour ce qui est de la perspective.

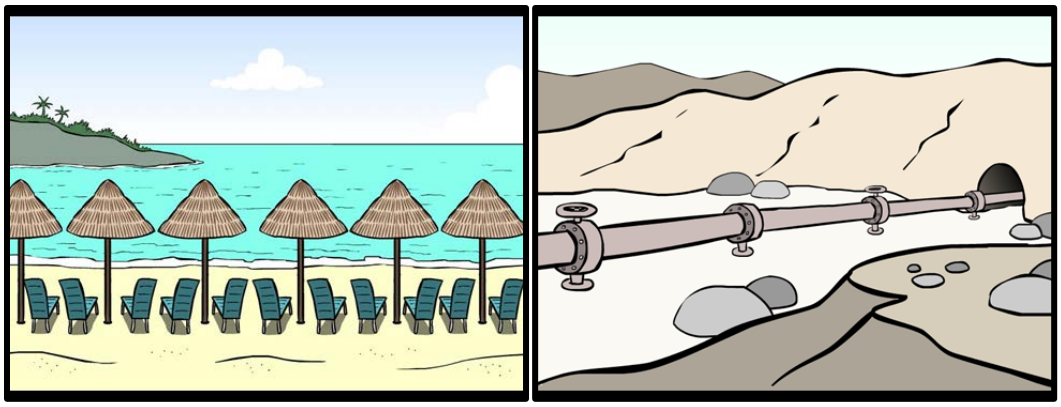


Graphiques (14a-b) (numérotés de gauche à droite) : Images de l'outil d'élicitation pour l'étude du mouvement fictif (Blomberg 2014) ; (a) clairement distal, (b) clairement proximal



Graphiques (14c-d) : Images de l'outil d'élicitation pour l'étude du mouvement fictif (Blomberg 2014) ; (c) moins nettement distal, (d) moins nettement proximal

Concernant le critère [PROPICE AU DEPLACEMENT], nous avons également repéré un biais inattendu : tandis que certaines images codées comme comme non propices au déplacement ne suggèrent effectivement pas naturellement le déplacement (15a), d'autres images de la même catégorie représentent en fait des éléments naturellement liés au déplacement, notamment les conduites (15b), qui ont effectivement suscité de nombreux énoncés de mouvement fictif.



Graphiques (15a-b) : Images de l'outil d'élicitation pour l'étude du mouvement fictif (Blomberg 2014) ; (a) image clairement non propice au déplacement, (b) image plus ambiguë

Nous avons par conséquent adopté un nouveau codage, abandonnant la perspective et retenant, pour le critère [PROPICE AU DEPLACEMENT], une typologie plus fine : [PROPICE AU DEPLACEMENT HUMAIN], [PROPICE AU DEPLACEMENT NON HUMAIN], [NON PROPICE AU DEPLACEMENT]. Les résultats obtenus avec ce nouveau codage sont, eux, significatifs, avec une fréquence moins élevée d'expressions de mouvement fictif dans le type [NON PROPICE AU DEPLACEMENT] (chi^2=231, p<.001), pour les 4 langues, et une fréquence élevée pour les deux autres, mais avec des différences qualitatives intéressantes (par exemple le recours aux verbes de mouvement causé comme CONDUIRE). Il reste à creuser la question des différences interlinguistiques, puisqu'il y a des différences significatives entre langues (de 30% d'énoncés de mouvement fictif en italien à 50% en allemand, par exemple).

Dans cette étude, nous avons également analysé les énoncés de mouvement fictif à partir d'un corpus parallèle de six langues (les quatre langues initiales plus le polonais et l'anglais, afin d'avoir également une image de la varition intratype), constitué à l'aide de deux bases de données : ParCoLab (Stosic 2015) et Parasol (von Waldenfels & Meyer 2006-). Dans ce corpus, nous avons recherché systématiquement un certain nombre de lemmes identifiés par Capelli (2013) comme liés au mouvement fictif, dans chaque langue source. Nous avons classé ces lemmes en suivant la typologie d'Aurnague (2004), et dégagé 9 types :

(1) voies de communication (routes, ponts, escaliers) ;
(2) objets allongés, liés au mouvement (tubes, conduites) ou non (câbles, fils) ;
(3) portions d'espace (immatériel), y compris les ouvertures fonctionnelles (portes, fenêtres) et les trous (trou, fissure) ;
(4) séries d'objets, incluant surtout des configurations linéaires (séries, lignes) ;
(5) endroits (pré, ville, zone, montagne) ;
(6) entités mixtes (maison, hall) ;
(7) parties du corps (tête, cou) et habits (robe, voile) ;
(8) autres objets (arbre, croix, colonne, table) ;
(9) autres entités perceptibles (lumière, son).

Pour chaque énoncé de mouvement fictif, nous avons cherché à voir si les traductions contenaient ou non du mouvement fictif. Le résultat le plus intéressant de cette étude sur corpus est que la proportion de mouvement fictif dans les traductions varie selon le type de lemme : les différents types identifiés ne sont pas tous aussi souvent traduits par du mouvement fictif. Pour les types les plus fréquents, le degré d'accord varie même assez nettement, de 67% à 84% d'accord (tableau 7).

Expressions de mouvement fictif	Type 8	Type 5	Type 6	Type 1	Type 9
non traduites par du mouvement fictif	23	14	5	20	12
traduites par du mouvement fictif dans au moins une autre langue	47	37	24	104	63
Proportion d'accord (partiel)	*67%*	*73%*	*83%*	*84%*	*84%*

Tableau 7 : Nombre d'expressions de mouvement fictif traduites ou non par du mouvement fictif, dans notre corpus, selon le type de lemme (Stosic et al. 2015 [V : 631 sqq.])

Ces résulats suggèrent qu'il y a à la fois des tendances interlinguistiques fortes – le type 1, correspondant à ce que nous avons testé à l'aide de l'outil d'élicitation, est ainsi fréquent dans toutes les langues, dans les deux modalités d'expérimentation (élicitations et corpus). Il suggère également que le mouvement fictif pourrait aussi être influencé par d'autres facteurs, qui rendent compte de constructions des types 5 et 8, et semblent présenter une variabilité interlinguistique plus grande. Ce ne sont là que des hypothèses à valider à l'aide d'études sur des corpus plus larges, étant donné la fréquence très faible du mouvement fictif.
Cela permet d'envisager une série de thèses en linguistique expérimentale (comme celle d'Emilia Tuuri sur l'expression du mouvement réel et fictif en finnois, à l'Université de Tampere, Finlande, pour laquelle je suis coencadrant), en linguistique de corpus (sur le modèle de la thèse de Capelli 2013 pour le français, mais dans d'autres langues), et contrastive.

4.4 L'asymétrie source/but

não te pergunto de onde chegas?,
porque sei para onde vais.
« je ne te demande pas 'd'où viens-tu ?',
parce que je sais où tu vas. »
(José Luís Peixoto,
A Criança em Ruínas,
2012, p. 28)

Dans les études actuelles sur l'espace dans les langues, un thème particulièrement intéressant au regard de la question des liens entre langue et cognition est celui de l'asymétrie entre source et but. Cette hypothèse, déjà présente chez Fillmore (1972) concernant l'asymétrie entre *come* et *go* en anglais, a été proposée comme possible universel par certains auteurs (cf. Ikegami 1987, Stefanowitsch & Rohde 2004, Lakusta & Landau 2005) et étudiée encore récemment (voir par exemple Kopecka & Ishibashi 2011, Ishibashi 2012). Ce phénomène pourrait être lié à l'attention plus importante accordée au point final des trajectoires, donc être dicté par des contraintes cognitives. J'en ai trouvé des traces assez frappantes en diachronie (Fagard 2010a). Ainsi, dans les extensions sémantiques que j'ai pu observer au sein du domaine de l'espace, pour les prépositions romanes médiévales, seul le point final de la trajectoire est fréquemment associé à la localisation en général, comme pour *vers* (107-108) :

(107) (ancien français)
Vait s'en Brandan **vers** le grant mer U sout par Deu que dout entrer
« Brendan s'en va **vers** la haute mer, où il savait par Dieu qu'il devait entrer »
(BFM, *Le Voyage de Saint Brendan*, v. 157, 1112)

(108) (moyen français)
Le roy y envoya monsr l'admiral bastard de Bourbon, accompaigné de bon nombre de gens, lesquelz bruslèrent grand quantité de leurs villes, commençans **vers** Abeville jusques à Arras.
« Le roi y envoya monseigneur l'amiral, bâtard de Bourbon, accompagné d'une forte troupe ; ils brûlèrent grand nombre de leurs villes, commençant **vers** Abeville et allant jusqu'à Arras »
(BFM, Commynes, *Mémoires*, p. 19, 1490)

Typiquement, si la première attestation d'une préposition dans mon corpus a un sens de localisation, la première extension sémantique est le but – ou l'inverse : c'est le cas pour les prépositions *lez*, *coste*, *jouste* et leurs dérivés, pour les prépositions *aval*, *contreval*, *amont* et *contremont*, pour la préposition *enmi*, qui

> [peut] être employée dès ses premières attestations pour désigner le [BUT] (point final de la trajectoire). Au milieu du 12ème siècle, soit un siècle après les premières attestations, apparaissent les emplois de type 'à travers' où *enmi* désigne le point intermédiaire de la trajectoire. Il semble en revanche que la désignation de l'origine ne soit pas possible : il n'y en a en tout cas pas d'exemple dans notre corpus. (Fagard 2010a : 402)

L'expression d'autres points de la trajectoire (source et point intermédiaire) à partir d'un sens de but ou de localisation est bien moins fréquente, et n'apparaît généralement que dans des constructions morphologiquement plus complexes, où s'ajoute à la base un élément prépositionnel ou préfixal – *de* pour la source, *par* pour le point intermédiaire. C'est le cas par exemple pour *devers* qui se renforce en *par devers*, et pour certaines occurrences de *devers* qui doivent en fait être analysées comme une construction libre *[de [vers GN]]*.

Le point d'origine ne peut être indiqué par les prépositions *coste*, *lez*, *jouste* et leurs dérivés que si elles sont précédées du 'préfixe' *de-*. Le point intermédiaire ne peut être indiqué par les prépositions concernées que lorsqu'elles sont précédées du 'préfixe' *par-* ; on trouve quelques occurrences d'un tel emploi pour *par delez* et *par decoste*.

Comme je l'ai noté,

> il semble ainsi y avoir un marquage spécifique pour deux points de la trajectoire, la source et le point intermédiaire, plutôt qu'une simple *asymétrie source/but*, puisque le point final de la trajectoire n'est jamais marqué par un préfixe. (Fagard 2010a : 403-404)

J'ai repris ce thème avec Anetta Kopecka, en travaillant sur les données trajectoire de l'allemand et du polonais (communication, Kopecka & Fagard 2014 ; Fagard & Kopecka, en préparation). Nous avons pour l'instant analysé 16 scènes présentant une trajectoire simple : 8 sont orientées vers la source, c'est-à-dire que la source est saillante – ce sont les scènes SOURCE et 8 orientées vers le but, c'est-à-dire que le but est saillant – ce sont les scènes BUT), soit un total de 304 descriptions en allemand (16 scènes, 19 participants) et 224 en polonais (16 scènes, 14 participants).

Les résultats montrent que, dans la grande majorité des cas, les fonds exprimés correspondent à ce qui est saillant dans la vidéo, et que la différence entre l'expression du but et de la source

est faible. Elle est d'ailleurs non significative d'un point de vue statistique, que l'on prenne en compte la fréquence d'expression du fond attendu dans les deux langues (en comparant scènes source et scènes but), ou la fréquence d'expression du fond attendu dans les deux types de scène (en comparant allemand et polonais) (tableau 8). Cependant, d'un point de vue qualitatif, on peut noter que la part de fonds indiquant la source dans les scènes BUT est plus faible que l'inverse, pour les deux langues.

langue	scènes	(au moins) un *fond* indiquant la source	(au moins) un *fond* indiquant le but	ni l'un ni l'autre	Total
allemand	SOURCE	82,5 % (118)	**1,4 % (2)**	21,1 % (32)	100 % (152)
	BUT	**0 % (0)**	83,2 % (124)	18,4 % (28)	100 % (152)
polonais	SOURCE	88,5 % (92)	**5,8 % (6)**	12,5 % (14)	100 % (112)
	BUT	**2,7 % (3)**	87,5 % (98)	8 % (9)	100 % (112)

Tableau 8 : Expression de la source et du but par un fond dans les données trajectoire (16 scènes) (Kopecka & Fagard 2014)

L'asymétrie source/but apparaît également dans la complexité et la diversité de l'expression. C'est le cas en polonais : avec la même complexité morphologique, l'expression de la source est plus limitée sémantiquement que celle du but (deux distinctions contre trois) (12).

Contexte	Construction	Exemple
source + franchissement de frontière	*wy.V z N.GEN*	109
source + éloignement	*od.V od N.GEN*	110
but + franchissement de frontière	*w.V do N.GEN*	111
but + franchissement de frontière spécifique	*w.V w N.GEN*	112
but + rapprochement	*pod.V do N.GEN*	113

Tableau 9 : Expression typique de la source et du but en polonais, dans les données analysées

(109) (polonais)

kobieta **wy-** chodzi **z** pola kukurydzy
'femme.NOM hors.de- marcher.3SG.PRS de champ.ACC maïs.GEN'
« la femme sort d'un champ de maïs avec un panier »
(Traj38_pl_05)

(110) (polonais)

kobieta **od-** chodzi **od** drzewa
'femme.NOM depuis- marcher.3SG.PRS depuis arbre.ACC'
« la femme s'éloigne de l'arbre »
(Traj32_pl_07)

(111) (polonais)

chłopiec **w-** biega **do** morza
'garçon.NOM dans- courir.3SG.PRS vers mer.GEN'
« le garçon entre dans la mer en courant»
(Traj59_pl_02)

(112) (polonais)

kobieta	**we-**	szła	**w**	pole kukurydzy
‘femme.NOM	dans-	marcher.3SG.F.PST	dans	champ.ACC maïs.GEN’

« la femme entre dans le champ de maïs »
(Traj69_pl_03)

(113) (polonais)

kobieta	**pode-**	szła	**do**	wysokiego drzewa
‘femme.NOM	sous-	marcher.3SG.F.PST	vers	haut.GEN arbre.GEN’

« la femme s’est approchée de l’arbre »
(Traj61_pl_01)

Ce n’est pas le cas en allemand, apparemment, où il y a le même niveau de détail pour la source et le but (tableau 10) :

Contexte	Construction	Exemple
source + franchissement de frontière	*(hin/her/r)aus.V aus N.DAT*	114
source + éloignement	*weg.V von N.DAT*	114
but + franchissement de frontière	*((hin/her/r)ein.)V in N.ACC*	115
but + rapprochement	*zu.V auf N.ACC*	116

Tableau 10 : Expression typique de la source et du but en allemand, dans les données analysées

(114) (allemand)

eine	Frau	läuft	**aus**	der	Höhle	**raus**
‘IND.F	femme.NOM	marcher.3SG.PRS	hors	DET.F.DAT	grotte	DEI.hors

von	der	Kamera	**weg**
de	DET.F.DAT	caméra	loin’

« une femme sort de la grotte, en s’éloignant de la caméra »
(Traj25_gm_12)

(115) (allemand)

ein	kleiner	Junge	läuft	**ins**	Meer
‘IND	petit.M.NOM	garçon	marcher.3SG.PRS	dans.DET.N.ACC	mer’

« un petit garçon entre dans la mer »
(Traj59_ gm_10)

(116) (allemand)

eine	Frau	läuft	**auf**	einen	Baum	**zu**
‘IND.F	femme.NOM	marcher.3SG.PRS	vers	DET.ACC	arbre	vers’

« une femme s’approche d’un arbre »
(Traj61_gm_04, Traj61_gm_06, Traj61_gm_11, Traj61_gm_12[27])

[27] La même construction exactement est utilisée par quatre participants différents pour la même scène.

Dans les deux langues, on trouve d'autres constructions moins fréquentes pour l'expression de la source du but, par exemple des verbes comme *s'éloigner* (all. *sich entfernen* (117), pol. *oddalać się*), *quitter* (all. *verlassen*, pol. *opuszczać*, exemple 118), *s'approcher* (all. *sich nähern*[28], pol. *zbliżać się*) :

(117) (allemand)

eine	Frau	die	sich	**vom**	Baum	**entfernt**
'IND.F	femme.NOM	qui.F.NOM	se	de.DET.M.DAT	arbre	éloigner.3SG.PRS'

« une femme qui s'éloigne de l'arbre »
(Traj32_gm_18)

(118) (polonais)

Kobieta	**opuszcza**	swoją	jaskinię
'femme	quitter.3SG.PRS	sa.ACC	grotte.ACC'

« la femme quitte sa grotte »
(Traj23_pl_08)

On peut noter que ces constructions restent très minoritaires, comme le montrent pour l'allemand les tableaux suivants (11-12) :

Construction	0	*aus*	*von*	*aus* + *von*	Total
0	18	39	5	1	63
*aus**	6	57	1		64
weg	1		15		16
*aus** + *weg*				3	3
verlassen	3				3
sich entfernen			3		3
Total	28	96	24	4	152

Tableau 11 : Expression de la source dans les données analysées, en allemand (* *aus* = *hinaus*, *heraus*, *raus*, *aus* ; les cases de couleur indiquent les degrés de fréquence)

Construction	0	*auf*	*in* + *auf*	*in*	*in* + *nach*	*nach*	*zu*	*zwischen*	Total
0	10			31		1	7		49
*ein**	4			43					47
zu	1	29	1						31
zu + *ein**		1							1
zu + *verschwinden*			1						1
verschwinden				8				2	10
zurück	3			4	1		1		9
sich nähern + *DAT*	2								2
sich erretten				1					1
betreten	1								1
Total	21	30	2	87	1	1	8	2	152

Tableau 12 : Expression du but dans les données analysées, en allemand (* *ein* = *hinein*, *herein*, *rein*, *ein* ; les cases de couleur indiquent les degrés de fréquence)

[28] Je n'ai pas trouvé d'occurrences de ce verbe dans les 16 scènes sélectionnées, mais il est utilisé ailleurs dans le corpus, pour l'expression du but.

Mais la caractéristique la plus intéressante concernant la source est l'existence d'énoncés où elle est implicite, comme dans l'exemple (119) :

(119) (allemand)

eine	Frau	steht	vor	einem	großen	Baum
'IND.F	femme	être.debout.3SG.PRS	devant	IND.DAT	gros.M.DAT	arbre

und	geht	dann
et	aller.3SG.PRS	ensuite'

« une femme se tient devant un grand arbre puis elle part »
(Traj32_gm_07)

Ce type d'énoncé pourrait expliquer que nos résultats ne soient pas plus marqués : dans un certain nombre de cas, le fond indiquant la source est présent, mais il n'est pas indiqué linguistiquement par une préposition l'identifiant comme source. Plusieurs pistes s'ouvrent à nous pour poursuivre ces recherches : continuer l'analyse sur l'ensemble des données trajectoire (autres scènes pour l'allemand et le polonais, autres langues : communication, Kopecka, Fagard & Stosic 2015) ; inclure dans les calculs non seulement l'expression ou non du fond désignant source et but mais aussi les marques grammaticales ; déterminer de manière plus fine le lien entre saillance dans les vidéos et fréquence d'apparition dans les descriptions (cf. Stosic et al. 2015 [V : 631 sqq.]) ; comparer avec d'autres données (notamment des corpus, cf. Stosic et al. 2015 [V : 631 sqq.]).

5 La grammaticalisation

Grammaticalisation et localisme

Une partie au moins des travaux sur la grammaticalisation est clairement liée à la théorie localiste. D'une part, le localisme préfigure la 'théorie' de la grammaticalisation[29], avec l'hypothèse d'un glissement systématique ESPACE > TEMPS > ABSTRACTION et celle d'une origine concrète de tout élément grammatical (cf. la présence du terme « grammatically » dans la citation (tardive) de Lyons (1977) donnée plus haut : « the hypothesis that spatial expressions are more basic, grammatically and semantically, than various kinds of non spatial expressions »). D'autre part, la théorie de la grammaticalisation fait écho au localisme, avec une attention très particulière à tout ce qui relève de l'espace, comme le note Heine (2003) :

> The third phase of grammaticalization studies started in the 1970s and was initially connected with the paradigm of localism (Anderson 1971, 1973). (Heine 2003 : 576)

D'ailleurs, le bien-fondé de l'approche localiste (du moins de l'approche localiste revisitée, mâtinée d'*embodiment*) apparaît clairement à la lumière des travaux sur la grammaticalisation, en particulier ceux de typologues comme Bernd Heine. Ces travaux ont en effet montré l'importance des chaînes de grammaticalisation, ou chaînes sémantiques, qui montrent bien la fréquence et l'importance des glissements [CONCRET > ABSTRAIT] dans les langues :

> Lexical items from concrete domains are pressed into service to express concepts in more abstract domains. The transition from source to target structures is described as metaphorical abstraction. (Mithun 1993 : 518)

De plus, depuis, les recherches sur les universaux ont permis d'établir qu'il ne s'agit pas d'un phénomène limité à quelques groupes linguistiques (Svorou 1994, Haspelmath 1997, entre autres). Cela fait de la grammaticalisation un cadre théorique de choix pour mes recherches.

La grammaticalisation est un phénomène fondamental dans l'évolution du langage (au point qu'on peut bien parler d'ubiquité, Mithun 1993 : 517, 521), bien connu et largement décrit, sur lequel on dispose à la fois de manuels (voir par exemple Hopper & Traugott 2003 ou encore Narrog & Heine 2011), de descriptions théoriques fines (Lehmann 2002, Marchello-Nizia 2006), d'approches typologiques de surface (Heine & Kuteva 2002) et d'un certain

[29] Il n'est pas certain en effet qu'on puisse réellement parler de *théorie* (Prévost 2003), cf. la remarque de Heine : « grammaticalization theory is not a theory of language change and (…) constitutes merely one of the factors that determine the history and future development of grammar » (Heine 2003 : 599).

nombre de numéros de revue concernant des points de détail (*Language Sciences* 23/2-3, 2001, Prévost & Fagard 2007 [VI : 973 sqq.]).

Il me semble qu'il reste tout de même un certain nombre de questions. La question de base – qu'est-ce que la grammaticalisation ? – peut se scinder en plusieurs autres : La grammaticalisation est-elle un processus univoque, identique quelle que soit la langue, la partie du discours, la période ? Quelle est l'importance respective des différentes composantes de la grammaticalisation, et en particulier de l'évolution sémantique par rapport aux autres (syntaxe, morphologie, phonétique…) ? Quelles sont ses relations avec des phénomènes proches (lexicalisation, subjectivation, intersubjectivation) ? Chacune de ces questions a, à son tour, mené à d'autres interrogations.

On pourrait se demander à ce stade quel est l'intérêt d'étudier un thème présent dans la littérature scientifique depuis au moins un siècle. Et qui, de surcroît, est devenu à la mode récemment (dans les années 90), mais semble en perte de vitesse depuis quelques années[30]… La réponse tient au fait que le phénomène de grammaticalisation semble bien être un des phénomènes clefs de l'évolution linguistique, représentatif de la tension qui existe au sein du langage – des langues –, entre l'arbitraire et l'universel. Il s'agit bien, apparemment, d'un phénomène universel :

> Les nombreuses études réalisées dans les vingt dernières années sur les phénomènes de grammaticalisation dans de très nombreuses langues semblent montrer que ce type de changement est universellement représenté, et particulièrement fréquent. (Marchello-Nizia 2006 : 107)

Mais les instanciations de ce phénomène dans chaque langue peuvent être spécifiques, voire uniques : les nombreux exemples rassemblés par Heine & Kuteva (2002) montrent bien qu'il y a des tendances marquées, comme la grammaticalisation bien connue de verbes de mouvement ou de volonté pour exprimer le futur, à côté d'autres chaînes de grammaticalisation apparemment moins répandues.

Si l'on repart de la définition donnée par Meillet il y a plus d'un siècle, c'est-à-dire de la grammaticalisation comme l'« attribution du caractère grammatical à un mot jadis autonome » (Meillet 1912 : 131, repris d'assez près par Kuryłowicz 1965)[31], il est assez frappant de voir à quel point elle reste valable. Il y a un assez large consensus, au moins dans la communauté des linguistes diachroniciens[32], sur l'existence de la grammaticalisation comme phénomène linguistique, résultant d'une conjonction de facteurs – qui eux-mêmes se

[30] Avec notamment un cycle de conférences internationales, « New Reflections on Grammaticalisation », qui étaient dédiées au phénomène et semblent s'essouffler : NRG à Potsdam en 1999, NRG2 à Amsterdam en 2002, NRG3 à Santiago de Compostela en 2005, NRG4 à Leuven en 2008, NRG5 à Edinburgh en 2012, mais la suite n'est à ma connaissance pas encore annoncée.

[31] Il faut noter, comme cela a déjà été remarqué, que la théorie n'est pas nouvelle à l'époque, mais inspirée des travaux de von der Gabelentz, Whitney, ou encore Condillac (cf. Coseriu 1967, Lehmann 1995, De Mulder 2001).

[32] Il y a un semblant de débat (qui paraît à vrai dire assez artificiel), avec notamment le numéro de *Language Sciences* mentionné plus haut, dédié à une critique en règle des travaux sur la grammaticalisation. Richard Janda, notamment, y reproche aux théoriciens de la grammaticalisation la circularité de leur raisonnement, indiquant : « a reconstruction initially justified by invoking a certain principle (grammaticalization) cannot later be argued to provide independent confirmation for that same principle » (Janda 2001 : 271 ; voir aussi Newmeyer 1998, 2001).

retrouvent par ailleurs dans d'autres phénomènes. Les facteurs principaux sont le figement morpho-syntaxique, l'érosion phonétique et l'affaiblissement ou la généralisation du sens ; d'autres facteurs, comme l'élargissement de la portée et la subjectivation, sont plus discutés. Quoi qu'il en soit, il me semble bien plus difficile de rendre compte du changement linguistique sans faire appel à ce phénomène. Car c'est bien là le but que je me suis fixé : rendre compte du changement linguistique, et plus précisément tenter d'expliquer pourquoi, s'il est possible de rendre compte des changements linguistiques a posteriori, il est impossible de les prévoir. Le changement linguistique obéit à un certain nombre de règles, mais demeure imprévisible tant la part d'arbitraire est grande. On en revient donc à la question initiale de la variation : dans quelle mesure l'évolution linguistique est-elle contrainte, et par quoi ? Pourquoi les langues romanes sont-elles à la fois si différentes qu'un locuteur non averti de l'une d'entre elles est bien souvent incapable de comprendre les autres, et si proches que leur origine commune serait indéniable même en l'absence de preuves historiques ?

Ces recherches m'ont amené à poser des questions reliées, avec différents angles d'attaques : quelles sont les contraintes sémantiques portant sur l'évolution linguistique, dans le cadre de la grammaticalisation ? Quelles sont les contraintes systémiques ? Quel est le rythme de l'évolution ? Par ailleurs, ces contraintes sont-elles les mêmes pour tous les phénomènes de grammaticalisation ? Sont-elles valables de manière générale, pour toutes les langues ? J'ai approfondi ces questions au travers d'une série d'études portant sur divers types de marque linguistique (prépositions, locutions prépositionnelles, conjonctions, subordonnants, adverbes locatifs et déictiques, marqueurs discursifs, verbes de mouvement, particules verbales), principalement dans les langues romanes (français, occitan, italien, catalan, espagnol, portugais, galicien, sarde, roumain) mais également dans des langues germaniques (anglais, suédois, allemand, néerlandais), slaves (polonais, russe, serbe) et autres (hongrois, finnois, thaï).

5.1 Panorama des études actuelles et antérieures – Définition et état de l'art

Un des problèmes bien connus concernant la grammaticalisation est la difficulté qu'il y a à la définir, en particulier dans son rapport à d'autres phénomènes importants dans l'évolution linguistique : réanalyse (voir Hopper & Traugott 1993 : 48-50, Heine, Claudi & Hünnemeyer 1991 : 215-220), lexicalisation (voir Prévost & Fagard 2007 [VI : 973 sqq.], Trousdale 2012), subjectivation, intersubjectivation, etc. Mes recherches m'ont amené régulièrement à étudier ces problématiques, avec un certain nombre de résultats dont je rends compte plus bas (§5.2-5.4) ; je commencerai, ici, par proposer une définition du phénomène. Il existe dans la littérature scientifique dédiée à la grammaticalisation un certain nombre de définitions qui se recoupent sur l'essentiel, mais sont plus ou moins strictes (voir Lehmann 2002 pour une définition particulièrement restrictive). Je prends plutôt le parti d'une définition lâche, qui s'accomode bien de l'idée que la grammaticalisation recouvre un certain nombre de phénomènes, constituant une catégorie de nature prototypique. Ce type d'approche reconnaît donc les limites d'une définition en termes de conditions nécessaires et suffisantes, qui

excluerait de fait un grand nombre de phénomènes linguistiques qui, de mon point de vue, sont également des cas de grammaticalisation. Un phénomène exemplaire de grammaticalisation présente, dans cette approche, les caractéristiques notées dans le tableau (13) ci-dessous.

Les fondamentaux de la grammaticalisation
1. PHONETIQUE : réduction, coalescence
2. MORPHOLOGIE : perte de l'accord, décatégorisation
3. SYNTAXE : perte de la complexité interne (lexicalisation, constructionnalisation), perte du statut autonome (cliticisation), figement, réduction ou augmentation de portée, réanalyse
4. SEMANTIQUE : perte de référentialité, gain d'un sens grammatical
5. PRAGMATIQUE : subjectivation, intersubjectivation
→ *évolution possible – non nécessaire mais unidirectionnelle – sur chacun de ces plans, accompagnée de (a) diversification des contextes d'emploi et (b) augmentation de fréquence*

Tableau 13 : Définition *a minima* de la grammaticalisation

Les évolutions phonétiques sont plus difficiles à aborder en diachronie : on ne peut que reconstruire la **coalescence** qui s'est produite, à partir de la graphie et de son évolution. Ainsi, j'ai pu observer la réduction de *parce que* à */pask(ə)/*, liée à un nouveau stade de grammaticalisation en français moderne (Fagard & Degand 2008 [III : 389 sqq.]). La perte 'sèche' d'éléments constitutifs d'une construction peut également être indicative d'une grammaticalisation en cours, comme l'évolution de *du côté de* à *côté* (Fagard 2008a [II : 211 sqq.]), de *conciossiacosacché* à *conciosiaché* (communication, Fagard 2010), de *au fond de X* à *au fond* (Fagard & Sarda 2014 [IV : 435 sqq.]), etc. Les caractéristiques phonétiques sont pour moi un élément à prendre en compte comme symptôme d'un processus de grammaticalisation en cours. Notons pour finir que, comme cela a déjà été remarqué par tous les spécialistes de la grammaticalisation, les aspects phonétiques, comme les autres, ne sont pas des caractéristiques spécifiques de la grammaticalisation, et peuvent donc se retrouver par ailleurs dans l'évolution linguistique (ce qui vaut, répétons-le, pour l'ensemble des facteurs mentionnés).

Les évolutions liées à la morphologie se manifestent de manière plus repérable à l'écrit. Elles sont particulièrement évidentes dans le cas de la grammaticalisation d'éléments simples, par opposition à la grammaticalisation d'éléments complexes, qui suppose une première phase de lexicalisation, faisant plutôt intervenir des processus de figement syntaxique. Par exemple, au cours de la grammaticalisation de l'adjectif *sauf* en préposition (communication, Fagard 2008a), on voit bien que le glissement sémantique (assez net si l'on compare les exemples 120 et 121) est accompagné d'un figement syntaxique (*sauf* étant de plus en plus systématiquement préposé au nom, exemple 122) mais aussi à une perte d'accord (exemple 123) :

Phase 0 – *sauf* est adjectif, placement libre, accord systématique
(120) (ancien français)
Si recevrat la nostre lei plus **salve** ; Chrestïens ert, de mei tendrat ses marches
« Il recevra notre loi, plus salutaire ; il sera chrétien, et tiendra de moi ses terres »
(BFM, *Chanson de Roland*, v. 189, 1100)

Phase I – *sauf* est ajdectif antéposé, qui s'accorde
(121) (ancien français)
Biaus sire, **sauve** vostre grasce, ne l'ai pas mort
« Beau sire, sans offense (litt. *sauve votre grâce*), je ne l'ai pas tué »
(BFM, *Eracle*, v. 1797, 1176-1184)

Phase II – emplois ambigus
(122) (moyen français)
et y a peu de roys, **sauf** celuy de France, plus puissans de luy
« et il y a peu de rois, sauf celui de France, qui soient plus puissants que lui »
(BFM, Commynes, *Mémoires*, t. 1, p. 38, 1489-1491)

Phase III – manque d'accord
(123) (moyen français)
(…) Chatart de Rochedegoux, escuier, fut receu en l'office de seneschal de Rouergue ou lieu de messire Raoulet de Laire, chevalier, **sauf** l'opposition d'icellui messire Raoulet, à la dire ainsi bien après que devant.
« Chatart de Rochedegoux, écuyer, fut reçu comme sénéchal de Rouergue à la place de Raoulet de Laire, chevalier, [sauf l'opposition] de messire Raoulet, à exprimer après ou avant »
(BFM, Nicolas de Baye, *Journal*, p. 196, 1400-1417)

La **perte de variabilité** peut être partielle, avec des restes liés à la catégorie d'origine. C'est le cas, par exemple, des marqueurs discursifs formés à partir d'impératifs de verbes de perception dans les langues romanes : on trouve ainsi *mira/mire/mirad* en espagnol, *mira/miri/mirin* en catalan, *uite/uitaţi* en roumain, etc. (Fagard 2010b [IV : 587 sqq.], 2012b) ; il faut noter cependant qu'une de ces formes est en général prépondérante, et tend à s'imposer aux dépens des autres. Quoi qu'il en soit, le résultat de ce processus lorsqu'il arrive à son terme est une **décatégorisation** : ce qui était en l'occurrence un adjectif, partant variable en genre et en nombre, ou un verbe, variable en nombre et en personne, est désormais devenu une préposition, ou un marqueur discursif, donc dans les deux cas un élément invariable.

Les **évolutions syntaxiques** sont elles aussi plus repérables à l'écrit, et donc plus accessibles pour les études sur corpus en diachronie. Elles sont de plusieurs ordres. Un des phénomènes marquants est la perte de complexité interne : on observe ainsi, pour les prépositions ou conjonctions complexes, ou encore pour certains adverbiaux, un phénomène de **lexicalisation** (que l'on pourrait également appeler **constructionnalisation**, cf. Koch 2012b) qui fait passer la construction d'une liberté syntaxique importante – et généralement repérable sur corpus – à un figement plus ou moins net. Ce figement implique la disparition progressive des modifieurs. Une autre conséquence est la **perte du statut autonome** : on passe ainsi d'un élément lexical, autonome (accentuable, contrastable, etc.) à un élément de plus en plus dépendant, jusqu'à la **cliticisation**. Il s'agit là semble-t-il du dernier stade avant la disparition totale (Hopper & Traugott 2003 : 202) ; on peut en voir un exemple assez avancé avec l'émergence d'emplois fonctionnels pour les prépositions, comme l'emploi de *à* pour le MDO (Fagard & Mardale 2014 [I : 3 sqq.]). Je laisserai de côté la question de la réduction ou de l'augmentation de portée, qui fait débat (voir, par exemple, Fischer 2010) et est particulièrement peu claire pour les marqueurs discursifs. Enfin, la question de la **réanalyse** a

été largement débattue, mais il semble maintenant assez clair (cf. Marchello-Nizia 2006 : 45 sqq.) qu'elle est à la fois fréquente dans les processus de grammaticalisation, non-nécessaire, et parallèlement présente dans d'autres processus. Autrement dit, il y a des grammaticalisations sans réanalyse et inversement, mais les deux vont tout de même souvent de pair. Ainsi, en français médiéval, la grammaticalisation de la construction *de part N* « de la part (de) N » en une préposition complexe *de par(t)* suppose une étape au cours de laquelle le substantif *part* a été réanalysé comme la préposition *par* : on est passé de *[je viens] [de [(la) part [du roi]]]* à *[je viens] [de par [le roi]]*.

Les **évolutions sémantiques** sont plus difficiles à faire émerger, du moins de manière automatique (et en l'absence de lemmatisation, cf. Blumenthal 2008), car il faut analyser les contextes des occurrences pour mettre au jour l'évolution du sens. On peut dire, de manière générale, qu'il y a lors de la grammaticalisation une perte de sens référentiel : ainsi, lorsque l'on utilise la préposition complexe *face à* dans un contexte comme *face à ce problème*, il n'est pas question de *visage*, ce à quoi renvoie bien l'étymologie de *face*. L'évolution sémantique suit, lors de processus de grammaticalisation, ce que l'on appelle des **chaînes de grammaticalisation**, une sorte de parcours sémantique préférentiel qui obéit généralement aux règles du localisme révisé (cf. § 3).

Les **évolutions pragmatiques** posent problème, dans la mesure où elles semblent plus marginales en tant que composantes des processus de grammaticalisation. Leur statut est doublement débattu : certains auteurs considèrent que la subjectivation ne fait pas partie des symptômes de la grammaticalisation, qu'il s'agit d'un processus à part entière (cf. la notion de pragmaticalisation chez Erman & Kotsinas 1993, Aijmer 1997, et notamment Dostie 2004 qui considère la pragmaticalisation comme une sorte de grammaticalisation du deuxième type[33]), et enfin qu'il peut y avoir à la fois subjectivation (ou pragmaticalisation) sans grammaticalisation aussi bien que l'inverse (voir le § 4.2).

L'ordre des changements

Une question majeure qui reste cependant à régler est celle de l'ordre des changements dans les processus de grammaticalisation : parmi toutes les modifications signalées ci-dessus, lesquelles se produisent en premier ? Y a-t-il un ordre fixe ? Je développe cette question, abordée notamment dans Fagard & Sarda (2014 [IV : 435 sqq.]), en § 4.3.

On peut, pour conclure cette brève introduction des problématiques liées à la grammaticalisation, revenir sur un point essentiel – peut-être la plus âprement débattue des caractéristiques de la grammaticalisation : **l'unidirectionnalité**. Une grande partie des critiques contre la (théorie de la) grammaticalisation concernent ce point (cf. *Language Sciences* 23/2-3, 2001 ; Fischer et al. (eds), 2004 ; Norde 2009 ; Nuyts 2013). L'idée avancée

[33] Il s'agit de cas dans lesquels « une unité lexicale / grammaticale [développe] des emplois où elle ne joue pas un rôle sur le plan référentiel, mais bien, sur le plan conversationnel » (Dostie 2004 : 27). L'auteure distingue cela de la *grammaticalisation*$_1$, liée à la « zone grammaticale$_1$ » qui comprend « les unités qui donnent des informations liées à la quantité et aux repères spatio-temporels », « les 'spécifieurs' nominaux, verbaux, adjectivaux et adverbiaux », « les morphèmes qui lient des propositions à l'intérieur d'un énoncé (coordonnants, subordonnants) et les prépositions ». A l'inverse, « les unités pragmatiques ne participent que rarement au contenu propositionnel des énoncés, et leur portée n'est pas intraphrastique » ; elles jouent un rôle « sur le plan macro-textuel » (Dostie 2004 : 30).

par les détracteurs est que l'on trouve des contre-exemples. Il est vrai en effet que certains éléments grammaticaux peuvent être lexicalisés, les plus célèbres étant le *–s* du génitif anglo-saxon, ou encore l'emploi verbal de *up* dans *to **up** the ante* (cf. Hopper & Traugott 2003 : 58). Cependant, comme le note judicieusement Prévost (2006), il ne s'agit pas alors du processus inverse d'une grammaticalisation : les mécanismes vus ci-dessus ne sont pas appliqués « à rebours » à une forme grammaticalisée. Il s'agit généralement de lexicalisation (dans le cas cité par Nuyts 2013, il me semble que l'on pourrait parler avec plus de profit de subjectivation). L'unidirectionnalité n'est d'ailleurs pas une caractéristique exclusive des processus de grammaticalisation, et on la trouve pour des changements sémantiques (Heine 1997b : 145, 2000 : 13, Traugott & Dasher 2002 : 26). Une étude menée sur corpus m'a permis de vérifier qu'une évolution semblant à première vue relever de la dégrammaticalisation n'en est pas nécessairement une, comme dans le cas du remplacement de *coste* et de ses dérivés (prépositions simples) par des prépositions complexes formées sur *côté* (Fagard 2008a [II : 211 sqq.]). Ce type d'évolution est plutôt dû au caractère récurrent et à la finesse des « chaînes sémantiques ».

5.2 Etudes menées sur la grammaticalisation

Je présente ci-dessous les principales études que j'ai menées sur le sujet, en les regroupant en fonction des problèmes théoriques qu'elles permettent d'aborder : rôle de la lexicalisation, de la subjectivation, ordre des changements, sources des grammaticalisations, vitesse des changements.

5.2.1 Grammaticalisation et lexicalisation

Certains phénomènes sont décrits comme des cas de grammaticalisation par les uns, de lexicalisation par les autres (Prévost 2003) – c'est le cas, par exemple, pour la formation des adverbes en *–ment(e)* dans les langues romanes (grammaticalisation pour Hopper & Traugott 2003, lexicalisation pour Anttila 1989). De manière plus générale, pour la grammaticalisation d'expressions complexes, on peut hésiter dans la terminologie à adopter : s'agit-il de lexicalisation, au sens où il y a création d'une nouvelle entrée dans le répertoire des formes linguistiques, ou déjà de grammaticalisation – et comment faire la différence ? J'ai entamé des réflexions sur ce problème dans le cadre du projet « Lexicalisation et grammaticalisation – Changement de classes de mots »[34], et abordé cette question avec elle dans l'introduction d'un numéro de *Langue française* (Prévost & Fagard 2007 [VI : 973 sqq.]), ainsi qu'avec Walter de Mulder (Fagard & De Mulder 2007 [II : 229 sqq.]) dans un article du même numéro, portant sur l'évolution des prépositions complexes. J'y suis revenu par la suite dans un certain nombre d'articles (Fagard 2008a [II : 211 sqq.], Fagard 2009b [II : 191 sqq.], Fagard 2012a [II : 167 sqq.]), et une publication plus ambitieuse est en projet avec un groupe de collègues

[34] Un projet financé par l'ILF (CNRS) et dirigé par Sophie Prévost (2005-2007).

(José Pinto de Lima, Dejan Stosic, Elena Smirnova – nous avons déjà reçu des offres de la part de deux éditeurs) sur les phénomènes de grammaticalisation des prépositions complexes dans les langues d'Europe. L'idée de ce dernier projet est de proposer, sur le modèle des publications Eurotyp, une typologie réduite aux langues d'Europe pour un phénomène spécifique. Partir de langues relativement homogènes, bien décrites et bien connues permet en effet d'affronter la diversité linguistique de manière gérable, et peut servir à étudier, par la suite, des systèmes linguistiques différents (cf. la condition (ii) énoncée p. 36).

Plusieurs problèmes se posent pour déterminer si un changement linguistique [A > B] constitue un phénomène de grammaticalisation ou de lexicalisation, mais la première question est de savoir si l'élément B est grammatical ou lexical. De ce point de vue, les prépositions sont une catégorie hautement problématique, puisqu'elles sont classées tantôt dans le lexique, tantôt dans la grammaire[35]. Il faut aussi déterminer si les prépositions complexes sont des structures figées ; étudier la nature du figement (lorsqu'il a lieu) et les processus impliqués. Dans le cas des prépositions complexes, j'ai étudié sur corpus un certain nombre de constructions du type {Préposition (+ article) + Base + Préposition}, sur une période allant de l'ancien au moyen français (Fagard & De Mulder 2007 [II : 229 sqq.]). Ces constructions sont classées comme semi-figées par Hoffmann (2005) et Cifuentes-Honrubia (2003), qui notent leur affinité importante avec les prépositions simples. On peut reconstruire, pour la plupart des formes, une évolution qui suit (au moins) une partie de la chaîne suivante :

> [forme libre > 1 locution prépositionnelle > 2 préposition complexe > 3 préposition simple > 4 préposition fonctionnelle]

Les étapes 1 et 2 constituent probablement une lexicalisation au sens de Brinton & Traugott (2005 : 96) :

> lexicalization is the change whereby in certain linguistic contexts speakers use a syntactic construction or word formation as a new contentful form with **formal and semantic properties that are not completely derivable or predictable** from the constituents of the construction or the word formation pattern. (je souligne)

Les constructions étudiées dans Fagard & De Mulder (2007 [II : 229 sqq.]) sont en effet conformes à cette définition : leur figement morpho-syntaxique s'accompagne d'un glissement sémantique. Ainsi, en portugais médiéval, les prépositions complexes *por amor de* 'en raison de', 'à cause de' et *a pesar de* 'malgré' se sont formées sur les bases nominales *amor* 'amour' et *pesar* 'tristesse'. Ces sens nouveaux, qui apparaissent en premier lieu comme inférences contextuelles, finissent parfois par se retrouver dans des contextes excluant un sens compositionnel.

En revanche, les étapes suivantes relèvent selon moi de la grammaticalisation. Si on reprend l'exemple de la préposition *du côté de*, on peut noter d'une part un glissement sémantique, avec l'introduction de topiques de discours (cf. Le Querler 2001), comme en (123-125) :

[35] Ainsi, il s'agit de morphèmes grammaticaux pour Riegel, Pellat & Rioul (2001 : 537), mais l'adéquation de ce classement dépend des critères retenus. Parmi les critères généralement retenus, *avoir un sens grammatical, non descriptif* est d'application difficile (et on pourrait sans doute répondre par la négative même pour un certain nombre de prépositions simples) ; *appartenir à une classe fermée* pose également problème, dans la mesure où la classe des prépositions est certainement plus ouverte que d'autres classes comme les déterminants (voir par exemple Noailly 2006).

(124) (français contemporain)
et donc euh + je reviens à mes questions sur la mondialisation + donc ça n'a pas donné grand chose **du côté des** copains mais **du côté d'**la cuisine
(CFPP, 12_04)

(125) (français contemporain)
donc en gros ça change plutôt en mieux **du côté des** + des infrastructures de façon générale ?
(CFPP, KB_01)

(126) (français contemporain)
y a un mieux **du côté du** chien + euh + en sachant que + euh + y a encore des tas d'gens qui laissent leur chien faire pipi sur vos pieds enfin j'veux dire + ça m'est arrivé encore l'autre jour
(CFPP, 07_05)

D'autre part, *côté* perd certaines de ses caractéristiques nominales : il ne peut plus être modifié, par exemple, par un possessif et on dira *Puis-je m'asseoir à côté de vous ?* (cf. Lehmann 2002 : 70) plutôt que *Puis-je m'asseoir à votre côté ?* – la survivance de *à vos côtés*, avec un sens spécifique, est d'ailleurs un bon indice en ce sens. Le nom *côté* tend donc à changer de catégorie grammaticale au sein de la préposition complexe, ce qui est bien une des caractéristiques de la grammaticalisation (Wischer 2000 : 365, Rostila 2007). Ce processus de grammaticalisation s'est d'ailleurs poursuivi, dans le cas de *côté*, avec une simplification formelle progressive de *du côté de* (127) à *du côté* (128) puis *côté* (129), qui finit ainsi par avoir un statut de préposition à lui seul :

(127) (français classique)
de sorte que nos pensées, qui devaient être incorruptibles **du côté de leur principe**, deviennent périssables du côté de leur objet
(Frantext, J.-B. Bossuet, *Oraison funèbre de Henriette Anne d'Angleterre*, p. 668, 1670)

(128) (français moderne)
Mon Dieu, comme sous les rois, depuis pas mal de temps, on n'a pas été très gâté **du côté gloire**, ces guerriers qui rapportaient tant de couronnes qu'ils en mettaient jusque sur les bras des fauteuils, je trouve que ça a un certain chic !
(Frantext, Marcel Proust, *Du côté de Guermantes*, p. 519, 1921)

(129) (français moderne)
je me fais l'effet d'un naufragé qui ne peut aborder ni sur la rive droite, côté roman, ni sur la rive gauche, **côté théâtre**, et qui finirait par se dire : « mais je suis bien, là, au milieu ! »
(Frantext, Jules Renard, *Journal*, p. 927, 1910)

La forme simple est attestée en français parlé (130), à côté de formes plus complexes (131) :

(130) (français contemporain)
non non moi je vais **côté** Cler rue Cler parce que parce que y a une animation
(CFPP, 07_06)

(131) (français contemporain)
ça **du côté des** ventes + les artisans sont partis maintenant et ça se + ça se vend cher aussi
(CFPP, 11_01)

On peut considérer qu'il y a des prépositions lexicales et des prépositions non-lexicales, également appelées fonctionnelles (Rauh 1993, Tremblay 1996, Goyens, Lamiroy & Melis

2002), mais même les prépositions simples les plus grammaticalisées conservent des emplois lexicaux ; le poids respectif de ces emplois peut être différent d'une préposition à l'autre.

Pour revenir à la distinction entre grammaticalisation et lexicalisation, je reprends le texte de Fagard & De Mulder (2007 : 27-28 [II : 249]) :

> Notons toutefois que ce changement de statut grammatical n'a été possible qu'au sein de l'unité lexicale *du côté de* et que *côté* peut s'employer dans une classe moins restreinte de contextes, conformément à la définition de la grammaticalisation de Brinton et Traugott.
> Il s'ensuit que, dans l'évolution de cette préposition, lexicalisation et grammaticalisation sont imbriquées l'une dans l'autre : la construction *du côté de* devient une nouvelle unité lexicale, mais lors de cette évolution même, le nom *côté* est décatégorisé, ce qui constitue plutôt un processus de grammaticalisation. Cette imbrication des deux types d'évolution n'est pas surprenante : il est bien connu que la grammaticalisation et la lexicalisation impliquent les mêmes mécanismes phonétiques et syntaxiques : réduction phonétique, réanalyse, démotivation, conventionnalisation, etc. (Wischer 2000 : 364). En fin de compte, ce qui distingue les deux types de processus, c'est leur résultat : la lexicalisation crée de nouvelles unités lexicales, la grammaticalisation crée de nouvelles formes grammaticales. Ainsi, la formation de la préposition complexe *du côté de* et même la transformation de *côté* en préposition est une lexicalisation, mais le changement de catégorie impliquée par cette évolution peut être considéré comme un phénomène de grammaticalisation. De même, la création de la préposition *à* est un phénomène de lexicalisation, mais lorsque cette préposition assume des tâches grammaticales, on assiste à un phénomène de grammaticalisation. L'idée sous-jacente est évidemment que la grammaire et le lexique ne constituent pas deux « modules » séparés : pour autant que les morphèmes grammaticaux sont des unités de forme et de sens, ils font partie du lexique ; ils se distinguent toutefois des autres unités lexicales par le fait qu'ils expriment (aussi) des relations grammaticales.

5.2.2 Grammaticalisation et subjectivation

La question des limites du phénomène de grammaticalisation se pose également vis-à-vis de la subjectivation et de l'intersubjectivation. C'est d'ailleurs le point central du projet Gramis, auquel j'ai participé en tant que chercheur invité à l'UCL, sous la direction de Liesbeth Degand. La question se pose de manière particulièrement aiguë pour deux classes de mots : les conjonctions et les marqueurs discursifs.

Le cas des marqueurs discursifs

Une invention récente

La nomenclature traditionnelle des parties du discours prend sa source dans les écrits des philosophes grecs, notamment Aristote et Platon, qui distinguaient deux parties majeures (noms et verbes) et une partie mineure (conjonctions). La liste des parties du discours s'est enrichie au fil des siècles : adjectifs, participes, adverbes, prépositions, etc.

Les marqueurs discursifs, s'ils sont désormais inclus dans certaines grammaires de référence (par exemple RAE 2009 : 2344 sqq.), ne semblent pourtant pas (encore ?) des candidats

consensuels au statut de partie du discours à part entière (dans la RAE, ils constituent d'ailleurs un sous-chapitre des adverbes). On lit bien souvent qu'il s'agit d'éléments trop hétérogènes, d'origines trop diverses et dont la seule unité est fonctionnelle (cf. Fagard 2010b [IV : 587 sqq.], 2012b, Sarda et al. 2014 [IV : 411 sqq.]). Le flottement terminologique que l'on peut observer à leur égard est assez révélateur.

Vers une définition

La principale difficulté d'une définition des marqueurs discursifs tient selon moi à l'absence de consensus sur leurs fonctions. Selon qu'on prend en considération l'emploi des MD comme modalisateurs ou comme marques d'interaction, comme typiques de l'oral ou comme structurateurs de tout discours y compris à l'écrit, ou à l'oral formel, on fait appel à des notions tout à fait différentes.

Si l'on part d'une définition maximale, les MD sont l'ensemble des éléments linguistiques qui marquent la présence du locuteur dans son discours. Cela fait intervenir plusieurs types de fonctions[36] :

a) La fonction modale
   Elle indique le point de vue du locuteur sur ce qu'il dit.
   Relèvent de cette fonction les marques épistémiques, les modalisateurs, les marques d'emphase.
   Exemples : *bon*, *eh bien* dans leurs emplois évaluatifs ; les particules modales des langues germaniques, comme l'allemand *ja* (cf. Diewald 2013).

b) La fonction interactive
   Elle joue sur l'échange dialogique.
   Relèvent de cette fonction les marques de fin de tour, les interrupteurs, les marques visant à conserver la parole (*floorholders*), les marques d'emphase.
   Exemples : *vous voyez*, *hein*.

c) La fonction métalinguistique
   Elle porte sur le dire, c'est-à-dire sur la forme de l'énonciation plutôt que sur son contenu (ce qui la différencie de la fonction modale).
   En relèvent les marques d'hésitation, de reprise, de précaution oratoire, portant sur l'énonciation.
   Exemples : *on pourrait dire*, *c'est-à-dire*, *à vrai dire*.

d) La fonction connective
   Elle porte sur les relations entre différents éléments ou portions du discours.
   En relèvent les marques de changement ou de continuité du topique, les marques de relations logiques, les marques thématiques, les cadratifs (Charolles 1997, Diewald 2011, Degand, Cornillie & Pietrandrea 2013).
   Exemples : *donc*, *alors* ; *enfin*, *quoi* reformulatifs.

[36] On pourrait relier ces fonctions, notamment les trois premières, aux fonctions de Jakobson (1963-1973) : fonction expressive pour (a), conative et phatique pour (b), métalinguistique, évidemment, pour (c).

Des distinctions similaires peuvent être trouvées chez d'autres auteurs, avec des variations plus ou moins importantes. Un élément fondamental de mon point de vue est que de nombreux MD combinent des fonctions reliées à deux ou même trois de ces macro-fonctions (modale, interactive, métalinguistique, connective). C'est précisément ce qui justifie une approche large des MD. Les différentes nomenclatures des MD renvoient donc en partie seulement à une différente appréciation de leurs fonctions. On peut ainsi souligner le lien entre particule discursive et les emplois modaux ou interactifs, tandis que le terme de connecteur renvoie davantage à la fonction de cohésion discursive. Cependant, il me semble justifié de garder un terme englobant – je prendrai donc marqueur discursif au sens large – pour souligner la proximité formelle entre les éléments participant aux trois fonctions.

Une approche purement fonctionnelle et onomasiologique des MD amènerait le linguiste à identifier, pour une langue ou un groupe de langues, une série d'éléments plus ou moins figés pour chacune de ces fonctions. Il existe en effet d'une part des éléments très fréquents et présentant une variabilité interne limitée (par exemple *car*, *bon*, *puis*, …), d'autre part des tournures complexes et présentant une variabilité interne plus importante (par exemple *en ce qui me* (*nous*, *vous*) *concerne*, *pour autant que je sache*, *que nous sachions*), l'ensemble des MD ainsi identifiables se répartissant le long d'un continuum de figement entre ces deux extrêmes.

La question est alors celle de la cohérence des paradigmes ainsi mis en évidence. On pourrait quelque peu naïvement prendre pour acquise la cohérence des autres parties du discours, et il est sûr qu'elles ont une certaine cohérence. Cependant, toute étude dédiée à l'une d'entre elles permet de mettre en évidence les limites de cette cohérence. À la suite de Company Company (2002) et dans un esprit proche de Van Lier & Rijkhoff (2013), je propose donc une approche prototypique des parties du discours. Cela signifie que chaque partie du discours présente à la fois un noyau, avec des éléments plus représentatifs de la catégorie, mais aussi un certain nombre d'éléments qui sont plus ou moins éloignés de ce noyau.

De ce double point de vue, on peut donc remarquer que les éléments généralement classés comme MD présentent tout de même une certaine homogénéité (tableau 14) :

a) Classes morphologiques de MD	
Formées sur un verbe à l'impératif	*écoute*, *regarde*, *dis*, *tiens*, *voyons*, *allons*, *allez*… : P2, présent (moins souvent P4, P5)
Formées sur un verbe à l'indicatif	*tu vois*, *tu entends*, *tu dis*, *tu crois*, *tu penses*, *vous voyez* : P1, P2
Formées sur un adverbe, un adverbial	*bien*, *alors*, *toujours*, *parallèlement* ; *en effet*, *en réalité*, *au fond*, *d'un côté*, *en même temps*, *vite fait*
Formées sur un adjectif	*bon*, *super*, *grave*
Formées sur une conjonction	*parce que, comme quoi*

b) Traits syntaxiques et distributionnels communs	
Position	Libre, avec une association partielle entre position (initiale, médiane ou finale) et fonction (reprise, interruption, etc.)
Phonétique	Presque toujours des unités intonationnelles isolées, sur un plan différent du reste
Répétition et cumul	Tendance marquée des MD à se retrouver à plusieurs dans une même unité intonationnelle (elle-même isolée du reste) : « *bon ben tu vois*, il est en retard. *Quand même alors*, il vient ou non ? »
Syntaxe	Pas d'intégration à la structure syntaxique de base de la phrase (les MD ne sont pas des arguments du verbe, par exemple) ; coordination et mise en focus impossibles

Tableau 14 : Cohérence des marqueurs discursifs

On ne peut donc pas dire que le manque d'homogénéité formelle ou distributionnelle soit un obstacle. Il me semble qu'il y a deux pistes privilégiées pour mettre en évidence des paradigmes plus resserrés de MD : l'une serait d'étudier les similarités formelles et syntaxiques, ou distributionnelles, d'un groupe fonctionnel, l'autre d'étudier les similarités fonctionnelles et syntaxiques, ou distributionnelles, d'un groupe de marqueurs issus du même type de formation. J'ai suivi principalement la seconde approche, sans exclure la première, dans mes recherches sur les MD.

Grammaticalité et grammaticalisation

Si l'on admet que les MD forment un paradigme – même si sa cohérence est limitée, comme voire plus que les autres parties du discours – il faut maintenant évaluer leur caractère grammatical ou non. Il s'agit là en effet d'une question régulièrement posée dans la littérature (cf. Traugott 2007a, 2007b, 2010, Waltereit 2002, Waltereit & Detges 2007) – en réalité, d'une série de questions auxquelles il faut maintenant répondre :

- Est-ce que les MD sont grammaticaux ?
- Est-ce qu'ils sont obligatoires ?
- Est-ce qu'ils sont le résultat d'un processus de grammaticalisation ?

Je ne souhaite pas répondre à ces questions de manière définitive et argumentée, mais plutôt suggérer des pistes de réflexion auxquelles je reviendrai dans le fil du texte. La question de la grammaticalité reçoit bien souvent une réponse négative (Waltereit 2002), parfois positive (Traugott 2007a, Lewis 2011, Prévost 2011c), parfois intermédiaire, avec la notion de pragmaticalisation, de même que celle du caractère obligatoire des MD. Je propose de partir des postulats suivants – en insistant sur leur statut de postulats :

- les MD constituent une catégorie grammaticale ;
- il s'agit d'éléments obligatoires, au sens où leur absence est significative ;
- ils sont donc le résultat de processus de grammaticalisation (*contra* Erman & Kotsinas 1993 : 79 « it is possible (but not necessary) for a lexical element to develop directly

into a discourse marker without an intermediate stage of grammaticalization » ; Heine (2013) apporte lui aussi une réponse négative à cette question).

On pourrait ajouter, après Brinton (1996), qu'il s'agit d'éléments qui se renouvellent (elle donne l'exemple de l'ancien anglais *hwaet → you know* et *hweat pa → so*, on pourrait ajouter par exemple l'ancien espagnol *cata → mira* et le remplacement de l'ancien français *si*) : il y a « continuité dans la langue pour disposer d'une large variété de sens textuels et interpersonnels » (Brinton 1996 : 278, rapporté dans Dostie 2004 : 23), à travers des processus de grammaticalisation.

La définition que j'adopte ici est que les MD

a. d'un point de vue syntaxique,
    a. sont des éléments extra-prédicatifs,
    b. sont assez libres,
b. présentent un degré de paradigmatisation limité mais certain,
c. d'un point de vue sémantique,
    a. n'ont pas une fonction descriptive, c'est-à-dire qu'ils n'apportent pas de contribution directe (on passe à un sens non propositionnel, cf. Dostie 2004 : 21, 44[37], Hansen 1996, 1998),
    b. ont une fonction procédurale (Trousdale 2012) et non descriptive ou référentielle, c'est-à-dire qu'ils jouent un rôle sur le plan énonciatif, sur celui de la construction du sens ou du discours (Nølke 1990 : 44, Bertin 2002).

Il s'agit là cependant d'une définition idéale, prototypique : cela signifie donc que seuls les MD les plus typiques remplissent ces critères, tandis que d'autres s'en éloignent plus ou moins nettement. Un des buts de mes recherches sur les MD a été de voir dans quelle mesure ces postulats sont tenables, pour un certain nombre de marqueurs – issus principalement d'impératifs, d'adverbes et d'adverbiaux des langues romanes médiévales et modernes : *alors* (Degand & Fagard 2011 [IV : 525 sqq.]), *au fond*, *parallèlement*, *d'un côté* (Fagard & Sarda 2014 [IV : 435 sqq.]), *en effet* (Charolles & Fagard 2012 [IV : 471 sqq.]) et ses équivalents romans (Fagard 2011b [IV : 499 sqq.]), *en fait* et ses équivalents romans, germaniques et polonais (communication, Mortier & Fagard 2008), *regarde* et ses équivalents romans (Fagard 2010b [IV : 587 sqq.], 2012b).

Résultats

Je m'intéresse au développement de marqueurs discursifs parce qu'il s'agit d'un cas limite de grammaticalisation (Fagard 2010a : 131), comme le suggérait déjà Pons Bordería (1998 : 218) ; on pourrait aussi dire, avec Traugott and Dasher (2002 : 88), que la grammaticalisation de marqueurs discursifs s'accompagne généralement d'une (inter)subjectivation. Il me semble tout de même préférable de ne pas distinguer fondamentalement ce phénomène de celui de la grammaticalisation en général. Cette composante de subjectivation et/ou d'intersubjectivation peut être illustrée par le parcours de *au fond*, qui signifie, dans une première phase de son histoire, 'au fond (d'un espace donné)', comme dans l'exemple ci-dessous :

[37] Ce qui explique que « leur présence ou leur absence ne modifie pas, en règle générale, la valeur de vérité des énoncés auxquels ils sont joints » (Dostie 2004 : 44).

(132) (moyen français)
il dit que en demy jour ne venist pas une pierre **au fons**.
« il dit que même en une demi-journée, une pierre n'arriverait pas **au fond** [du trou] »
(BFM, *Bérinus*, t. 1, p. 129, 1350)

Dans une deuxième phase, avec un premier mouvement de subjectivation, on peut gloser *au fond* par « à ce que je conçois comme l'élément essentiel (d'une réalité complexe) », comme dans l'exemple ci-dessous :

(133) (moyen français)
pour venir **au fons de la vraye hystoire et matiere**
« pour arriver **au fond** [= à l'essence, au cœur] de la vraie histoire »
(BFM, Jean Froissart, *Chroniques*, t. 3, p. 142, 1390)

Les emplois intersubjectifs apparaissent plus tard ; *au fond* peut alors se gloser par 'en fait, pour moi', et a pour fonction d'indiquer à l'interlocuteur que ce qui suit (ici) représente le fond de la pensée du locuteur :

(134) (français prémoderne)
Il y a dans notre position des choses embarrassantes, mais il faut les supporter et nous en tirer ensemble. **Au fond,** j'ai eu tort de la laisser exposée à une situation dont elle-même m'avait averti.
(Frantext, Benjamin Constant, *Journaux intimes*, p. 420-421, 1816)

On peut reconstruire une évolution semblable pour les autres marqueurs que j'ai étudiés, avec un glissement vers plus de subjectivité ou d'intersubjectivité qui intervient en même temps que la grammaticalisation (par exemple *parallèlement*, ibid.).

Le cas des conjonctions de coordination

On peut également observer une simultanéité des processus de subjectivation ou d'intersubjectivation et de grammaticalisation dans le cas de certaines conjonctions de coordinations. J'ai abordé ces questions dans une série d'articles, en travaillant avant tout sur les conjonctions et locutions prépositionnelles romanes, en diachronie (Degand & Fagard 2008 [III : 371 sqq.], Fagard 2008b [VI : 929 sqq.], Fagard & Degand 2008 [III : 389 sqq.], Fagard 2009c [III : 343 sqq.], Fagard & Degand 2010 [III : 313 sqq.], Fagard 2011a [III : 293 sqq.], Evers-Vermeul et al. 2011 [VI : 895 sqq.], Degand & Fagard 2012 [III : 269 sqq.]).

Un cas illustre particulièrement bien cette approche, celui du remplacement de *car* par *parce que*. Il constitue un bon exemple, parce qu'il y a, pour les deux conjonctions, une grammaticalisation accompagnée d'un mouvement de subjectivation et d'intersubjectivation, à partir d'un sens de base très proche, mais à plusieurs siècles de distance. Tandis que *car* a à peu près disparu du français moderne[38], *parce que* semble être devenu la conjonction causale par défaut (Simon and Degand 2007, Debaisieux 2002). J'ai ainsi pu montrer que l'évolution de *car* et celle de *parce que* sont bien des cas de subjectivation et d'intersubjectivation ; pour *parce que*, en particulier, sa fréquence élevée à l'oral est un bon indice de ce changement, comme son emploi comme MD en début de tour de parole (Degand & Fagard 2012 [III : 280-282]), avec l'émergence de formes raccourcies (*parce*, cf. Wiesmath 2002 : 396 ; *paske*, cf. Moeschler 1986).

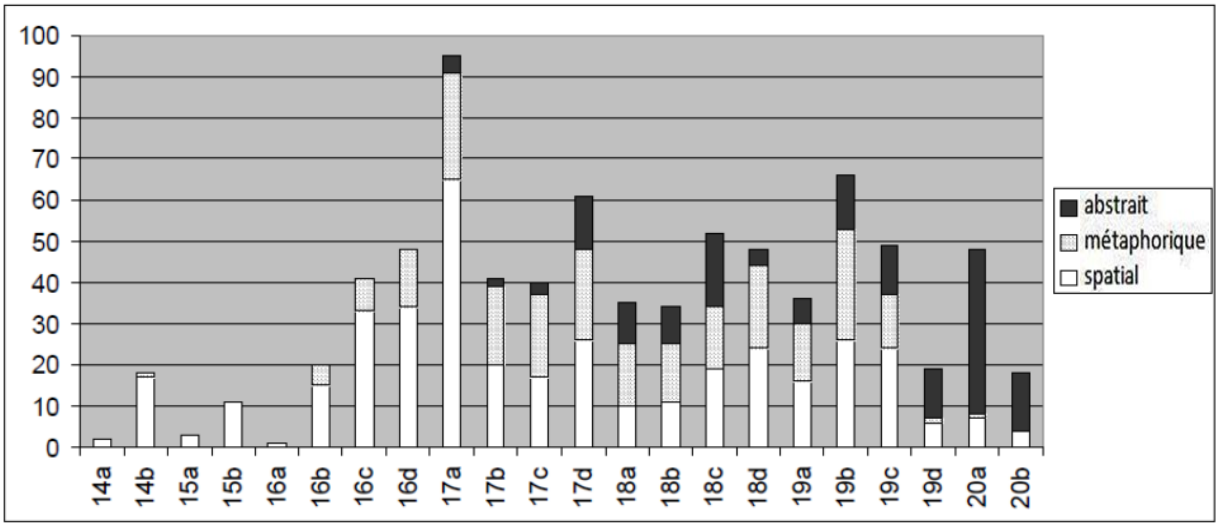
[38] Véronis & Guimier De Neef (2006) notent que *car* refait surface dans les SMS et les forums internet, "sans doute pour des raisons de facilité de saisie", ibid. : 241).

5.3 Grammaticalisation et syntaxe

Etant donné la complexité du phénomène de grammaticalisation, qui implique un ensemble de processus, comme je l'ai montré plus haut, une question fondamentale est celle de l'ordre de ces processus. Une interrogation, notamment, revient souvent (cf. Lewis 2014) : le glissement sémantique précède-t-il le figement syntaxique ? J'ai abordé cette question dans le cadre de plusieurs études, notamment concernant l'évolution des locutions prépositionnelles, des locutions conjonctives et des marqueurs discursifs. Je décris ci-dessous les résultats obtenus pour les MD (Fagard & Sarda 2014 [IV : 435 sqq.]) ; ces résultats valent, dans une certaine mesure, pour les autres phénomènes de grammaticalisation. Ce qui apparaît, c'est que la première étape est formelle : il s'agit d'une réduction de la construction initiale (1) ; cette étape est suivie d'une évolution sémantique, avec un glissement vers des sens métaphoriques (2) ; la phase suivante est à nouveau formelle, avec un glissement en début de phrase (3). J'illustre ci-dessous ce triple mouvement pour trois MD, *au fond*, *parallèlement* et *d'un côté* (ibid). Il faut noter, cependant, que les étapes (2) et (3) sont presque simultanées.

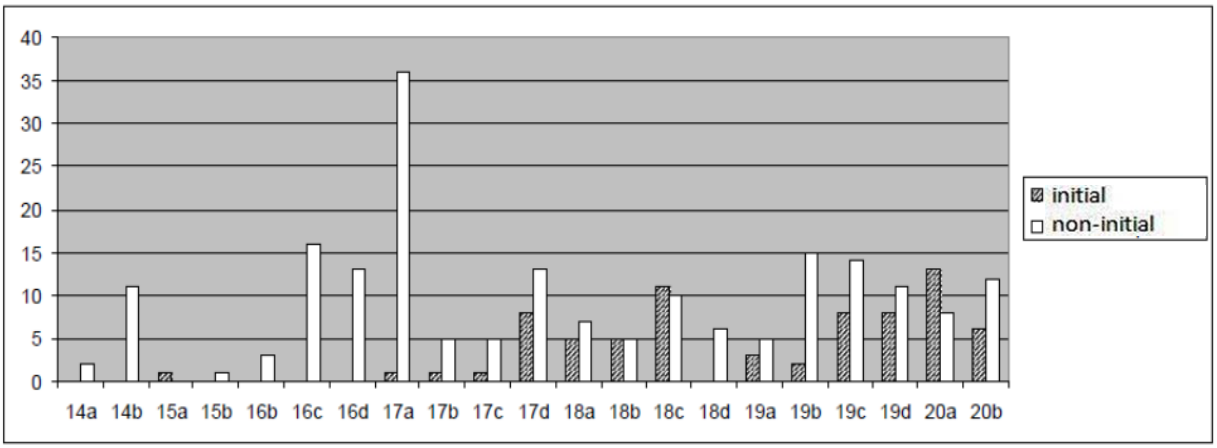
(1) Perte des compléments : *au fond de N*, *parallèlement à N*, *d'un côté de N* > *au fond*, *parallèlement*, *d'un côté*
(2) Changement sémantique : *au fond* « en bas de » > « au cœur de » ; *parallèlement* « formant une ligne parallèle » > « en même temps » ; *d'un côté* « sur un côté de » > « ceci dit »
(3) Mouvement syntaxique : d'une position interne vers le début de phrase puis vers une position détachée à l'initiale (évolution illustrée dans le graphique 16 ci-dessous)

Ainsi, l'hypothèse de Heine selon laquelle la sémantique joue un rôle prépondérant dans la grammaticalisation[39] ne peut être conservée que si l'on part du principe qu'il y a avant la phase 1 une sélection sur la base du sémantisme (ce qui est vraisemblable, cf. la section §3).



Graphique 16 : Evolution sémantique de *au fond* (Fagard & Sarda 2014 [IV : 451])

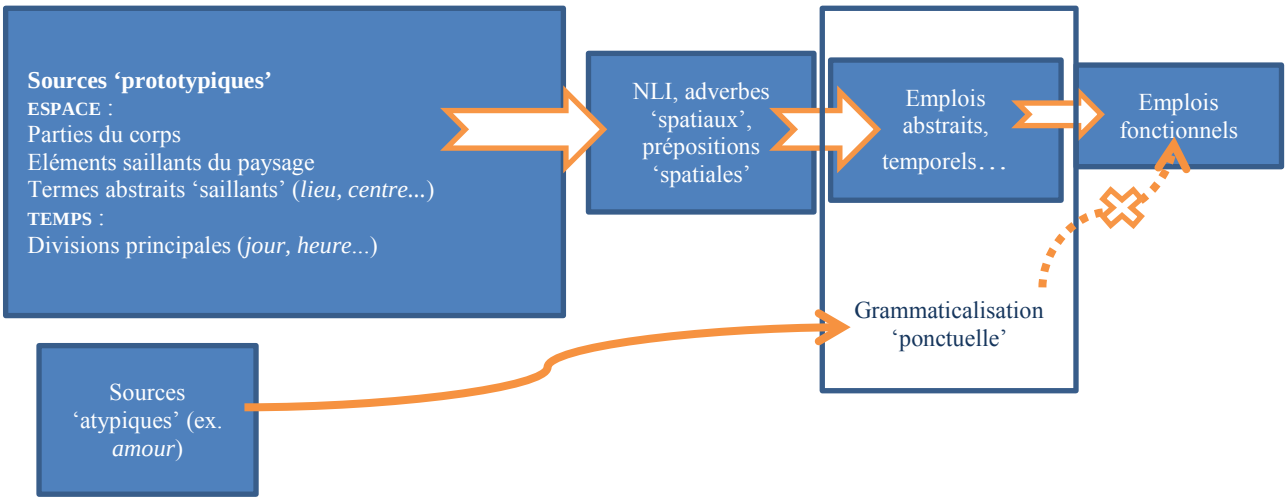
[39] Il écrit en effet : « grammaticalization, as conceived here, is above all a semantic process. (…) desemanticization precedes and is immediately responsible for decategorialization and erosion » (Heine 2003 : 583).



Graphique 17 : Mouvement syntaxique de *au fond* vers la position initiale (Fagard & Sarda 2014 [IV : 451])

5.4 Sources de la grammaticalisation et fin anticipée du processus

Un certain nombre de constructions présentent une évolution sémantique assez atypique, par exemple avec l'émergence de sens causaux à partir de lexèmes très spécifiques comme *amour* : (français classique) *pour l'amour de*, *pour l'amour que*, (galicien) *por mor de*, etc. Une autre caractéristique – probablement liée à cette dernière – de ces phénomènes est qu'ils ne donnent en général pas lieu à des grammaticalisations subséquentes. On pourrait ainsi distinguer d'une part les sources prototypiques, qui peuvent donner lieu par étapes à une série de grammaticalisations (I), et les sources atypiques, qui ne donnent pas lieu à de telles séries (II) :



Graphique 18 : Grammaticalisation(s)

Il faut noter, cependant, que le succès ou non d'un processus de grammaticalisation ne dépend pas que de sa source. La question des processus de grammaticalisation qui semblent n'avoir pas abouti, et que j'appellerais des grammaticalisations ratées, est complexe. Si l'on pose

comme base les quelques traits récurrents des phénomènes de grammaticalisation comme attrition, fréquence, perte des indices morphologiques, perte de liberté syntaxique, on retrouve un autre trait connu qui est que la grammaticalisation peut s'arrêter en route (Hopper & Traugott 2003) ; il est en fait constitutif de la grammaticalisation dans la mesure où bien souvent la grammaticalisation implique un choix dans un paradigme de formes.

Pour tenter d'élucider les conditions de la réussite d'une grammaticalisation, j'ai étudié quelques cas spécifiques comme le renforcement de constructions déjà grammaticalisées (esp. méd. *como* > *en como*, fr. méd. *puis que* > *de puis que*…), qui augmente en quelque sorte mécaniquement les pertes – puisque le nombre de formes du paradigme augmente et que le nombre de formes d'un paradigme grammatical est par définition limité.

Il est aisé d'illustrer l'importance numérique de ces pertes. Prenons par exemple les conjonctions complexes en français classique : celles que j'indique ci-dessous en gras ont toutes disparu :

> (135) (français préclassique)
> Ouy veritablement : Et ***combien qu'***il avient souvent que l'opinion est vraye, ce n'est point toutesfois science, ***à raison qu'***elle ne procede pas par les premiers principes, ny par les causes necessaires : comme, *quand* j'ay opinion *que* mon cousin vient par derriere parler à moy, *parce qu'*il m'a promis qu'il y viendroit, ***aveq ce que*** j'oy quelqu'un marcher : car il pourra estre que mon opinion sera vraye, et que ce sera mon cousin, nonobstant je ne l'ay pas cognu par les propres causes, et il pouvoit avenir que celuy que je pensois estre mon cousin estoit mon serviteur, ou bien quelque autre chose, *veu que* je n'avois pas veu que c'estoit. (…) Je le fay, ***partant que*** cela n'est point hors de propos, ***joint que*** j'en veux sçavoir ton avis, ***moienant que*** je ne t'importune de rien.
> (Frantext, Guy de Brués, *Les Dialogues de Guy de Brués contre les nouveaux académiciens*, p. 175, 1557)

Une liste complète serait difficile à établir, mais elle compterait au moins les constructions suivantes :

> à cause que, à ce que, à raison que, avec ce que, ainçois que, au moyen que, avec ce que, combien que, comme ainsi soit que, comme que, comment que, considéré que, de mode que, en façon que, en manière que, ensemble que, fors que, ja soit que, joint que, mais que, ne fût que, moienant que, neanmoins que, nonobstant que, ore que, orendroit que, outre ce que, partant que, pour l'amour que, quand ainsi seroit que, quelque part que, si est-ce que, si que, sinon que, somme que… (cf. entre autres Gougenheim 1974)

On peut faire les mêmes remarques sur d'autres langues, comme le montrent les listes suivantes pour l'espagnol et l'italien, respectivement :

> de commo, en commo, commo quier(a) que, por causa que, por esto que, por razón que, pues que… (Abreu 1986)
> conciossiacosacché, poscià che, in perché, inper ciò che, acciò che, avvegna che… (Fagard 2009 [III : 302-303])

À partir d'une étude sur corpus[40], j'ai tenté d'en savoir plus sur ce phénomène de disparition, en suivant l'évolution de quelques-unes des constructions notées ci-dessus :

> français : *si est-ce que (si esse que)*, *ore(s) que*, *partant que*, *quand ainsi seroit que*

[40] J'ai utilisé, pour le français : NCA / BTMF / Frantext / Valibel ; pour l'espagnol : Corde, Crea ; pour l'italien : OVI / Repubblica / Badip (Fagard 2010. Le renouvellement grammatical ou les ratés de la grammaticalisation. Colloque Dia, Gand).

italien : *con ciò sia cosa che (conciossiacosacché)*
espagnol : *por tal que*, *com(m)o quier(a) que*

Ce qui est remarquable, pour ces constructions, c'est qu'elles présentent toutes au moins certains indices d'une grammaticalisation en cours : augmentation de fréquence, glissement sémantique et/ou attrition. Ainsi, *si est-ce que* se trouve également sous la forme *si esse que*, dont la graphie peut être trompeuse, mais aussi indiquer un début d'attrition. Dans le même ordre d'idées, *ore(s) que* présente des emplois causaux ou au moins compatibles avec une interprétation causale, comme le suivant :

(136) (moyen français)
Dont furent moult esmerveilliez pour ce que, quant ilz estoient en l'autre cave avant les portes de mettail, ilz avoient oy de tresgrans bruis et murmuremens de gens, se leur sembloit; et **ores que** ilz estoient dedens, n'en ooyent, tant fust peu.
« C'est pourquoi ils furent très étonnés, parce que, quand ils étaient dans l'autre cave, avant les portes métalliques, ils avaient entendu de très grands bruits et ce qui leur semblait être des gens qui murmuraient ; et **maintenant qu'**ils étaient dedans, ils n'entendaient rien du tout. »
(Frantext, Antoine de La Sale, *La Salade*, t. 4, p. 90-91, 1442)

Toutes les constructions étudiées présentent aussi des signes de figement incomplet, indiquant que la grammaticalisation n'a pas atteint son terme. Par exemple, *comme ainsi soit que* permet l'insertion d'éléments entre la forme verbale et la conjonction, comme en attestent les exemples suivants :

(137) (français préclassique)
Mais, quand il en est parlé à la male part, le mot hérésie se prend pour « mauvaise secte », et hérétique pour un qui est d'une mauvaise secte ; ainsi qu'en françois, quand il est parlé d'une garse en la mauvaise part, il s'entend d'une garse impudique, c'est-à-dire d'une putain, **comme ainsi soit** *toutesfois* **que** le mot de garse signifie une fille.
(Frantext, Sébastien Castellion, *Conseil à la France désolée*, p. 57, 1562)

(138) (français préclassique)
Comme ainsi soit, *disoit-il,* **que** je fusse mortel, aussi devoit mourir, ce que j'avois engendré.
(Frantext, Jacques de Lavardin, *La Celestine*, p. 253, 1578)

(139) (français préclassique)
J'adjousteray encores, qu'en remuant la terre et dessous les pierres, en nostre contrée du Bresil, on trouve des scorpions lesquels, combien qu'ils soyent beaucoup plus petits que ceux qu'on voit en Provence, neantmoins pour cela ne laissent pas, comme je l'ay experimenté, d'avoir leurs pointures venimeuses et mortelles. **Comme ainsi soit** *doncques* **que** cest animal cerche les choses nettes, advint qu'apres que j'eu un jour fait blanchir mon lict de cotton, l'ayant rependu en l'air, à la façon des sauvages, il y eut un scorpion qui s'estant caché dans le repli : ainsi que je me voulu coucher, et sans que je le visse, me picqua au grand doigt de la main gauche, laquelle fut si soudainement enflée, que si en diligence je n'eusse eu recours à l'un de nos Apothicaires (lequel en tenant de morts dans une phiole, avec de l'huile, m'en appliqua un sur le doigt), il n'y a point de doute que le venin ne se fust incontinent espanché par tout le corps.
(Frantext, Jean de Léry, *Histoire d'un voyage faict en la terre du Brésil*, p. 294, 1578)

On peut également repérer d'autres indices : ainsi, la fréquence n'est pas toujours, à elle seule, un élément suffisant pour affirmer qu'il y a grammaticalisation, même avec une variabilité réduite. C'est ce que montre assez bien l'exemple de *con ciò sia cosa che*

(conciossiacosacché) « puisque, bien que » en italien médiéval : il a une fréquence relativement élevée avec 2 000 occurrences environ dans la base OVI, comparable donc avec les formes *poiché*, *da poi ché*, etc. (7 000 occurrences) et même *perché* (plus de 10 000 occurrences). La forme apparaît avec un sens abstrait (causal, concessif, voir (140-141) ci-dessous), dans plusieurs centaines de documents et dans plusieurs dialectes (émilien, toscan, vénitien, ombrien, sicilien), avec une variation formelle très marginale : les principales variantes sont *già sia cosa che* (2 occurrences), *pur che sia cosa che* (2 occurrences), *con ciò dunque sia cosa che* (3 occurrences) et *como fosse cosa che* (1 occurrence).

(140) (italien médiéval)
XV. **Cum ço sia cosa che**'l bono amigo scia meglio ca lo reo parente, la vostra amistade voglio tenere cara, cognoscando inutile essere lo stranio parentado.
« **Etant donné** (= si l'on accepte l'idée) qu'un bon ami est mieux qu'un mauvais parent, votre amitié me sera très chère, puisque je connaît l'inutilité des parents qui me sont étrangers »
(OVI, Guido Faba, *Gemma*, p. 8, 1239-48)

(141) (italien médiéval)
Questo sangue ci mostra anche l'amore e la carità tua, perché solo per foco d'amore ci ricomprasti, **conciosiacosacché** tu non hai bisogno di noi.
« Ce sang nous montre aussi ton amour et ta charité, parce que tu nous as rachetés grâce au seul feu de ton amour, **alors que** tu n'as pas besoin de nous »
(Santa Caterina Da Siena, *Orazione Quinta*, 14ème siècle)

Cependant, le fait que plus de 95 % des occurrences soient toscanes (alors que les documents toscans constituent moins des deux tiers du total) semble limiter son potentiel de grammaticalisation.

On pourrait faire la même analyse pour l'espagnol médiéval *por tal que* « puisque, bien que » (142) qui, malgré une fréquence élevée et un sens clairement abstrait, reste cantonné à des genres spécifiques – surtout le genre historique : 12 % du corpus, 44 % des occurrences de *por tal que*.

(142) (espagnol médiéval)
el deudor, non queriendo fazer testament, daua todos sus bienes a sus fillos en razón et en uotz de uéndida, **por tal que** podiesse engaynnar ad aqueillos que querían mouer alguna demanda contra eill
« Le débiteur, ne voulant pas faire de testament, donnait tous ses biens à ses enfants sous forme de vente, **pour** pouvoir tromper ceux qui voudraient faire une requête contre lui »
(Corde, *Vidal Mayor*, p. 46, 1250)

Il est très difficile d'identifier des causes à l'absence de grammaticalisation, mais on peut noter que dans tous les cas que j'ai étudiés certains des éléments fondamentaux du processus de grammaticalisation se sont produits mais pas d'autres, comme l'attrition, la montée en fréquence, la généralisation des emplois. J'ai ainsi proposé une série d'hypothèses dont chacune pourrait expliquer certains des cas d'échec du processus de grammaticalisation :

i. Facteurs systémiques :
    a. La disparition d'un élément de la construction (pour *si est-ce que*, le fait que *si* ait disparu comme connecteur ; pour *ore que*, la disparition de *ore* au sens de « maintenant »)

b. La disparition d'un paradigme entier (les subordonnées « d'accompagnement et de disjonction » (Gougenheim 1974) *ensemble que*, *avec ce que*, *outre ce que*, *joint que…*)
c. L'écrémage d'un paradigme de formes (disparition d'une variante parmi d'autres : *par autant que*, *d'autant que*, etc. > *d'autant que*)

ii. Facteurs sémantiques :
a. Le manque d'extension sémantique (par exemple, pour *ore que*, la rareté du sens causal)
b. Un sens de départ trop marqué pour permettre une généralisation rapide à d'autres contextes : ce serait ainsi le cas de *pour l'amour que*, malgré un début d'extension sémantique que l'on peut observer en (143) ; pareil pour *somme que*.

(143) (français préclassique)
– aimé vous bien l'Infante ?
– non.
– pourquoy ?
– **pou l'amou qu'**ale est Espagnole je n'en veu point.
(*Journal d'Héroard*, 1605, Ernst 1985 : 145, cité par Marchello-Nizia 1999 : 122)

c. Une variation sémantique trop importante (*partant que* – sens causal et concessif, cf. la perte du sens concessif pour *perché*, *porque…*)

iii. Autres facteurs :
a. La limitation à un genre marqué (la chancellerie pour *conciossiacosacché*, le genre historique pour *por tal que*) ou à un registre marqué (*à cause que*, *malgré que*)
b. Le gabarit même du terme (c'est-à-dire sa longueur ; pour *conciossiacosacché*, par exemple)

Pour certains de ces facteurs, il me semble qu'il s'agit d'éléments non spécifiques, qui caractérisent simplement l'absence de grammaticalisation (ainsi de la limitation à un genre marqué, iii-a) ou bien la sélection d'un terme parmi d'autres, conséquence inévitable et bien connue de la grammaticalisation (i-c) – c'est ce que Hopper & Traugott (1993 : 113) appellent la spécialisation (*specialization*) ; pour d'autres, il pourrait s'agir de caractéristiques empêchant la grammaticalisation – ainsi de la nécessité (ou plutôt du rôle facilitateur) d'un sens non spécialisé (ii-a) également notée par d'autres auteurs (Trousdale 2012, Hopper & Traugott 1993 : 97).

En résumé, on pourrait dire que le **marquage** est à la fois moteur de sélection et d'échec pour la grammaticalisation : les nouveaux éléments grammaticaux doivent être plus marqués que les anciens, mais ce caractère marqué n'est pas un élément suffisant pour garantir la diffusion de l'innovation.

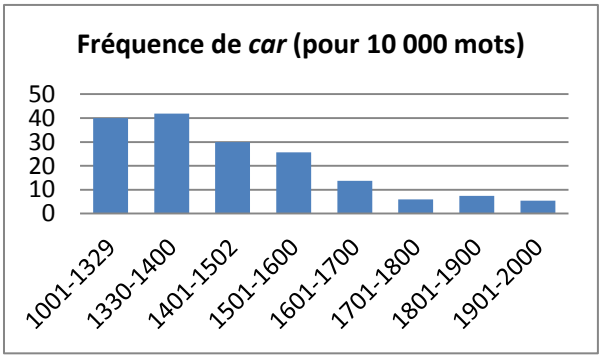
5.5 Conclusions sur la vitesse des changements

Outre les résultats spécifiques à la grammaticalisation, les études dont je rends compte dans les sections précédentes permettent de tirer quelques conclusions sur la vitesse des changements linguistiques – principalement l'évolution sémantique. On peut noter en particulier qu'il semble y avoir plusieurs vitesses, conformément à l'hypothèse de la courbe en S (voir entre autres Kroch 1989, Blythe & Croft 2012), à savoir des changements très rapides et d'autres très lents, ou plutôt des phases de changement plus ou moins rapides. Ces deux modalités sont peut-être à mettre en rapport avec la question du temps apparent ou celle, reliée, de l'âge des innovateurs, que j'ai abordée dans une étude déjà mentionnée (Fagard & Combettes 2013 [VI : 819 sqq.]), et celle de la propagation des changements, que j'ai commencé à modéliser à l'aide de collègues physiciens (Feltgen et al. 2015 [VI : 793 sqq.], Feltgen et al. soumis).

L'ensemble de ces recherches m'ont donc permis de montrer que, du point de vue de la vitesse du changement, on trouve deux types de phase : des périodes longues (préparation d'un changement, ou survivances d'un état antérieur pouvant durer plusieurs siècles) et des accélérations (apparition d'un nouveau sens, disparition d'un sens préexistant en un demi-siècle ou moins, ou encore remplacement partiel ou complet d'un mot par une variante : *en* > *dans*, *puis* > *depuis*, etc.). Un élément fondamental est l'âge des innovateurs : c'est apparemment une variable complexe, qui affecte l'évolution de plusieurs manières.

Le changement sémantique est parfois si rapide que les premières attestations d'une nouvelle construction ou d'un nouvel emploi (par exemple un adverbe devenant préposition ou conjonction, ou marqueur de discours, une préposition complexe qui émerge…) sont déjà polysémiques, même avec des tranches de 25 ans. J'ai déjà fourni plusieurs illustrations de ces points (§ 3.2).

À l'inverse, j'ai trouvé un certain nombre d'indices montrant la lenteur de certaines évolutions. Ces lenteurs de l'évolution linguistique peuvent être regroupés en deux types : persistance d'emplois très peu fréquents, et persistance au moins à l'écrit de lexèmes très rares. La persistance d'emplois très peu fréquents – entre un et trois pour cent des occurrences du corpus – s'observe pour de nombreuses constructions que j'ai étudiées, et on peut illustrer ce phénomène par la persistance de *car* (graphique 19) :



Graphique 19 : Fréquence de *car* dans un corpus diachronique (Fagard & Degand 2008 [III : 393])

Les recherches sur la question de l'âge des innovateurs (cf. § 1.2) nous ont permis de montrer (Fagard & Combettes 2013 [VI : 819 sqq.]) que, du moins pour une évolution spécifique – le remplacement de *en* par *dans* – il semblait y avoir chez plusieurs auteurs une évolution parallèle à l'évolution générale (tableau 15), avec pour tous trois une baisse de la fréquence relative de *en* et une hausse de celle de *dans*[41] :

auteur	date de l'œuvre	Fréquence relative (pour 10 000 mots)	
		en	*dans*
Chapelain	1623	198	35
	1656	119	53
Balzac	1620-1624	115	21
	1625-1629	112	21
	1630-1634	110	39
	1650-1654	111	52
Sorel	1623	171	19
	1627	143	30
	1664	141	54

Tableau 15 : Evolution de la fréquence de *en* et *dans* chez trois auteurs (base Frantext ; données de Fagard & Combettes 2013 [VI : 837])

On ne peut pas tirer de ces données des conclusions générales – il faudrait pour cela avoir un corpus beaucoup plus large, qui ne soit pas limité à quelques parties du discours (prépositions, conjonctions, adverbiaux et marqueurs discursifs) ni à quelques langues. Il faudrait, de plus, être tout à fait sûrs que l'évolution constatée pour la fréquence de *en* et *dans* ne soit pas le résultat artificiel de l'action des imprimeurs du 16ème siècle, ou des éditeurs des siècles suivants. Une question en particulier demeure en suspens : quel est le rapport entre l'évolution de la langue et ce que l'on peut reconstruire à travers les corpus ? Plus précisément, les changements rapides ne pourraient-ils pas être simplement des changements qui se sont propagés très lentement mais sont passés à l'écrit dans un laps de temps limité ? Inversement, les périodes longues ne pourraient-elles pas s'expliquer par des archaïsmes littéraires, bref par un artefact lié au caractère nécessairement écrit des corpus diachroniques ? Pour aller plus loin dans ces réflexions, il me semble que des collaborations avec des sociologues, notamment, sont indispensables.

La modélisation peut être une autre piste intéressante, et les recherches entamées avec Elisa Omodei (communication, Fagard & Omodei 2012) et poursuivies avec Jean-Pierre Nadal et Quentin Feltgen (Feltgen et al. 2015 [VI : 793 sqq.], à paraître) sont une occasion de remettre à plat les mécanismes de la grammaticalisation et de l'évolution linguistique. À la différence d'un certain nombre de modèles d'évolution des langues, qui restent souvent éloignés de la réalité linguistique (Nowak & Krakauer 1999, Kirby 2001, Briscoe 2002, Niyogi 2002, Jäger 2003, Victorri 2004b, Steels 2005, Choudhury et al. 2006, Puglisi et al. 2008, Pawlowitsch et al. 2011), le but de nos recherches – et de la thèse en cours de Quentin Feltgen, que je suis de près – est de comparer systématiquement les résultats du modèle avec des données de corpus.

[41] Notons en passant que ces différences sont significatives d'un point de vue statistique, pour les trois auteurs : Chapelain (chi deux = 14,51, $p<0,01$), Balzac (chi deux = 10,89, $p<0,01$) et Sorel (chi deux = 19,6, $p<0,01$) (en prenant uniquement, pour les deux derniers auteurs, la première et la dernière œuvre).

III Suites

> – We need six units programmed with separate languages as far removed from each other as possible.
> – Units?
> – We need working brains competent in six linguistically diverse languages. Six is an adequate statistical sample.
> (Ian Watson, *The embedding*, 1973)

> He was breaking apart from the stress of a lifetime spent building tenuous bridges between alien concepts that had only one thing in common – neither could possibly occur in the other language. Sometimes the pressure was unbearable and a lingster's mind fragmented, caught in the malign vortex of interface.
> (Sheila Finch, *The Guild of Xenolinguists*, 2007 : 153)

6 Projet de recherche

6.1 De la synthèse au programme

J'ai décrit dans ce qui précède mon parcours depuis la thèse. J'ai commencé par présenter (section 1) l'approche méthodologique suivie dans mes recherches – il s'agit pour moi d'observer la parole, en somme, pour arriver à la langue –, et (section 2) leur cadre théorique – incluant des concepts issus de divers courants : linguistique diachronique, contrastive, typologique et cognitive. J'ai ensuite illustré plus en détail trois pans de mes recherches : localisme (section 3), espace (section 4) et grammaticalisation (section 5). Pour chacun de ces trois pans, j'ai pris appui sur les études que j'ai menées et présentées ou publiées, seul ou en collaboration, pour indiquer la spécificité de mon approche dans le domaine en question, par rapport aux principaux courants actuels. Cette spécificité réside dans la combinaison des méthodes et des approches, qui est une tendance de fond.
À ce stade, il est temps de faire le bilan et de poser des jalons pour la suite. Je laisse le lecteur juger de la qualité du bilan, et je me bornerai à indiquer ici que je referais différemment un bon nombre d'études, que je les aborderais tout à fait d'un autre œil. C'est même là ce qui me permet d'envisager pour la suite de continuer à travailler en partie sur ces mêmes domaines. Je décris donc dans cette dernière section mon programme de recherches pour les années à venir. Il ne se limite pas à une thématique, ni à une méthode : **mon projet est de mettre une approche méthodologique renouvelée, que j'appelle la microtypologie**, **au service d'une ligne de recherche précise, la question du rapport entre impact cognitif des catégories linguistiques et degré de grammaticalisation.**
Comme l'ont montré les résultats inégaux des expériences de Gennari et al. (2002), Marian & Fausey (2006), Billman et al. (2000), Soroli & Hickmann (2010), Papafragou et al. (2002,

2007, 2008) pour les évidentiels et le mouvement, von Stutterheim et al. (2012) pour l'aspect, il est possible en effet que les catégories grammaticalisées aient un impact cognitif plus fort que les catégories lexicalisées – le *thinking for speaking* de Slobin serait ainsi, pour les catégories grammaticalisées, un *seeing for speaking* (von Stutterheim et al. 2012 : 835). Mais tout cela reste à étayer largement, comme le suggère la citation suivante :

> One of the main issues in research on language and cognition centers on the methods required in investigating the interplay of these different factors in cognitive and language processing. We need fine-grained experimental studies, embedded in an interdisciplinary approach that takes into account relevant findings in language processing, as well as visual and auditory processing in neuroscience and cognitive psychology. (von Stutterheim et al. 2012 : 862)

L'idée est d'aborder des questions ayant trait à l'interface entre langue et cognition, sans pour autant abandonner l'ambition descriptive, autrement dit sans perdre en finesse d'analyse linguistique : s'intéresser au système linguistique et cognitif sans pour autant lâcher prise avec les détours de la parole. Ce projet se prête bien à un passage en revue des thèmes que j'ai déjà parcourus. Ainsi, une réanalyse des enjeux de la sémantique prépositionnelle (à la Landau, par exemple) croisée avec des études sur corpus ciblées produirait selon moi des résultats intéressants. Il serait également utile de poursuivre dans la voie des études engagées sur l'espace, en mettant davantage en rapport les différents systèmes abordés : en quoi l'expression de la deixis, de la trajectoire, de la manière sont-ils reliés ? Dans quelle mesure dépendent-ils du (reste du) système linguistique ? Y a-t-il un lien avec la capacité ou la propension des locuteurs à recruter l'espace pour l'expression d'autres domaines ? Peut-on faire la même analyse pour des domaines tout à fait différents, comme la couleur, les nombres, les parties du corps ? Ma thèse m'a ouvert, à travers de nouvelles collaborations, sur d'autres horizons de l'étude de la sémantique prépositionnelle à celle des conjonctions, des marqueurs discursifs ; de la même manière, j'aimerais maintenant passer à un ensemble de domaines plus vaste, impliquant à son tour un élargissement du réseau de collaborations.

6.2 Une nouvelle linguistique

Le nom de microtypologie[42] doit suggérer à la fois une vocation typologique et la volonté d'attention portée aux détails de chaque langue. Il ne s'agit pas de rendre compte de l'ensemble de la diversité linguistique, tâche dont l'énormité me paraît rédhibitoire, mais de fournir un pont entre typologues chargés d'identifier les systèmes et phénomènes existant dans des langues très variées d'une part, et linguistes spécialisés dans l'étude d'une langue d'autre part. Cet entre-deux est nécessairement incommode, parce qu'il exclut aussi bien la prétention universaliste des typologues que celle de connaissance parfaite de la langue à laquelle peuvent prétendre les spécialistes.

Il me semble pourtant qu'il pourrait constituer un domaine de recherche indépendant, où le regard porté vers la diversité des langues est tempéré par la connaissance d'au moins un

[42] Ce nom m'a été suggéré par Alain Lemaréchal, que je tiens à remercier ici.

certain nombre d'entre elles, et où l'analyse systématique et en profondeur d'une langue donnée est informée par la connaissance d'autres systèmes linguistiques. Le but de cette entreprise reste de rendre compte à la fois de la diversité des langues et du fait que cette diversité, bien que considérable (cf. Evans & Levinson 2009), n'est pas infinie ; de démêler l'universel de l'arbitraire, c'est-à-dire d'isoler d'une part ce qui, dans la langue, est le fait du hasard, c'est-à-dire d'un simple jeu de probabilités, de ce qui est dû aux contraintes intrinsèques à tout système linguistique. Il s'agit en fin de compte d'identifier et d'expliquer les contraintes qui limitent cette diversité, ce qui correspond, pour certains, au but central de la linguistique :

> the central goal of linguistics is to describe the diversity of human languages and explain the constraints on that diversity. (Dunn et al. 2011 : 79)

L'évolution linguistique est-elle le fruit d'un choix arbitraire parmi toutes les possibilités offertes par le système, ou bien est-elle guidée par des règles strictes ? Les règles d'évolution phonétique ont été largement décrites, et celles qui gouvernent la grammaticalisation commencent à être bien connues. Peut-on avancer dans la description des règles qui valent plus spécifiquement pour le sens, dans la lignée des travaux de Darmesteter, Bréal ou, plus récemment, Blank et Koch ?

La microtypologie peut servir d'intermédiaire entre le linguiste spécialiste d'une langue d'un côté, et le typologue de l'autre. Le microtypologue, spécialiste d'un groupe de langues, devra connaître plusieurs voir toutes les variétés du groupe, et être capable d'en faire une typologie. L'intérêt est d'avoir une vue d'ensemble sans perdre la granularité nécessaire selon moi au travail du linguiste ; également, de s'intéresser à un groupe suffisamment restreint de langues pour rendre possible le travail de détail, c'est-à-dire le traitement de corpus, l'étude diachronique, etc. Nul doute que nombre de collègues ont déjà mis en pratique une telle approche : l'idée est aussi d'unifier ces efforts sous une bannière commune, et de redonner un souffle à un champ comme celui de la romanistique, en perte de vitesse dans bien des pays (particulièrement en France, mais aussi en Belgique, et y compris en Allemagne, son berceau, cf. Koch 2003 : 41-42), en lui associant des approches comparables. La différence principale entre romanistique traditionnelle et microtypologie est principalement que le regard sur la variation doit s'ouvrir encore sur la richesse typologique.

Le but de la microtypologie est ainsi, entre autres, de donner un cadre et un intermédiaire à la collaboration entre 'descripteurs' et typologues – collaboration dont Lazard, par exemple, déplore l'absence, en particulier en France (Lazard 2013 : 52-53)[43]. Il ne s'agit pas seulement

[43] « Ajoutons, pour compléter le tableau, quelques mots sur la situation en France, qui paraît assez particulière. Il semble que les linguistes s'y partagent, depuis assez longtemps, en deux groupes quantitativement inégaux : ceux qui s'intéressent à la diversité des langues, qui sont très minoritaires, et ceux, de loin les plus nombreux, qui s'appliquent à scruter la langue française et éventuellement les quelques langues européennes voisines. Les premiers contribuent à la grande oeuvre de description des langues dans les diverses parties du monde et à la comparaison typologique et tiennent une place honorable dans le concert international. Les seconds, sans ignorer complètement les structures, s'attachent surtout à l'étude des divers aspects de la mise en oeuvre de la langue dans le discours : la pragmatique, l'analyse du discours, l'information et la communication, la construction du sens, l'interaction, l'acquisition du langage, la didactique, etc. Ces deux groupes semblent n'avoir presque pas de contacts. Cette situation, qui s'explique probablement par l'organisation des cursus dans nos universités, est évidemment déplorable. »

d'améliorer grâce aux travaux typologiques la qualité des descriptions de langues individuelles en un cercle vertueux :

> As Dryer (2006:210) notes, 'typological work had a tremendous impact on descriptive work' over the last few decades, and typology is relying on better and better descriptions becoming available. Thus, prospects for a fruitful interplay in the future are excellent, and neither descriptivists nor typologists have to content themselves with a secondary role in this common enterprise of understanding human languages. (Haspelmath 2010 : 698)

Il ne me semble pas en effet que les linguistes doivent se limiter à être des descripteurs. Et en retour, les linguistes n'ont pas que des descriptions à offrir ; ils peuvent également servir de garde-fou à l'enthousiasme soulevé par la création de bases de données gigantesques :

> as more large linguistic databases become available, the approach developed here could be used to explore the dependency relationships between a wide range of linguistic features. (Dunn et al. 2011 : 81-82)

Certes, les exploitations numériques de données linguistiques sont désormais possibles ; mais la qualité des résultats obtenus n'est pas nécessairement proportionnelle à la quantité de langues, ou de données, analysées (voir par exemple Dyen et al. 1992, Gray et al. 2003, Bickel 2011), notamment si les typologues en viennent à accepter de revoir à la baisse leur exigence de qualité pour inclure davantage de langues :

> If we confine ourselves to high quality information, we will not be able to sample more than a small fraction of the 5,000 or so languages of the world. (Goddard 2001 : 1191)

De plus, ces approches de surface sacrifient nécessairement la diachronie, qui permet d'expliquer les motivations historiques du système linguistique vu en synchronie. Je suis en cela – par exemple – l'avis d'un éminent typologue et diachronicien :

> Language is a historical product and must be explained first of all with reference to the forces that have shaped it. (…) Explanations of language in terms of its synchronic structure are (…) likely to account for only a small part of why language is structured the way it is. (Heine 1997b : 4-6)

> linguistic explanations are of necessity incomplete unless they are supported by appropriate historical reconstructions. (Heine 1997b : 148)

Les typologues peuvent, dans certains contextes, tenter de reconstruire des états antérieurs et en déduire des évolutions possibles (cf. François 2004). Mais il s'agit toujours de reconstructions invérifiables. La beauté des langues romanes est qu'elles permettent de faire de telles reconstructions, puis d'en vérifier la validité à l'aide d'études sur les états anciens. La romanistique permet ainsi parfois de tester des hypothèses formulées sur d'autres groupes linguistiques. C'est que la linguistique romane est en fait, par essence, un moyen terme entre linguistique descriptive et typologie, comme le soulignait le regretté Peter Koch :

> In einem programmatischen Artikel stellte Brigitte Schlieben-Lange (1999) fest, dass sich die als Kind der deutschen Frühromantik entstandene Romanistik auch heute noch in einer fruchtbaren und zukunftsträchtigen Position des *entre deux* befindet, insofern sich ihr Blickfeld zwischen der universalen Ebene der Sprechtätigkeit und der historischen Ebene der Einzelsprache im Sinne Coserius (1981, 269-286) bewegt. Einerseits steht die

> Romanistik von dieser höheren Warte aus über den Verstrickungen, in denen bornierte Nationalphilologien allzu oft befangen sind. Andererseits bleibt sie dank ihrer «Bodenhaftung» auf der historisch-einzelsprachlichen Ebene gefeigt gegen die Verlockungen eines unbedachten Universalismus. (…) Dieses «komparatistische Potential», diese «Mehrperspektivität», die «bereits in der Disziplin angelegt ist» (Schlieben-Lange 1999, 852), prädestiniert die romanische Sprachwissenschaft zu einem Vorgehen, das Divergenz und Konvergenz abwägt und den begründeten Mut zu vorsichtigen Generalisierungen beinhaltet. In gewisser Weise stellt die Romania einen «Mikrokosmos» dar, der für übereinzelsprachliche Betrachtungen und Sprachvergleich wie geschaffen ist. (Koch 2006 : 101)

Cela fait même, d'après lui, de la romanistique un modèle pour la linguistique en général :

> Schon öfter ist die Romanistik als *praeceptrix linguisticae* bezeichnet worden (…) Sie sollte sich dieses Selbstverständnis nicht nehmen lassen. (Koch 2006 : 128 ; voir aussi Koch 2003 : 44-45)

Les avancées de la typologie, depuis Greenberg au moins, pourraient être intégrées davantage encore dans la pratique des romanistes. De plus, il me semble indispensable que la microtypologie ne soit pas limitée aux langues romanes, mais s'attaque, si possible, à tout groupement de langues apparentées. La question fondamentale est celle de la microvariation, et de son impact en diachronie (quelle que soit l'échelle).

Il s'agit à la fois de reconnaître l'apport immense des recherches typologiques à grande échelle, et de pointer leurs limites : il ne peut y avoir de microtypologie sans une (très) bonne connaissance des langues à l'étude ; inversement, il ne peut y avoir de typologie avec une bonne connaissance des langues à l'étude, puisque personne ne peut se targuer de connaître 50 langues sans lien génétique entre elles (à ma connaissance). En clair, la typologie est riche, mais certains travaux menés dans ce domaine ne sont guère plus que des introductions très éloignées de la réalité des données linguistiques.

6.3 Un exemple

Au-delà de la combinaison de méthodes, le but est d'obtenir des résultats qui surclassent ceux que l'on pourrait obtenir avec n'importe laquelle des méthodes en question. On peut reprendre la question des préfixes verbaux, en la précisant quelque peu : « comment fonctionnent les verbes à particule ? ». Il faut déjà noter qu'on ne peut pas la poser pour n'importe quelle langue, puisque cette structure n'existe pas partout. Il faut noter, de plus, que la définition même de « verbe à particule » dépendra en partie de la langue choisie, puisque, comme le dit très justement Haspelmath (2010), les catégories ne sont pas indépendantes de la langue, et cela n'a pas grand sens de parler d' « adverbe » dans une centaine de langues en faisant comme s'il s'agissait de la même chose. Cela justifie pleinement les approches de type prototypique, ou encore l'idée de cristallisation préférentielle autour de certaines notions invariantes (Lazard 1992 : 431-432) : pour des raisons cognitives et fonctionnelles, la plupart des langues se trouvent finalement avoir des catégories grammaticales assez proches, comme

si elles recrutaient dans l'espace des possibles des portions spécifiques. Ainsi, de nombreuses langues ont des mots de type verbal pour exprimer des actions, qui peuvent être déterminées pour un certain nombre de critères dont l'agent, sa nature, son nombre, son genre, et des mots de type nominal pour faire référence à différents types d'entités, avec là encore un certain nombre de marquages récurrents. Ce n'est pas systématique ou universel ; il s'agit de tendances, que l'expression de Lazard rend très bien.
Cependant, il faut bien admettre que la réponse à la question pour une langue donnée ne peut qu'être plus intéressante si l'on a une idée du fonctionnement de structures proches dans d'autres langues ; et inversement, que l'étude du fonctionnement de ce type de structure dans une série de langues sera plus intéressante si l'on a une idée très précise de leur fonctionnement dans (au moins) une langue donnée. On peut en effet noter d'une part que les approches monolingues tendent parfois à ignorer les données provenant d'autres langues, d'autre part que les approches typologiques larges courent le risque de rester trop superficielles. Il faut noter en outre le risque de distorsion entre motivations sémantiques générales et structure de chaque langue, comme je l'ai noté par exemple pour le mouvement fictif (§ 5.3).

6.4 Interdisciplinarité

Je conçois la microtypologie comme une approche ancrée dans les sciences cognitives, qui ne s'arrête pas à l'interrogation du système linguistique et de ses transformations mais interroge ses fondements cognitifs. Parmi les nombreuses contraintes qui pèsent sur l'évolution linguistique – contraintes fonctionnelles liées à la situation d'interaction, à la structure sociale, contraintes liées au système lui-même – les contraintes cognitives sont en effet fondamentales. Ces contraintes sont, semble-t-il, en partie liées à la structure neuronale du cerveau, au fait que le langage n'est pas un module isolé des autres fonctions cognitives (Stowe et al. 2004). Cela peut s'expliquer par la proximité et à la complexité des aires du cerveau, dont une conséquence est que des stimulus différents peuvent être traités par les mêmes neurones, aires neuronales ou zones neuronales :

> Le modèle phrénologique « une fonction = une région » est ainsi définitivement abandonné au profit d'un modèle en larges réseaux (large-scale network), une région donnée pouvant participer aux réseaux de différentes fonctions mentales. La mise en évidence de ces réseaux constitue une remise en cause assez fondamentale des modèles « modularistes » du fonctionnement cognitif. (Mazoyer 2002 : 606 ; je souligne[44])

Bien au contraire, on sait maintenant que certains types de neurones (dont les fameux neurones miroir, ou multimodaux, pour reprendre le terme de Gallese) sont sensibles à plusieurs stimulus (perception, action et langage) : ces phénomènes ont donc lieu non seulement au niveau des aires du cerveau, mais encore au niveau des neurones eux-mêmes. Il

[44] Plus récemment, et plus spécifiquement, on se reportera avec profit aux recherches de Vittorio Gallese et ses collègues – entre autres – sur le cerveau (de l'homme et du macaque), cf. Uithol et al. (2015), Murata et al. (1997).

me semble inévitable, dans la suite de mes recherches, de faire a minima un travail de veille scientifique sur ce domaine, en suivant des travaux en particulier en neurologie (notamment ceux des équipes de Vittorio Gallese et Giacomo Rizzolatti), neurolinguistique (par exemple D'Angiulli et al. 2015, Montoro et al. 2015), psycholinguistique (par exemple Flecken et al. 2013, Majid et al. 2014), ainsi que d'autres pistes comme celle de Bergen & Chang (2013).
Mais s'intéresser à ces contraintes cognitives suppose également que l'on fasse la part des autres. Il existe en effet des contraintes liées à l'histoire externe, notamment au contact linguistique, avec des phénomènes de substrat, superstrat, adstrat. D'autres phénomènes externes incluent l'effort de régulation et standardisation, qu'il soit formalisé par une académie ou par la scolarisation par exemple, ou qu'il soit plus diffus. On pourra penser également aux contraintes phonétiques (Martinet 1955), et il ne faut pas oublier les contraintes fonctionnelles, liées à l'usage que font les êtres humains du langage : situation de dialogue, oralité, société humaine, corps humain, existence sur terre. Enfin, les contraintes systémiques ne sont pas négligeables, comme l'existence d'une grammaire, si on peut la relier à des effets de fréquence (cf. Hopper 1987, Bybee 2007, Diessel 2007). Le défi est de profiter des avancées de chaque domaine, de suivre ces différentes perspectives, sans perdre de vue la question centrale du lien entre grammaticalisation et conceptualisation.

6.5 Applications

Le programme de recherches tel que je l'ai défini ci-dessus a vocation à être appliqué, bien sûr, aux domaines que j'ai déjà étudiés : un bon point de départ serait ainsi d'analyser l'expression du mouvement et de la localisation dans les langues en croisant corpus, approche typologique ou contrastive et études expérimentales, afin de voir dans quelle mesure chaque approche permet de confirmer ou non les résultats des autres. J'ai déjà commencé à le faire, notamment dans Evers-Vermeul et al. (2011 [VI : 895 sqq.]) et Stosic et al. (2015 [V : 631 sqq.]), mais il reste encore du chemin, que ce soit du point de vue de la théorie, de la méthode ou des phénomènes à analyser. Il serait de plus intéressant d'étendre ces analyses à des phénomènes nouveaux, ou que je n'ai que peu étudiés pour l'instant, comme les parties du corps, les couleurs, les évidentiels ; autant de domaines qui sont souvent au cœur des débats entre relativistes et universalistes, et constituent donc un terrain privilégié pour étudier la variation interlinguistique.
Les domaines les plus indiqués pour commencer sont ceux qui ont trait à la question de l'universalité et sont à la marge de processus de grammaticalisation, puisqu'ils permettraient ainsi de continuer à affiner nos connaissances sur ce phénomène complexe. Il s'agirait en particulier d'étudier la question de la constructionnalisation : l'approche constructionnelle constitue-t-elle réellement un gain pour la compréhension des phénomènes de grammaticalisation ? La réponse n'est pas évidente, et sera conditionnée par un certain nombre d'études de cas, sur différents types de phénomènes. Le phénomène de grammaticalisation, malgré le nombre impressionnant d'études scientifiques qui lui ont été dédiées ces trente dernières années, conserve une part de son mystère ; si, intuitivement, on

est tenté d'identifier différents types de grammaticalisation (lexicalisation d'une construction, pragmaticalisation d'un marqueur discursif, grammaticalisation avancée d'un élément déjà grammatical), cela n'exclut pas une analyse unitaire (cf. §5). Beaucoup reste à faire dans ce domaine, qu'il s'agisse de l'étude des marqueurs spatiaux (prépositions simples et complexes, cas, adverbes déictiques) ou non spatiaux (marqueurs discursifs, modaux, évidentiels).
Dans la lignée des travaux de Wierzbicka (1996), une réflexion sur les universaux « sémantiques » serait également intéressante, car elle permettrait de voir les spécificités de la grammaire ; l'articulation entre sémantique et grammaire est d'ailleurs une des pistes de ses recherches (Wierzbicka 1988). Dans la lignée du projet EoSS, il serait tout à fait intéressant d'aborder des questions comme celle des dénominations des couleurs ou des parties du corps (ces dernières, également abordées dans le cadre du projet DECOLAR à Tübingen, sont un terrain de choix pour étudier les questions de conceptualisation) dans une perspective microtypologique.
Les questions plus proprement cognitives restent également ouvertes, comme celles de l'asymétrie source-but, du statut cognitif de constructions comme le mouvement fictif ou les structures existentielles. Une approche expérimentale « à la MPI », pour reprendre Slobin et al. (2014), semble fondamentale pour pouvoir prétendre aller plus loin que des affirmations sur la structure linguistique. La modélisation permet elle aussi d'aller au-delà du fait linguistique pour interroger les dynamiques du changement. De ce point de vue, ces deux approches me semblent prometteuses, sans pour autant devoir remplacer les études purement linguistiques.

La triangulation, le croisement des données de plusieurs époques, de plusieurs langues, tout cela n'est pas une garantie de scientificité, et de justesse des résultats. On pourrait même dire que l'exercice devient plus périlleux à mesure qu'on ajoute des perspectives nouvelles, car le risque grandit de mal comprendre ou mal maîtriser les techniques utilisées et les enjeux. Une chose est sûre : un tel programme de recherche n'est pas humain s'il est confié à une seule personne. Si je ne l'envisage sérieusement, ce n'est qu'au travers de collaborations multiples, en allant chercher des spécialistes pour travailler avec eux sur tel ou tel point de détail. Cette approche suppose en outre la direction de thésards ou de post-doctorants, avec une répartition des tâches par langue ou, à la limite, par groupe de langues.
S'agissant de thèses par exemple, je pourrais, en repartant de la liste de thèmes donnée plus haut, diriger des thèses portant sur des sujets comme les suivants :

Emergence des marqueurs discursifs dans les langues romanes
La deixis statique et dynamique en typologie
L'asymétrie source-but en diachronie (langues romanes, slaves ou germaniques)
Localisme et lexique dans les langues européennes
La grammaticalisation (un lexique universel pour les langues romanes)
Les évidentiels (langues européennes et approche typologique)
Les parties du corps (diachronie, romanistique)
Les couleurs (approche expérimentale, langues d'Europe)

Ces thèses, comme les projets de recherche, se feraient de préférence en collaboration avec des collègues d'autres disciplines (notamment des physiciens, des TAListes et des psycholinguistes), et des linguistes spécialistes d'autres langues ou domaines linguistiques.

Afin de mettre en place ce type de collaboration, je pourrai compter sur celles dont j'ai déjà bénéficié ou sur les nombreux contacts établis lors de conférences, au fil des ans : en France (à Lille, Lyon, Toulouse, Aix-Marseille, Strasbourg et Paris notamment : laboratoires Modyco, Lilpa, CLLE-ERSS, SeDYL, DDL, ICAR, LPS), en Belgique (Louvain-la-Neuve : Liesbeth Degand, Leuven : Béatrice Lamiroy, Anvers : Walter de Mulder), en Allemagne (Tübingen : Sarah Dessi, Stuttgart : Achim Stein), en Angleterre (Cambridge : Wendy Ayres-Bennett, Oxford : Martin Maiden), en Italie (Turin : Massimo Cerruti), au Portugal (Lisbonne : José Pinto de Lima, Ana Margarida Abrantes), en Suède (Lund : Jordan Zlatev), en République Tchèque (Prague : Jaroslav Štichauer).

Bibliographie

Aarts, Bas & Charles F. Meyer (eds). 1995. *The Verb in Contemorary English: Theory and Description.* Cambridge : Cambridge University Press.

Aboh, Enoch O., Elisabeth van der Linden, Josep Quer & Petra Sleeman (eds). 2009. *Romance languages and linguistic theory 2007*. Amsterdam / Philadelphie : John Benjamins.

Abreu, Josefa María Mendoza. 1986. El sistema conjuntivo causal en textos no literarios de la Edad Media (siglos XIII-XV), *Cauce* 9, 5-26.

Ahlqvist, Anders (ed). 1982. *Papers from the Fifth International Conference on Historical Linguistics*. Amsterdam / Philadelphie : John Benjamins.

Aijmer, Karin & Bengt Altenberg. 2013. *Advances in Corpus-Based Contrastive Linguistics: Studies in honour of Stig Johansson*. Amsterdam / Philadephie : John Benjamins.

Aijmer, Karin. 1997. *I think* – an English modal particle. In Swan & Westvik (eds), 1-47.

Alturo, Núria, Òscar Bladas, Marta Payà & Lluís Payrató (eds). 2004. *Corpus oral de registres. Materials de treball.* Barcelone : Publicacions i Edicions de la Universitat de Barcelona.

Ameka, Felix, Alan Dench & Nicholas Evans. 2006. *Catching Language: Issues in Grammar Writing.* Berlin / New York : Mouton de Gruyter.

Anderson, John M. 1971. *The Grammar of Case: Towards a Localistic Theory*. Cambridge : Cambridge University Press.

Anderson, John M. 1973. Maximi Planudis in memoriam. In Kiefer & Ruwet (eds), 20-47.

Anschütz, Susanne (ed). 1992. *Texte, Sätze, Wörter und Moneme*. Heidelberg : Heidelberger Orientverlag.

Anttila, Raimo. 1989 [1972]. *Historical and Comparative Linguistics*. Amsterdam / Philadelphie : John Benjamins.

Archaimbault, Sylvie, Jean-Marie Fournier & Valérie Raby (eds). 2013. *Penser l'histoire des savoirs linguistiques. Etudes épistémologiques, historiques et linguistiques en hommage à Sylvain Auroux*. Lyon : ENS Editions.

Artstein, Ron & Massimo Poesio. 2008. Inter-Coder Agreement for Computational Linguistics, *Computational Linguistics* 34(4), 555-596.

Aske, Jon. 1989. Path Predicates in English and Spanish: A Closer Look. In *Proceedings of the Fifteenth Annual Meeting of the Berkeley Linguistics Society*, 1-14.

Aurnague, Michel. 2004. Les structures de l'espace linguistique : regards croisés sur quelques constructions spatiales du basque et du français. Louvain / Paris : Peeters.

Aurnague, Michel. 2011. *Quittant tout, nous partîmes* : *quitter* et *partir* à la lumière des changements de relation locative, *Journal of French Language Studies* 21(3), 285-312.

Aurnague, Michel. 2012. Quand la routine s'installe : remarques sur les emplois de *à* de type 'routine sociale', *Revue Romane* 47(2), 189-218.

Ayres-Bennett, Wendy. 2004. *Sociolinguistic Variation in Seventeenth-Century France: Methodology and Case Studies.* Cambridge : Cambridge University Press.

Banniard, Michel. 1993. Les deux vies de saint Riquier : du latin médiatique au latin hiératique. *Médiévales* 25, 45-52.

Barsalou, Lawrence W. 1999. Perceptual symbol systems, *Behavioral and Brain Sciences* 22, 577-660.

Beavers, John, Beth Levin & Shiao Wei Tham. 2010. The typology of motion expressions revisited. *Journal of Linguistics* 46, 331-377.

Bennett, Paul, Martin Durrell, Silke Scheible & Richard J. Whitt (eds). 2013. *New Methods in Historical Corpora*. Tübingen : Narr.

Bergen, Benjamin K. & Nancy Chang. 2013. Embodied Construction Grammar. In Hoffmann & Trousdale (eds).

Bergen, Benjamin K., Carl Polley & Kathryn Wheeler. 2010. Language and inner space. In Evans & Chilton (eds), 79-92.

Bergen, Benjamin K., Shane Lindsay, Teenie Matlock & Srini Narayanan. 2007. Spatial and Linguistic Aspects of Visual Imagery in Sentence Comprehension, *Cognitive Science* 31, 733-764.

Berthele, Raphael. 2013. Disentangling manner and path. Evidence from varieties of German and Romance. In Goschler & Stefanowitsch (eds), *Variation and Change in the Encoding of Motion Events*, 55-75.

Berthoz, Alain. 2013. *Le sens du mouvement*. Paris : Odile Jacob.

Bertin, Annie. 2002. L'émergence du connecteur *en effet* en moyen français, *Linx* 46, 37-50.

Biber, Douglas, Susan Conrad & Randi Reppen. 1998. *Corpus linguistics – Investigating Language Structure and Use*. Cambridge : Cambridge University Press.

Biber, Douglas. 1990. Methodological issues regarding corpus-based analyses of linguistic variation, *Literary and linguistic computing* 5, 257-269.

Bickel, Balthasar. 2011. Statistical modeling of language universals, *Linguistic Typology* 15(2), 401-413.

Billman, Dorrit, Angela Swilley & Meredyth Krych. 2000. Path and manner priming: verb production and event recognition. In *Proceedings of the 22nd Annual Conference of the Cognitive Science Society*. Hillsdale : Lawrence Erlbaum.

Blanche-Benveniste, Claire. 1997. *Approches de la langue parlée en français*. Paris : Ophrys.

Blank, Andreas & Peter Koch (eds). 1999. *Historical Semantics and Cognition*. Berlin / New York : Mouton de Gruyter.

Blank, Andreas. 1997. *Prinzipien des lexikalischen Bedeutungswandels am Beispiel der romanischen Sprachen*. Tübingen : Niemeyer.

Blomberg, Johan & Jordan Zlatev. 2014. Actual and non-actual motion: why experientialist semantics needs phenomenology (and vice versa), *Phenomenology and the Cognitive Sciences* 13(3), 395-418.

Blomberg, Johan. 2014. *Motion in Language and Experience : Actual and Non-actual motion in Swedish, French and Thai*. Thèse. Lund : Travaux de l'Institut de linguistique de Lund, 53. The Faculties of Humanities and Theology.

Bloom, Paul, Mary A. Peterson, Lynn Nadel & Merrill F. Garrett (eds). 1996. *Language and Space*. Cambridge : MIT Press.

Blumenthal, Peter. 2008. Combinatoire des prépositions : approche quantitative, *Langue Française* 157, 37-51.

Blythe, Richard A. & William Croft. 2012. S-curves and the mechanisms of propagation in language change, *Language* 88(2), 269-304.

Boas, Hans C. (ed). 2010. Contrastive Studies in Construction Grammar. Amsterdam / Philadelphie : John Benjamins.

Boas, Hans C. 2013. Cognitive Construction Grammar. In Hoffmann & Trousdale (eds), 233-254.

Boix-Fuster, Emili, Marina Àlamo Sala, Mireia Galindo Solé & Francesc Xavier Vila i Moreno (eds). 2007. *Corpus de Varietats Socials. Materials de treball.* Barcelona: Publicacions i Edicions de la Universitat de Barcelona.

Bonami, Olivier, Danièle Godard & Brigitte Kampers-Manhe. 2003. Adverb Classification. In Corblin & De Swart (eds), 143-184.

Bonvini, Emilio, Joëlle Busuttil & Alain Peyraube (eds). 2011. *Dictionnaire des langues*. Paris : PUF.

Bopp, Franz. 1833[-1852]. *Vergleichende Grammatik des Sanskrit, Zend, Griechischen, Lateinischen, Litthauischen, Altslawischen, Gothischen und Deutschen*. Berlin : Königliche Akademie der Wissenschaften.

Borillo, Andrée. 1998. *L'espace et son expression en français*. Paris : Ophrys.

Boroditsky, Lera & Michael Ramscar. 2002. The Roles of Body and Mind in Abstract Thought, *Psychological Science* 13(2), 185-189.

Boroditsky, Lera. 2001. Does Language Shape Thought? Mandarin and English Speakers' Conceptions of Time, *Cognitive Psychology* 43, 1-22.

Bowerman, Melissa & Eric Pederson. 1992. *Topological relations picture series*. In Levinson (ed).

Branca-Rosoff, Sonia, Serge Fleury, Florence Lefeuvre & Matthew Pires. 2012. *Discours sur la ville. Présentation du Corpus de Français Parlé Parisien des années 2000 (CFPP2000)*. [cfpp2000.univ-paris3.fr/CFPP2000.pdf]

Bréal, Michel. 1924 [1897]. Essai de sémantique : science des significations. Paris : Gérard Monfort.

Bréal, Michel. 1995 [1866]. De la forme et de la fonction des mots (leçon faite au Collège de France pour la réouverture du cours de grammaire comparée). In Desmet & Swiggers (eds), 89-96. [*Revue des cours littéraires de la France et de l'étranger*, 4^{e} année, n° 25 (29 déc.), p. 65-71].

Bresnan, Joan. 2007. Is syntactic knowledge probabilistic? Experiments with the English dative alternation. In Featherston & Sternefeld (eds), 75-96.

Brinton, Laurel J. & Elizabeth C. Traugott. 2005. *Lexicalization and language change*. Cambridge : Cambridge University Press.

Brinton, Laurel J. 1996. *Pragmatic Markers in English: Grammaticalization and Discourse Functions*. Berlin / New York : Mouton de Gruyter.

Briscoe, Ted (ed). 2002. *Linguistic evolution through language acquisition*. Cambridge : Cambridge University Press.

Burnett, Heather & Mireille Tremblay. 2009. Variable-behaviour Ps and the Location of PATH in Old French. In Aboh et al. (eds), 25-50.

Bybee, Joan L. & Dan I. Slobin. 1982. Why small children cannot change language on their own: suggestions from the English past tense. In Ahlqvist (ed), 29-38.

Bybee, Joan L. 2007. *Frequency of use and the organization of language*. Oxford : Oxford University Press.

Capelli, Fabien. 2013. *Etude du mouvement fictif à travers un corpus d'exemples du français : perspective sémantique du lexique au discours*. Thèse de doctorat, Université Toulouse 2.

Carapezza, Marco, Daniele Gambarara & Franco Lo Piparo (eds). 1997. *Linguaggio e cognizione. Atti del XXVIII Congresso*. Rome : Bulzoni.

Carrera-Sabaté, Josefina & Joaquim Viaplana (ed). 2008. *Corpus Oral Dialectal (COD). Textos orals del nord-occidental*. Barcelone : Dipòsit Digital de la UB.

Chanet, Anne-Marie. 1985. Maxime Planude, localiste ?, *Histoire Epistémologie Langage* 7(1), 127-148.

Charolles, Michel. 1997. L'encadrement du discours : Univers, champs, domaines et espaces, *Cahiers de Recherche Linguistique* 6, 1-73.

Chilton, Paul. 2010. *Introduction*. In Evans & Chilton (eds), 1-18.

Choudhury, Monojit, Anupam Basu & Sudeshna Sarkar. 2006. Multi-agent Simulation of Emergence of the Schwa Deletion Pattern in Hindi, *Journal of Artificial Societies and Social Simulation* 9(2). [jasss.soc.surrey.ac.uk/9/2/2.html]

Chuwicha, Y. 1993. *Clausehood in Serial Verb Constructions in Thai.* Thèse de doctorat, Université Chulalongkorn, Bangkok.

Cifuentes-Honrubia, José Luis. 2003. *Lengua y Espacio. Introducción al problema de la deíxis en español.* Alicante : Universidad de Alicante.

Col, Gilles & Daniel Roulland (eds). 2001. *La grammaticalisation, Travaux linguistiques du Cerlico* 14(2). Rennes : Presses universitaires de Rennes.

Colombat, Bernard. 2007. Les premières descriptions des langues du monde. [htl.linguist.univ-paris-diderot.fr/bernard/langues_monde.pdf]

Combettes, Bernard & Christiane Marchello-Nizia. 2008. La périodisation en linguistique : problèmes théoriques et méthodologiques. In Durand, Habert & Laks (eds), Actes du CMLF 2008. [www.linguistiquefrancaise.org/articles/cmlf/pdf/2008/01/cmlf08322.pdf]

Combettes, Bernard & Christiane Marchello-Nizia. 2010. La périodisation en linguistique historique : le cas du français préclassique. In Combettes et al. (eds), 129-142.

Combettes, Bernard, Céline Guillot, Sophie Prévost, Evelyne Oppermann-Marsaux & Amalia Rodríguez Somolinos (eds). 2010. *Le changement en français. Etudes de linguistique diachronique*. Berne : Peter Lang.

Company Company, Concepción. 2002. Grammaticalization and category weakness. In Wischer & Diewald (eds), 201-215.

Comrie, Bernard, Matthew S. Dryer, David Gil & Martin Haspelmath. 2013. *Introduction*. In Dryer & Haspelmath (eds). [wals.info/chapter/s1]

Condamines, Anne (ed). 2005. *Sémantique et corpus*. Paris : Hermès / Lavoisier.

Corbett, Greville G. 2000. *Number*. Cambridge : Cambridge University Press.

Corbett, Greville G. 2005. The canonical approach in typology. In Frajzyngier, Hodges & Rood (eds), 25-49.

Corblin, Francis & Henriëtte De Swart (eds). *Handbook of French semantics*. Stanford : CSLI.

Coseriu, Eugenio. 1958. *Sincronía, diacronía e historia. El problema del cambio lingüístico*, Montevideo : Universidad de la República.

Coseriu, Eugenio. 1967. Georg von der Gabelentz et la linguistique synchronique, *Word* 23(1-3), 74-100.

Coseriu, Eugenio. 1970. Über Leistung und Grenzen der kontrastiven Grammatik. In *Probleme der kontrastiven Grammatik. Jahrbuch 1969 des IDS*. Düsseldorf. 9-30.

Coseriu, Eugenio. 1997 [1981]. Typologie des langues romanes. In Rabassa & Roché (eds), 217-251. [*Cahiers Ferdinand de Saussure* 35, 3-16].

Croft, William, Jóhanna Barðdal, Willem Hollmann, Violeta Sotirova & Chiaki Taoka. 2010. Revising Talmy's typological classification of complex event constructions. In Boas (ed), 201-236.

Croft, William. 2000. Parts of speech as language universals and as languageparticular categories. In Vogel & Comrie (eds), 65-102.

Croft, William. 2003 [1990]. *Typology and Universals*. Cambridge : Cambridge University Press.

Cumming, Geoff. 2014. The New Statistics: Why and How, *Psychological Science* 25(1), 7-29.

Cuyckens, Hubert, Walter De Mulder & Tanja Mortelmans (eds). 2005. *Adpositions of movement*. Amsterdam / Philadelphie : John Benjamins.*

D'Achille, Paolo, Livio Gaeta, Michela Cennamo, Claudio Iacobini & Davide Ricca. 2014. Corpus *Midia* : *Morfologia dell'Italiano in Diacronia*. www.corpusmidia.unito.it

D'Angiulli, Amedeo, Gordon Griffiths & Fernando Marmolejo-Ramos. 2015. Neural correlates of visualizations of concrete and abstract words in preschool children: a developmental embodied approach, *Frontiers in Psychology* 6, 856.

Dahmen, Wolfgang, Günter Holtus, Johannes Kramer, Michael Metzeltin, Wolfgang Schweickard & Otto Winkelmann (eds). 2006. *Was kann eine vergleichende romanische Sprachwissenschaft heute (noch) leisten? Romanistisches Kolloquium XX*. Tübingen : Narr.

Dascălu Jinga, Laurenţia. 2002. *Corpus de română vorbită (CORV): eşantioane*. Bucarest : Oscar Print.

Davidse, Kristin, Lieven Vandelanotte & Hubert Cuyckens (eds). 2010. *Subjectification, Intersubjectification and Grammaticalization*. Berlin / New York : Mouton de Gruyter.

Davies, Mark & Michael Ferreira. 2006. *Corpus do Português: 45 million words, 1300s-1900s*. [www.corpusdoportugues.org]

Davies, Mark. 2002. *Corpus del Español: 100 million words, 1200s-1900s.* [www.corpusdelespanol.org]

De Mauro, Tullio, Federico Mancini, Massimo Vedovelli & Miriam Voghera. 1993. Lessico di frequenza dell'italiano parlato, *Italica* 71(4), 577-579.

De Mulder, Walter. 2001. La linguistique diachronique, les études sur la grammaticalisation et la sémantique du prototype : présentation, *Langue française* 130, 8-32.

Debaisieux, Jeanne-Marie. 2002. Le fonctionnement de *parce que* en français parlé : étude quantitative sur corpus. In Pusch & Raible (eds), 349-376.

Degand, Liesbeth, Paola Pietrandrea & Bert Cornillie (eds). 2013. *Discourse markers and modal particles. Categorization and Description*. Amsterdam / New York : John Benjamins.

Dejours, Christophe. 2014. Evaluation individualisée des performances et tournant gestionnaire. France culture, La grande table (2ème partie), 3/5. Comment refaire Histoire ? Le collectif à l'épreuve de l'individualisme. [www.franceculture.fr/player/export-reecouter?content=4769410]

Deleuze, Gilles. 1986. *Foucault*. Paris : Editions de minuit.

Denzin, Norman K. 2012. Triangulation 2.0, *Journal of Mixed Methods Research* 6(2), 80-88.

Desmet, Piet & Pierre Swiggers. 1995. *De la grammaire comparée à la sémantique. Textes de Michel Bréal publiés entre 1864 et 1898.* Leuven / Paris : Peeters.

Di Girolamo, Costanzo (ed). 1999. *Repertorio informatizzato dell'antica lirica catalana*. [www.rialc.unina.it]

Di Sciullo, Anna Maria (ed). 1996. *Configurations. Essays on Structure and Interpretation*. Somerville : Cascadilla Press.

Diessel, Holger. 2007. Frequency effects in language acquisition, language use, and diachronic change, *New Ideas in Psychology* 25, 108-127.

Diewald, Gabriele. 2011. Pragmaticalization (defined) as grammaticalization of discourse functions, *Linguistics* 49(2), 365-390.

Diewald, Gabriele. 2013. Same same but different – Modal particles, discourse markers and the art (and purpose) of categorization. In Degand, Pietrandrea & Cornillie (eds), 19-46.

Dirven, René & Günter Radden (eds). 1987. *Concept of Case*. Tübingen : Narr.

Dister, Anne, Dominique Longrée & Gérald Purnelle (eds). 2012. *Actes des 11es Journées internationales d'Analyse statistique des Données Textuelles (JADT)*. Liège / Bruxelles : Université de Liège.

Dister, Anne, Francard, Michel, Hambye, Philippe & Simon, Anne Catherine. 2009. Du corpus à la banque de données. Du son, des textes et des métadonnées. L'évolution de banque de données textuelles orales VALIBEL (1989-2009), *Cahiers de Linguistique* 33(2), 113-129.

Dostie, Gaétane (ed). 2006. *CFPQ (Corpus de français parlé au Québec)*. Sherbrooke : CATIFQ, Université de Sherbrooke. [recherche.flsh.usherbrooke.ca/cfpq]

Dostie, Gaétane. 2004. *Pragmaticalisation et Marqueurs Discursifs. Analyse Sémantique et Traitement Lexicographique*. Bruxelles : De Boeck, Duculot.

Dryer, Matthew S. & Martin Haspelmath (eds). 2011. *The World Atlas of Language Structures Online*. Leipzig : Max Planck Institute for Evolutionary Anthropology. [wals.info]

Dryer, Matthew S. 1998. Why Statistical Universals are Better than Absolute Universals. *Papers from the 33rd Annual Meeting of the Chicago Linguistics Society*. Chicago : Chicago Linguistic Society.

Dryer, Matthew S. 2006. Descriptive theories, explanatory theories, and basic linguistic theory. In Ameka, Dench & Evans (eds), 207-234.

Duchesne, Sophie & Alain Trautmann. 2014. Non à l'instrumentalisation de la connaissance. *Le Monde*, mercredi 3 septembre.

Dunn, Michael, Simon J. Greenhill, Stephen C. Levinson & Russell D. Gray. 2011. *Evolved structure of language shows lineage-specific trends in word-order universals*, *Nature* 473, 79-82.

Durand, Jacques, Bernard Laks & Chantal Lyche (eds). 2009. *Phonologie, variation et accents du français*. Paris : Hermès.

Durand, Jacques, Bernard Laks & Chantal Lyche. 2002. La phonologie du français contemporain : usages, variétés et structure. In Pusch & Raible (eds), 93-106.

Durand, Jacques, Bernard Laks & Chantal Lyche. 2009. Le projet PFC : une source de données primaires structurées. In Durand, Laks & Lyche (eds), 19-61.

Dutta, Sujay & Chris Pullig. 2015. A commentary on reporting effect size and confidence intervals : Response to Palmer and Strelan (2014), *Journal of Business Research* 68, 1082-1085.

Dyen, Isidore, Joseph B. Kruskal & Paul Black. 1992. An Indoeuropean Classification: A Lexicostatistical Experiment, *Transactions of the American Philosophical Society New Series* 82(5), iii-iv & 1-132.

Emmorey, Karen (ed). 2003. *Perspectives on Classifier Constructions in Sign Language*. Mahwah : Erlbaum.

Erman, Britt & Ulla-Britt Kotsinas. 1993. Pragmaticalization: The case of *ba'* and *you know*, *Studier i Modern Språkvetenskap* 10, 76-93.

Evans, Nicholas & Stephen Levinson. 2009. The myth of language universals: Language diversity and its importance for cognitive science, *Behavioral and Brain Sciences* 32, 429-492.

Evans, Vyvyan & Paul Chilton (eds). 2010. *Language, Cognition and Space*. Londres / Oakville : Equinox.

Evans, Vyvyan. 2010. From the spatial to the non-spatial: the 'state' lexical concepts of *in*, *on* and *at*. In Evans & Chilton (eds), 215-248.

Featherston, Sam & Wolfgang Sternefeld (eds). 2007. *Roots: linguistics in search of its evidential base*. Berlin / New York : Mouton de Gruyter.

Fillmore, Charles J. 1992. 'Corpus linguistics' or 'Computer-aided armchair linguistics'. In Svartvik (ed), *Directions in Corpus Linguistics. Proceedings of Nobel Symposium 82.* Stockholm, 4-8 August 1991, 35-59.

Fillmore, Charles. 1972. How to know whether you're coming or going. In Hyldgaard-Jensen (ed), *Linguistik 1971 : Referate des 6.linguistischen Kolloquiums, 11.-14. August 1971 in Kopenhagen*, 369-379.

Fillmore, Charles. 1997. *Lectures on deixis*. Stanford : CSLI Publications.

Fischer, Olga, Anette Rosenbach & Dieter Stein (eds). 2000. *Pathways of change. Grammaticalization in English*. Amsterdam / Philadelphie : John Benjamins.

Fischer, Olga, Muriel Norde & Harry Perridon (eds). 2004. *Up and Down the Cline*. Amsterdam / Philadelphie : John Benjamins.

Fischer, Olga. 2010. On Problem Areas in Grammaticalization: Lehmann's Parameters and the Issue of Scope. In van Linden et al. (eds), 17-42.

Flecken, Monique, Christiane Von Stutterheim & Mary Carroll. 2013. Principles of information organization in L2 use: Complex patterns of conceptual transfer, *International Review of Applied Linguistics in Language Teaching* 51(2), 229-242.

Fort, Karën. 2012. *Les ressources annotées, un enjeu pour l'analyse du contenu : vers une méthodologie de l'annotation manuelle de corpus*. Thèse, Université Paris 13 – Sorbonne Paris Cité.

Fortis, Jean-Michel. 2010. De l'hypothèse de Sapir-Whorf au prototype : sources et genèse de la théorie d'Eleanor Rosch, *Corela* 8(2).

Fortis, Jean-Michel. 2013. Localisme et théorie des cas. In Archaimbault, Fournier & Raby (eds), 75-85.

Frajzyngier, Zygmunt, David Hodges & Adam Rood (eds). 2005. *Linguistic Diversity and Language Theories*. Amsterdam / Philadelphie : John Benjamins.

Francard, Michel, Régine Geron & Geneviève Wilmet. 2002. La banque de données VALIBEL : des ressources textuelles orales pour l'étude du français en Wallonie et à Bruxelles. In Pusch & Raible (eds), 71-80.

François, Jacques. 2004. Le fonctionnalisme linguistique et les enjeux cognitifs. In Fuchs (ed), 99-133.

Frank, Roslyn M., René Dirven, Tom Ziemke & Enrique Bernárdez (eds). 2008. *Body, Language and Mind*, vol. 2: *Sociocultural Situatedness*. Berlin / New York : Mouton de Gruyter.

Fuchs, Catherine & Stéphane Robert (eds). 1999. *Language Diversity and Cognitive Representations*. Amsterdam / Philadelhie : John Benjamins.

Fuchs, Catherine (ed). 2004. *La linguistique cognitive*. Paris : Presses de la Maison des Sciences de l'Homme.

Fuchs, Catherine. 1999. Diversity in linguistic representations: a challenge for cognition. In Fuchs & Robert (eds), 3-20.

Fürbeth, Frank (ed). 1999. *Zur Geschichte und Problematik der Nationalphilologien in Europa*. Tübingen : Niemeyer.

Ganten, Detlev, Volker Gerhardt, Jan-Christoph Heilinger & Julian Nida-Rümelin (eds). 2008. *Was ist der Mensch?* Berlin / New York : De Gruyter.

Gauchat, Louis. 1905. L'unité phonétique dans le patois d'une commune. In *Aus Romanischen Sprachen und Literaturen: Festschrift Heinrich Mort*, Halle : Max Niemeyer, 175-232.

Geeraerts, Dirk & Gitte Kristiansen. 2014. Cognitive Linguistics and language variation. In Littlemore & Taylor (eds), 202-218.

Gennari, Silvia, Steven Sloman, Barbara Malt & Tecumseh Fitch. 2002. Motion events in language and cognition, *Cognition* 83, 49-79.

Gibbs, Raymond W. Jr & Nicole L. Wilson. 2002. Bodily action and metaphorical meaning, *Style* 36(2), 524-540.

Gibbs, Raymond W. Jr. 2003. Embodied experience and linguistic meaning, *Brain and Language* 84, 1-15.

Goddard, Cliff. 2001. Universal units in the lexicon. In Haspelmath, König, Österreicher & Raible (eds), 1190-1023.

Goldberg, Adele E. 1995. *Constructions: A construction grammar approach to argument structure.* Chicago / Londres : University of Chicago Press.

Goschler, Juliana & Anatol Stefanowitsch (eds). 2013. *Variation and Change in the Encoding of Motion Events*. Amsterdam / Philadelphie : John Benjamins.

Goyens, Michèle, Béatrice Lamiroy & Ludo Melis. 2002. Déplacement et repositionnement de la préposition *à* en français, *Lingvisticae Investigationes* 25, 275-310.

Gray, Russell D. & Quentin D. Atkinson. 2003. Language-tree divergence times support the Anatolian theory of Indo-European origin, *Nature* 426(6965), 435-439.

Greenberg, Joseph H. (ed). 1966a. *Universals of Language.* Cambridge : MIT Press.

Greenberg, Joseph H. 1966b. Some universals of grammar with particular reference to the order of meaningful elements. In Greenberg (ed), 73-113.

Gries, Stefan T. & Martin Hilpert. 2008. The identification of stages in diachronic data: variabilitybased neighbor clustering, *Corpora* 3(1), 59-81.

Gries, Stefan T. & Martin Hilpert. 2010. Modeling diachronic change in the third person singular: a multifactorial, verb- and author-specific exploratory approach, *English Language and Linguistics* 14(3), 293-320.

Grinevald, Colette. 2011. On constructing a working typology of the expression of *path*. *Cahiers de Faits de Langues* 3, 43-70.

Groussier, Marie-Line. 1997. Prépositions et primarité du spatial : de l'expression de relations dans l'espace à l'expression de relations non spatiales, *Faits de Langues* 9, 221-234.

Guillaume, Gustave. 1973. *Leçons de Gustave Guillaume, 1948-49, Série C.* Paris : Klincksieck-Laval.

Guillot, Céline, Bernard Combettes, Alexei Lavrentiev, Evelyne Oppermann-Marsaux & Sophie Prévost (eds). 2012. *Le changement en français. Etudes de linguistique diachronique*. Berne : Peter Lang.

Guillot, Céline, Christiane Marchello-Nizia & Alexei Lavrentiev. 2007. La Base de Français Médiéval (BFM) : états et perspectives. In Kunstmann & Stein (eds), 143-152.

Gumperz, John J. & Stephen C. Levinson (eds). 1996. *Rethinking linguistic relativity*. Cambridge : Cambridge University Press.

Guo, Jiansheng, Elena Lieven, Susan Ervin-Tripp, Nancy Budwig, Seyda Özçaliúkan & Keiko Nakamura (eds). 2009. *Crosslinguistic approaches to the psychology of language. Research in the tradition of Dan Isaac Slobin*. New York : Taylor & Francis.

Hampe, Beate (ed). 2005a. *From Perception to Meaning. Image Schemas in Cognitive Linguistics*. Berlin / New York : Mouton de Gruyter.

Hampe, Beate. 2005b. Image schemas in Cognitive Linguistics: Introduction. In Hampe (ed), 1-12.

Hansen, Anita Berit & Isabelle Malderez. 2004. Le *ne* de négation en région parisienne : une étude en temps réel, *Langage et société* 107, 5-30.

Hansen, Maj-Britt Mosegaard. 1996. *The function of discourse particles. A study with special reference to spoken standard French*. Thèse de doctorat. Université de Copenhague.

Hansen, Maj-Britt Mosegaard. 1998. *The function of discourse particles.* Amsterdam / Philadelphie : John Benjamins.

Haroche, Claudine. 2010. L'inévaluable dans une société de défiance, *Cahiers internationaux de sociologie* 1(128-129) : 53-78.

Hartmann, Iren, Martin Haspelmath & Michael Cysouw. 2014. Identifying semantic role clusters and alignment types via microrole coexpression tendencies. *Studies in Language* 38(3), 463-484.

Haspelmath, Martin, Ekkehard König, Wulf Österreicher & Wolfgang Raible (eds). 2001. *Language typology and language universals : an international handbook*. Berlin / New York : De Gruyter.

Haspelmath, Martin, Matthew S. Dryer, David Gil & Bernard Comrie (eds). 2005. *The World Atlas of Language Structures*. Oxford : Oxford University Press.

Haspelmath, Martin. 1997. *From space to time. Temporal adverbials in the world's languages*. München / Newcastle : Lincom Europa.

Haspelmath, Martin. 2010. The interplay between comparative concepts and descriptive categories (Reply to Newmeyer), *Language* 86(3), 696-699.

Heiden, Serge & Alexei Lavrentiev. 2012. The TXM Portal Software giving access to Old French Manuscripts Online. In *Proceedings of the 1st Workshop on Adaptation of Language Resources and Tools for Processing Cultural Heritage Objects, Seventh International Conference on Language Resources and Evaluation (ELRA), Istanbul, Turkey*, 29-35.

Heiden, Serge, Jean-Philippe Magué & Bénédicte Pincemin. 2010. TXM : Une plateforme logicielle open-source pour la textométrie – conception et développement. In Bolasco, Chiari & Giuliano (eds), *Statistical Analysis of Textual Data-Proceedings of 10th International Conference JADT 2010, Rome, 9-11 juin 2010*, 1021-1032.

Heiden, Serge. 2010. The TXM Platform: Building Open-Source Textual Analysis Software Compatible with the TEI Encoding Scheme. In Otoguro et al. (eds), *Proceedings of the 24th Pacific Asia Conference on Language, Information and Computation (PACLIC 24). 4-7 November 2010, Sendai*, 389-398.

Heine, Bernd & Tania Kuteva. 2002. *World Lexicon of Grammaticalization.* Cambridge : Cambridge University Press.

Heine, Bernd, Ulrike Claudi & Friederike Hunnemeyer (eds). 1991. *Grammaticalization: A Conceptual Framework*. Chicago : University of Chicago Press.

Heine, Bernd. 1989. Adpositions in African languages, *Linguistique Africaine* 2, 77-127.

Heine, Bernd. 1997a. *Possession. Cognitive sources, forces and grammaticalization*. Oxford : Oxford University Press.

Heine, Bernd. 1997b. *Cognitive foundations of grammar*. Oxford : Oxford University Press.

Heine, Bernd. 2002. On the role of context in grammaticalization. In Wischer & Diewald (eds), 83-101.

Heine, Bernd. 2003. Grammaticalization. In Joseph & Janda (eds), 575-601.

Heine, Bernd. 2013. On discourse markers: Grammaticalization, pragmaticalization, or something else?, *Linguistics* 51(6), 1205-1247.

Herdan, Gustav. 1964. *Quantitative linguistics*. Oxford : Butterworth.

Hermann, Eduard. 1929. Lautveränderungen in der Individualsprache einer Mundart, *Nachrichten der Gesellschaft der Wissenschaften zu Göttingen* 9, 195-214.

Herschensohn, Julia (ed). 2010. *Romance Linguistics 2010. Selected papers from the 40th Linguistics Symposium on Romance Languages (LSRL), Seattle, Washington March 2010*. Amsterdam / Philadelphie : John Benjamins.

Herskovits, Annette. 1982. *Space and the prepositions in English: regularities and irregularities in a complex domain*. Thèse de doctorat, Université de Stanford.

Herskovits, Annette. 1986. *Language and spatial cognition*. Cambridge : Cambridge University Press.

Herskovits, Annette. 1988. Spatial expressions and the plasticity of meaning. In Rudzka-Ostyn (ed), 271-298.

Hickey, Raymond (ed). 2003. *Motives for language change*. Cambridge : Cambridge University Press.

Hickmann, Maya & Stéphane Robert. 2006. *Space in languages: linguistic systems and cognitive categories.* Amsterdam / Philadelphie : John Benjamins.

Hilpert, Martin & Stefan T. Gries. 2009. Assessing frequency changes in multistage diachronic corpora: Applications for historical corpus linguistics and the study of language acquisition, *Literary and Linguistic Computing* 24(4), 385-401.

Hilpert, Martin. 2013 : *Constructional change in English: Developments in Allomorphy, Word Formation, and Syntax*. Cambridge : Cambridge University Press.

Hjelmslev, Louis. 1935-1937. Catégorie des cas, *Acta Jutlandica* VII/1, xii-184 & IX/2, viii-78.

Hockett, Charles F. 1950. Age-grading and linguistic continuity, *Language* 26, 449-459.

Hoffmann, Sebastian. 2005. *Grammaticalization and English Complex Prepositions. A corpus-based analysis*. Londres / New York : Routledge.

Hoffmann, Thomas & Graeme Trousdale (eds). 2013. *The Oxford Handbook of Construction Grammar*. Oxford : Oxford University Press.

Hopper, Paul J. & Elizabeth C. Traugott. 2003 [1993]. *Grammaticalization*. Cambridge : Cambridge University Press.

Hopper, Paul J. 1987. Emergent Grammar. In *Proceedings of the Thirteenth Annual Meeting of the Berkeley Linguistics Society*, 139-157.

Houdé, Olivier, Bernard Mazoyer & Nathalie Tzourio-Mazoyer (eds). 2002. *Cerveau et psychologie*. Paris : PUF.

Hurford, James R. 1991. The Evolution of the Critical Period for Language Acquisition, *Cognition* 40(3), 159-201.

Huumo, Tuomas. 2014. Path Settings: How dynamic conceptualization permits the use of path expressions as setting adverbials. In Sarda et al. (eds), 69-96.

Hyldgaard-Jensen, Karl (ed). 1972. *Linguistik 1971: Referate des 6. Linguistischen Kolloquiums 11.-14. August 1971 in Kopenhagen*. Francfort : Athenäum.

Iacobini, Claudio & Federica Masini. 2006. The emergence of verb-particle constructions in Italian: locative and actional meanings, *Morphology* 16, 155-188.

Ibarretxe-Antuñano, Iraide. 2009. Path salience in motion events. In Guo et al. (eds), 403-414.

Ibarretxe-Antuñano, Iraide. 2010. Static topological relations in Basque. In Evans & Chilton (eds), 251-265.

Ikegami, Yoshihiko. 1987. Source *vs.* Goal: A Case of Linguistic Dissymmetry. In Dirven & Radden (eds), 292-308.

Ishibashi, Miyuki. 2012. Expression of 'putting' and 'taking' events in Japanese: 'Asymmetry of Source and Goal' revisited. In Kopecka & Narasimhan (eds), 253-272.

Ishibashi, Miyuki, Anetta Kopecka & Marine Vuillermet. 2006. *Trajectoire : Matériel visuel pour élicitation des données linguistiques*. Laboratoire Dynamique du Langage (CNRS / Université Lyon 2) – Fédération de Recherche en Typologie et Universaux Linguistiques, CNRS, France.

Itkonen, Esa. 2005. *Ten Non-European Languages: An Aid to the Typologist*. Turku : Université de Turku.

Iverson, Jana M. & Esther Thelen. 1999. Hand, mouth, and brain: The dynamic emergence of speech and gesture, *Journal of Consciousness Studies* 6, 19-40.

Jackendoff, Ray. 1983. *Semantics and Cognition*. Cambridge : MIT Press.

Jäger, Gerhard. 2003. Simulating language change with Functional OT. In Kirby (ed), 52-61.

Jakobson, Roman. 1963-1973. *Essais de linguistique générale* (1-2). Paris : Editions de Minuit.

Janda, Richard D. 2001. Beyond 'pathways' and 'unidirectionality': on the discontinuity of language transmission and the counterability of grammaticalization, *Language Sciences* 23(2-3), 265-340.

Jick, Todd D. 1979. Mixing qualitative and quantitative methods: triangulation in action, *Administrative Science Quarterly* 24, 602-662.

Johnson, Mark. 1987. *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason*. Chicago : University of Chicago Press.

Johnston, Judith R. & Dan I. Slobin. 1979. The development of locative expressions in English, Italian, Serbo-Croatian and Turkish, *Journal of Child Language* 6, 529-545.

Joseph, Brian D. & Richard D. Janda (eds). 2003. *The Handbook of Historical Linguistics*. Oxford : Blackwell.

Kapuściński, Ryszard. 2009. *Cet Autre*. Paris : Plon.

Kelle, Udo. 2006. Combining qualitative and quantitative methods in research practice: purposes and advantages, *Qualitative Research in Psychology* 3, 293-311.

Kennedy, Graeme. 2002. Variation in the distribution of modal verbs in the British National Corpus. In Reppen, Fitzmaurice & Biber (eds), 73-90.

Kerswill, Paul. 1996. Children, adolescents, and language change, *Language Variation and Change* 8(2), 177-202.

Kiefer, Ferenc & Nicolas Ruwet (eds). 1973. *Generative grammar in Europe*. Dordrecht : Reidel.

Kilgariff, Adam. 2005. Language is never, ever, ever, random. *Corpus Linguistics and Linguistic Theory* 1(2), 263-275.

Kilgarriff, Adam, Pavel Rychly, Pavel Smrz & David Tugwell. 2004. The Sketch Engine. In *Proceedings of EURALEX*, 110-122.

Kirby, Simon. 2001. Spontaneous evolution of linguistic structure-an iterated learning model of the emergence of regularity and irregularity, *IEEE Transactions on Evolutionary Computation* 5(2), 102-110.

Kirby, Simon. 2003 (ed). *Language Evolution and Computation*. Vienne : ESSLLI Workshop.

Koch, Peter. 1997b. La diacronia quale campo empirico della semantica cognitiva. In Carapezza, Gambarara & Lo Piparo (eds), 225-246.

Koch, Peter. 2003. Historical Romance Linguistics and the Cognitive Turn, *La corónica: A Journal of Medieval Hispanic Languages, Literatures, and Cultures* 31(2), 41-55.

Koch, Peter. 2006. Romanische Sprachwissenschaft und diachronische kognitive Linguistik – eine Wahlverwandtschaft? In Dahmen et al. (eds), 101-136.

Koch, Peter. 2012a. Location, existence, and possession: A constructional-typological exploration, *Linguistics* 50(3), 533-603.

Koch, Peter. 2012b. Changement lexical et constructionnalisation dans le domaine du verbe. In *Actes du CMLF 2012*, SHS Web of Conferences 1, 43-80. [dx.doi.org/10.1051/shsconf/20120100341]

Koehn, Philipp. 2005. Europarl : A Parallel Corpus for Statistical Machine Translation. In *Proceedings of the 10th Machine Translation Summit. Phuket, Thailand*, 19-86.

Kopecka, Anetta & Bhuvana Narasimhan (eds). 2012. *Put and Take events: A crosslinguistic perspective*. Amsterdam / Philadelphie : John Benjamins.

Kopecka, Anetta & Miyuki Ishibashi. 2011. L'(a)symétrie dans l'expression de la Source et du But : perspective translinguistique. *Cahiers de Faits de Langues* 3, 131-149.

Kopecka, Anetta. 2006. The semantic structure of motion verbs in French : Typological perspectives. In Hickmann & Robert (eds), 83-101.

Kopecka, Anetta. 2009. L'expression du déplacement en français : l'interaction des facteurs sémantiques, aspectuels et pragmatiques dans la construction du sens spatial, *Langages* 173(1), 54-75.

Kopecka, Anetta. À paraître. From a satellite- to a verb-framed pattern: A typological shift in French. In Cuyckens, De Mulder & Mortelmans (eds), *Variation and Change in Adpositions of Movement*.

Kroch, Anthony. 1989. Reflexes of grammar in patterns of language change, *Language Variation and Change* 1, 199-244.

Kunstmann, Pierre & Achim Stein (eds). 2007. *Le Nouveau Corpus d'Amsterdam. Actes de l'atelier de Lauterbad, 23-26 février 2006*. Stuttgart : Steiner.

Kuryłowicz, Jerzy. 1965. The evolution of grammatical categories, *Diogenes* 13, 55-71.

Labov, William. 1963. The social motivation of a sound change, *Word* 18, 1-42.

Labov, William. 1994. *Principles of linguistic change. Vol. 1. Internal Factors*. Oxford : Blackwell.

Lakoff, George & Mark Johnson. 1999. *Philosophy in the Flesh: The embodied mind and its challenge to Western thought*. New York : Basic Books.

Lakoff, George. 1987. *Women, fire and dangerous things: What categories reveal about the mind*. Chicago : University of Chicago Press.

Lakoff, George. 1990. The invariance hypothesis: Is abstract reason based on image schemas?, *Cognitive Linguistics* 1(1), 39-74.

Lakoff, George. 1993. The contemporary theory of metaphor. In Ortony (ed), 202-251.

Lakusta, Laura & Barbara Landau. 2005. Starting at the end: The importance of goals in spatial language, *Cognition* 96, 1-33.

Landau, Barbara & Ray Jackendoff. 1993. 'What' and 'where' in spatial language and spatial cognition, *Behavioral and Brain Sciences* 16, 217-265.

Landau, Barbara, Banchiamlack Dessalegn & Ariel M. Goldberg. 2010. Language and space: momentary interactions. In Evans & Chilton (eds), 51-77.

Landau, Barbara. 2009. Spatial language, spatial cognition: Origins, development, and interaction. Keynote address. In *Proceedings of the Boston University Conference on Language Development* 33. Somerville : Cascadilla Press, 1-23.

Landau, Barbara. 2012. Space in semantics and cognition: Three themes. In Maienborn, von Heusinger & Portner (eds), 2805-2824.

Langacker, Ronald W. 1987. *Foundations of cognitive grammar*, vol. 1: *Theoretical prerequisites.* Stanford : Stanford University Press.

Langacker, Ronald W. 1991a. Concept, image and symbol. The cognitive basis of grammar. Berlin / New York : Mouton de Gruyter.

Langacker, Ronald W. 1991b. *Foundations of cognitive grammar*, vol.2: *Descriptive application.* Stanford : Stanford University Press.

Langacker, Ronald W. 2010. Reflections on the Functional Characterization of Spatial Prepositions, *Corela, HS-7*. [corela.revues.org/999]

Langacker, Ronald. W. 1987. *Foundations of cognitive grammar*, vol. 1: *Theoretical perspectives.* Stanford : Stanford University Press.

Langacker, Ronald. W. 1991a. *Concept, Image, and Symbol. Cognitive Basis of Grammar.* Berlin / New York : Mouton de Gruyter.

Langacker, Ronald. W. 1997. The contextual basis of cognitive semantics. In Nuyts & Pederson (eds), 229-252.

Lanly, André. 1999, Compte-rendu de Banniard 1997 *Du latin aux langues romanes*, *Cahiers de civilisation médiévale* 42, 199-201.

Lawrence, Peter. 2011. The heart of research is sick. *Lab Times* 2, 24-31.

Lazard, Gilbert. 1992. Y a-t-il des catégories interlangagières ? In Anschütz (ed), 427-434.

Lazard, Gilbert. 2008. Vers une typologie scientifique, *La Linguistique* 44(2), 13-24.

Lazard, Gilbert. 2013. Réflexions séculaires, *La Linguistique* 49(1), 49-65.

Le Querler, Nicole. 2001. La grammaticalisation de *côté* introducteur de topicalisation. In Col & Roulland (eds), 155-180.

Lebel, Marie-Elaine. 2003. Le morphème *par* en ancien français. Thèse de doctorat, Université du Québec à Montréal.

Lehmann, Christian. 1985. Grammaticalization: Synchronic Variation and Diachronic Change, *Lingua E Stile* 20, 303-318.

Lehmann, Christian. 1995 [1982]. *Thoughts on Grammaticalization*. Unterschleissheim : Lincom Europa.

Lehmann, Christian. 2002 [1982]. *Thoughts on grammaticalization*. Erfurt : Seminar für Sprachwissenschaft der Universität.

Levinson, Stephen C. (ed). 1992. *Space stimuli kit* 1.2: November 1992, 51. Nijmegen: Max Planck Institute for Psycholinguistics.

Levinson, Stephen C. & David P. Wilkins (eds). 2006. *Grammars of Space: Explorations in Cognitive Diversity*, Cambridge : Cambridge University Press.

Levinson, Stephen C. 2003. *Space in language and cognition: explorations in cognitive diversity*. Cambridge : Cambridge University Press.

Levinson, Stephen C., Sérgio Meira & The Language and Cognition Group. 2003. 'Natural concepts' in the spatial topological domain – adpositional meanings in crosslinguistic perspective: An exercise in cognitive typology, *Language* 79, 485-516.

Lewandowska-Tomaszczyk, Barbara (ed). 2008. *Asymmetric Events*. Amsterdam / Philadelphie : John Benjamins.

Lewis, Diana. 2011. A discourse-constructional approach to the emergence of discourse markers in English, *Linguistics* 49(2), 415-443.

Lewis, Diana. 2014. Word order and the development of connective markers in English. In Sarda et al. (eds), 229-254.

Liszkowski, Ulf, Penny Brown, Tara Callaghan, Akira Takada & Conny de Vos. 2012. A Prelinguistic Gestural Universal of Human Communication, *Cognitive Science* 36(4), 698-713.

Littlemore, Jeannette & John R. Taylor. 2014. The Bloomsbury Companion to Cognitive Linguistics. Londres : Bloomsbury.

Loiseau, Sylvain (ed). 2015. *Langages* 197, *La fréquence textuelle : Bilan et perspectives*. Paris : Larousse.

Lupyan, Gary & Rick Dale. 2010. Language Structure Is Partly Determined by Social Structure, *PLoS ONE* 5(1): e8559.

Lyons, John. 1977. *Semantics*, vol. 1. Cambridge : Cambridge University Press.

Maienborn, Claudia, Klaus von Heusinger & Paul Portner (eds). 2012. *Semantics: An international handbook of natural language meaning*. Berlin / New York : Mouton de Gruyter.

Mair, Christian, 1995. Changing Patterns of Complementation, and Concomitant Grammaticalisation, of the Verb *Help* in Present-Day British English. In Aarts & Meyer (eds), 258-272.

Majid, Asifa, Fiona Jordan & Michael Dunn. 2014. Semantic systems in closely related languages, *Language sciences* 49, 1-18.

Majid, Asifa, Melissa Bowerman, Sotaro Kita, Daniel B. M. Haun & Stephen C. Levinson. 2004. Can language restructure cognition? The case for space, *Trends in Cognitive Sciences* 8(3), 108-114.

Makel, Matthew C. & Jonathan A. Plucker. 2014. Creativity is more than novelty: Reconsidering replication as a creativity act, *Psychology of Aesthetics, Creativity, and the Arts* 8 : 27-29.

Marchello-Nizia, Christiane. 1995. *L'évolution du français : ordre des mots, démonstratifs, accent tonique*. Paris : Colin.

Marchello-Nizia, Christiane. 2006. *Grammaticalisation et changement linguistique*. Bruxelles : De Boeck.

Marchello-Nizia, Christiane. 2012. L'oral représenté : un accès construit à une face cachée des langues 'mortes'. In Guillot et al. (eds), 247-264.

Marian, Viorica & Caitlin Fausey. 2006. Language-dependent memory in bilingual learning, *Applied Cognitive Psychology* 20(8), 1025-1047.

Marotta, Giovanna, Alessandro Lenci, Linda Meini & Francesco Rovai (eds). 2010. *Space in language*. Pise : Editrice Testi Scientifici.

Martín Zorraquino, Maria Antonia & Estrella Montolío Duran (eds). 1998. *Marcadores del Discurso. Teoría y Análisis*. Madrid : Arco-Libros.

Martinet, André. 1955. *Economie des changements phonétiques. Traité de phonologie diachronique* (Bibliotheca Romanica, Manualia et Commentationes, t. X). Berne : A. Francke.

Marzo, Stefania, Kris Heylen & Gert De Sutter. 2012. *Corpus Studies in Contrastive Linguistics*. Amsterdam / Philadelphie : John Benjamins.

Matlock, Teenie. 2004. Fictive motion as cognitive simulation, *Memory & Cognition* 32(8), 1389-1400.

Matlock, Teenie. 2010. Abstract motion is no longer abstract, *Language and Cognition* 2(2), 243-260.

Matsumoto, Yo. 1996. Subjective Motion and English and Japanese Verbs, *Cognitive Linguistics* 7(2), 183-226.

Mazoyer, Bernard. 2002. *Conclusion – L'imagerie cérébrale fonctionnelle : apports, limites et perspectives.* In Houdé, Mazoyer & Tzourio-Mazoyer (eds), 605-609.

Mazziotta, Nicolas. 2012. Le Syntactic Reference Corpus of Medieval French : Structure, outils et exploitation. In Dister, Longrée & Purnelle (eds), 701-713.

McDonough, Laraine, Soonja Choi & Jean Mandler. 2003. Understanding spatial relations: Flexible infants, lexical adults, *Cognitive Psychology* 46, 229-259.

Meillet, Antoine. 1912. L'évolution des formes grammaticales, *Scientia, Rivista di scienza* 12(26). [= Meillet, Antoine. 1958. *Linguistique Historique et Linguistique générale*. Paris : Champion, 130-149].

Mertens, Donna M. & Sharlene Hesse-Biber. 2012. Triangulation and Mixed Methods Research: Provocative Positions, *Journal of Mixed Methods Research* 6(2), 75-79.

Michaelis, Susanne Maria, Philippe Maurer, Martin Haspelmath, & Magnus Huber. 2013. *The atlas of pidgin and creole language structures*. Oxford : Oxford University Press.

Mithun, Marianne. 1993. Bernd Heine, Ulrike Claudi & Friederieke Hühnemeyer, *Grammaticalization: a conceptual framework*. *Journal of Linguistics* 29(2), 517-521.

Mithun, Marianne. 2012. Tags: Cross-linguistic diversity and commonality, *Journal of Pragmatics* 44, 2165-2182.

Moeschler, Jacques. 1986. Connecteurs pragmatiques, lois de discours et stratégies interprétatives : *parce que* et la justification énonciative, *Cahiers de Linguistique Française* 7, 149-167.

Montoro, Pedro R., María José Contreras, María Rosa Elosúa & Fernando Marmolejo-Ramos. 2015. Cross-modal metaphorical mapping of spoken emotion words onto vertical space, *Frontiers in Psychology* 6, 1205.

Morgenstern, Aliyah, Christophe Parisse & Martine Sekali. 2010. Les premières prépositions chez l'enfant : grammaticalisation de l'espace relationnel, *Faits de Langue* 34, 95-108.

Murata, Akira, Luciano Fadiga, Leonardo Fogassi, Vittorio Gallese, Vassilis Raos & Giacomo Rizzolatti. 1997. Object representation in the ventral premotor cortex (Area F5) of the monkey, *Journal of Neurophysiology* 78(4), 2226-2230.

Murru, Furio. 1979. Sull'origine della teoria localistica di Massimo Planude, *L'antiquité classique* 48(1), 82-97.

Narrog, Heiko & Bernd Heine (eds). 2011. *The Oxford handbook of grammaticalization*. Oxford : Oxford University Press.

Nerlich, Brigitte. 2014. Michel Bréal : mettre l'homme dans la langue. In Archaimbault et al. (eds), 611-619.

Netanda, Rendani S. 2012. Mixed methods-triangulation war: Hidden challenges to their conceptual survival, *Journal of Applied Global Research* 5(14), 45-55.

Nettle, Daniel. 1999. Using Social Impact Theory to simulate language change, *Lingua* 108, 95-117.

Newmeyer, Frederick J. 1998. *Language Form and Language Function*. Cambridge : MIT Press.

Newmeyer, Frederick J. 2001. Deconstructing grammaticalization, *Language Sciences* 23(2-3), 187-229.

Niyogi, Partha. 2002. Theories of culture evolution and their application to language change. In Briscoe (ed), 205-234.

Noailly, Michèle. 2006. Quoi de neuf *côté* préposition ?, *La préposition en français, Modèles linguistiques* 27(1), vol. 53, 75-90.

Nølke, Henning. 1990. Il y a connecteurs et connecteurs : la connexion analytique et synthétique en français moderne. In Stammerjohan (ed), 41-48.

Norde, Muriel. 2009. *Degrammaticalization*. Oxford : Oxford University Press.

Nowak, Martin A. & David C. Krakauer. 1999. The evolution of language, *Proceedings of the National Academy of Sciences* 96(14), 8028-8033.

Nuyts, Jan & Eric Pederson (eds). 1997. *Language and conceptualization*. Cambridge : Cambridge University Press.

Nuyts, Jan. 2013. De-auxiliarization without de-modalization in the Dutch core modals: a case of collective degrammaticalization?, *Language Sciences* 36, 124-133.

O'Keefe, John. 1996. The spatial prepositions in English, Vector grammar and the cognitive map theory. In Bloom, Peterson, Nadel & Garrett (eds), 277-316.

Oakey, David. 2002. Formulaic language in English academic writing. In Reppen, Fitzmaurice & Biber (eds), 111-129.

Ortony, Andrew (ed). 1993 [1979]. *Metaphor and Thought*. Cambridge : Cambridge University Press.

Palmer, Matthew A. & Peter Strelan. 2014. Commentary on Dutta and Pullig: Corrective action is more effective than downplaying harm for restoring brand equity. *Journal of Business Research* 68(6), 1271-1272.

Panther, Klaus-Uwe & Günter Radden (eds). 2004. *Motivation in grammar*. Berlin / New York : Mouton de Gruyter.

Papafragou, Anna, Christine Massey & Lila Gleitman. 2002. Shake, rattle, 'n' roll: The representation of motion in language and cognition, *Cognition* 84, 189-219.

Papafragou, Anna, Justin Hulbert & John Trueswell. 2008. Does language guide event perception? Evidence from eye movements, *Cognition* 108(1), 155-184.

Papafragou, Anna, Peggy Li, Youngon Choi & Chung-hye Han. 2007. Evidentiality in language and cognition, *Cognition* 103, 253-299.

Parret, Herman. 2009. Spatialiser haptiquement, *Actes Sémiotiques* 112. [epublications.unilim.fr/revues/as/2570]

Pashler, Harold & Eric-Jan Wagenmakers. 2012. Editor's Introduction to the Special Section on Replicability in Psychological Science: A Crisis of Confidence?, *Perspectives on Psychological Science* 7(6), 528-530.

Pavlenko, Aneta. 2003. Eyewitness memory in late bilinguals: Evidence for discursive relativity, *The International Journal of Bilingualism* 7(3), 257-281.

Pawlowitsch, Christina, Panayotis Mertikopoulos & Nikolaus Ritt. 2011. Neutral stability, drift, and the diversification of languages, *Journal of theoretical biology* 287, 1-12.

Perea, Maria-Pilar & Joaquim Viaplana. (2008) *Corpus Oral Dialectal (COD). Selecció de textos.* Dipòsit Digital de la UB [Arxiu obert]

Peyraube, Alain. 2004. Grammaire diachronique et cognition : l'exemple du chinois. In Fuchs (ed), 135-154.

Pick, Herbert L. Jr & Linda P. Acredolo (eds). 1983. *Spatial Orientation: Theory, Research, and Application*. New York : Plenum Press.

Pitts-Taylor, Victoria. 2015. A Feminist Carnal Sociology? Embodiment in Sociology, Feminism, and Naturalized Philosophy, *Qualitative Sociology* 38(1), 19-25.

Pons Bordería, Salvador. 1998. Los apelativos *oye* y *mira* o los límites de la conexión. In Martín Zorraquino & Montolío Duran (eds), 213-228.

Pottier, Bernard. 2001. *Représentations mentales et catégorisations linguistiques.* Louvain / Paris : Peeters.

Prévost, Sophie. 2003. La grammaticalisation : unidirectionnalité et statut, *Le Français Moderne* LXXI(2), 144-166.

Prévost, Sophie. 2005. Exploitation d'un corpus de français médiéval : enjeux, spécificités et apports. In Condamines (ed), 147-176.

Prévost, Sophie. 2006. Grammaticalisation, lexicalisation et dégrammaticalisation : des relations complexes, *Cahiers de praxématique* 46, 121-140.

Prévost, Sophie. 2011a. *Français médiéval en diachronie : du corpus à la langue*. Habilitation à Diriger des Recherches – Mémoire de synthèse.

Prévost, Sophie. 2011b. *Expression et position du sujet pronominal du 12ème au 14ème siècle : une approche quantitative*. Habilitation à Diriger des Recherches – Mémoire de recherche inédite.

Prévost, Sophie. 2011c. *À propos* from verbal complement to discourse marker: a case of grammaticalization?, *Linguistics* 49(2), 391-413.

Prosser, James I. 2010. Replicate or lie. *Environmental Microbiology* 12(7), 1806-1810.

Puglisi, Andrea, Andrea Baronchelli & Vittorio Loreto. 2008. Cultural route to the emergence of linguistic categories, *Proceedings of the National Academy of Sciences* 105(23), 7936-7940.

Pusch, Claus D. & Wolfgang Raible (eds). 2002. *Romanistische Korpuslinguistik – Korpora und gesprochene Sprache*, *Romance Corpus Linguistics, Corpora and Spoken Language*. Tübingen : Narr.

Queller, Kurt. 2008. Toward a socially situated, functionally embodied lexical semantics: the case of *(all) over*. In (eds), 265-300.

Rabassa, Lídia & Michel Roché (eds). 1997. *Variation linguistique et enseignement des langues. Langue parlée, langue écrite.* Toulouse : Cerclid.

RAE 2009 : Real Academia Española, 2009. *Nueva gramática de la lengua española*. Madrid : Espasa.

Raible, Wolfgang. 2001. Language universals and language typology. In Haspelmath et al. (eds), 1-25.

Rainsford, Tom, Céline Guillot, Alexei Lavrentiev & Sophie Prévost. 2012. La zone préverbale en ancien français : apport des corpus annotés. In Neveu et al. (eds), *Actes du 3e Congrès Actes du 3e Congrès Mondial de Linguistique Française (CMLF), Lyon, July 2012*, Paris: Institut de Linguistique française, 159-176.

Ramsay, John. 2007. The dangers of methodological extremism, *Journal of Purchasing & Supply Management* 13, 202-203.

Rauh, Gisa. 1993. On the grammar of lexical and non-lexical prepositions in English. In Zelinsky-Wibbelt (ed), 99-150.

Reppen, Randi, Susan M. Fitzmaurice & Douglas Biber. 2002. *Using Corpora to Explore Linguistic Variation*. Amsterdam / Philadelphie : John Benjamins.

Rheinländer, Kathrin. 2011. Triangulation. Wissenschaftshistorische und methodologische Aspekte aus der Perspektive der sozialwissenschaftlichen Forschung, *Zeitschrift für Qualitative Forschung* 12(1), 111-123.

Ricca, Davide. 1993. *I verbi deittici di movimento in Europa: una ricerca interlinguistica.* Florence : La Nuova Italia.

Richardson, Daniel C. & Teenie Matlock. 2007. The integration of figurative language and static depictions: An eye movement study of fictive motion, *Cognition* 102, 129-138.

Riegel, Martin, Jean-Christophe Pellat & René Rioul. 2001. *Grammaire méthodique du français*. Paris : PUF.

Rohrer, Tim. 2005. Image Schemata in the Brain. In Hampe (ed), 165-196.

Rudzka-Ostyn, Brygida (ed). 1988. *Topics in Cognitive Linguistics*. Amsterdam / Philadelphie : John Benjamins.

Ruxăndoiu, Liliana. 2002. *Interacţiunea verbală în limba română actuală: corpus (selectiv) : schiţă de tipologie*. Bucarest : Ed. Univ. Bucureşti.

Sabah, Gérard (ed). 2006. *Compréhension des langues et interaction*. Paris : Hermès-Lavoisier.

Sarda, Laure & Michel Charolles. 2009. *Parallèlement* : de l'espace au temps puis à l'énonciation. In Servet & Boissiéras (eds), 127-156.

Sarda, Laure & Michel Charolles (eds). 2012. *Les adverbiaux prépositionnels : position, fonction et portée. Travaux de linguistique*, Numéro thématique, 64(1).

Savalei, Victoria & Elizabeth Dunn. 2015. Is the call to abandon *p*-values the red herring of the replicability crisis?, *Frontiers in Psychology* 6, 245.

Schlieben-Lange, Brigitte. 1999. Die deutsche Romanistik – ein Modell für die Zukunft? In Fürbeth (ed), 847-854.

Schmid, Hans-Jörg (ed). 2012. *Handbooks of Pragmatics*, vol. 4. Berlin / New York : Mouton De Gruyter.

Schmid, Hans-Jörg (ed). 2012. *Handbooks of Pragmatics,* vol. 4. Berlin / New York : Mouton De Gruyter.

Schøsler, Lene. 2008. L'expression des traits manière et direction des verbes de mouvement. Perspectives diachroniques et typologiques. In Stark, Schmidt-Riese & Stoll (eds), 113-132.

Schultze-Berndt, Eva. 2000. *Simple and complex verbs in Jaminjung. A study of event categorisation in an Australian language*. Thèse de doctorat, Université de Nimègue.

Schultze-Berndt, Eva. 2006. Sketch of a Jaminjung grammar of space. In Levinson & Wilkins (eds), 63-114.

Servet, Marie-Hélène & Fabienne Boissiéras (eds). 2009. *Hiérarchisation, énonciation*. Louvain/Paris : Peeters.

Shibatani, Masayoshi & Sandra A. Thompson (eds). 1996. *Grammatical constructions: Their form and meaning*. Oxford : Clarendon Press.

Shopen, Timothy (ed). 1985. *Language typology and syntactic description*, vol. 3 : *Grammatical categories and the lexicon*. Cambridge : Cambridge University Press.

Siewierska, Anna. 2004. Language Typology and Language Universals: An International Handbook by Martin Haspelmath; Ekkehard König; Wulf Oesterreicher; Wolfgang Raible. Review, *Journal of Linguistics* 40(3), 683-687.

Simon, Anne-Catherine & Liesbeth Degand. 2007. Connecteurs de causalité, implication du locuteur et profils prosodiques. Le cas de *car* et de *parce que*, *Journal of French Language Studies* 17, 323-341.

Slobin, Dan I. & Nini Hoiting. 1994. Reference to movement in spoken and signed languages: typological considerations. In *Proceedings of the Twentieth Annual Meeting of the Berkeley Linguistic Society*, 487-505.

Slobin, Dan I. 1985. *The crosslinguistic study of language acquisition*, vol. 2. Hillsdale : Erlbaum.

Slobin, Dan I. 1996a. Two ways to travel : verbs of motion in English and Spanish. In Shibatani & Thompson (eds), 195-220.

Slobin, Dan I. 1996b. From "thought and language" to "thinking for speaking". In Gumperz & Levinson (eds), 70-96.

Slobin, Dan I. 2004. The many ways to search for a frog : linguistic typology and the expression of motion events. In Strömqvist & Verhoeven (eds), 219-257. [ihd.berkeley.edu/linguistictypologyofmotionevents.pdf]

Smith, Jeffrey K. & Lisa F. Smith. 2014. Replicability? Not That Again! *Psychology of Aesthetics, Creativity, and the Arts* 8(1), 21-23.

Soroli, Efstathia, & Maya Hickmann. 2010. Language and spatial representations in French and in English: Evidence from eye-movements. In Marotta et al. (eds), 581-597.

Stammerjohan, Harro (ed). 1990. *Analyse et synthèse dans les langues slaves et romanes*. Tübingen : Narr.

Stark, Elisabeth, Roland Schmidt-Riese & Eva Stoll (eds). 2008. *Romanische Syntax im Wandel*. Tübingen : Narr.

Steels, Luc. 2005. The emergence and evolution of linguistic structure: from lexical to grammatical communication systems, *Connection science* 17(3-4), 213-230.

Stefanowitsch, Anatol & Rohde, Ada. 2004. The goal bias in the encoding of events. In Panther & Radden (eds), 249-268. [www-user.uni-bremen.de/~anatol/publications.html]

Stein, Achim & Sophie Prévost. 2013. Syntactic annotation of medieval texts: the Syntactic Reference Corpus of Medieval French (SRCMF). In Bennett et al. (eds), 275-282.

Stein, Achim & Sophie Prévost. 2013. Syntactic annotation of medieval texts: the Syntactic Reference Corpus of Medieval French (SRCMF). In Bennett et al. (eds), 275-282.

Stolova, Natalya I. 2015. *Cognitive Linguistics and Lexical Change – Motion Verbs from Latin to Romance*. Amsterdam / Philadelphie : John Benjamins.

Stowe, Laurie A., Marco Haverkort & Frans Zwarts. 2004. Rethinking the neurological basis of language, *Lingua* 115, 997-1042.

Strömqvist, Sven & Ludo Verhoeven (eds). 2004. *Relating events in narratives*, vol. 2 : *Typological and contextual perspectives*. Mahwah : Erlbaum.

Svorou, Soteria. 1994. *The grammar of space*. Amsterdam / Philadelphie : John Benjamins.

Swan, Toril & Olaf Jansen Westvik (eds). 1997. *Modality in Germanic languages. Historical and comparative perspectives*. Berlin / New York : Mouton de Gruyter.

Sweetser, Eve. 1990. *From etymology to pragmatics. Metaphorical and cultural aspects of semantic structure*. Cambridge : Cambridge University Press.

Talmy, Leonard. 1983. How language structures space. In Pick & Acredolo (eds), 225-282.

Talmy, Leonard. 1985. Lexicalization patterns: semantic structure in lexical forms. In Shopen (ed), 57-149.

Talmy, Leonard. 2000. *Toward a cognitive semantics*, 2 vol. Cambridge : MIT Press.

Talmy, Leonard. 2003. The representation of spatial structure in spoken and signed language. In Emmorey (ed), 169-195.

Talmy, Leonard. 2010. How spoken language and signed language structure space differently. In Evans & Chilton (eds), 319-350.

Tashakkori, Abbas & Charles Teddlie (eds). 2003. *Handbook of mixed methods in social & behavioral research*. Thousand Oaks : Sage.

Teddlie, C, & Abbas Tashakkori. 2003. Preface. In Tashakkori & Teddlie (eds), ix-xv.

Tenbrink, Thora. 2007. *Space, time, and the use of language: an investigation of relationships.* Berlin / New York : Mouton de Gruyer.

Thelen, Esther, Gregor Schöner, Christian Scheier & Linda B. Smith. 2001. The dynamics of embodiment: A field theory of infant perseverative reaching, *Behavioral and Brain Sciences* 24(1), 1-34.

Thepkanjana, Kingkarn. 2008. Verb Serialization as a Means of Expressing Complex Events in Thai. In Lewandowska-Tomaszczyk (ed), 103-120.

Thompson, Joe, Nehdia Sameen, Maximilian Bibok & Timothy Racine. 2013. Agnosticism gone awry: Why developmental robotics must commit to an understanding of embodiment and shared intentionality, *New Ideas in Psychology* 31(3), 184-193.

Trabant, Jürgen. 2008. Der Mensch ist nur Mensch dur Sprache. In Ganten et al. (eds), 240-243.

Trabant, Jürgen. 2014. Pourquoi Humboldt ? In Archaimbault et al. (eds), 573-585.

Traugott, Elizabeth C. & Bernd Heine (eds). 1991. *Approaches to Grammaticalization*, 2 vol. Amsterdam / Philadelphie : John Benjamins.

Traugott, Elizabeth C. & Ekkehard König. 1991. The semantics-pragmatics of grammaticalization revisited. In Traugott & Heine (eds), vol. 1, 189-218.

Traugott, Elizabeth C. & Richard B. Dasher. 2002. *Regularity in Semantic Change*. Cambridge : Cambridge University Press.

Traugott, Elizabeth C. 2003. From subjectification to intersubjectification. In Hickey (ed), 124-139.

Traugott, Elizabeth C. 2007a. Discussion article: Discourse markers, modal particles, and contrastive analysis, synchronic and diachronic, *Catalan Journal of Linguistics* 6, 139-157.

Traugott, Elizabeth C. 2007b. (Inter)subjectification and unidirectionality, *Journal of Historical Pragmatics* 8(2), 295-309.

Traugott, Elizabeth C. 2010. (Inter)subjectivity and (inter)subjectification: a reassessment. In Davidse, Vandelanotte & Cuyckens (eds), 29-69.

Tremblay, Mireille. 1996. Lexical and non-lexical prepositions in French. In Di Sciullo (ed), 77-98.

Tristram, Anna & Wendy Ayres-Bennett. 2012. From Negation to Agreement: Revisiting the Problem of Sources for Socio-Historical Linguistics, *Neuphilologische Mitteilungen* CXIII(3), *Special Issue on Comparative Historical Sociolinguistics*, 365-402.

Troberg, Michelle. 2011. Directed motion in Medieval French. In Herschensohn (ed), 117-136.

Trousdale, Graeme. 2012. Grammaticalization, lexicalization and constructionalization from a cognitive-pragmatic perspective. In Schmid (ed), 533-558.

Tyler, Andrea & Vyvyan Evans. 2003. *The semantics of English prepositions: spatial scenes, embodied meaning and cognition*. Cambridge : Cambridge University Press.

Uithol, Sebo, Michele Franca, Katrin Heimann, Daniele Marzoli, Paolo Capotosto, Luca Tommasi & Vittorio Gallese. 2015. Single-pulse Transcranial Magnetic Stimulation Reveals Contribution of Premotor Cortex to Object Shape Recognition, *Brain Stimulation* 8, 953-956.

Valdman, Albert & Kevin J. Rottet (eds). 2009. *Dictionary of Louisiana French: As spoken in the Cajun, Creole, and Native American Communities*. Jackson : University Press of Mississippi.

Van der Auwera, Johan. 1998. *Adverbial constructions in the languages of Europe*. Berlin / New York : Mouton de Gruyter.

Van Lier, Eva & Jan Rijkhoff. 2013. Flexible word classes in linguistic typology and grammatical theory. In Van Lier & Rijkhoff (eds), 1-30.

Van Lier, Eva & Jan Rijkhoff. 2013. *Flexible Word Classes: Typological studies of underspecified parts of speech*. Oxford : Oxford University Press.

van Linden, An, Jean-Christophe Verstraete & Kristin Davidse (eds). 2010. *Formal Evidence in Grammaticalization Research*. Amsterdam / Philadelphie : John Benjamins.

Vandeloise, Claude. 1986. *L'espace en français*. Paris : Seuil.

Vandeloise, Claude. 1987. La préposition *à* et le principe d'anticipation, *Langue française* 76, 77-111.

Vandeloise, Claude. 1999. Quand *dans* quitte l'espace pour le temps, *Approches sémantiques des prépositions, Revue de Sémantique et Pragmatique* 6, 145-163.

Véronis, Jean & Emilie Guimier De Neef. 2006. Le traitement des nouvelles formes de communication écrite. In Sabah (ed), 227-248.

Viaplana, Joaquim & Maria Pilar Perea (eds). 2003. *Textos orals dialectals del català sincronitzats. Una selecció*. Barcelone : Promociones y Publicaciones Universitarias (PPU).

Viaplana, Joaquim, Maria-Rosa Lloret, Maria-Pilar Perea & Esteve Clua. 2007. *COD. Corpus Oral Dialectal*. Barcelone : Promociones y Publicaciones Universitarias (PPU).

Victorri, Bernard. 2004a. *Les grammaires cognitives.* In Fuchs (ed), 73-98.

Victorri, Bernard. 2004b. Continu et discret en sémantique lexicale, *Cahiers de praxématique* 42, 75-94.

Victorri, Bernard. 2010. Le localisme à l'épreuve du verbe *aller*, *Corela*, *HS-7*. [corela.revues.org/1009]

Vogel, Petra M. & Bernard Comrie (eds). 2000. *Approaches to the Typology of Word Classes*. Berlin / New York : Mouton de Gruyter.

von Stutterheim, Christiane, Martin Andermann, Mary Carroll, Monique Flecken & Barbara Schmiedtová. 2012. How grammaticized concepts shape event conceptualization in language production: Insights from linguistic analysis, eye tracking data, and memory performance, *Linguistics* 50(4), 833-867.

Waltereit, Richard & Ulrich Detges. 2007. Different functions, different histories. Modal particles and discourse markers from a diachronic point of view, *Catalan Journal of Linguistics* 6, 61-80.

Waltereit, Richard. 2002. Imperatives, Interruption in Conversation, and the Rise of Discourse Markers: A Study of Italian *guarda*, *Linguistics* 40(5), 987-1010.

Waters, Lindsay. 2004. *Enemies of Promise: Publishing, Perishing and the Eclipse of Scholarship*. Chicago : University of Chicago Press.

Wierzbicka, Anna. 1988. *The Semantics of Grammar*. Amsterdam / Philadelphie : John Benjamins.

Wierzbicka, Anna. 1996. *Semantics: Primes and Universals*. Oxford : Oxford University Press.

Wiesmath, Raphaële. 2002. Présence et absence du relatif et conjonctif *que* dans le français acadien : tendances contradictoires ? In Pusch & Raible (eds), 393-406.

Wischer, Ilse & Gabriele Diewald (eds). 2002. *New reflections on grammaticalization*. Amsterdam / Philadelphie : John Benjamins.

Wischer, Ilse. 2000. Grammaticalization versus lexicalization. 'Methinks' there is some confusion. In Fischer, Rosenbach & Stein (eds), 355-370.

Wüllner, Franz. 1827. *Die Bedeutung des sprachlichen Casus und Modi*. Münster : Coppenrath.

Xue, Jin, Fernando Marmolejo-Ramos & Xuna Pei. 2015. The linguistic context effects on the processing of body–object interaction words: An ERP study on second language learners, *Brain Research* 1613, 37-48.

Zarka, Yves-Charles. 2009. L'évaluation : un pouvoir supposé savoir, *Cités* 37, Paris : PUF.

Zelinsky-Wibbelt, Cornelia. 1993. *The Semantics of Prepositions: From Mental Processing to Natural Language Processing*. Berlin / New York : Mouton de Gruyter.

Ziemke, Tom, Jordan Zlatev & Roslyn M. Frank (eds). 2007. *Body, Language and Mind*. Berlin / New York : Mouton de Gruyter.

Zlatev, Jordan & Caroline David. 2003. Motion event construction in Swedish, French and Thai : three different language types, *Manusya : journal of humanities* 6, 18-42.

Zlatev, Jordan & Peerapat Yangklang. 2004. A Third Way to Travel: The Place of Thai in Motion Event Typology. In Strömqvist & Verhoeven (eds), 159-190.

Zlatev, Jordan. 2005. What's in a schema? Bodily mimesis and the grounding of language. In Hampe (ed), 313-343.

Zlatev, Jordan. 2007. Language, embodiment and mimesis. In Ziemke, Zlatev & Frank (eds), 297-338.

Zlatev, Jordan. 2008. Levels of Meaning, Embodiment, and Communication, *Cybernetics & Human Knowing* 16(3-4), 149-174.

Corpus et bases de données

Badip. Banca dati dell'italiano parlato. Karl-Franzens-Universität Graz. [www.badip.uni-graz.at]

BFM. Base de Français Médiéval. Lyon : ENS de Lyon, Laboratoire ICAR, 2012. [bfm.ens-lyon.fr]

Frantext. ATILF-CNRS & Université de Lorraine. [www.frantext.fr]

Grup d'Estudi de la Variació. 2008. Corpus de Català Contemporani de la Universitat de Barcelona. Barcelone : Departament de Filologia Catalana, Universitat de Barcelona.

OVI. *Opera del Vocabolario Italiano*. Consortium Italnet. [artfl-project.uchicago.edu/content/ovi]

Payrató, Lluís & Núria Alturo (eds). 2002. *Corpus oral de conversa col·loquial. Materials de treball*. Barcelone : Publicacions de la Universitat de Barcelona.

PFC. Corpus du projet « Phonologie du français contemporain ». [www.projet-pfc.net]

Pons, Clàudia & Joaquim Viaplana (ed). 2009. *Corpus oral dialectal (COD). Textos orals del balear.* Dipòsit Digital de la UB.

Real Academia Española : Banco de datos (CORDE). Corpus diacrónico del español. [www.rae.es]

Real Academia Española : Banco de datos (CREA). Corpus de referencia del español actual. [www.rae.es]

Ricketts, Peter T. 2001. *Concordance de l'occitan médiéval* (COM1). Turnhout : Brepols.

Ricketts, Peter T. 2005. *Concordance de l'occitan médiéval* (COM2). Turnhout : Brepols.

Stein, Achim et al. (ed). 2011. *Nouveau Corpus d'Amsterdam. Corpus informatique de textes littéraires d'ancien français (ca 1150-1350), établi par Anthonij Dees (Amsterdam 1987), remanié par Achim Stein, Pierre Kunstmann et Martin-D. Gleßgen*. Stuttgart : Institut für Linguistik/Romanistik.

Stosic, Dejan. 2015. *ParCoLab (beta), a Parallel Corpus of French, Serbian and English.* CLLE-ERSS, CNRS & Université de Toulouse 2, Toulouse, France. [parcolab.univ-tlse2.fr]

Valdman, Albert et al. 2003. *À la découverte du français cadien à travers la parole. Discovering Cajun French through the spoken word*. Indiania University Creole Institute.

von Waldenfels, Ruprecht & Roland Meyer. 2006-. *ParaSol, a Corpus of Slavic and Other Languages*. Bern : Regensburg. [parasolcorpus.org]

Caractéristiques des corpus

acronyme	RIALC
nom complet	Repertorio informatizzato dell'antica letteratura catalana
langue & période	catalan médiéval
nombre de mots (en millions)	0,15
degré d'accessibilité	accès libre (site www.rialc.unina.it), textes uniquement (pas d'interface de recherche)
Auteur / Publications de référence	Di Girolamo (1999)
Type d'encodage	aucun

acronyme	CCCUB
nom complet	Corpus de Català Contemporani de la Universitat de Barcelona
langue & période	catalan moderne (oral)
nombre de mots (en millions)	0,08
degré d'accessibilité	accès libre : www.ub.edu/cccub
Auteur / Publications de référence	Grup d'Estudi de la Variació (2008) [45]
Type d'encodage	aucun

acronyme	CdE
nom complet	Corpus del Español
langue & période	espagnol (du 13ème au 20ème s.)
nombre de mots (en millions)	100
degré d'accessibilité	accès libre (sur inscription)
Auteur / Publications de référence	Davies (2002-)
Type d'encodage	lemmatisation
acronyme	Corde

nom complet	Corpus Diacrónico del Español
langue & période	espagnol avant 1975
nombre de mots (en millions)	250
degré d'accessibilité	accès libre

[45] Alturo et al. 2004, Boix-Fuster et al. (2007), Carrera-Sabaté & Viaplana (eds, 2008), Payrató & Alturo (eds, 2002), Perea & Viaplana (2008), Pons & Viaplana (2009), Viaplana & Perea (eds, 2003), Viaplana et al. (2007).

Auteur / Publications de référence	Real Academia Española
Type d'encodage	aucun

acronyme	CREA
nom complet	Corpus del Español Actual
langue & période	espagnol moderne (à partir de 1975), écrit et oral
nombre de mots (en millions)	160
degré d'accessibilité	accès libre
Auteur / Publications de référence	Real Academia Española
Type d'encodage	aucn

acronyme	Frantext
nom complet	Frantext
langue & période	français classique et moderne
nombre de mots (en millions)	210
degré d'accessibilité	sur abonnement
Auteur / Publications de référence	ATILF-CNRS & Université de Lorraine
Type d'encodage	lemmatisation et étiquetage morpho-syntaxique partiels

acronyme	BFM
nom complet	Base du français médiéval
langue & période	français médiéval (ancien et moyen)
nombre de mots (en millions)	7
degré d'accessibilité	sur inscription
Auteur / Publications de référence	Laboratoire ICAR, ENS de Lyon ; Guillot et al. (2007)
Type d'encodage	étiquetage morpho-syntaxique pour une partie des textes (corpus SRCMF) ; balisage du discours direct et indirect

acronyme	SRCMF
nom complet	Syntactic Reference Corpus of Medieval French
langue & période	français médiéval (ancien et moyen)
nombre de mots (en millions)	0,245
degré d'accessibilité	accès libre
Auteur / Publications de référence	Stein & Prévost (2013)
Type d'encodage	étiquettes morpho-syntaxiques et structure syntaxique

acronyme	DMF
nom complet	Base du Dictionnaire du moyen français
langue & période	français médiéval (moyen français)
nombre de mots (en millions)	7
degré d'accessibilité	accès libre
Auteur / Publications de référence	ATILF-CNRS & Université de Lorraine
Type d'encodage	lemmatisation

acronyme	Valibel
nom complet	Valibel
langue & période	français moderne (Belgique), oral

nombre de mots (en millions)	4
degré d'accessibilité	accès libre (sur inscription)
Auteur / Publications de référence	Dister et al. (2009)
Type d'encodage	aucun

acronyme	CLAPI
nom complet	Corpus de LAngue Parlée en Interaction
langue & période	français moderne (France), oral
nombre de mots (en millions)	2,5
degré d'accessibilité	accès libre (pour 90% des données)
Auteur / Publications de référence	Laboratoire ICAR, ENS de Lyon
Type d'encodage	notation des caractéristiques phonétiques (prononciation) et dialogiques (tours de parole, chevauchements, etc.)

acronyme	DCF
nom complet	À la découverte du français cadien à travers la parole *Discovering Cajun French through the spoken word*
langue & période	français moderne (Louisiane), oral
nombre de mots (en millions)	0,08
degré d'accessibilité	commercial
Auteur / Publications de référence	Valdman et al. (2003)
Type d'encodage	aucun

acronyme	CFPP2000
nom complet	Corpus du français parlé parisien
langue & période	français moderne (Paris), oral
nombre de mots (en millions)	0,35
degré d'accessibilité	accès libre
Auteur / Publications de référence	Branca-Rosoff et al. (2012)
Type d'encodage	aucun

acronyme	PFC
nom complet	Phonologie du français contemporain
langue & période	français moderne (monde), oral
nombre de mots (en millions)	0,2
degré d'accessibilité	accès libre
Auteur / Publications de référence	Durand et al. (2002, 2009)
Type d'encodage	notation des caractéristiques phonétiques (prononciation, liaisons) et morpho-syntaxiques

acronyme	CFPQ
nom complet	Corpus du français parlé au Québec
langue & période	français moderne (Québec), oral
nombre de mots (en millions)	0,5
degré d'accessibilité	accès libre
Auteur / Publications de référence	Dostie (2006-)
Type d'encodage	notation des caractéristiques phonétiques (prononciation, intonation, etc.), dialogiques (chevauchements), gestuelles

acronyme	OVI
nom complet	Opera del Vocabolario Italiano
langue & période	italien médiéval (avant 1375)
nombre de mots (en millions)	23
degré d'accessibilité	accès libre (sur inscription)
Auteur / Publications de référence	Italnet
Type d'encodage	aucun

nom complet	Morfologia dell'Italiano in Diacronia
langue & période	italien en diachronie
nombre de mots (en millions)	7,5
degré d'accessibilité	accès libre
Auteur / Publications de référence	D'Achille et al. (2014)
Type d'encodage	notation morpho-syntaxique

acronyme	BADIP (LIP)
nom complet	BAnca Dati dell'Italiano Parlato (Lessico di frequenza dell'italiano parlato)
langue & période	italien moderne, oral
nombre de mots (en millions)	0,5
degré d'accessibilité	accès libre (sur inscription)
Auteur / Publications de référence	Di Mauro et al. (1993), Karl-Franzens-Universität Graz
Type d'encodage	aucun

acronyme	COM
nom complet	Concordance de l'occitan médiéval
langue & période	occitan médiéval
nombre de mots (en millions)	7
degré d'accessibilité	commercial
Auteur / Publications de référence	Ricketts (2001, 2005)
Type d'encodage	aucun

acronyme	CdP
nom complet	Corpus do Português
langue & période	portugais (du 14ème au 20ème s.)
nombre de mots (en millions)	45
degré d'accessibilité	accès libre (sur inscription)
Auteur / Publications de référence	Davies & Ferreira (2006-)
Type d'encodage	lemmatisation

acronyme	CIPM
nom complet	Corpus Informatizado do Português Medieval
langue & période	portugais médiéval
nombre de mots (en millions)	2,2
degré d'accessibilité	accès libre (sur inscription)
Auteur / Publications de référence	Centro de Linguística, Universidade Nova de Lisboa
Type d'encodage	aucun

acronyme	Ruxândoiu
nom complet	Interacţiunea verbală în limba română actuală

langue & période	roumain moderne, oral
nombre de mots (en millions)	0,05
degré d'accessibilité	livre (pas de version consultable)
Auteur / Publications de référence	Ruxândoiu (2002)
Type d'encodage	aucun

acronyme	CORV
nom complet	Corpus de română vorbită
langue & période	roumain moderne, oral
nombre de mots (en millions)	0,05
degré d'accessibilité	livre
Auteur / Publications de référence	Dascălu Jinga (2002)
Type d'encodage	

acronyme	Sketch Engine
nom complet	The Sketch Engine
langue & période	multiples (plus de 200 corpus)
nombre de mots (en millions)	à voir pour chaque corpus
degré d'accessibilité	commercial
Auteur / Publications de référence	Kilgariff et al. (2004)
Type d'encodage	aucun

acronyme	SFA
nom complet	Base des Syntagmes Prépositionnels
langue & période	français moderne
nombre de mots (en millions)	2,4 (corpus de base), 0, 02 (éléments annotés)
degré d'accessibilité	accès libre sur demande
Auteur / Publications de référence	Sarda & Charolles (2012)
Type d'encodage	morpho-syntaxique et sémantique

Tableaux (16) : Principaux corpus utilisés pour mes recherches

Table des matières

0 PROLOGUE **5**

I PROLEGOMENES METHODOLOGIQUES **7**

1 QUESTIONS DE METHODE : LE CORPUS COMME OUTIL DE BASE **9**

1.1 Avantages des corpus **10**
1.1.1 Le corpus, outil indispensable pour le diachronicien 10
1.1.2 Compenser les limites des grammaires 12
1.1.3 Analyse quantitative et subjectivité 12
1.1.4 Phénomènes rares et structure du langage 13

1.2 Conditions et limites du recours aux corpus **15**
1.2.1 Problèmes liés à l'étude sur corpus 15
1.2.2 Précautions d'emploi (1) : Constitution du corpus 18
1.2.3 Précautions d'emploi (2) : Utilisation des corpus 19

2 ECLAIRAGES **23**

2.1 L'approche diachronique **23**

2.2 Approche contrastive et linguistique romane **26**
2.2.1 Approche onomasiologique, approche sémasiologique 27
2.2.2 Le codage en contrastif 28
2.2.3 La question de la maîtrise de la langue 29
2.2.4 La question de la granularité 29
2.2.5 Les corpus en linguistique contrastive 30

2.3 L'approche typologique **31**
2.3.1 Apports et limites de la typologie 31
2.3.2 Conditions d'une typologie scientifique 33
2.3.3 La question de la diversité 34

2.4 La linguistique cognitive **35**

II QUESTIONNEMENTS (ET QUELQUES DEBUTS DE REPONSE) **39**

3 LE LOCALISME **41**

3.1 Panorama des études actuelles et antérieures – Le localisme : introduction **41**
3.1.1 Définition préliminaire du localisme 41
3.1.2 Bref historique des théories localistes 42
3.1.3 Etat de l'art – le localisme aujourd'hui 42

3.2 Etudes sur le localisme **44**
3.2.1 Du spatial au comitatif 45

3.2.2 Du spatial au temporel 46
3.2.3 Du spatial à la possession 48
3.2.4 Du spatial à l'exclusion 49
3.2.5 Du spatial au comparatif 50
3.2.5.1 Les évolutions reconstruites 51
3.2.5.2 Les évolutions romanes 53
3.2.6 Du spatial au discursif 54
3.2.6.1 Arguments forts 54
3.2.6.2 Illustrations 55
3.2.6.3 Reformulations 56
3.2.7 Après l'espace 58
3.2.8 Des contre-exemples ? 60

3.3 Résultat : un localisme revisité 62

4 L'ESPACE DANS LA LANGUE 65

4.1 Les relations spatiales statiques 67

4.2 L'expression du déplacement 72
4.2.1 Expérimentation « Trajectoire » : Récolte des données 73
4.2.2 Expérimentation « Trajectoire » : Analyse des données 74
4.2.3 Etude de l'expression du mouvement : approches diachroniques 78
4.2.3.1 Limites et pistes pour des recherches futures 83

4.3 Le mouvement fictif 85
4.3.1 Données diachroniques sur le mouvement fictif 85
4.3.2 Mouvement fictif : données élicitées 87

4.4 L'asymétrie source/but 93

5 LA GRAMMATICALISATION 99

5.1 Panorama des études actuelles et antérieures – Définition et état de l'art 101

5.2 Etudes menées sur la grammaticalisation 105
5.2.1 Grammaticalisation et lexicalisation 105
5.2.2 Grammaticalisation et subjectivation 108
Le cas des marqueurs discursifs 108
Une invention récente 108
Vers une définition 109
Grammaticalité et grammaticalisation 111
Résultats 112
Le cas des conjonctions de coordination 113

5.3 Grammaticalisation et syntaxe 114

5.4 Sources de la grammaticalisation et fin anticipée du processus 115

5.5 Conclusions sur la vitesse des changements 120

III SUITES **123**

6 PROJET DE RECHERCHE **125**

6.1 De la synthèse au programme **125**

6.2 Une nouvelle linguistique **126**

6.3 Un exemple **129**

6.4 Interdisciplinarité **130**

6.5 Applications **131**

BIBLIOGRAPHIE **135**

Corpus et bases de données **156**
Caractéristiques des corpus 157

TABLE DES MATIERES **163**