Biases in phonological processing and learning

Alexander Martin

THÈSE DE DOCTORAT

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

Préparée à
L'École normale supérieure

Biases in phonological processing and learning

Les biais dans le traitement et l'apprentissage phonologiques

École doctorale n°158

ED3C – CERVEAU COGNITION COMPORTEMENT

Spécialité SCIENCES COGNITIVES

COMPOSITION DU JURY :

Mme CHITORAN Ioana
Université Paris Diderot, Rapporteur,
Présidente du jury

M. PELLEGRINO François
Université Lyon 2, Rapporteur

Mme JACQUEMOT Charlotte
École normale supérieure, Membre du jury

M. VAN DE VIJVER Ruben
Heinrich-Heine-Universität, Membre du jury

Mme PEPERKAMP Sharon
École normale supérieure, Directrice de thèse

Soutenue par Alexander Martin
le 30 juin 2017

Dirigée par **Sharon Peperkamp**

PSL RESEARCH UNIVERSITY PARIS

ENS

Université de recherche Paris Sciences et Lettres
École Normale Supérieure
Département d'Études Cognitives

Les biais dans le traitement et l'apprentissage phonologiques

Biases in phonological processing and learning

Alexander Martin

Thèse de doctorat dirigée par Sharon Peperkamp
Laboratoire de Sciences Cognitives et Psycholinguistique
UMR 8554 CNRS–ENS–EHESS
Soutenue le 30 juin 2017 à Paris

Abstract

During speech perception, listeners are biased by a great number of factors, including cognitive limitations such as memory and attention and linguistic limitations such as their native language. This thesis focuses on two of these factors: processing bias during word recognition, and learning bias during the transmission process. These factors are combinatorial and can, over time, affect the way languages evolve.

In the first part of this thesis, we focus on the process of word recognition. Previous research has established the importance of phonological features (e.g., voicing or place of articulation) during speech processing, but little is known about their weight relative to one another, and how this influences listeners' ability to recognize words. We tested French participants on their ability to recognize mispronounced words and found that the manner and place features were more important than the voicing feature. We then explored two sources of this asymmetry and found that listeners were biased both by bottom-up acoustic perception (manner contrasts are easier to perceive because of their acoustic distance compared to the other features) and top-down lexical knowledge (the place feature is used more in the French lexicon than the other two features). We suggest that these two sources of bias coalesce during the word recognition process to influence listeners.

In the second part of this thesis, we turn to the question of bias during the learning process. It has been suggested that language learners may be biased towards the learning of certain phonological patterns because of phonetic knowledge they have. This in turn can explain why certain patterns are recurrent in the typology while others remain rare or unattested. Specifically, we explored the role of learning bias on the acquisition of the typologically common rule of vowel harmony compared to the unattested (but logically equivalent) rule of vowel disharmony. We found that in both perception and production, there was evidence of a learning bias, and using a simulated iterated learning model, showed how even a small bias favoring one pattern over the other could influence the linguistic typology over time, thus explaining (in part) the prevalence of harmonic systems. We additionally explored the role of sleep on memory consolidation and showed evidence that the common pattern benefits from consolidation that the unattested pattern does not, a factor that may also contribute to the typological asymmetry.

Overall, this thesis considers a few of the wide-ranging sources of bias in the individual and discusses how these influences can over time shape linguistic systems. We demonstrate the dynamic and complicated nature of speech processing (both in perception and learning) and open the door for

future research to explore in finer detail just how these different sources of bias are weighted relative to one another.

Résumé

Pendant la perception de la parole, les locuteurs sont biaisés par un grand nombre de facteurs. Par exemple, il existe des limitations cognitives comme la mémoire ou l'attention, mais aussi des limitations linguistiques comme leur langue maternelle. Cette thèse se concentre sur deux de ces facteurs : les biais de traitement pendant la reconnaissance des mots, et les biais d'apprentissage pendant le processus de transmission. Ces facteurs peuvent se combiner et, au cours du temps, influencer l'évolution des langues.

Dans la première partie de cette thèse, nous nous concentrons sur le processus de la reconnaissance des mots. Des recherches antérieures ont établi l'importance des traits phonologiques (p. ex. le voisement ou le lieu d'articulation) pendant le traitement de la parole. Cependant, nous en savons peu sur leur poids relatif les uns par rapport aux autres, et comment cela peut influencer la capacité des locuteurs à reconnaître les mots. Nous avons testé des locuteurs français sur leur capacité à reconnaître des mots mal prononcés et avons trouvé que les traits de mode et de lieu sont plus importants que le trait de voisement. Nous avons ensuite considéré deux sources de cette asymétrie et avons trouvé que les locuteurs sont biaisés et par la perception acoustique ascendante (les contrastes de mode sont plus facile à percevoir à cause de leur distance acoustique importante) et par la connaissance lexicale descendante (le trait de lieu est plus exploité dans le lexique français que les autres traits). Nous suggérons que ces deux sources de biais se combinent pour influencer les locuteurs lors de la reconnaissance des mots.

Dans la seconde partie de cette thèse, nous nous concentrons sur la question d'un biais d'apprentissage. Il a été suggéré que les apprenants peuvent être biaisés vers l'apprentissage de certains patrons phonologiques grâce à leurs connaissances phonétiques. Cela peut alors expliquer pourquoi certains patrons sont récurrents dans la typologie, tandis que d'autres restent rares ou non-attestés. Plus spécifiquement, nous avons exploré le rôle d'un biais d'apprentissage sur l'acquisition de la règle typologiquement commune de l'harmonie vocalique comparée à celle de la règle non-attestée (mais logiquement équivalente) de la disharmonie vocalique. Nous avons trouvé des preuves d'un biais d'apprentissage aussi bien en perception qu'en production. En utilisant un modèle d'apprentissage

itéré simulé, nous avons ensuite montré comment un biais, même petit, favorisant l'un des patrons, peut influencer la typologie linguistique au cours du temps et donc expliquer (en partie) la prépondérance de systèmes harmoniques. De plus, nous avons exploré le rôle du sommeil sur la consolidation mnésique. Nous avons montré que seul le patron commun bénéficie d'une consolidation et que cela est un facteur supplémentaire pouvant contribuer à l'asymétrie typologique.

Dans l'ensemble, cette thèse considère certaines des sources de biais possibles chez l'individu et discute de comment ces influences peuvent, au cours du temps, faire évoluer les systèmes linguistiques. Nous avons démontré la nature dynamique et complexe du traitement de la parole, à la fois en perception et dans l'apprentissage. De futurs travaux devront explorer plus en détail comment ces différentes sources de biais sont pondérées les unes relativement aux autres.

Dedication

To my grandmother Jeanne, whose life-long pursuit of knowledge inspired me from the youngest age.

Acknowledgements

Warning: these acknowledgements are filled with horrible clichés and obnoxious heartfelt emotion. J'assume entièrement and I mean every word.

Those who do a PhD get a rare opportunity to enshrine their gratefulness in black and white for eternity. While I have tried to say all of these things at some point or another, I know I have failed to express the true impact that you have all had on me. Please accept these words as a thank you for all you have given me.

To begin, I must recognize the person who has guided me through this process from beginning to end, my director **Sharon**. You have never hesitated to challenge and push me, call me out, and encourage me. Our direct and honest communication have ensured our success as a team, and I thank you for every bit of personal and professional investment that you have made on my behalf.

To the **researchers of the LSCP**, I say thank you for creating an incredibly stimulating scientific environment. It doesn't take much discussion with others to realize just how good we have it, thanks to the infrastructure and ambiance that you have developed over the years.

I would also like to thank the support staff in their various roles, without whom the realization of this thesis would have been impossible. Du LSCP, **Isabelle Brunet**, qui a su supporter mes demandes incessantes de nouveaux sujets, **Radhia Achheb**, qui m'a aidé bien plus d'une fois à régler des problèmes administratifs compliqués, **Anne-Caroline Fiévet**, qui m'a dépanné dans toutes sortes de situations, **Vireack Ul** et **Michel Dutat**, qui ensemble forment une équipe qui définit la true LSCP experience. From Utrecht, **Iris Mulders** and **Chris van Run** for welcoming me to the phonetics lab and making me feel most at home. Du DEC, **Samira Boujidi**, qui a géré mes demandes quasi quotidiennes de réservation de salles, **Nathalie Marcinek**, qui était toujours prête à fournir un effort supplémentaire pour m'aider, et bien évidemment **Clémentine Eyraud**, qui, en plus d'avoir géré toutes mes questions de communication, internes comme externes, était une vraie source d'inspiration pour me rappeler que vouloir really is pouvoir.

I would additionally like to thank the following people (in no particular order) for having supported me in various ways during my PhD:

Eva D'hondt, for her calming help in my panic before my trip to Utrecht, and every pint of Delerium Tremens before and since.

Brigitta Keij, for taking my Dutch experience up a notch with a bike and bitterballen.

Merel van Goch, for endless bubbles, because "superdoei" never means goodbye for good.

Isabelle Dautriche, for her endless patience in explaining and re-explaining even the most basic statistics.

Marieke van Heugten, for being an inspiring role model and challenging co-author.

Mathilde Fort, pour m'avoir fait découvrir la science sans jamais me faire sentir moindre quand je ne savais pas quelque chose.

Maria Giavazzi, for her enthusiasm and dynamism, for listening to me and guiding me when things were hard. Teaching will never be the same without you!

My cohort-mates **Alex de Carvalho** and **Julia Carbajal**. We challenged, and supported each other every step of the way. Alex, you've been there, a thorn in my side, and the wind beneath my wings since the first days we arrived at the LSCP. Julia, I'd like to pretend that alfajores were less of a big deal than they are, but alas…

My aunts, **Gayle Martin**, **Robyne**, and **Sheri Martin**, who each in their own way at their own time was there for me, to listen, to inspire.

Page Piccinini, for challenging me, for teaching me, for listening to me. Everyone who's met Page knows that our studies not only would suffer but probably would not exist if it weren't for Page's knowledge, but mostly patience while we ask the most basic questions over and over and over. I really couldn't sum up everything that Page has done, so thank you for being there, every day, and always saying "yes" to Not Buns. Nems?

Auréliane Pajani, besides being the best stimuli recorder anyone has ever met, tu m'as inspiré avec ta persévérance, même face à des situations difficiles. Par contre, la fenêtre reste ouverte.

Annika Dean, you've listened to me whine over and over, and no matter how many kilometers we run, I never seem to run out of breath to complain; you take in every earful and end with a perfectly timed "c'est clair"… I needed them all.

Yue Sun, for every stimulating conversation, and every new board game.

Jeff Gustafson, for happily listening to me complain Skype after Skype.

Katie Devine, for proving that distance means nothing compared to friendship.

Magda Rosińska, I've never appreciated Lebanese food more than with you at Maison de Solenn. Those breaks, however frequent or rare, were always just what I needed.

Lisa Marie, thank you for being there, always, no matter what.

Margaux Romand-Monnier, pour m'avoir soutenu dans toutes sortes de situations, sans aucune hésitation.

Clémence Alméras, pour avoir supporté mes interruptions incessantes.

My poor cats, **Philomène** and **Suzie**, who were smothered with love on days good and bad, and who so bravely put up with wearing crocheted hats for my viewing pleasure.

Alya Vlassova, for knowing that asking to bum a piece of gum really meant "let's go drink five beers", блядь, as if that could somehow sum up our relationship…

Sári Zsuzsinak, hogy ott voltál (majdnem) minden pénteken, és hogy egy óráig, minden rendben volt, hogy ami alatt beszélgettünk, nem volt szakdolgozatom, nem volt határidőm, csak magyarul tanulni kellet, és ça suffisait largement comme défi.

Maggie Zander, for being there to celebrate the good times, and for making me food during the bad times. I may not remember the first time we met, but I sure won't forget the rest.

Dorothée Arzounian, ton investissement m'a beaucoup inspiré. C'est grâce à des gens comme toi que les choses avancent.

Andy King, for dealing with my constant berating like a champ, and for making the dark days of dissertating quite a bit more fun.

Ewan Dunbar, for taking over the position of noodle tsar upon Yue's departure.

Leonardo Barbosa, for caring about the bigger picture.

Mathilde Marié et **Audrey Champeau**, pour tous nos repas conviviaux assaisonnés au karaoké.

Maureen McFadden, for being there pour une petite mousse dès que ça n'allait pas. This has been a reel!

Charlotte Van den Driessche, for understanding more than anyone the way my mind works, and reminding me that it's ok.

My running buddies, **Mora Maldonado**, **Jeremy Kuhn**, and **Milica Denić**, few activities have had such a strong positive impact on my health, both mental and physical, as running, and it won't be quite the same without you.

Lorna Le Stanc, pour chaque bisou, chaque câlin, ton énergie et ta patience.

My #barbarplots buddies who I haven't mentioned yet, **Christina Bergmann** jaaaa, **Sho Tsuji**, can't wait to read all about Chantal, **Rory Turnbull**, I shall anxiously await my invitation to present at your department in Hawai'i , **Adriana Guevara-Rukoz**, [insert awkward joke here]. This project was a perfect example of how far teamwork can take you, and it really was a lot of fun too!

Finally, to anyone who has given me advice, that I have then ignored, only to learn a lesson the hard way and realize you were right…**my parents** are only the beginning of this very long list.

TL;DR: Thank you.

Publications

Journal Articles

- **Martin**, **A**., & S. Peperkamp (2017). Assessing the distinctiveness of phonological features in word recognition: prelexical and lexical influences. *Journal of Phonetics*, *62*, 1–11.
- **Martin**, **A**., & S. Peperkamp (2015). Asymmetries in the exploitation of phonetic features for word recognition. *The Journal of the Acoustical Society of America*, *137*(4), EL303–EL314.
- Fort, M., **A. Martin**, & S. Peperkamp (2015). Consonants are more important than vowels in the bouba-kiki effect. *Language and Speech*, *58*(2), 247–266.

Conference Proceedings

- Fort, M., A. Weiss, **A. Martin**, & S. Peperkamp (2013). Looking for the bouba-kiki effect in prelexical infants. In: S. Ouni, F. Berthommier & A. Jesse (eds.) *Proceedings of the 12th International Conference on Auditory-Visual Speech Processing*, INRIA, 71–76.

Submitted manuscripts

- **Martin**, **A**., & S. Peperkamp (in revision). Sensitivity to phonetic naturalness in phonological rules: evidence from learning and consolidation with sleep.
- **Martin**, **A**., A. Guevara-Rukoz, T. Schatz, & S. Peperkamp (resubmitted). Phonetic naturalness and the shaping of sound patterns: the role of learning bias and its transmission across generations.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

A question that has been at the heart of the scientific study of language for decades is, despite the vast diversity present in the world's languages, why are they nonetheless so similar to one another? Indeed we find recurrent patterns in languages that are geographically and genetically quite distant from each other. All languages have consonants and vowels, and the possible combinations are actually restricted in some universal ways. For example, the CV syllable is a language universal. It is present in every language on Earth, while its counterpart VC is not. This asymmetry can be attributed to the way consonants are articulated and perceived in these positions, with hyper-articulation of the onset consonant leading to better perception of its transitions into the vowel compared to the coda consonant (for a review, see Côté, 2011).

Continuing within the domain of speech sounds, this thesis is focused on two angles. First, we consider the way humans process the sounds of their native language, and how this influences the way they recognize words. Second, we consider the way humans *learn* sound patterns and how this might lead to similarities between languages. Overall, we explore the role of human cognition on the shaping of linguistic systems.

This chapter is divided into two sections. First, we will consider past research that has explored two main questions regarding speech processing: 1) How similar are speech sounds perceived as being to each other? 2) How do humans go about recognizing words from their phonetic structure. We will then briefly discuss previous attempts at understanding how these two things combine during

speech processing and allow listeners to distinguish words from one another. Second, we will go over past work that has attempted to provide evidence that learners show a preference for certain phonological patterns over others and discuss how this can over time lead to asymmetries in the linguistic typology. Each section concludes with a brief overview of the empirical work that will be presented in the various chapters of this thesis.

1.1 Bias in phonological processing

1.1.1 Acoustic perception

Phonological processing may be performed at a relatively high cognitive level, but it relies directly on a physical stimulus. This stimulus can be broken down into two entities: the visual signal received by the eye (i.e., what our interlocutor's mouth looks like during speech), and the acoustic signal received by the ear (i.e., what sounds our interlocutor produces). We focus here on the latter.[1]

If a listener's goal is to identify the phonemes being produced by their interlocutor in order to recognize words, which are in turn part of sentences, etc., their task begins with the encoding of acoustic signals. We can then ask questions like, "What makes a /p/ sound like a /p/?" Most research has attempted to answer this question by comparing a phone associated with /p/, e.g., [p], with a phone associated with a different phoneme, e.g., [d] and comparing that with a phone of yet a different phoneme, e.g., [b]. Phonologists would say that [p] and [b] are more similar to each other than [p] and [d] because the former pair shares more phonological features (here, [p] and [b] differ only in voicing, while [p] and [d] differ both in voicing and place of articulation). But what are these features like relative to one another? Are certain featural contrasts more important than others?

In 1955, Miller and Nicely addressed this question by looking at the sounds of English and designing a series of experiments asking participants to identify nonsense syllables. Based on participants' responses, they were able to trace confusion matrices describing how often a given sound, say [p] was misidentified as another sound associated with a different phoneme, say [b]. A confusion matrix is

[1] Though note that Chapter 2 reports on a study that also considered the role of visual information in word recognition.

organized into rows and columns of phonemes. The number of times a phoneme x (row) is identified as phoneme y (column)[2] is counted in each cell. Correct performance lies along the diagonal, and all other values are considered confusions. For example, if a participant mistakenly identifies /b/ as /p/, this is considered a confusion between these two phonemes.

Miller and Nicely specifically examined 16 English consonants in the set {p, t, k, f, θ, s, ʃ, b, d, g, v, ð, z, ʒ, m, n} presented in nonsense syllables of the form C/ɑ/. Participants in their tasks were required to report which consonant they thought they heard when the syllable was presented. The key manipulation in their study was the modulation of signal-to-noise ratios (SNR) and the introduction of low-pass filtering. The consonants they tested, which included stops, fricatives, and nasals, all have very different spectral qualities, which might be affected differentially by the addition of noise or filters.

On the one hand, their first finding was straightforward. As the SNR increased (i.e., as there was more signal compared to noise), the number of confusions was reduced (i.e., they observed higher values in each confusion matrix along the diagonal). The authors then performed an analysis considering phonological features. All of the consonants they tested differ along five phonetic dimensions: voicing, nasality, affrication, duration, and place of articulation. What the authors observed in their data was that confusions were much more likely to concern affrication, duration, or place than voicing or nasality. It seems that the latter two features withstand the introduction of random masking noise better than the former three.

When establishing the perceptual similarity of speech sounds, this approach seems promising. The issue, however, is that the question being answered is really how robust different sounds are to different kinds of noise. We cannot yet say that, all things being equal, [p] sounds more similar to [t] (difference in place) than it does to [b] (difference in voicing).

Further research has attempted to address this question using experiments performed in ideal listening conditions. For instance, Plauché, Delogu, and Ohala (1997) studied the diachronic asymmetry between /pi/ and /ki/ evolving to /ti/, but not the other way around. Their hypothesis was that the spectral qualities of /pi/ and /ki/ allow them to be confused more easily as /ti/, but that the reverse is

[2]Note that x and y may be the same phoneme. This is a correct identification.

not true. While they did find that /ki/ was confused by their participants as /ti/ 20% of the time, /pi/ was only confused with /ti/ 3% of the time. In accordance with their hypothesis, /ti/ was correctly identified 100% of the time. Once they filtered certain spectral qualities of /ki/ and /pi/, however, confusion rates skyrocketed. Indeed, under ideal listening conditions, participants were very accurate in identifying the consonants, and only began confusing sounds with explicit manipulation of the acoustic signal. This is bolstered by considering the results reported by Wang and Bilger (1973). Similarly to Miller and Nicely (1955), they presented participants with English nonsense syllables and asked them to identify the consonant they contained.[3] They found that a feature's perceptual weight depended on listening conditions (i.e., whether or not there was masking noise present), but that overall, performance in quiet was incredibly high, and few confusions were made.

Hall and Hume (2013) conducted a similar identification experiment, but tested French vowels rather than English consonants. Participants were presented with a series of /aDVDa/ (where /D/ is a voiced stop consonant) stimuli and were asked to identify the middle vowel by clicking on a word in French that contained the corresponding vowel, for example *nid* for the vowel /i/. Average performance was high, with one caveat. They specifically considered common, robust vowels such as /i/ and /õ/ and compared them to currently merging vowel pairs such as /e/~/ɛ/ and /ø/~/œ/. Confusions on the merging vowel pairs were high, despite the unaltered stimuli being presented in silence. Crucially though, the "control" vowel pairs were identified correctly at a rate of 93%.

While altering the conditions under which stimuli are presented cannot provide an accurate baseline for perceptual similarity of speech sounds, in quiet, participants are clearly too good at identifying the sounds of their native language when asked to perform an explicit identification task. Recall that all the aforementioned studies required participants to respond by choosing from a set of predetermined answers, thus allowing the measurement of *confusions*, that is, how often participants mistook one sound for another. Given participants' performance though, this methodology seems less than ideal to measure the relative importance of phonological features for word recognition.

Other methods have been proposed to study this question. Hahn and Bailey (2005), for instance, used an explicit similarity task where they asked participants which of two pairs of non-words they

[3]Unlike Miller and Nicely (1955), they tested both CV and VC syllables.

found to be more similar. For example, participants might hear /pʌsp/ - /bʌsp/, /pʌsp/ - /gʌsp/, and be asked which pair, A or B, was most similar. Note that they did not focus their study on phonological features relative to one another, but rather on the number of featural differences within the pair. In this example, the first pair differs only in voicing, while the second pair differs in both voicing and place. By aggregating participant responses along the "number of featural differences" dimension, the authors demonstrated that the more phonological features that two (non-)words share, the more similar they are perceived as being. This falls in line with previous research on word recognition showing that words that share more phonological features with each other more strongly prime each other in cross-modal priming experiments (e.g., Connine, Blasko, & Titone, 1993).

Bailey and Hahn (2005) considered the question of word-level phonological similarity in a slightly different way. They compared a handful of corpora each proposing a measure of similarity. Specifically, they considered data from Wickelgren (1969)'s confusions in short-term memory, Luce (1986)'s perceptual confusions of non-words, and the MIT speech errors corpus. The first data consider how many times participants misremembered a sound *x* as a sound *y*; this was considered a "confusion". Perceptual confusions were measured as in Miller and Nicely (1955). Speech errors provide a measure of confusion by considering how often a sound *x* was mis*produced* as a sound *y*. They found when comparing all of the data that confusion patterns were not straightforward, and that no derived similarity metric could easily explain all of the variance observed. In fact, none of the quantitative measures they tested out-performed theoretical principles based on phonological features. That is, simply counting the number of feature differences better predicted overall performance in the data, though the authors note that there is still plenty of unaccounted for variance. Crucially, they argue that measuring confusability between sounds does not equate to measuring their perceived similarity. They sum up their findings eloquently: "Confusability data do not offer a shortcut which solves, at least on a practical level, the problem of measuring the similarity between phonemes" (Bailey & Hahn, 2005, p. 364).

In **??**, we address this question and propose a new way of testing the perception of native speech sounds in quiet without altering the acoustic properties of the stimuli.

1.1.2 Word recognition

Beyond research on the extent to which individual sounds are perceived as more or less similar to each other, there has been extensive work on how similar real words and mispronunciations of those words are to each other. We focus here on a few seminal studies that have considered how listeners recover meaning from the acoustic form of a stimulus.

In a classic study, Connine et al. (1993) considered the importance of different parts of words for word recognition. They specifically were interested in whether phonemes at the beginnings of words carry a special status, in line with certain models of speech perception, but not others. They used a series of cross-model priming experiments to respond to their question. In this paradigm, participants are presented a word over headphones. This word is referred to as the prime. There is no task associated with this word. Participants then see a different word written on the screen (referred to as the target) and are asked to perform a lexical decision task. Naturally, in a certain number of the trials, the word on screen is actually a nonce word. Crucially, in trials where the target is a real word, the prime may be semantically related or not. For example, participants might hear *salt* and then see the word PEPPER. In such trials, response times are lower, since the semantics associated with the prime pre-activate the target word. Connine et al. added an additional manipulation. The primes they used were sometimes mispronounced, and the nature of these mispronunciations allowed them to measure perceived similarity between a canonical pronunciation and a mispronunciation. Their main result was that the number of phonological features shared by a canonical pronunciation and a mispronunciation was key to observing priming effects. That is, mispronunciations in one or two phonological features were considered close enough to the base word to present significant priming effects, but mispronunciations in more than two features yielded no priming effect.

Milberg, Blumstein, and Dworetzky (1988) used the same paradigm to test to what extent the relationship between phonetic distance and perceived similarity are in a linear relationship to each other. They found that one-feature different mispronunciations primed targets better than multi-feature different mispronunciations which themselves primed targets better than non-word primes. Clearly the perceived similarity to the base word is crucial in recognizing mispronunciations. So what about features relative to one another?

Cole, Jakimik, and Cooper (1978) used a mispronunciation detection task to test listeners' sensitivity to different phonological features. Participants listened to short stories that contained mispronunciations of various kinds and were asked to press a button as soon as they identified a mispronounced word. On average, changes to word-initial stops were more readily identified than changes to word-initial fricatives. Importantly, changes in place of articulation were much more readily identified than changes in voicing. This result does not align with those of Miller and Nicely (1955), who found that voicing was more robust in the confusion patterns they observed. There are two crucial differences between the studies that may explain the contradicting results. Recall that Miller and Nicely presented nonsense syllables to participants under difficult listening conditions (i.e., with masking noise). Cole et al. on the other hand presented participants with short stories read fluently in ideal listening conditions. Participants in their task therefore were performing lexical access during the task, whereas participants in Miller and Nicely's task were simpling doing phoneme identification. Thus, the different results could be explained by the presence or absence of masking noise. Indeed, Wang and Bilger (1973) acknowledged that different types and levels of noise affect phonological contrasts in different ways. It may be harder to identify differences in place of articulation in the presence of masking noise than differences in voicing, but this may not hold in silence. Alternatively (or additionally), the different results could be explained by considering the task participants were asked to do. If listeners are sensitive to statistical patterns in their native lexicon, they may pay more attention to contrasts that are commonly used to distinguish words, and less attention to more infrequent contrasts. Of course this is only a useful strategy when performing lexical access. Therefore, it could be that the voicing feature is relatively unimportant compared to the place feature in English, so participants in Cole et al.'s study focused their attention on the acoustic cues associated with the place feature while attempting to recognize the words in the short story. The participants in Miller and Nicely's study, however, had no need to recognize any words, and therefore may have simply focused on the acoustic properties of the phonemes they were presented.

1.1.3 Coalescing sources of information

So far, we have examined two sources of bias during speech processing; here we ask how these sources of bias may combine to jointly influence the way listeners recognize words. On the one hand,

listening is biased by the perceptual ability of the human auditory system. In Section 1.1.1, we saw how previous work has established some of the sensitivities of the ear with regards to phonological features. On the other hand, recognizing sounds in a lexical context is a slightly different question, and may be biased in different directions. In Section 1.1.2, we saw that the amount of phonetic content shared between a word and a mispronunciation was key to their perceived similarity to one another. But how may we understand the influence of the ear versus some more top-down language-specific influence on the process of word recognition? Some recent work has addressed the issue of combining these two sources of bias and understanding word recognition from a more holistic point of view.

Indeed, it is known that perception is modulated by knowledge of one's native language. For example, Goto (1971) showed that Japanese listeners had a hard time perceiving the contrast /l/~/ɹ/ which exists in English, but not Japanese. Even those participants who had relatively good English ability and were able to produce the contrast had a hard time hearing the contrast, even when produced by native English speakers. This language-specific perceptual experience must obviously impact the way listeners recognize words, and this has become the focus of a few recent studies.

For instance, Ernestus and Mak (2004) considered the impact of native language knowledge on word recognition by testing listeners of Dutch. Dutch has a phonological process whereby initial fricatives tend to be devoiced. This process presents a great deal of variability, with speakers from the North almost never producing initial voicing in words such as *vijf*, and speakers from the South almost always producing voicing. This variability means that listeners cannot be sure that an initial voiceless fricative is truly underlyingly voiceless. Ernestus and Mak (2004) used a lexical decision task to measure Dutch listeners' accuracy when presented with mispronunciations. They showed that participants were more likely to mistake the "mispronounced" *farken* as the real word *varken* (a change in voicing) than the mispronounced *zarken* (change in place). Crucially, the participants in this task were recruited from areas where fricative voicing is normally contrastive; thus, *farken* is not an acceptable production of *varken*. The authors conclude that exposure to other dialects of Dutch has rendered the feature voicing less relevant for Dutch listeners, regardless of their own productions, and that their weighting of cues in word recognition is thus affected by their linguistic experience.

Johnson and Babel (2010) considered the impact of language-specific tuning and how it might interact with more universal acoustic perception. They compared English- and Dutch-speaking listeners' abilities to perceive a series of phonological contrasts. They designed VCV stimuli where the consonant was drawn from the set of voiceless fricatives {f, θ, s, ʃ, x, h}. This set is particularly interesting given the tested populations; while some of the sounds exist in both languages (f, s, h), others exist phonemically only in English (θ, ʃ), or only in Dutch (x). They compared two levels of processing to examine to what extent bottom-up acoustic perception and top-down language-specific knowledge interact.

In a first task, participants were required to rate how similar various pairs of the VCV stimuli were to one another. The Dutch-speaking listeners overall rated [s] and [ʃ] and [s] and [θ] as more similar to each other than did the English-speaking participants (for whom these contrasts are phonemic). Then, the same participants took part in an AX discrimination experiment using the same pairs of stimuli. Log response times were analyzed on correct trials, and the authors found that both groups of participants generally patterned together. That is, some fricative pairs were harder to discriminate than others (e.g., [f]~[θ] and [h]~[x]), but this pattern was general, and not biased by the status of these sounds in the native language of the listener.

The authors argued that the results from their similarity judgment task indicate that "phonetic similarity is to some degree shaped by language-specific phonological patterning". On the other hand, acoustic distance clearly plays a role in phonetic similarity as well, they argue, as both linguistic groups had similar performance patterns in the AX discrimination task. Even outside of lexical context, it seems that tapping into different levels of processing (i.e., more conscious, higher level processing such as in a similarity judgment vs less conscious, lower level processing such as in an acoustic comparison [AX]) demonstrates that the perceptual similarity of speech sounds is a multidimensional measure that takes into account bias from different sources.

1.1.4 Summary

All in all, previous research has clearly shown that listeners are biased by language-specific knowledge in addition to baseline constraints on hearing. The nature of this bias is yet unclear. Johnson

and Babel (2010) showed how an absolute presence or absence of a sound from a listener's linguistic system influences their perception of that sound, while Ernestus and Mak (2004) showed how exposure to a more gradient phonological process can also influence perception. It should then follow that listeners may be gradiently biased according to statistical patterns in their native language. Thus, although voicing and manner may both be active contrasts in a given listener's native language, they may be differentially weighted according to their contrastiveness in that language. Recent work has highlighted the role of "marginal contrasts" in linguistic systems (Hall, 2013). Classical phonology considers that a contrast is distinctive in a system or is not. But there are many examples of contrasts that are used in certain positions, but not others. The Spanish distinction between /r/ and /ɾ/ is a classic example. They contrast inter-vocalically in words like *perro* versus *pero*, but never initially or finally. On the level of the feature, we can return to French. There is a distinctive difference between rounded and unrounded vowels (e.g., /y/ versus /i/), so it would seem that this feature is active in the phonological system, but there is no rounding contrast in the French back vowels. Does that make rounding less distinctive than, say, height? If we consider phonological contrast not as absolute, but as gradient, we should be able to observe gradient patterns in perception. Chapter 2 focuses on this question, first by examining the relative weight of phonological features in French consonants for the purposes of word recognition, and then by exploring two sources of bias: acoustic perception and lexical knowledge. We propose that these two sources coalesce during word recognition to bias listeners to pay attention to more "important" features, be they acoustically salient, or lexically relevant.

1.2 Bias in phonological learning

In addition to biases that influence the way we process sounds, there may be general cognitive biases that influence the way we *learn* sound patterns. Over time these biases can end up shaping linguistic systems. Of course if we assume that certain biases will be shared by all human beings (e.g., perceptual bias due to the design of the human auditory system), we can predict that even genetically distinct languages will share certain characteristics.

This thesis takes as an example a phonological pattern referred to as vowel harmony. Vowel harmony

is a co-occurrence restriction that requires all vowels within a certain domain (typically the word) to share the same value of some phonological feature. Hungarian, for example, has a productive system of palatal harmony such that all vowels within a word should be either front or back. This means that suffixes typically come in two forms, depending on the vowels of the stem. For instance, there are two forms of the dative suffix: *-nek* and *-nak*, where a word like *ember*, which contains only front vowels, takes the former and a word like *barát*, which contains only back vowels, takes the latter. This type of pattern is not uncommon, and can be found in varying language families, from Finno-Ugric to Turkic to Bantu. Interestingly, though, the reverse pattern, where vowels within a word must be mixed with regards to some phonological feature (we will refer to this throughout this dissertation as vowel *dis*harmony), is virtually unattested. Of course many languages allow the mixing of vowel types within the domain of a word, but systematic pressures for vowels to be mixed within a word are nearly unheard of.

We call vowel harmony a phonetically "natural" rule, because it facilitates articulation by promoting similarity between segments. This lowers the burden on the speaker (though potentially increases the burden on the perceiver, but see Finley (2012) for a case of perceputally-driven vowel harmony).

The question that we will attempt to address is the following: why do we observe such recurrent phonetically natural rules in the typology, but rarely if ever encounter their logically equivalent, phonetically unnatural counterparts? Two main propositions have been put forth in the literature: bias in the transmission process (a.k.a. "channel" bias), and bias on the part of the learner. Our experimental work focuses on the latter, but we begin by discussing some of the hypotheses put out in the former.

1.2.1 Channel bias

There are multiple possible sources for the typological asymmetry between phonetically natural and phonetically unnatural rules. Specifically for vowel harmony, it has been proposed that the pattern results from a re-encoding of natural vowel-to-vowel coarticulation (Ohala, 1994). Languages differ in the amount and direction of vowel-to-vowel coarticulation (Beddor, Harnsberger, & Lindemann, 2002), and this could, over time, lead to the development of vowel harmony systems for some,

but not all languages. In such a view, listeners hear a word such as /uti/ pronounced [u̟ti], with a fronted /u/ due to anticipatory coarticulation with /i/, and wind up over time re-encoding the underlying representation as /yti/. Should the trigger of the coarticulation wind up being lost over time, a new vowel category is born, creating a contrast between /u/ and /y/. Naturally, this re-encoding process is more likely to happen the stronger the coarticulation, such that languages with lower degrees of coarticulation are less likely to show such patterns. This idea has been explored with a computational model, which showed that hypothetical languages with higher degrees of coarticulation did indeed lead simulated agents to over-exaggerate this coarticulation to the extent that the lexicons of these languages became more and more harmonic (i.e., fewer and fewer words contained vowels of multiple categories) (Mailhot, 2013).

Such an explanation is often referred to as "channel bias" as it regards an error during the transmission process. The hypo- hyper-correction model (Ohala, 1993a) is an example of this. According to the hypotheses laid out in that work, variation in production can lead to sound change through two mechanisms: hypo-correction where a listener fails to correct for an error in pronunciation (which itself would be due to some phonetic bias), or hyper-correction where a listener corrects for a phonetic event that was actually intended by the speaker. For example if a speaker produces the word /an/ with heavy nasalization and little or no articulation of the consonant, the listener may perceive simply [ã]. If the listener fails to "correct" for this pronunciation, they may encode the word as /ã/ rather than the intended /an/. This small change can compound with others and lead to the emergence of, in this case, nasal vowels. On the other hand, listeners can also "correct" productions erroneously. Ohala proposed that the "unnatural" process of dissimilation is a case of hyper-correction. For example, original Latin /kwiŋkwe:/ was at some point re-encoded as /kiŋkwe:/, and the labialization on the initial consonant was lost. From the point of view of hyper-correction, the listener would have heard labialization at two places within the word (on the same kind of segment, namely /k/), and assumed a speech production error. In this case, the initial labialization would be assumed to be an anticipatory error and "corrected" for by encoding the underlying representation /kiŋkwe:/, leading to a sound change.

Another example of a model of channel bias and its effects in diachrony is Evolutionary Phonology (Blevins, 2004, 2006). Blevins lays out a framework that she calls the CCC-model describing how

these kinds of errors can alone lead to the shaping of typology over time. She proposes three types of transmission errors, explained in (1).

(1) a. CHANGE: where a listener misperceives an input stimulus, such as [anpa] perceived as [ampa]. This type of error is a perceptual bias, such that sounds that are more confusable in certain contexts are more likely to change over time. In this case, the listener "changed" the representation of the word from /anpa/ to /ampa/. Once this happens enough in a population, the word will have changed permanently.

b. CHANCE: where a listener has to come up with a representation from ambiguous input. If a listener hears [ʔa̰ʔ], there are a number of possible underlying representations they could settle on: /ʔa/, /aʔ/, /a̰/, etc. Any one of these underlying representations could then be the subject of phonological processes that yield the output [ʔa̰ʔ]. If the speaker intends /ʔa/, produces [ʔa̰ʔ], and then the listener encodes /aʔ/, a phonological shift has occurred, simply because the input to the listener was ambiguous with regards to its underlying form.

c. CHOICE: when a listener hears multiple input forms for the same word and has to choose how to represent that word. For example, if a listener hears [kakáta], [kăkáta], and [kkáta] all for the word /kakata/, the listener might assume that /kkata/ rather than /kakata/ is the real underlying form (especially if it is more frequent in the input).

In the case of vowel harmony, the re-encoding of the co-articulation might be considered an example of *chance*, since the fronted [u̟] could be interpreted either as underlying /y/ or as underlying /u/.

All of these examples of channel bias can lead to change over time that looks similarly cross-linguistically. One of the main tenants of such theories is that sound change is never "goal driven", but rather occurs due simply to phonetic pressures that are shared by all human beings.

1.2.2 Learning bias

Beyond the phonetic pressures that may drive sound change and shift typology over time is bias situated in the individual learner. We refer to such bias as "learning bias", and this will be the focus

of this section. It has been proposed that certain phonological patterns may be more prevalent in the typology due to the fact that they are easier to learn than other patterns (C. Wilson, 2006). Note that this goes beyond the phonetic pressures proposed by channel bias, as it concerns a sort of preference in the *learning* of certain patterns over others, rather than in their simple transmission within the adult population. That is, a listener is more likely to assume pattern *a* than pattern *b*, based on some internal (cognitive) bias. For example, pattern *a* might be articulatorily easier or featurally simpler, and thus assumed by the learner to be more likely than pattern *b*. Note that this requires either concrete phonetic or abstract phonological knowledge (or both) on the part of the learner.

This hypothesis has been put to the test by teaching different phonological patterns to learners to see if certain patterns are learnt better or more quickly than others. For example, C. Wilson (2006) tested the learning and generalization of a typologically common pattern of velar palatalization, where velar stops affricate when followed by certain vowels. In the typology, this process is attested for high vowels, and also for mid vowels, with the caveat that if a language has velar palatalization triggered by mid vowels, it also has palatalization for high vowels. C. Wilson introduced a paradigm dubbed the *poverty of the stimulus method* (PSM), where input in the experimental task withholds critical information. For example, in one condition of his experiment, participants were exposed to alternations of the form /ge/ → /d͡ʒe/, and were then tested on alternations of the form /gi/ → /d͡ʒi/. Crucially, participants in that condition had never seen the high vowel /i/, and were thus being tested on their generalization of the pattern they learned on the mid vowel to the high vowel. Indeed, participants extended the palatalization pattern from mid to high vowels, but not from high to mid vowels, in line with linguistic typology. Other typologically recurrent phenomena have been shown to be learnt better or faster than unattested ones. Schane, Tranel, and Lane (1974), for example, showed that a rule akin to French liaison was learnt better than the opposite rule of intervocalic consonant deletion. Phonetically motivated natural classes are learnt more easily than arbitrary collections of segments (Peperkamp, Skoruppa, & Dupoux, 2006). And alternations that are phonetically closer are learnt more quickly than those that are more phonetically distinct (Skoruppa, Lambrechts, & Peperkamp, 2011). These studies showed a preference for natural rules or patterns compared to unnatural ones using an artificial language learning paradigm. In fact, learning biases for typologically common constraints have also been tested outside the domain of

phonology. Culbertson, Smolensky, and Legendre (2012), for example, considered a word order preference that is attested cross-linguistically, and showed that English-speaking learners preferred the cross-linguistically attested pattern when learning an artificial grammar, even though it mismatched with the surface frequency of the same word types in their own language. The implications of such work reach beyond assuming phonetic knowledge on the part of the learner, and actually indicate that learners might also have encoded some kind of structural preferences in their linguistic system.

Vowel harmony has often been a test case when exploring learning bias using artificial language learning. For example, Finley and Badecker (2008) tested the learning of the attested pattern of directional vowel harmony (agreement spreads from a segment positioned at the edge of the domain to all other segments to the left or right of it) compared to the learning of the unattested pattern of majority-rules vowel harmony (whichever feature is more present in the domain spreads to the other, minority, segments). Interestingly, constraint-based theories of phonology are able to predict the presence of both kinds of rules in phonological grammars, but only the former are present in the typology. The authors exposed their participants to alternations of the type [pidego] → [pidege]. Such alternations are ambiguous with regards to direction versus majority-rules based vowel harmony, as they adhere simultaneously to both. Learners may therefore infer either rule. During the test phase, the authors then showed participants alternations of the form [pumite] → [pimite] (majority-rules) versus [pumite] → [pumuto] (direction), and asked them which alternation they thought was part of the language they had just been exposed to. Participants overwhelmingly inferred direction-based vowel harmony. The authors proposed that such a preference reveals a learning bias that may influence typology.

Specifically regarding the question of harmony versus disharmony, we are faced with an interesting test case for learning bias. Experimental work has shown that logically more complex patterns are harder to learn than simpler patterns (e.g., Moreton, 2008; Pycha, Nowak, Shin, & Shosted, 2003; Skoruppa & Peperkamp, 2011; Skoruppa et al., 2011),[4] but both vowel harmony and vowel disharmony are equivalent with regards to their complexity. Specifically, they operate on the same feature and refer to either a + or a − value for that feature (e.g., Hungarian's ±*front*). If a learning bias for the phonetically natural rule can be demonstrated, it would suggest that above and beyond the

[4]See Moreton and Pater (2012a) for a review.

complexity of a rule, phonetic grounding influences the learning process. Thus, rules with stronger phonetic grounding are more likely to be typologically prevalent, because they are easier to learn than rules with weaker or no phonetic grounding. This hypothesis is in line with substance-based theories of phonology (e.g., Archangeli & Pulleyblank, 1994; Donegan & Stampe, 1979; B. Hayes, Kirchner, & Steriade, 2004), which assume phonetic knowledge on the part of the learner.

In a landmark study, Pycha et al. (2003) tested the learning of vowel harmony compared to vowel disharmony. They exposed English-speaking learners to CVC stimuli, where the vowel was either a front or a back vowel, and a morphophonological alternation where the plural suffix was either /-ɛk/ or /-ʊk/. They tested three conditions: one where the suffixes were harmonic with the stem, one where they were disharmonic with the stem, and one where the correspondence was arbitrary. While performance in the arbitrary co-occurrence condition was near chance level—showing more difficulty on the part of the participants in learning a more complex over a simple pattern—both the harmony and disharmony pattern showed learning by most participants. Overall, the harmony pattern was learnt slightly better than the disharmony pattern, but not significantly so. It is important to note, however, that the authors only tested ten participants per condition, so the numerical trend in the direction of harmony may not have reached significance, but could represent a small learning bias nonetheless.

Skoruppa and Peperkamp (2011) also tested the learning of vowel harmony compared to disharmony, but using a novel methodology. Rather than teach their French participants a new language, the authors exposed them to short stories of accented French. The accent was consistently either harmonic or disharmonic.[5] This means that a word like /pydœʁ/, which contains two rounded vowels, would be presented as is in the harmonic condition, but as /pydɛʁ/ in the disharmonic condition. Conversely, a word like /likœʁ/, which contains one rounded and one unrounded vowel, would be presented as is in the disharmonic condition, but as /likɛʁ/ in the harmonic condition. At test, participants were above chance level both for harmony and disharmony, but did not show better learning of the harmony pattern. The authors conclude that a bias favoring the learning of vowel harmony does not operate in a perception-only situation, and that a production constraint may be responsible for the typological asymmetry.

[5]Note that they tested rounding harmony rather than palatal harmony, but the principle is the same.

1.2.3 Summary

While many of the studies discussed in this section have shown evidence for the existence of learning bias, others have failed to show strong preferences in line with the typology. Chapter 3 re-evaluates the existence of a learning bias favoring vowel harmony over disharmony by testing a series of experimental parameters. We specifically explore the effects of modality (production versus perception) and sleep-related memory consolidation, and implement a computational model of transmission over many generations of learners. We focus our discussion on other influences present in artificial language learning experiments, specifically the native language of the learners being tested.

Chapter 2

Phonological similarity and the lexicon

2.1 Introduction

This chapter examines the phonological characteristics that bias the recognition of words. We will be specifically concerned with the concept of phonological features that distinguish phonemes from one another. Are certain features more or less important than others during the process of word recognition? We will explore this question from a couple of angles. We will begin by testing whether or not certain mispronunciations are more disruptive for word recognition than others (i.e., if a word is mispronounced along the voicing dimension, is the base form harder or easier to recover than if the same word is mispronounced along a different featural dimension?). Section 2.2 details an experiment aimed at probing this specific question.

Any differences that are observed may of course find their source in various phenomena. Notably, it will be important to disentangle language-specific usage-based explanations from language-independent phonetic explanations. Section 2.3 proposes a two-fold approach. First, we explore to what extent the different featural contrasts are perceived as distinct outside of lexical context. The idea behind this approach is that performance should be based solely on participants' ability to perceive *acoustic* differences between the contrasts, without being overly biased by the status of the contrast in the lexicon. Second, we consider the usage of the different contrasts in the lexicon. Are there contrasts that are used more often to distinguish words? The hypothesis underlying this ques-

tion is that contrasts that are used more frequently would yield better performances, as participants lend more attention to cues associated with these highly important contrasts.

2.2 Asymmetries in the exploitation of phonetic features for word recognition in French

This section is a reprint of the following article: Martin, A., & Peperkamp, S. (2015). Asymmetries in the exploitation of phonetic features for word recognition. *The Journal of the Acoustical Society of America*, *137*(4), EL307-EL313. The content has been re-typeset, but is otherwise identical to the original article, unless stated otherwise.

2.2.1 Introduction

The phonetic features that speech sounds are composed of have long been known to play a role in both speech production and speech perception. In speech production for instance, speech errors can target individual features (Fromkin, 1971), and featurally similar sounds are more likely to interact in speech errors than featurally dissimilar sounds (Stemberger, 1991). Concerning speech perception, identification errors in noise tend to preserve features of misheard sounds (Miller & Nicely, 1955), and consonant identification in dichotic listening is impaired when the two consonants are featurally dissimilar compared to when they are similar (Studdert-Kennedy, Shankweiler, & Pisoni, 1972).

In the realm of word recognition, previous research has investigated the limits of listeners' capacity to recognize words with deformed featural information, using stimuli with deliberate mispronunciations and a variety of tasks. This research has shown first and foremost that the number of feature changes between a word's correct pronunciation and a mispronounced variant is crucial in its recognizability. For instance, written words are primed by auditorily presented mispronunciations of semantically related words, but only if the mispronunciations differ in at most two features (Connine et al., 1993). This is hardly surprising, as the more a word's pronunciation is altered, the more difficult it should become to recognize. But what about the differential perceptual weight of the features themselves?

Everyday experience suggests that the voicing feature is not particularly important for word recognition. Indeed, whispered speech, which is characterized by the absence of vocal fold vibration and hence contains little information about the voicing of speech sounds, poses no specific problem for listeners. Based on experimental evidence, Cole et al. (1978) argue that voicing is indeed less important compared to place of articulation. They asked participants to detect mispronunciations of words in 20-minute-long stories. Words could be mispronounced in one or more features. In one experiment, mispronunciations concerned either the voicing or the place of the initial consonant. Changes in place of articulation elicited higher detection rates than changes in voicing. This effect was robust across the different consonants, indicating that the features were processed similarly regardless of the consonant to which they were associated. Note, though, that as participants could rely on the sentential context, the results might partly reflect the predictability of the particular words used in the experiment.

More recently, Ernestus and Mak (2004) used mispronounced words in isolation, avoiding biases introduced by sentential context. In their study, Dutch listeners performed a lexical decision task in which they had to reject words whose initial stop or fricative was mispronounced in voicing, place or manner. No differences in error rates according to the type of feature change were observed for mispronunciations of stop-initial words. For fricative-initial words, however, error rates were higher for voicing mispronunciations than for place or manner mispronunciations. Ernestus and Mak (2004) argued that these results reflect the fact that in Dutch, the voicing feature is relatively uninformative in word-initial fricatives. Indeed, fricatives (but not stops) are subject to phonological processes that change voicing word-initially, and in some varieties of Dutch all word-initial fricatives are realized as voiceless. Therefore, listeners would pay less attention to voicing than to place and manner, but only in fricatives.

In this article we investigate the relative weight of consonantal features for word recognition in French . Like Ernestus and Mak (2004), we focus on word-initial obstruents. The French obstruent inventory provides a particularly good test case to explore the various weights of phonological features, as its members are divided evenly over two manners of articulation, three places of articulation, and two voicing values Table 2.1. Thus, a change in any of the three features of an obstruent yields another obstruent.

	plosive		*fricative*	
	−v	+v	−v	+v
labial	p	b	f	v
coronal	t	d	s	z
dorsal	k	g	ʃ	ʒ

Table 2.1: The twelve French obstruents arranged vertically by place (in bold) and horizontally by manner (italics) and voicing.

Contrary to Dutch, French has no phonological processes affecting the class of word-initial obstruents or any of its subsets. Thus, if listeners weight their attention to individual features as a function of their informativity in the sense of Ernestus and Mak (2004) , we should observe no differences in French listeners among voicing, place, and manner mispronunciations. Furthermore, we explore the role that visual cues may play in word recognition, by presenting the stimuli in two modalities: audio-only and audiovisual. To the extent that the place features has salient cues in the visual signal, we expect that compared to audio-only input, audiovisual input should make mispronunciations of place – but not of manner and voicing – more disruptive for word recognition.

We report on an experiment in which participants heard a series of real, correctly produced words as well as mispronounced words interspersed among clear non-words. Their task was to detect both correctly pronounced and mispronounced words; the latter differed from their correctly produced counterparts in one feature of their initial phoneme, which was always an obstruent.

2.2.2 Methods

2.2.2.1 Stimuli

For all of the French obstruents but /z/, we selected three disyllabic French words which 1. contained no other obstruent, 2. yielded a non-word if any one of the features (voicing, manner, or place) was modified in the initial obstruent (e.g., the real word /deli/ 'misdemeanor', is turned into the non-words /teli/ through a voicing change, /zeli/ through a manner change, and /beli/ and /geli/, though place changes), and 3. had a higher frequency than all of their phonological neighbors, according to the Lexique 3.80 database (New, Pallier, Ferrand, & Matos, 2001). For /z/, only two such words were

selected, as the French lexicon does not have a third one satisfying the selection criteria.

For each of these 35 base items, we created mispronunciations that were non-words by changing one feature of the initial obstruent. Each base item thus yielded four mispronunciations, one with a voicing change, one with a manner change, and two with a place change. An additional 127 non-word fillers were randomly generated which were also disyllabic, contained only one, initial, obstruent, and had no real word phonological neighbors.

All base items, mispronunciations, and fillers were recorded individually by a female native speaker of French in a soundproof booth with an M-Audio Micro Track II digital recorder and an M-Audio DMP3 pre-amplifier in 16-bit mono at a sampling rate of 44.1 kHz. Video of the speaker including her whole face and stopping at her shoulders was simultaneously recorded at 60 frames per second in 720p HD resolution. The speaker was positioned in the center of the frame with two spotlights cross-positioned to eliminate shadows on her face. The average audio stimulus lasted 517 ms. Video was recorded both before and after the production of the word by our speaker such that each video lasted exactly 1,500 ms.

2.2.2.2 Procedure

Two versions of the experiment were prepared, one with audio-only stimuli and the other with audiovisual stimuli. Half of the participants were tested with the audio-only stimuli, the other half with the audiovisual stimuli. During the experiment, participants sat in front of a computer screen in a sound-attenuated room while stimuli were played binaurally through a headset. In the audio-only version, participants were presented with a black screen for the entire duration of the experiment. In the audiovisual version, videos were played in synchrony with the audio, depicting the woman producing the word being presented through the headset; in between stimuli, a gray box was displayed over the area of the screen where the speaker was portrayed. Participants were told that a list of items would be read to them by a stroke patient. The patient was said to have reading difficulties and, more specifically, to produce mostly unintelligible words when reading aloud individual nouns, while occasionally producing intelligible words or very close mispronunciations. Participants were asked to press a key (in the audio-only version) or a button (in the audiovisual version) whenever

they recognized a noun—whether it was pronounced correctly or incorrectly—(go response) and to do nothing otherwise (no-go response). If a go response was recorded, a dialogue box prompted the participant to report the recognized word by typing it on a computer keyboard. The next stimulus was then played after 1,000 ms.

Participants were presented with target stimuli (i.e., the 35 base items presented in the control condition [correct pronunciation] or in one of the three test conditions [voicing, manner, or place mispronunciation]), interspersed among filler stimuli (i.e., clear non-words). Participants were randomly assigned to one of six counterbalanced groups such that they only heard each base item in one condition; they thus did not hear any given sound manipulation twice. As a consequence, subjects heard on average 29 target stimuli (the number varied across participants, due to the absence of a third /z/-initial base word), and all 127 filler stimuli. Stimuli were presented semi-randomly, such that target stimuli never directly followed one another, with an ISI of 2,500 ms.

The experiment started with a short training phase, containing stimuli recorded by a different speaker than was used in the main task. In this phase, three target stimuli (one correctly pronounced noun and two mispronunciations) were mixed into a dozen filler stimuli. The mispronunciations concerned sonorants rather than obstruents so as not to interfere with the main task. Participants received feedback about their responses; if they failed to identify one of the mispronunciations, a message alerted them to their error and indicated the noun they were meant to identify. They had to correctly identify two out of the three target stimuli before moving on to the main task. If necessary, the training phase was repeated until this criterion was met.

2.2.2.3 Participants

Twenty-four native speakers of French participated in each version (audio and audiovisual), for a total of forty-eight. None of the participants reported any history of hearing problems and they all had corrected-to-normal vision.

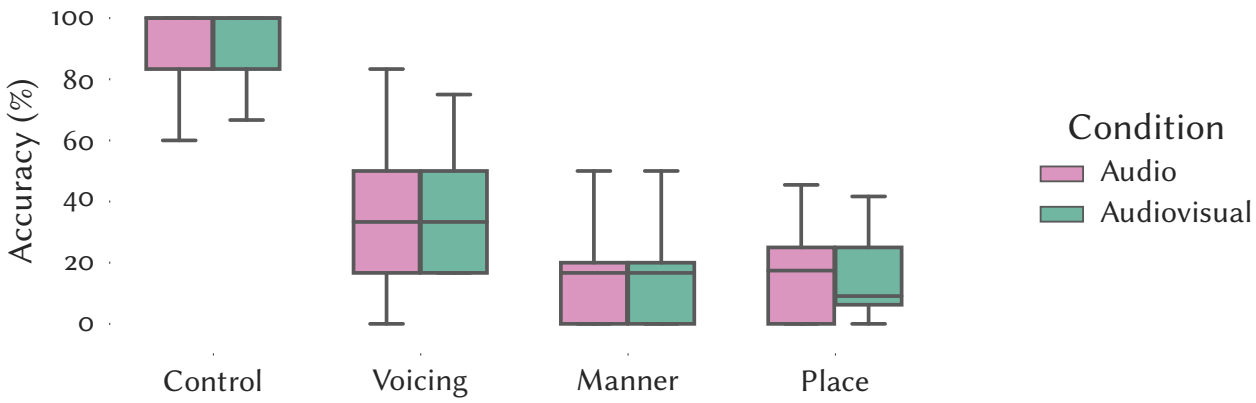

Figure 2.1: Participants' average accuracy by condition and modality.

2.2.3 Results

Six participants (two in the audio-only version and four in the audiovisual version) made more than 30% errors on control stimuli (correctly produced real words); their data were excluded from the analyses. Data from the forty-two remaining participants were analyzed using generalized mixed models in R (Bates, Maechler, Bolker, & Walker, 2014) with a declared binomial distribution given that the dependent variable was binary (hit or miss).

We analyzed individual hit rates on both test items (mispronunciations) and control items (correct pronunciations). For the former, a hit was defined as a go response with reporting of the correct target word (e.g., identification of mispronounced /teli/ as target /deli/ 'misdemeanor'). The mean hit rates per condition are shown in Fig. 2.1.

An original model was created with Modality (audio-only or audiovisual), Condition (correct, voicing mispronunciation, manner mispronunciation, or place mispronunciation), Frequency of base item, as well as all interactions as fixed factors, and Subject and Base Item as random factors. By-subject random slopes for Condition and Frequency were included in the model, as each participant was exposed to all four conditions and all frequency values. For the random factor Base Item, however, only a random slope for Condition was included given that Frequency perfectly correlates with Base Item. As neither Modality nor Frequency, nor any of the interactions was a significant predictor in this model (both $z < 1$), both were excluded as factors from subsequent models.

Our final model included Condition as a fixed factor and Subject and Base Item as random factors, each with a random slope for Condition. All three mispronunciation conditions were significantly different from the control condition (i.e., correct pronunciation): voicing (β = 4.44, SE = 0.72, z = 6.12, $p < 0.0001$), manner (β = 6.07, SE = 0.79, z = 7.69, $p < 0.0001$), place (β = 6.32, SE = 0.76, z = 8.33, $p < 0.0001$). Furthermore, the voicing mispronunciation condition differed significantly from both the place (β = 1.92, SE = 0.40, z = 4.76, $p < 0.0001$) and manner (β = 1.65, SE = 0.42, z = 3.92, $p < 0.0001$) mispronunciation conditions. However, there was no such difference between the place and manner mispronunciation conditions ($z < 1$).

2.2.4 Discussion

Overall performance in the mispronunciation conditions was low, indicating that participants had a hard time recognizing words with even a one-feature change. Nonetheless, clear differences were found between the types of feature changes, indicating that different features have different degrees of importance for word recognition. Specifically, participants were able to identify words that were mispronounced with a voicing change significantly more often than ones that were mispronounced with a manner or a place change. Thus, voicing mispronunciations are less detrimental to word recognition, indicating that listeners give less weight to the voicing feature compared to the place and manner features.

This finding is in line with Cole et al. (1978), who explored mispronunciations in context. Recall that in their study, English-speaking participants were presented with stories in which they had to detect mispronounced words. Performance was better on words with a mispronunciation of the place than of the voicing feature, indicating that the former hamper lexical access more than the latter. The predictability of the particular words manipulated in the stories might have biased the results, though. In the present experiment, we used isolated words and, moreover, ruled out item-specific effects, as each item was used in all mispronunciation conditions and had a higher frequency than all of its phonological neighbors. This, then, attests to the robustness of the larger weight of both place and manner compared to voicing.

The present results are also similar to the ones obtained by Ernestus and Mak (2004) with Dutch lis-

teners. Using a lexical decision task, they also focused on word-initial obstruents. Recall that they observed no difference between place and manner mispronunciations, and higher error rates for voicing mispronunciations in fricatives. Thus, Dutch listeners pay less attention to voicing word-initially, but only in fricatives. The number of trials in our experiment does not allow us to properly analyze mispronunciations in stop- and fricative-initial words separately. Note, though, that our finding that voicing is overall less important than place and manner is unexpected given Ernestus and Mak (2004) hypothesis that a feature's relative weight is determined by its informativity within the language. Indeed, word-initially, all obstruent features are equally informative in French, at least according to the definition of Ernestus and Mak (2004), as none of the French obstruent features is modified by a phonological process in word-initial position.[1] Thus, their hypothesis can explain the results for Dutch, a language in which voicing is relatively uninformative in word-initial fricatives, but not for French. This does not imply that the idea of a feature's informativity determining its weight should be abandoned altogether. Other factors might influence the informativity and hence the weight of individual features. In particular, a feature's functional load (i.e., the extent to which it is used to distinguish words from one another in the lexicon), might play a role: the higher a feature's functional load, the more we would expect listeners to pay attention to it during word recognition. Calculations of functional load have traditionally focused on segment pairs rather than individual features (Hockett, 1967), and although more recent formalisms have been adapted to explore featural comparisons (Surendran & Niyogi, 2003), different measures of functional load are still being debated (cf. Wedel, Kaplan, & Jackson, 2013). This current state of affairs makes testing the hypothesis difficult. As a first step, however, it would be interesting to compare the relative importance of phonological features in prelexical versus lexical processing, since lexical factors such as functional load should not affect prelexical perception. Of course, confusion studies such as the one by Miller and Nicely (1955) have already examined the relative importance of features in prelexical perception, and shown that – contrary to all findings with lexical tasks – place is more likely to be misperceived than voicing. However, these studies used stimuli presented in noise. As noise masks the spectral properties of sounds differentially, this finding is uninformative for the present research question. Future studies addressing this question should therefore use a prelexical task with stimuli

[1]Word-finally, the voicing feature is subject to assimilation, but based on their data from Dutch listeners, Ernestus and Mak (2004) argue that a feature's informativity is specific to its position within the word.

presented in clear speech to determine, all things being equal, which features are more perceptually salient.

Finally, we did not observe a difference between the audio-only and the audiovisual versions of the task. To our knowledge, no previous studies have explored multimodal interaction in the processing of mispronunciations, but given that visual information contains cues to place but only secondarily to manner[2] and not at all to voicing, we expected that in the audiovisual version participants' ability to recognize words with a mispronunciation of the place feature would be even more reduced. It is possible we did not obtain such a difference because of a floor effect; indeed, even in the audio-only version performance in the place condition was low. Alternatively, it has been argued that visual cues are especially relied upon under difficult listening conditions, for instance when the auditory signal is presented in noise (e.g., Sumby & Pollack, 1954). Yet another potential explanation comes from recent research suggesting that visual input might be involved in prelexical but not in lexical processing (Samuel & Lieblich, 2014). More research is necessary to tease apart these possibilities and to explore the role of visual input in the processing of mispronunciations.

2.3 Assessing the distinctiveness of phonological features in word recognition: prelexical and lexical influences

This section is a reprint of the following article: Martin, A., & Peperkamp, S. (2017). Assessing the distinctiveness of phonological features in word recognition: prelexical and lexical influences. *Journal of Phonetics*, *62*, 1–11.

2.3.1 Introduction

What makes two words sound similar to each other? Consider the English word pin - /pɪn/. Intuitively, we can understand how a word like shin - /ʃɪn/ sounds more similar to pin than a word like train - /tɹe͡ɪn/ does. Indeed pin and shin form a minimal pair; the two words are minimally different,

[2] In particular, labial stops and fricatives visibly differ in that the former are bilabial and the latter labio-dental. We do not have enough trials to analyze an effect of modality on labial obstruents only.

in that they share all but one phoneme. Yet cross-modal priming experiments have shown that a word like bin - /bɪn/, which also forms a minimal pair with pin, more strongly activates pin than shin does (e.g., Connine et al., 1993; Milberg et al., 1988). This is because the segments that distinguish pin from shin share fewer phonological features than those that distinguish pin from bin. Now consider the word tin - /tɪn/. Both the /t/ in tin and the /b/ in bin are one feature different from the /p/ in pin (a difference in place and voicing[3] respectively). Is the nature of the featural difference pertinent for the notion of similarity?

Research on lexical perception has demonstrated that featural differences in one's native language are not all perceived as equally distinct. In both English (Cole et al., 1978) and Dutch (Ernestus & Mak, 2004), mispronunciations have been shown to be less disruptive for word recognition (i.e., easier to recognize) if they involve a change in voicing than if they involve a change in place or in manner. This indicates that a difference in voicing is perceived as less stark than a difference in another major class feature in these languages. More recently, Martin and Peperkamp (2015) exposed French listeners to a series of auditorily- (or audiovisually-) presented nouns supposedly produced by a stroke patient. These included correctly pronounced words, mispronounced words, and non-words that did not resemble any real word. Participants were asked to press a button when they recognized a word - whether it was correctly pronounced or mispronounced - and report it. All mispronunciations involved a change in one of the major class features: voicing, manner, or place on a word-initial obstruent. The results from the audio-only version of that experiment, reported as the proportion of correctly identified mispronounced words, are reproduced in Fig. 2.2[4]. Similar to the previous findings for English and Dutch (Cole et al., 1978; Ernestus & Mak, 2004), words with a voicing mispronunciation were more likely to be recognized than those with a manner or a place mispronunciation. For example, the word sommet, /sɔmɛ/ - "summit" was more likely to be recognized when it was presented as /zɔmɛ/, with a mispronunciation in voicing, than when it was presented as /fɔmɛ/ or /tɔmɛ/ (a place or manner mispronunciation, respectively). Thus, the voicing feature's role in contrasting words from one another is perceived as different than that of the other features.

[3]Note that throughout this paper we will refer to any two-way laryngeal contrast as "voicing", although the phonetic realization of this contrast may vary across languages.

[4]The results were not significantly different by modality.

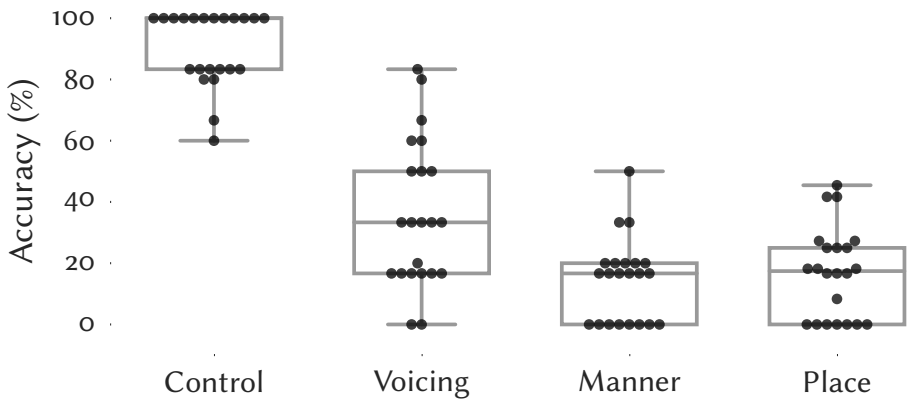

Figure 2.2: Boxplot of participant means from the audio-only version of the mispronunciation detection task reported in Martin and Peperkamp (2015) by condition. The central line in the boxplot represents the median; the space between the central line and the bottom or top of the box represents the second and third quartile spread; and the distance from the bottom or top of the box to the tip of the whiskers represents the first and fourth quartile spread. In the dotplots, each dot represents an individual's score.

The sources of this asymmetry remain unclear, however, and could be due to a number of factors. Most obvious is the acoustic proximity of the sounds being considered. Some sounds are acoustically closer, and will thus be perceived as more similar than other, distant, sounds. A further source of bias is language-specific knowledge. Listeners may use knowledge of their native language for the purposes of efficient word recognition. That is, they may preferentially attend to cues associated with featural contrasts which are more informative in their language. Indeed, listeners are influenced both by acoustic information and by language-specific knowledge (Ernestus & Mak, 2004; Johnson & Babel, 2010). Ernestus and Mak (2004), for example, argued that Dutch listeners are influenced by a process of initial fricative devoicing in their language, which renders voicing information on these segments uninformative. This would explain why these listeners ignore voicing mispronunciations more often than manner mispronunciations in a lexical decision task. Similarly, Johnson and Babel (2010) found language-specific influence using a similarity judgment task. They had native English- and Dutch-speaking participants rate the similarity of pairs of VCV non-words containing English fricatives, and showed that Dutch listeners rated [s], [ʃ], and [θ] as more similar to each other than English listeners did. They argued that this is due to the phonological status of these sounds in the respective languages. While all three sounds are distinctive in English, [ʃ] and [θ] are not phonologically distinctive in Dutch; the former is a contextual allophone of /s/ and the

latter does not occur at all. However, in an AX discrimination experiment, Dutch listeners' response times were not shown to differ from English listeners'; both groups discriminated the same pairs of sounds equally well. The authors argued that their discrimination task reveals low-level acoustic differences between the stimuli, with some of the contrasts yielding longer response times because of their acoustic proximity (e.g., [f]~[θ] and [h]~[x]), regardless of the native language of the listener, while their similarity judgment task reveals language-specific influences, with Dutch listeners being perturbed by the absence of [θ] and [ʃ] as phonemes in their native language.

Note, though, that this reasoning does not explain why in English and French, voicing mispronunciations are also harder to detect (Cole et al., 1978; Martin & Peperkamp, 2015), because the voicing feature is fully distinctive in these languages (voicing contrasts can be neutralized in English and French but never word-initially). These results do not necessarily imply that listeners are not influenced by lexical patterns during word recognition. Indeed, following, inter alia, Hall (2013), we argue that a more gradient understanding of "distinctiveness" is necessary to properly address this issue. If, for example, there were fewer voicing minimal pairs than place and manner minimal pairs in English and French, this could explain why words presented with voicing mispronunciations were perceived as closer to the target word. Here, we further explore gradient distinctiveness using a combination of experimental and computational techniques.

The specific aim of our research is to disentangle low-level, prelexical influences from top-down, lexical ones in word recognition. To this end, we take French obstruents as a case study, allowing for a direct comparison with the results on lexical perception from the mispronunciation detection task reported in Martin and Peperkamp (2015), which we take as our starting point. Building on those results, we start off with an examination of the way phonetic differences between features are perceived outside of lexical context, using a prelexical discrimination task. We then examine the French lexicon by measuring the functional load of various feature contrasts as a proxy for the lexical knowledge shared by speakers of French. This allows us to understand if there are asymmetries in the usage of these different features, even though they are not affected by any phonological process. Finally, we compare our results with the word recognition results reported in Martin and Peperkamp (2015), and propose that the relative weight of phonological features during word recognition is determined jointly by the role of these features in both bottom-up acoustic perception and top-down

lexical knowledge.

2.3.2 Prelexical perception

The perceptual similarity of speech sounds has been investigated for decades, focusing mostly on the effects of different types of noise on perceptual confusion (e.g., Bell, Dirks, & Carterette, 1989; Cutler, Weber, Smits, & Cooper, 2004; Miller & Nicely, 1955; Weber & Smits, 2003). For instance, Miller and Nicely (1955) presented a series of English syllables embedded in different kinds of noise (including low-pass filtering and white noise) at various signal-to-noise ratios (SNR) and asked participants to report what consonant the syllable began with. They found that place of articulation was more likely to be confused than voicing, across consonants and across different SNRs. While this line of research is important for understanding speech perception in noisy conditions, it cannot provide us with an accurate baseline of perceptual similarity of speech sounds, because noise affects individual features differentially (for discussion, see Bell et al., 1989; Cutler et al., 2004).

Some studies have addressed the question of perceptual similarity in silence. However, in the absence of noise, listeners are exceedingly good at identifying sounds in their native language, hence observing differences between different types of contrasts is difficult. For instance, Plauché et al. (1997) found that Spanish listeners correctly identify the initial consonant of the syllables /pi/, /ti/, and /ki/ in 95.4% of trials; only when the stimuli were artificially manipulated did participants begin to confuse them. Wang and Bilger (1973) similarly reported ceiling performance in syllable identification, unless the stimuli were presented at low volume. Finally, in a study on the perceptual confusability of currently merging vowels in Parisian French, Hall (2013) obtained identification scores at ceiling for control, non-merging vowels (an average of 93%).

Thus, in ideal listening conditions, native sounds are reliably identified. How, then, can we measure perceptual similarity without degrading the acoustic signal? Many studies have used explicit similarity judgments: participants are asked to compare two pairs of non-words that each differ in one segment, and explicitly state which pair they find to be more similar. This methodology has been used, for example, to explore the role of feature differences in word similarity (Bailey & Hahn, 2005; Hahn & Bailey, 2005). These studies have revealed that the more phonological features the

two differing sounds share, the more likely the non-words are to be judged as similar, in line with research from the lexical perception literature mentioned above (Connine et al., 1993). Although this same methodology could be applied to our current research question, related to the nature of the feature differences themselves, the fact that explicit similarity judgments require participants to metalinguistically reflect on the stimuli makes this paradigm less than ideal. Johnson and Babel (2010) compared their similarity judgment results with those from a discrimination task and showed that language-specific influences were more readily reflected in the explicit judgment task. Results of their discrimination task, they argue, were more driven by low-level acoustic perception. In the present study, we will therefore likewise use a discrimination task to test perceptual similarity in listeners' native language.

Discrimination tasks are routinely used to assess the perception of non-native and second language sound contrasts. Here, we design an ABX discrimination paradigm that aims at avoiding ceiling effects when used with native listeners. First, the stimuli are produced by multiple synthesized speakers, two male and one female, thus augmenting acoustic variability between tokens. Second, we use long (trisyllabic) stimuli, thereby increasing working memory load. Third, and most importantly, in each AB pair for which a given consonantal contrast is being tested, only one vowel and two consonants are used. For instance, for a trial with the contrast /p/-/b/, A and B might be /pa**b**aba/ and /pa**p**aba/. The fact that the crucial consonants occur multiple times should make the task particularly difficult. Note that in the given example the difference between the two items lies in the second syllable (in bold); by randomly varying this position across trials we make the task even harder, since participants cannot predict where the crucial difference will appear in a given trial.

We use this methodology to test the perceptual similarity of French obstruents that differ in only one phonological feature.

2.3.2.1 Methods

We follow Martin and Peperkamp (2015) in studying "one-feature" (major class features) obstruent contrasts in French. French has twelve obstruents that are defined by these three featural contrasts:

	/a/	/i/	/u/
/p/ ×1, /b/ ×2	/pababa/	/pibibi/	/pububu/
	/bapaba/	/bipibi/	/bupubu/
	/babapa/	/bibipi/	/bubupu/
/p/ ×2, /b/ ×1	/bapapa/	/bipipi/	/bupupu/
	/pabapa/	/pibipi/	/pubupu/
	/papaba/	/pipibi/	/pupubu/

Table 2.2: Non-word items used for the contrast /p/~/b/.

voicing (voiced vs voiceless), manner (stop vs fricative), and place (labial vs coronal vs post-coronal). This yields twenty-four one-feature contrasts (Table 2.1).

2.3.2.1.1 Stimuli

For each of the 24 one-feature consonant contrasts, we constructed 18 non-word items, for a total of 432 items. Each item had the structure CVCVCV. Its vowels were identical and were drawn from the French point vowels (i.e., /a/, /i/, and /u/). The consonants were the ones from the contrast under consideration; one of them occurred once and the other one twice. By way of example, the complete set of 18 items for the contrast /p/~/b/ is shown in Table 2.2. Note that it is either one or the other making up the contrast, and that it occurs either in the first, the second or the third syllable.

All stimuli were synthesized using the Apple Say program's diphone synthesizer; each was produced by three of the European French voices: two male (*Thomas* and *Sébastien*) and one female (*Virginie*). This gave us a total of 1,296 unique tokens (432 items × 3 voices). We chose to use synthesized stimuli given the large number of items and their tongue twister-like construction. The stimuli had a mean duration of 727 ms (± 85 ms), and sounded relatively natural. They may be downloaded from the first author's website.

2.3.2.1.2 Procedure

A total of 1,728 unique trials were created by combining the stimuli into ABA, ABB, BAA, and BAB trials. These were counterbalanced into twelve 144-trial-long lists (with each participant seeing

only one list), such that each list contained a total of six trials for each of the twenty-four obstruent contrasts, including two trials for each vowel /a, i, u/.

Participants sat in front of a computer screen in a sound-attenuated room while stimuli were played binaurally through a headset. They read instructions presented on screen that described the ABX paradigm: In every trial, the first two non-words they heard would be different and the third one would always correspond to either the first or second one; they should indicate whether the third stimulus matched the first or the second; and they should give their response by pressing one of two buttons on a response box.

In each trial, participants heard a sequence of three stimuli, each produced by a different voice, with an ISI of 300 ms. Thus, X was acoustically different from both A and B. The position of the difference between A and B changed from one trial to another, making it impossible for participants to know where specifically to attend upon hearing the A stimulus of a given trial. The position of the difference was counterbalanced across trials. Voice order was randomized. For example, a participant might hear /papaba/$V_{irginie}$ - /pababa/T_{homas} - /papaba/$Sé_{bastien}$, to which they should respond 'A'. In another trial, they might hear /putupu/T_{homas} - /tutupu/$V_{irginie}$ - /tutupu/$Sé_{bastien}$, to which they should respond 'B'. Following a given trial, the next was presented 1,250 ms after the participant had given a response, or after a timeout of 2,000 ms after stimulus offset, whichever came first. If participants failed to give a response before the timeout, this was counted as an error. No feedback was given to participants during the experiment.

The experiment lasted about 20 minutes.

2.3.2.1.3 Participants

Forty-eight native speakers of French participated (34 women, 14 men) were randomly assigned to one of the twelve counterbalanced lists. They were aged between 18 and 35 (mean: 25.1). None of them reported any history of hearing problems.

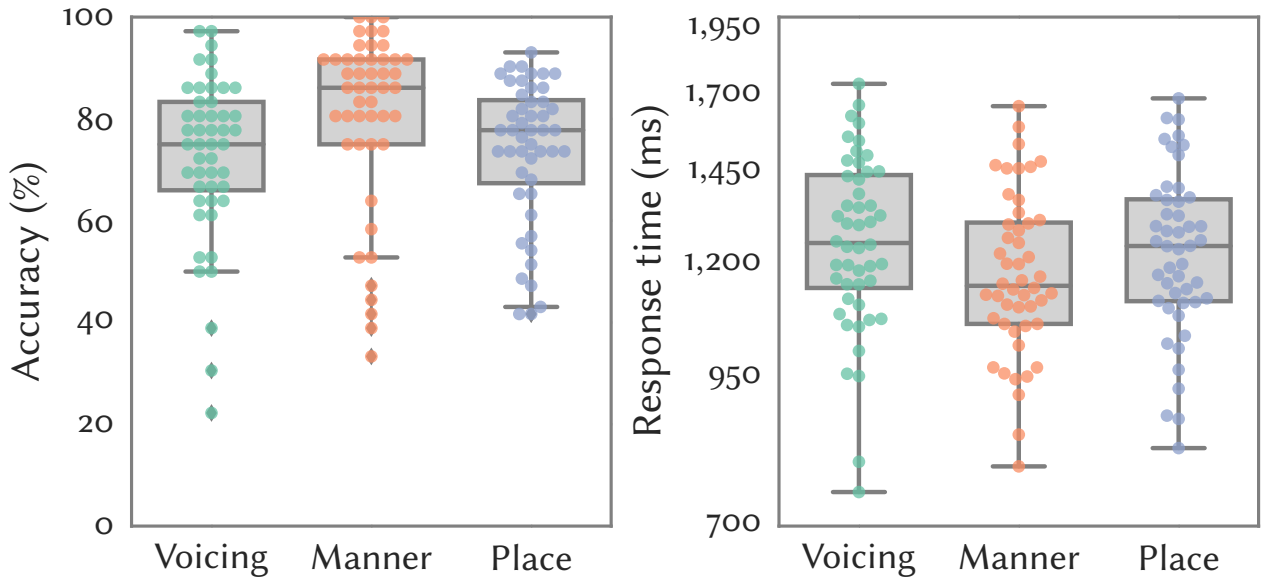

Figure 2.3: Box- and dotplots of participant means of accuracy (left) and response times (on a log scale) on correct trials (right) by feature.

2.3.2.2 Results

The mean accuracy scores, in addition to the response times for correct trials (measured from the onset of the third stimulus) per phonological feature, are shown in Fig. 2.3. All analyses were performed using mixed-effects models in R (Bates et al., 2014). Accuracy was analyzed with a logistic mixed-effects regression model; the log response times were analyzed using a linear mixed-effects regression model.

Both models included random intercepts for Participant and Contrast (the phoneme pair). Original models were constructed that also included a random slope for Feature by Participant, but as these models did not converge, we simplified the random effects structure. The final models included the factor Feature (voicing vs manner vs place) and random intercepts for Participant and Contrast. We took one of the values of the Feature factor as intercept and then releveled the data to assess the three-way comparison. We started by taking manner as the model intercept, comparing manner to voicing and manner to place. We then took place as the intercept, which allowed us to compare it to voicing (and redundantly to manner). The implementation of the analysis is available on the first author's website.

In both the accuracy scores and the response times, manner was found to be significantly different

from voicing (accuracy: $\beta = -0.50$, $SE = 0.15$, $z = -3.36$, $p < 0.001$; RT: $\beta = 0.06$, $SE = 0.03$, $t = 2.30$)[5], while manner was found to be different from place in accuracy ($\beta = -0.43$, $SE = 0.13$, $z = -3.30$, $p < 0.001$), but only marginally so in response times ($\beta = 0.04$, $SE = 0.02$, $t = 1.83$). Voicing did not differ from place in accuracy ($\beta = -0.07$, $SE = 0.13$, $z < 1$) or response times ($\beta = 0.02$, $SE = 0.02$, $t < 1$).

These results indicate that manner contrasts yielded more correct responses than both place and voicing contrasts, and that on correct trials, manner contrasts yielded faster responses than voicing contrasts. Furthermore, place and voicing contrasts did not yield different results from each other in either accuracy or RT. In other words, manner contrasts are generally perceived as being more distinct than the other two types of contrast.

2.3.2.3 Discussion

In this experiment, we tested the perceptual similarity of French obstruents that differ from one another in only one phonological feature, using an ABX discrimination task. We found significant differences in both accuracy and response times for manner contrasts compared to place and voicing contrasts. That is, participants were both more accurate and faster to discriminate obstruents differing in manner than obstruents differing in place or voicing. This result indicates that obstruents differing in manner are on average perceived as being more distinct than those differing in place or voicing, even though all of the contrasts we tested are phonologically distinctive in French.

From an acoustic point of view this makes sense, given that the manner contrasts we tested differentiated stops from fricatives. That is, stops are characterized by a period of silence followed by a burst, whereas fricatives involve aperiodic noise throughout their production; intuitively, the difference between a period of silence and a period of noise is easy to perceive. Unsurprisingly, our results do not mirror those mentioned above concerning the perception of speech in noisy conditions. Indeed, the difference between a period of silence (stops) and a period of noise (fricatives) can be easily masked when additional noise is superimposed onto the stimuli, thereby diminishing per-

[5]The `lme4` package does not provide p values for linear models; a t value greater than 2 is usually considered to be significant. Given our experimental design (specifically the number of participants and the number of items), this method has been shown to have the lowest Type I error of the common methods for assessing significance of mixed-effects models (Luke, 2016).

ceptual differences between the two. The presence or absence of low-frequency periodicity (voiced vs voiceless sounds respectively), and the formant transitions associated with distinguishing different places of articulation may resist such noise manipulation better. The present results, then, give credence to our claim that the study of speech in noise cannot provide us with an accurate baseline of perceptual similarity.

We verified that stark acoustic differences drive effects in the prelexical task by performing acoustic analyses on our stimuli. We used the spectral package in Python (Versteegh, 2015) to extract forty-dimensional Mel filterbanks coefficients with a cubic-root compression for each stimulus; these acoustic features are meant to roughly reflect the way the sounds are represented at the level of the cochlea. We then used the ABXpy package (Schatz, 2016; Schatz et al., 2013) to model predicted performance in an ABX task taking into account only the acoustic properties of the stimuli. This model uses Dynamic Time Warping (Rabiner & Juang, 1993) to measure the acoustic distance between the stimuli A and X on the one hand and B and X on the other hand, in order to predict a response based on which distance is shorter (i.e., if the A-X distance is shorter than the B-X distance, the predicted response is A). While the model's overall performance is lower than that of the participants in our experiment, the pattern is exactly the same: The model performs better on manner contrasts (mean: 69.6%) than it does on voicing (mean: 61.5%) or place (mean: 60.8%) contrasts, with performance on the latter two being nearly identical.

Previous attempts to measure perceptual similarity of speech sounds in silence have yielded ceiling performance, and it is only when noise is added that stimuli begin to be confused by participants. Our paradigm, on the other hand, allowed us to measure asymmetries in processing without degrading the speech sounds. The similarity of the A, B, and X tokens (including many repetitions of each of the target sounds) made the task sufficiently difficult that we did not observe ceiling performance. This crucially provides us with a baseline of perceptual similarity of the sounds we tested, and our task could also be used to study native listeners' perception of other consonantal, vocalic, or even tonal contrasts.[6]

Returning to the question of word-level similarity, the results from this experiment are only partially

[6]We actually used the experimental design reported in this study to test one-feature vowel contrasts in French as well. The results of that study are not reported here, but the data are available on the first author's website.

in line with Martin and Peperkamp (2015)'s lexical results, according to which words with either a manner or a place mispronunciation are harder to recognize as the intended real words than words with a voicing mispronunciation. Although the present result can explain why manner mispronunciations are more disruptive for the purposes of word recognition (i.e., they are perceived as being more different from the canonical pronunciations), it does not explain why place mispronunciations are equally disruptive. We hypothesize that the latter result is due to another bias, namely lexical knowledge. Indeed, lexical effects can be observed at many levels of speech processing, including phonological judgments (Hay, Pierrehumbert, & Beckman, 2004). In the following section, we quantify the relative weight of phonological features in the lexicon, using a new measure of functional load, which we propose as a proxy for lexical knowledge.

2.3.3 Functional load

The term functional load, in a broad sense, refers to the amount of work a phonemic contrast does in a language to distinguish words from one another. Consider the English distinction between /θ/ and /ð/ (the "th" sounds in think and that respectively). Although these sounds are distinctive in English, they actually only disambiguate a handful of words (e.g., ether~either in American English), and the contrast is therefore considered to have a low functional load. Compare this to the high functional load of the contrast /p/~/t/, which disambiguates a great many pairs of words (e.g., pack~tack, pin~tin, cope~coat).

Functional load has been proposed as a key factor in language change (Martinet, 1955). Specifically, contrasts that have low functional load are predicted to be more likely to merge over time than contrasts that have high functional load. This hypothesis has been put to the test by, for instance, Wedel et al. (2013), who showed that in many languages, contrasts that have low functional load are indeed more likely to merge over time. They compared two specific measures of functional load: minimal pair counts, and difference in information entropy. The first is rather straightforward. Reconsider the English examples from above: only seven pairs of words are disambiguated by the /θ/~/ð/ contrast, compared to well over 300 for the /p/~/t/ contrast. The second measure of functional load, based in information theory (Shannon, 1948), concerns information entropy (Hockett, 1955).

This is a quantification of the uncertainty of the system, with the lexicon considered to be a complex system. The higher the functional load of a given contrast, the more "uncertainty" is removed from the system, should that contrast be excised. The measure is therefore calculated as the difference in entropy of the system with or without the contrast; the greater the difference, the higher the functional load. For a detailed mathematical description of the calculation of this measure, including its application to different levels of analysis, see Surendran and Niyogi (2003, 2006). While the entropy measure has been widely used (e.g., Hume & Mailhot, 2013; Severen et al., 2013; Stevenson, 2015; Stokes, Klee, Carson, & Carson, 2005), Wedel et al. (2013) found that, overall, minimal pair counts are a more accurate predictor of sound change, and more recent work has backed up this finding (Wedel, 2015).

Here, we are interested in the functional load of phonological features (e.g., voicing), rather than of individual phonemic contrasts (e.g., /p/~/b/). Previous research has indicated that a handful of these individual contrasts tend to be responsible for distinguishing a disproportionate amount of words from one another in the lexicon of a given language (Oh, Coupé, Marsico, & Pellegrino, 2015), but has not examined whether these contrasts concern the same phonological feature. Thus, are high functional load phoneme pairs likely to contrast in the same feature? Minimal pair counts and entropy measures have been adapted to respond to this question, but the results are slightly problematic. For minimal pairs, calculating such a score equates to summing the minimal pair counts for all contrasts in a given feature; for entropy, calculating scores for features rather than for phoneme contrasts can be done by summing the scores for all phoneme contrasts within a given feature. For instance, for both minimal pair counts and entropy, in a four-phoneme system like /p, b, t, d/, the score for place would be the summed scores for /p/~/t/ and /b/~/d/. Likewise, the score for voicing would be the summed scores for /p/~/b/ and /t/~/d/; thus one value per feature. Surendran and Niyogi (2003) report a series of values obtained by applying the entropy measure in this way cross-linguistically, but their method raises the question of how to interpret an absolute difference of, say, 0.002 or 0.009 bits.[7] This problem pertains to all the existing implementations of both the entropy and the minimal pair methods. Are observed differences meaningful, or do they simply reflect some kind of noise? Assessing the significance of a difference typically relies on inferential statistics, for which

[7]Entropy is measured in "bits" of information.

distributions of scores rather than single numbers are required. Below, we propose a new method of measuring functional load that is based on minimal pair counts but that allows for the use of inferential statistics. Our method also differs from previous ones in another aspect. Existing minimal pair counts and entropy methods provide scores of the absolute functional load of a given contrast within a system. They are thus affected by the individual frequency of the phonemes that form the contrast. That is, the theoretically maximum functional load of a phoneme contrast depends upon the frequency of the phonemes in question. This issue has been reported in a previous study on the correlation between functional load and perceptual similarity (Hall, 2009). It is also problematic for our present purpose of measuring the functional load of phonological features, as the functional load of a feature (e.g., voicing) is a function of the functional load of the relevant phoneme contrasts (e.g., /p/~/b/, /t/~/d/, ...). We ask here to what extent each feature is used, all else being equal. We thus propose a relative measure of functional load, which abstracts away from individual phoneme frequency.

In the same vein, our measure abstracts away from phonotactics, that is, the constraints governing the combination of sounds in a language, which may make a given contrast impossible in certain positions. In French, for instance, the initial cluster /tl/ is not permitted (Dell, 1995). It is therefore impossible for /pl/-initial words, such as /plɥi/ (pluie, "rain"), to be changed into another word by replacing /p/ with /t/. We consider this to be uninformative regarding the distinctive weight of /p/~/t/ (or of the place feature for that matter). We specifically place phonotactic knowledge on a different level from lexical structure. Our question should instead be framed as: When /p/ can contrast with /t/, does it? And how does the frequency with which it does compare to the frequency with which /p/ contrasts with /b/ when replacing /p/ by /b/ is phonotactically legal?

Below, we detail our new method.

2.3.3.1 A new measure of functional load

Given the previous success of minimal pair counts in assessing the functional load hypothesis for sound change (Wedel et al., 2013), our measure of the functional load of phonological features is based on minimal pair counts. It consists of an observed-over-expected ratio (henceforth O/E ratio),

as defined in Eq. (2.1).

$$\text{O/E ratio}_{i,j} = \log\left(\frac{o_{i,j}}{e_{i,j}}\right) \tag{2.1}$$

For each phoneme *i*, and each feature *j*, an observed-over-expected score is calculated, where *e* represents the number of possible (or expected) minimal pairs and o the number of observed minimal pairs for that phoneme in that feature. This function is iterated over the lexicon for each feature, and each phoneme. For example, consider the French phoneme /p/ and the place feature (here, specifically the change from /p/ to /t/, although the change from /p/ to /k/ would also need to be included). Upon encountering the word /po/ (*peau*, "skin"), a minimal pair is theoretically possible (i.e., a change in place on the segment /p/ yields the phonotactically legal word /to/). We consider that if the lexicon maximally exploited all contrasts, then we should expect to find a minimal pair between /po/ and /to/. This is precisely what the value of *e* is meant to represent. Thus $e_{/\text{p}/,\text{PLACE}} = 1$. Furthermore, the word /to/ does exist (*taux*, "amount"). Thus $o_{/\text{p}/,\text{PLACE}} = 1$. If we next consider a case such as /pjɛʒ/ (*piège*, "trap"), we observe that the theoretical minimal pair it would form with a place change is possible (i.e., /tjɛʒ/ is phonotactically legal). Thus $e_{/\text{p}/,\text{PLACE}} = 2$ now (the scores are cumulative as we iterate over the lexicon). However, this word does not actually exist in French, and o/p/,PLACE therefore remains at 1. Now if we consider the case of /plɥi/ (*pluie*, "rain"), we know that the theoretical minimal pair it would form from a place change is not possible (i.e., /tlɥi/ is ruled out by the French phonotactic constraint that words cannot begin with /tl/), thus $e_{/\text{p}/,\text{PLACE}}$ remains at 2. Of course, if a minimal pair is not possible, it will not be observed, and indeed $o_{/\text{p}/,\text{PLACE}}$ remains at 1. This process is repeated over the entire lexicon for each combination of a phoneme and a feature (so we might next consider manner minimal pairs for the sound /p/, or place minimal pairs for the sound /t/, until all combinations were exhausted), yielding distributions of scores for each feature. These scores are log-transformed to ensure they follow a normal distribution; thus a score of zero represents the highest possible functional load. That is, if every possible minimal pair were attested, the score for that phoneme and that feature would be zero. Additionally, if no minimal pairs are observed at all, the score will be $-\infty$. Note further that the operation is performed over lemma forms, not over the unique set of phonological forms, so if two lemmata have the same phonological

form (i.e., they are homophones), they are both counted.

2.3.3.2 Applying O/E ratios to French nouns, and comparison with entropy

As our current question is about exploring different sources of influence on word recognition, so as to understand asymmetries reported in the literature, we focused our analysis on French nouns, the class of words tested by Martin and Peperkamp (2015). Using the Lexique database (New et al., 2001), we calculated the functional load of each phonological feature for each obstruent, using lemma forms as reported in Lexique. We chose the lemma forms, as functional load calculated on lemma rather than on surface forms is a better predictor of sound change (Wedel et al., 2013). The lemma forms of French nouns are simply the singular. Note that very few words in French are marked phonologically in the plural form (e.g., *journal* /ʒuʁnal/ - *journaux* /ʒuʁno/,"newspaper(s)"); these words would be considered only in the singular. We followed the method detailed above, respecting French phonotactic constraints. Unlike minimal pair counts and entropy differences, which are fairly straightforward to measure, our proposition requires knowledge of the language's phonotactic constraints, and further depends on their interpretation. For instance, while /tl/ is universally rejected as a possible onset in French, /pn/, which is a rare onset occurring exclusively in words of Greek origin (e.g., /pnø/ - *pneu*, "tire"), may or may not be accepted, depending on the speaker.[8] For the present purposes, we considered possible clusters those described to be well-formed according to Dell (1995), plus some of those described in the same study as rare, such as /sm/ as in *smiley* or /sn/ as in *snob* (for an exhaustive list, see appendix). We included only rare clusters that were deemed acceptable during an informal survey. We calculated the possible minimal pairs for obstruents in all positions (i.e., not just word-initially). The results of this calculation are shown in Fig. 2.4. The distributions represented are the scores of the twelve phonemes for each feature. For voicing, and manner, each phoneme is involved in one comparison (e.g., /p/ vs. /b/, or /p/ vs. /f/), while for place, each phoneme is involved in two comparisons (e.g., /p/ vs. /t/ *and* /p/ vs. /k/).

We performed a one-way ANOVA on the distributions of scores obtained using our functional load method. It should be noted that instead of sampling a distribution from a population, we are indeed

[8]This specific example is known to vary according to region, with an epenthesized version /pən/ being preferred in the south of France.

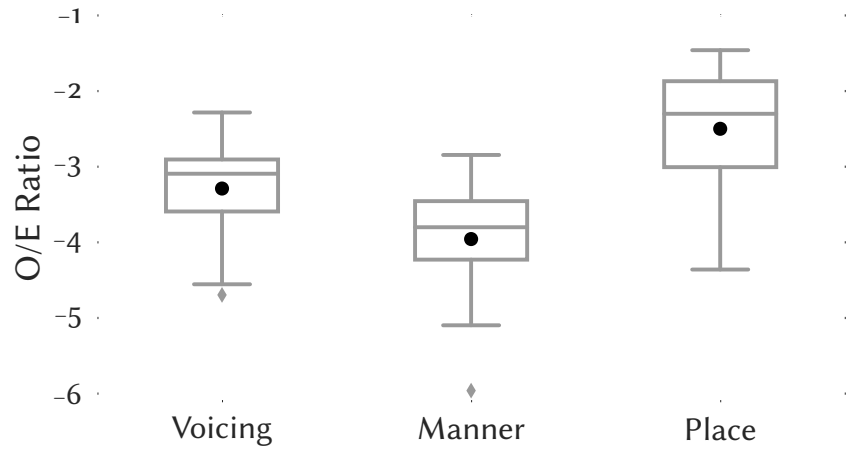

Figure 2.4: Boxplots of functional load as measured with O/E ratios for French nouns. Black dots represent the means of the distributions.

sampling the *entire* distribution. That is, we include all of the contrasts that each feature involves within the subset of sounds under consideration. A significant difference was observed across the phonological features ($F = 9.11$, $p < 0.001$). Post-hoc analyses using the Tukey HSD test showed that it was the place feature that was significantly different from manner ($p < 0.001$) and marginally significantly different from voicing (p = 0.068), but no difference was found between manner and voicing ($p > 0.05$). This indicates that the place feature has a higher functional load in French nouns than the other two tested features.

Next, we compared these results to ones obtained using a measure of the difference in entropy (Surendran & Niyogi, 2003, 2006). Although the entropy measure for features is calculated by summing the difference in entropy for each contrast within that feature (recall the mini-language with just /p, b, t, d/ described above, where the difference in entropy for voicing is equal to the differences for /p/~/b/ and /t/~/d/ combined), it is possible, for the purposes of performing inferential statistics, to consider each contrast as one data point in a distribution. The functional load of voicing, for example, then becomes a vector of entropy differences (in our example, /p/~/b/, /t/~/d/), allowing for statistical comparison across the features. We used the Phonological CorpusTools kit (Hall, Allen, Fry, Mackie, & McAuliffe, 2015a) to calculate the differences in entropy within French nouns extracted from the LEXIQUE database; the results can be seen in Figure 4. The distributions represented are made up of every contrast in that dimension (e.g., for voicing /b/~/p/, /z/~/s/, etc.).

Note that place has a seemingly higher functional load (mean = 0.0140) than the other two features

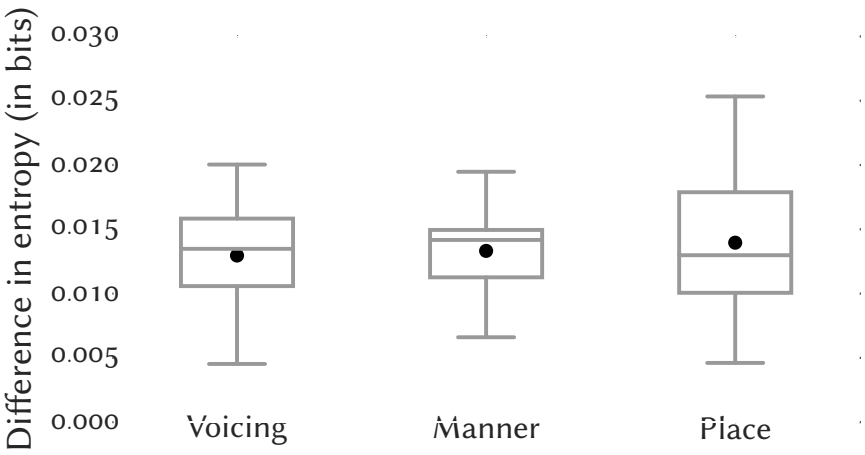

Figure 2.5: Boxplots of functional load as measured by differences in entropy for French nouns. Black dots represent the means of the distributions.

(manner: mean = 0.0133; voicing: mean = 0.0130), just as with our O/E ratio measure. A one-way ANOVA, however, revealed no significant difference amongst the features ($F < 1$). Thus, given the available data points, place's modestly higher score appears to be uninterpretable. Of course, given that the numerical pattern is similar to what we obtain with the O/E ratio method, it is likely that there is a true effect, which is simply masked (perhaps due to frequency or phonotactic effects as described above). Indeed, the O/E ratio and entropy methods are grossly measuring the same thing (use of a contrast within the lexicon), and are in fact highly correlated (Pearson's $r = 0.73$, $p = 0.001$).

2.3.3.3 Discussion

The two traditional measures of functional load[9] (minimal pair counts and information entropy) are inappropriate for comparing the distinctiveness of phonological features. Both can be biased by individual phoneme frequency, as well as by the presence of phonotactic constraints, which we would like to separate from the question of lexical distinctiveness. Indeed the traditional methods measure the *absolute* functional load of contrasts, while we are interested in their relative weight. The speaker may ask, "Given the constraints of my language, how distinctive is the place contrast compared to the manner contrast?" We therefore proposed a new measure, O/E ratios, that abstracts

[9]We focus here on functional load, and presume it to represent the knowledge of lexical organization shared by speakers of French. Another common lexical measure, neighborhood density, is not appropriate to answer our question about single features, since it includes minimal pairs differing in multiple features, as well as minimal pairs differing in the absence versus presence of a segment (e.g., the neighborhood of the word peau contains not only beau and other one-feature change words, but also vaut (a multi-feature change), eau (a deletion), and pôle (an insertion).

away from these properties, and additionally provides distributions of scores, allowing the use of inferential statistics to test the significance of observed differences.

When applying our method to obstruents in French nouns, we found that the place feature has a significantly higher functional load than the manner and voicing features. This means that in the French lexicon, nouns are more likely to be distinguished from one another by place than by manner or voicing, where possible. Because our score is based on an observed-over-expected ratio, it excludes effects due to frequency, or even due to the number of possible contrasts (French has twice as many place as manner or voicing contrasts within obstruents). Hence, any observed differences truly reflect the extent to which the contrasts are used in a distinctive way in the lexicon (at a type, rather than token, level). It would be pertinent in future applications to test to what extent weighting minimal pairs by token frequency would affect the patterns we observe.

Although our method yields results that are highly correlated with those of the difference in entropy measure, we observed significant differences within the French noun class that are not captured by the entropy measure. Moreover, it is more insightful even for results that are in line with the entropy measure. For instance, the contrasts /d/~/z/ and /s/~/z/ have very low functional load according to both measures. While observing this in the entropy measure alone might lead one to think that the effect is driven by the low frequency of /z/ (indeed, all four of the lowest entropy scores are contrasts involving this phoneme), the fact that we observe a similar pattern in the O/E ratios shows that even when French does use /z/, it rarely contrasts with other sounds, be it in voicing, manner, or place. This, then, shows that there is something going on beyond the simple distribution of sounds in the language.

The use of our method, though, requires language-specific knowledge. While we based our analysis on a phonological description of French syllable structure (Dell, 1995), a more bottom-up approach may be adopted by extracting phonotactic rules from the corpus being studied. For the case of French, Lexique contains certain words with very rare clusters (e.g., /ft/ as in *phtaléine*); it may therefore be prudent, when using such an approach, to set a frequency threshold, including only clusters which appear a certain number of times. It further presumes a specific representation of phonotactics. The current implementation of our method considers phonotactics categorically: ei-

ther a wordform is licit or it is not. However, as phonotactic acceptability is known to be gradient, and dependent on lexical statistics (e.g., Frisch, Large, & Pisoni, 2000), it may be interesting to incorporate such a notion in a future implementation of our method.

A further issue is that of phonological processes. For example, when considering a language with final devoicing, such as Dutch, the question arises as to whether to perform the calculation on underlying forms (e.g., /mud/ - "courage") or on surface forms (e.g., [mut]). If the underlying form is chosen, then /mud/ contrasts with /mut/ - "must". Of course, this is an issue for all calculations of functional load, as are the similar issues of lemma versus inflected forms and canonical versus reduced forms. For example, the English word "probably" is often produced as [pɹɑli]; as noted by (Hall, Allen, Fry, Mackie, & McAuliffe, 2015b), this common variant, but not the canonical pronunciation, contrasts with the word /tɹɑli/ - "trolley". Using the entropy measure, Hall et al. showed that patterns of results do not greatly differ according to whether the analysis is based on the most common pronunciation or on the canonical form. It may therefore be reasonable to likewise focus on the more abstract, underlying, level, where voicing would therefore be distinctive word-finally in a language such as Dutch. This will be an important consideration for future implementations of this measure, which did not arise in the French data, as the sounds tested are not affected by any neutralization process.

Finally, let us return to the question of relative weight of phonological features during word recognition. Based on the results from the ABX task, we argued above that prelexical perception accounts for the relative importance of manner compared to the other features, given its basis in a stark acoustic difference. The functional load results, then, provide an explanation for the relative importance of place in the results reported in Martin and Peperkamp (2015): Given that within nouns, the place feature has a higher functional load than the voicing feature, French listeners lend more importance to place than to voicing cues during word recognition, thereby making it harder to recognize a word with a changed place feature than one with a changed voicing feature.

2.3.4 General discussion

French listeners are more likely to recognize a mispronounced version of an obstruent-initial word if the mispronunciation concerns the voicing feature than if it concerns the place or manner features (Martin & Peperkamp, 2015). Thus, in French obstruents, both place and manner are more important for word recognition than voicing (at least in nouns), akin with findings in other languages (Cole et al., 1978; Ernestus & Mak, 2004). Where does this asymmetry come from? We examined two sources: prelexical acoustic perception, and lexical knowledge. In order to do so, we introduced two methodological novelties. First, we developed a version of the ABX discrimination paradigm that allows for assessing differences in the perception of native language sounds without presenting the stimuli in noise. Specifically, we increased the difficulty of the task (by using long, very similar non-words, a short ISI, and multiple voices), and showed that even among fully distinctive contrasts, some are more difficult to discriminate than others. Contrary to the two most-reported methods for assessing similarity, syllable identification and similarity judgments, our method neither yields ceiling performance, as is common in syllable identification in clear speech, nor requires participants to meta-linguistically reflect on the sounds themselves, as in explicit similarity judgments. Second, we developed a new method of measuring functional load, based on an observed-over-expected ratio of minimal pairs in the lexicon. This method is less vulnerable to language-specific tendencies that can bias traditional measures such as minimal pair counts and differences in entropy, and allows, moreover, for the use of inferential statistics to compare features amongst themselves. This makes it more appropriate for our current research question than the traditional methods of simply counting the number of minimal pairs, or of calculating difference in system entropy with and without the contrast.

Using these new methodologies, we examined the prelexical perception of obstruent features by French listeners on the one hand, and the functional load of these features in French nouns on the other hand. Results from the perception experiment showed that French listeners are better at discriminating French nonce words with obstruents that differ in manner of articulation than in place or voicing; this mirrors the fact that manner contrasts are acoustically more distant than place or voicing contrasts. Results from the functional load computation showed that within the class of

French nouns, place differences are more often used to distinguish words than voicing and manner differences.

We propose, then, that French listeners are biased by both of these phenomena during word recognition. First, the strong acoustic difference between stops and fricatives makes the manner contrast easy to perceive (note that we predict this type of effect to be observable cross-linguistically). This explains why participants in Martin and Peperkamp (2015) had great difficulty recognizing words with a manner mispronunciation (for instance, replacing /v/ with /b/ disrupted recognition of the word *voleur* - "thief"). Second, the fact that their language uses the place feature more often than the other features to distinguish words leads French listeners to preferentially pay attention to place cues during word recognition. This explains why participants similarly failed to recognize words with a place mispronunciation (for instance, replacing /v/ with /ʒ/ also strongly disrupted recognition of the word *voleur*). By contrast, as voicing stands out neither in prelexical perception nor in functional load, participants had less difficulty recognizing words with a voicing mispronunciation (replacing /v/ with /f/ was less disruptive for recognition of *voleur*).

Thus, the combination of our prelexical experiment and lexical analysis can explain the word recognition results reported in Martin and Peperkamp (2015). Our results indicate that listeners are sensitive to lexical structure, and that they recruit this knowledge during word recognition. They further demonstrate that, unsurprisingly, low-level acoustic information biases listeners at multiple levels of processing (i.e., in the higher-level lexical task as well as in the lower-level discrimination task). Our conclusion is that during word recognition, listeners' knowledge of the French lexicon coalesces with their low-level perceptual biases, yielding greater perceived distinctiveness for the manner and place compared to the voicing feature. This is in line with a vast literature on the integration of bottom-up and top-down influences (for a review, see Davis & Johnsrude, 2007).

It is, though, important to consider our specific definition of feature. We have been examining featural contrasts along the major class dimensions, without consideration of more specific features (binary or not). Of course the methodologies we have presented in this paper could be used to examine any set of features, but we do make certain assumptions that merit discussion. One major point is our consideration of the manner feature as concerning only stops and fricatives, since we

restrict ourselves to obstruents. Our experiment and argumentation focus on the stark acoustic difference between these two types of sounds, which, we argue, yields the importance of the manner feature within the obstruent class. This does not necessarily transfer to other types of manner contrasts. For example, the difference between nasals and voiced stops, although a manner difference, is not automatically predicted to behave in the same way as the manner contrasts we tested here. It is entirely possible that the manner feature has a different weight relative to place and voicing when considering nasals. Our conclusions therefore must be taken within the class of sounds we tested. Future implementations of this measure might consider different types of features and different distinctions, in addition to other types of contrast (vowels, tones, etc.).

Furthermore, although our functional load measure specifically abstracts away from the number of contrasts within a certain feature by using ratios (i.e., the fact that there are twelve place contrasts but only six voicing contrasts in French obstruents is corrected for by expecting more place minimal pairs in the lexicon), our results do not allow us to say definitively that it is functional load that drives the importance of place for French listeners during word recognition. While we found that French uses place more than manner or voicing to distinguish nouns from one another, it is also true that place, for example, is a three-way contrast, while voicing is strictly binary. It would be interesting to focus on a language with higher dimensionality in the voicing feature (e.g., Korean or Eastern Armenian). If the number of contrasts in both the place and voicing features is the same, and speakers of such a language still pay more attention to place than to voicing in a lexical task, then our claim that it is lexical knowledge rather than knowledge of the phonological inventory that is exploited during word recognition would be bolstered.

A further consideration regarding our functional load analysis is that it focuses on nouns, and presumes that listeners track statistical information pertaining to phonological contrasts within lexical classes. This is indeed supported by various studies (Farmer, Christiansen, & Monaghan, 2006, 2011; Heller & Goldrick, 2014; Strand, Simenstad, Cooperman, & Rowe, 2014). For instance, Strand et al. (2014) showed that listeners are sensitive to syntactic context during isolated word recognition. In their word identification task, accuracy was negatively affected by a measure of within-class grammatical density when syntactic category was constrained. Additionally, previous work on sound change has shown that within-category minimal pairs better predict mergers over time than across-

category pairs, further suggesting that contrast may be category-sensitive (Wedel et al., 2013). Future work on lexical influence during speech processing could further explore the role of lexical classes. In particular, we predict that any functional load differences found within the French verb class should similarly be reflected in bias during word recognition. Thus, if within verbs one feature has a higher functional load than others, we expect that mispronunciations of that feature in verbs will be more disruptive for word recognition than mispronunciations of other features.[10]

Finally, our conclusions hinge on a qualitative comparison of three results: place and manner were shown to be significantly more important for word recognition in Martin and Peperkamp (2015); the importance of manner can be explained by its acoustic saliency, as demonstrated by the prelexical experiment reported in the present study; the importance of place can be explained by its lexical status, as demonstrated by our functional load measure. However, because of the different ways that each of these results were obtained, making a *quantitative* comparison is rather difficult. For example, it may be tempting to compare the individual contrasts tested in the lexical and prelexical tasks, to examine whether performance on the manner contrasts in the prelexical task negatively correlates with performance on manner contrasts in the lexical task. This is not straightforward, though. In the ABX task, comparing X to A and B is symmetrical; if the contrast tested is /t/~/s/, participants compare, say, /bivibi/X to both /vivibi/A and /bivibi/B. By contrast, the lexical task is asymmetrical, as participants attempt to map a given mispronunciation onto an existing lexical representation. In attempting to map non-existent *boleur* to the real word *voleur*, the question can be very clearly stated: does /b/ activate /v/? Other trials using other words (e.g., non-existent *veignet* mapped to real *beignet* – "fritter") ask the reverse question. In order to directly compare the contrasts tested in the lexical task, it might be preferable to have an asymmetrical prelexical task, such as an oddball paradigm, where a deviant stimulus is compared unidirectionally to a standard.

To conclude, word recognition is a complex process that takes into account both low-level acoustic information and language-specific, phonological and lexical, knowledge. We have provided evi-

[10]It would be interesting to extend the present research to French verbs, but this is less straightforward than one would hope. Recall that we used lemma forms to calculate functional load. Lexique codes the lemma form of French verbs to be the infinitive, but French verbs invariably have infinitive morphology (/-e/, /-iʁ/, or /-ʁ/) that may influence the outcome of the calculation. For example, the French verb battre - /batʁ/ (to hit) does not form a minimal pair with the possible but nonexistent form /datʁ/. The stem of the same verb, /bat/, however, does contrast with the stem /dat/ of the verb /date/ (to be dated, old). Thus, the choice of the lemma form impacts the functional load of, in this case, the place feature.

dence that acoustic salience and lexical distinctiveness coalesce and bias listeners' weighting of phonological features. The two methodological tools that we developed can be used independently, one for assessing the prelexical discriminability of native phonemes without altering their acoustic properties, and one for assessing the relative functional load of phonological features.

2.4 Further questions regarding the prelexical disciminability of native phonemes

Section 2.3.2 detailed our implementation of the ABX paradigm to study the perception of native phoneme contrasts. In that study, we specifically focused on participants' ability to perceive featural contrasts affecting a subset of French consonants, namely obstruents. We did additionally test the paradigm's ability to assess performance on much easier contrasts on the one hand, and vowel contrasts on the other. This section reports the unpublished results of those experiments.

2.4.1 Many features

The obstruent contrasts tested in the previous study were all relatively small, both in terms of acoustics, and also in terms of features (recall that all contrasts tested were one-feature differences). It remains to be seen, though, if our paradigm is sensitive enough to test larger, and hence easier contrasts. One of the points we highlighted above was the paradigm's ability to capture differences in performance and without participants performing at ceiling.

2.4.1.1 Methods

We followed the methodology laid out in Section 2.3.2.1, but tested contrasts that are maximally different in French. We specifically chose six contrasts between an obstruent and a sonorant consonant, when possible at different places of articulation, namely: /f/~/l/, /g/~/w/, /k/~/m/, /p/~/n/, /t/~/ʁ/, and /s/~/j/.

2.4.1.1.1 Stimuli

For each of the six many-feature consonant contrasts, we constructed 18 non-word items, for a total of 108 items. The items were constructed in the same way as described in ?? 2.3.2.1.1. An example set can be seen in Table 2.2.

All stimuli were synthesized using the Apple Say program's diphone synthesizer; each was produced by three of the European French voices: two male (*Thomas* and *Sébastien*) and one female (*Virginie*). This gave us a total of 324 unique tokens (108 items × 3 voices).

2.4.1.1.2 Procedure

A total of 432 unique trials were created by combining the stimuli into ABA, ABB, BAA, and BAB trials. These were counterbalanced into four 108-trial-long lists (with each participant seeing only one list), such that each list contained a total of six trials for each of the twenty-four obstruent contrasts, including two trials for each vowel /a, i, u/. This made the experiment of similar (but shorter) duration to the one-feature experiment.

Participants sat in front of a computer screen in a sound-attenuated room while stimuli were played binaurally through a headset. They read instructions presented on screen that described the ABX paradigm: In every trial, the first two non-words they heard would be different and the third one would always correspond to either the first or second one; they should indicate whether the third stimulus matched the first or the second; and they should give their response by pressing one of two buttons on a response box.

In each trial, participants heard a sequence of three stimuli, each produced by a different voice, with an ISI of 300 ms. Thus, X was acoustically different from both A and B. The position of the difference between A and B changed from one trial to another, making it impossible for participants to know where specifically to attend upon hearing the A stimulus of a given trial. The position of the difference was counterbalanced across trials. Voice order was randomized. For example, a participant might hear /fafala/$V_{irginie}$ - /falala/T_{homas} - /fafala/$Sé_{bastien}$, to which they should respond 'A'. In another trial, they might hear /mumuku/T_{homas} - /kukumu/$V_{irginie}$ - /kukumu/$Sé_{bastien}$, to which

they should respond 'B'. Following a given trial, the next was presented 1,250 ms after the participant had given a response, or after a timeout of 2,000 ms after stimulus offset, whichever came first. If participants failed to give a response before the timeout, this was counted as an error. No feedback was given to participants during the experiment.

The experiment lasted about 12 minutes.

2.4.1.1.3 Participants

Thirty-two native speakers of French participated (23 women, 9 men) were randomly assigned to one of the four counterbalanced lists. They were aged between 20 and 35 (mean: 25.3). None of them reported any history of hearing problems.

2.4.1.2 Results

To analyze the results of this experiment, we collapsed the three conditions in the original experiment, calling the collapsed data "one feature contrasts". We compared these data to those from the present experiment, which we termed "many feature contrasts".

The mean accuracy scores, in addition to the response times for correct trials (measured from the onset of the third stimulus) from both the original one feature experiment and the present many features experiment are shown in Fig. 2.6. All analyses were performed using mixed-effects models in R (Bates et al., 2014). Accuracy was analyzed with a logistic mixed-effects regression model; the log response times were analyzed using a linear mixed-effects regression model.

Both models included random intercepts for Participant and Contrast (the phoneme pair) and the fixed factor Experiment (one feature vs many features). We included the factor Experiment using contrast coding and assessed significance by comparing these models with models excluding the factor Experiment (i.e., with models that only included an intercept).

In both the accuracy scores and the response times, the factor Experiment was found to be a significant predictor of performance (accuracy: $\beta = 0.78$, $SE = 0.26$, $\chi^2(1) = 8.50$, $p < 0.01$; RT: $\beta = -0.10$, $SE = 0.05$, $\chi^2(1) = 4.34$, $p < 0.05$).

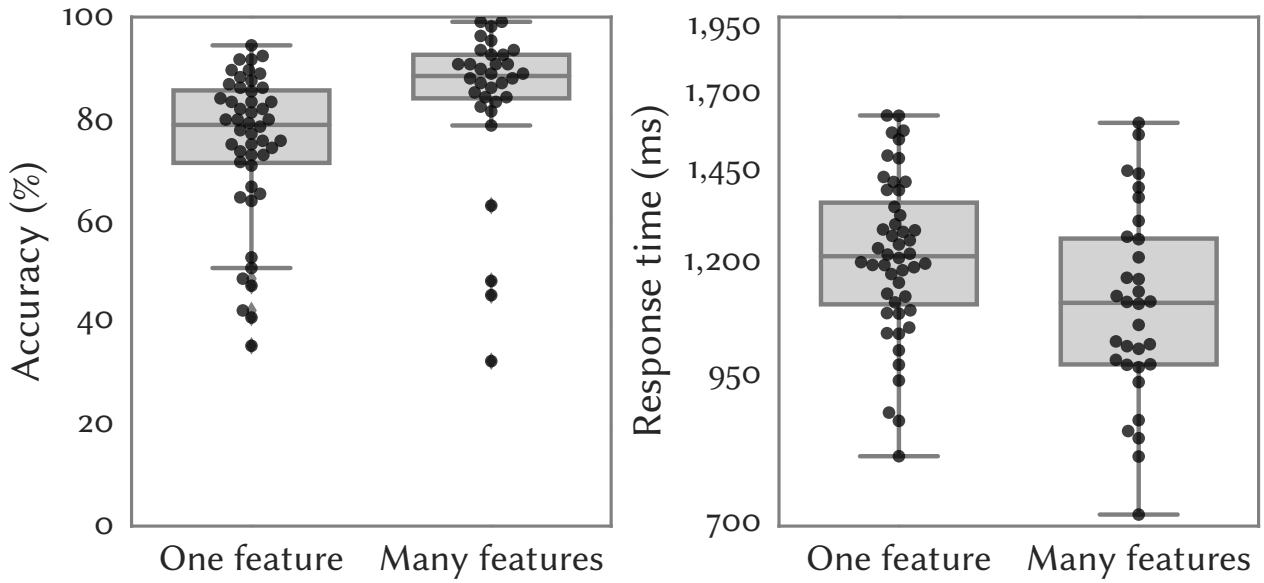

Figure 2.6: Box- and dotplots of participant means of accuracy (left) and response times (on a log scale) on correct trials (right) by experiment.

These results indicate that participants were both more accurate, and on correct trials faster, for contrasts involving multiple features than those involving only one feature. We therefore have evidence that this paradigm is able to distinguish between different degrees of difficulty, with harder contrasts yielding lower accuracy and slower response times.

2.4.2 Vowels

Now that we have established that different types of consonant contrasts may be examined using the paradigm, it is important to test whether or not it is able to be exploited for other kinds of phonological contrasts. We chose therefore to design a version of the experiment examining vowel contrasts, to see whether or not we could observe differences in one feature contrasts.

2.4.2.1 Methods

We again followed the methodology laid out in Section 2.3.2.1, but tested vowel contrasts that differ along one featural dimension. We chose a subset of the French vowels in order to have a relatively symmetric set of target contrasts. We were able to distinguish six French vowels forming seven contrasts along three dimensions: frontness, height, and rounding. These vowels can be seen in

	front		*back*
	unrounded	rounded	rounded
high	i	y	u
mid	e	ø	o

Table 2.3: The six French vowels we tested arranged vertically by height (in bold) and horizontally by frontness (italics) and rounding.

	/p/	/t/	/k/	/f/	/s/	/ʃ/
/u/ ×1, /y/ ×2	/pupypy/	/tutyty/	/kukyky/	/fufyfy/	/susysy/	/ʃuʃyʃy/
	/pypupy/	/tytuty/	/kykuky/	/fyfufy/	/sysusy/	/ʃyʃuʃy/
	/pypypu/	/tytytu/	/kykyku/	/fyfyfu/	/sysysu/	/ʃyʃyʃu/
/u/ ×2, /y/ ×1	/pypupu/	/tytutu/	/kykuku/	/fyfufu/	/sysusu/	/ʃyʃuʃu/
	/pupypu/	/tutytu/	/kukyku/	/fufyfu/	/susysu/	/ʃuʃyʃu/
	/pupupy/	/tututy/	/kukuky/	/fufufy/	/sususy/	/ʃuʃuʃy/

Table 2.4: Non-word items used for the contrast /u/~/y/.

Table 2.3.

2.4.2.1.1 Stimuli

For each of the seven vowel contrasts, we constructed thirty-six non-word items, for a total of 252 items. Each item had the structure CVCVCV. Its consonants were identical and were drawn from the set of French voiceless obstruents (viz. /p/, /t/, /k/, /f/, /s/, /ʃ/). We chose this set to provide the maximum range of coarticulation in French while maintaining stark acoustic contrast between the vowels and the frame consonants. The vowels were the ones from the contrast under consideration; one of them occurred once and the other one twice. By way of example, the complete set of thirty-six items for the contrast /u/~/y/ is shown in Table 2.4.

All stimuli were synthesized using the Apple Say program's diphone synthesizer; each was produced by three of the European French voices: two male (*Thomas* and *Sébastien*) and one female (*Virginie*). This gave us a total of 756 unique tokens (252 items × 3 voices).

2.4.2.1.2 Procedure

A total of 1,008 unique trials were created by combining the stimuli into ABA, ABB, BAA, and BAB trials. These were counterbalanced into eight 126-trial-long lists (with each participant seeing only one list). This made the experiment of similar duration to the original one-feature consonant experiment.

Participants sat in front of a computer screen in a sound-attenuated room while stimuli were played binaurally through a headset. They read instructions presented on screen that described the ABX paradigm: In every trial, the first two non-words they heard would be different and the third one would always correspond to either the first or second one; they should indicate whether the third stimulus matched the first or the second; and they should give their response by pressing one of two buttons on a response box.

In each trial, participants heard a sequence of three stimuli, each produced by a different voice, with an ISI of 300 ms. Thus, X was acoustically different from both A and B. The position of the difference between A and B changed from one trial to another, making it impossible for participants to know where specifically to attend upon hearing the A stimulus of a given trial. The position of the difference was counterbalanced across trials. Voice order was randomized. For example, a participant might hear /pupupy/$V_{irginie}$ - /pupypy/T_{homas} - /pupupy/$Sé_{bastien}$, to which they should respond 'A'. In another trial, they might hear /sesise/T_{homas} - /sisise/$V_{irginie}$ - /sisise/$Sé_{bastien}$, to which they should respond 'B'. Following a given trial, the next was presented 1,250 ms after the participant had given a response, or after a timeout of 2,000 ms after stimulus offset, whichever came first. If participants failed to give a response before the timeout, this was counted as an error. No feedback was given to participants during the experiment.

The experiment lasted about 20 minutes.

2.4.2.1.3 Participants

Thirty-two native speakers of French participated (27 women, 5 men) were randomly assigned to one of the four counterbalanced lists. They were aged between 18 and 35 (mean: 23.6). None of

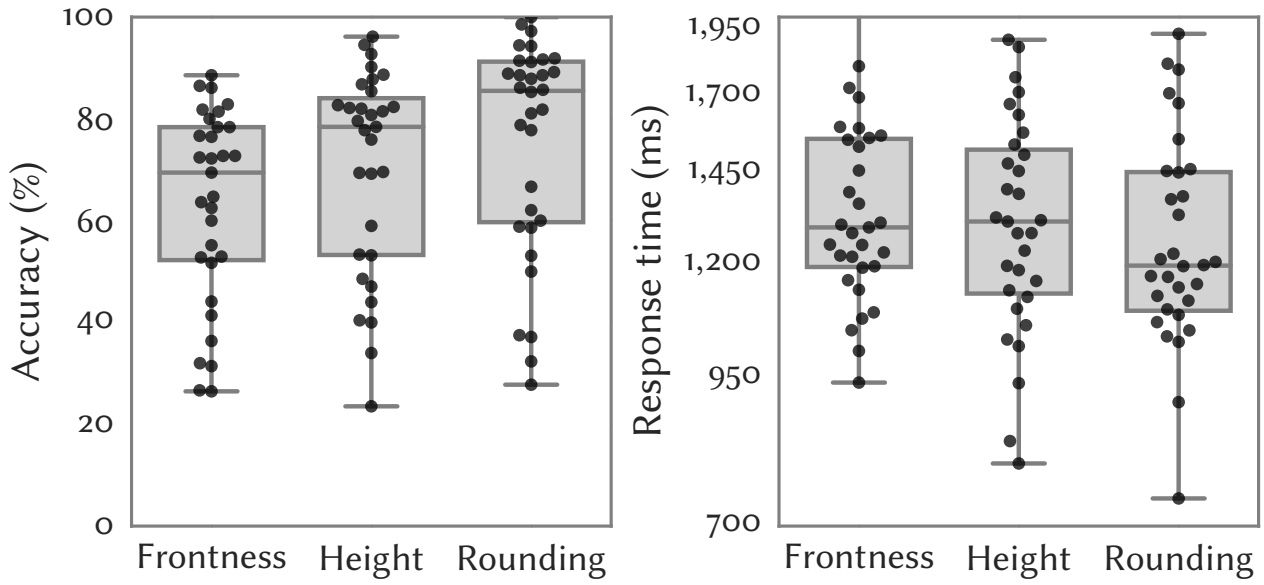

Figure 2.7: Box- and dotplots of participant means of accuracy (left) and response times (on a log scale) on correct trials (right) by feature.

them reported any history of hearing problems.

2.4.2.2 Results

The results from the present experiment are displayed by feature in Fig. 2.7. We performed the analyses in the same manner as for the one feature consonant experiment reported in Section 2.3.2.2. That is, we designed a logistical mixed effects model for the accuracy data and a linear mixed effects model for the response times.

Both models included random intercepts for Participant and Contrast (the phoneme pair). and the fixed factor Feature (frontness vs height vs rounding). We took one of the values of the Feature factor as intercept and then releveled the data to assess the three-way comparison. We started by taking rounding as the model intercept, comparing rounding to frontness and rounding to height. We then took frontness as the intercept, which allowed us to compare it to height (and redundantly to rounding).

In the accuracy scores, rounding was found to be significantly different from frontness ($\beta = 0.68$, $SE = 0.33$, $z = 2.08$, $p = 0.04$), though this was not found for response times ($\beta = -0.07$, $SE = 0.04$, $t = -1.77$). In neither the accuracy scores, nor the response times was rounding found to be different

from height (accuracy: $z < 1$; RTs: $t < 1$). Furthermore, frontness was not found to differ significantly from height (accuracy: $\beta = 0.40$, $SE = 0.30$, $z = 1.33$, $p = 0.18$; RTs: $\beta = -0.05$, $SE = 0.04$, $t = -1.29$).

Unlike the results from the previous two experiments, where clear-cut differences were observed between features and between single and multiple feature differences, the results of the present study are difficult to interpret. While it seems that rounding contrasts are perceived more accurately than frontness contrasts, no difference was observed in response times. Additionally, the height feature seems indistinguishable from the other two features.

2.4.3 Discussion

After having established the ability of our implementation of the ABX paradigm to assess differences between difficult contrasts, we tested the applicability of the paradigm to much easier contrasts, namely multi feature contrasts involving sonorants vs obstruents. We found that performance vastly improved, but that we were able to observe a difference between these contrasts and the one feature contrasts tested in the original experiment, with more distinct contrasts being perceived. We then tested its applicability outside of the domain of consonants by considering a subset of the French vowel inventory. We considered, as we did for consonants in the original experiment, vowel contrasts that differed in only one of three features: frontness, height, or rounding. We observed that rounding contrasts were perceived as more distinct than frontness contrasts, but that neither differed from height contrasts. These results suggest that while the paradigm is applicable to different types of phonological contrasts, it is unclear if it can capture the full range of possibly very subtle differences in performance.

Future research should apply our implementation of the ABX paradigm to even more kinds of phonological contrasts. For example, it would be interesting to see if perception of suprasegmental differences such as stress or tone could be assessed without overly modifying the experimental design. It has been shown for example that Mandarin Chinese speakers preferentially alter non-words into real words by changing their tone, rather than their vowels or consonants (Wiener & Turnbull, 2015), but are all tones created equal? Testing the relative perceived contrastiveness of the tones of Mandarin Chinese seems easily doable using our paradigm and would be an interesting avenue for

future work. It is unclear though if the current proposed structure of the stimuli is ideal for such contrasts. Are CVCVCV stimuli ideal/too hard/too easy for such contrasts? Further parameters of the experiment, such as ISI, could also be manipulated to find the optimal setup for the type of phonological contrast under consideration.

All in all, our novel methodology seems promising and is clearly able to assess the differences that our study targeted (i.e., one feature consonant contrasts), but more research will be required to test its applicability to other research questions.

2.5 Conclusion

This chapter presented evidence that phonological features are not all processed with the same weight during word recognition, and explored two possible sources for the asymmetries we observed. Section 2.2 reported on a mispronunciation detection task in which we saw that French participants were more likely to recognize a word if it began with a mispronunciation in voicing than if the mispronunciation was in manner or place of articulation. Section 2.3 considered one bottom-up and one top-down motivation for this. First, we used an ABX task to see how (dis-)similar the French obstruents tested in Section 2.2 are perceived as being to one another outside of lexical context. We found that manner contrasts were better perceived that voicing or place contrasts, indicating that a mispronunciation in manner might be perceived as further away from a canonical pronunciation. Second, we conducted an analysis of the functional load of the three features using a new measure. We found that place of articulation is more likely to be used in French to distinguish words from one another. If listeners use this information online during word recognition, it is likely that a place mispronunciation would be considered further from a canonical pronunciation since this feature is important in distinguishing words in the French lexicon.

Taken altogether, our results show that listeners process phonological features differently according to their status, either as easy to perceive contrasts or as important contrasts in the native language of the listener. Both bottom-up and top-down sources of information thus coalesce to inform the listener during word recognition.

Chapter 3

Phonetic naturalness and typology

3.1 Introduction

As presented in Section 1.2, this chapter will explore the question of phonological pattern learning bias, specifically with regards to the typologically common rule of vowel harmony versus the practically unattested rule of vowel disharmony. We will consider learning bias as it manifests itself in artificial language learning experiments.

In Section 3.2, we begin by testing learning of the two patterns when participants are exposed to exceptions. This study has two critical components. First, the task differs from the classical AGL experiments à la Pycha et al. (2003) by requiring participants to *produce* the plural forms of new words they have learned. This manipulation is important given vowel harmony's suggested source in a production constraint (Ohala, 1994). Second, it contains a modeling component, where we simulate the transmission of a learned pattern from one generation of learners to the next. We focus on the role that learning bias can have in shaping sound patterns over time.

In Section 3.3, we use the same stimuli to further explore the effect of modality on learning. In this study, we expose learners to the same patterns, but ask them to report their answers in a two-alternative forced choice task requiring participants to simply listen to and choose the plural form they prefer. This methodology does not implicate the production system in the same way that the production task described in Section 3.2 does. Differential results between the two studies could

pinpoint modality-related learning effects.

Finally in Section 3.4, we explore one final possible influence on the learning of phonological patterns: sleep. Sleep is known to play an important role in the consolidation of memories (for a review, see M. P. Walker & Stickgold, 2004). We explore the role of sleep on the consolidation of the natural pattern versus the unnatural pattern. If the natural pattern, in addition to being learned better to begin with, also benefits from increased consolidation with sleep, this could play a role in the shaping of typology as well, and add to our understanding of the absence of disharmonic patterns in vowel co-occurrence restrictions.

The chapter concludes with a discussion of an important source of bias present in all artificial language learning experiences: learners' native languages. We discuss how the statistical tendencies present in learners' native lexicon could influence their learning patterns.

3.2 Phonetic naturalness and the shaping of sound patterns: the role of learning bias and its transmission across generations

This section is an adaptation of the following manuscript: Martin, A., Guevara-Rukoz, A., Schatz, T., & Peperkamp, S. (resubmitted). Phonetic naturalness and the shaping of sound patterns: the role of learning bias and its transmission across generations. *Phonology*.

3.2.1 Introduction

Sound patterns tend to be phonetically 'natural': they reflect constraints on speech production and perception. For instance, many phonological alternations, such as consonant assimilation or vowel harmony, increase ease of articulation; they mirror low-level gradient phonetic effects due to automatic processes such as coarticulation and gestural overlap. It has long been observed that cross-linguistically, phonetically natural rules are much more prevalent than unnatural ones (e.g., Hooper, 1976). Over the past few decades, this typological asymmetry has often been explained by focusing on the diachronic aspect. Specifically, it has been argued that the typology is born out of general

phonetic pressures that yield categorical shifts (i.e., sound changes) over time. Both the hypo- and hyper-correction model (Ohala, 1993b) and Evolutionary Phonology (Blevins, 2004) are instances of this approach (for a comprehensive review, see Hansson, 2008). They explain the distribution of sound patterns through spontaneous changes that occur as words are passed from speakers to listeners, for instance due to misperception. These shifts are based in universal principles of perception and production that yield strikingly similar patterns cross-linguistically, but which may not be part of any individual's grammar.

As an example, consider vowel harmony. This common phonological phenomenon involves co-occurrence restrictions on vowels, such that all of the vowels within a word must share one or more phonological features. It often presents itself in the form of morphophonological alternations. Hungarian, for example, has a restriction on the backness of vowels within a word, such that most suffixes of the language have two allomorphs: one containing a back vowel, and one containing a front vowel. The data in (2) demonstrates this restriction, where (2a) contains only back vowels and (2b) contains only front vowels, though both contain the same dative suffix.

(2) a. [bɒraːtnɒk] 'friend-DAT'

b. [ɛmbɛrnɛk] 'person-DAT'

Vowel harmony has been proposed to arise out of vowel-to-vowel coarticulation (e.g., Ohala, 1994), in that the general phonetic bias for vowels to be coarticulated at a distance may cause listeners to re-encode vowels produced with coarticulation. For instance, an underlying /u/ that is fronted due to its proximity to an /i/ may be re-encoded as /y/. The plausibility of such an explanation has been demonstrated with an iterated learning model (Mailhot, 2013). This simulation showed that over time, the coarticulation of vowels within words, coupled with channel noise, can indeed yield harmonic lexicons. That is, the proportion of words containing vowels that share some phonetic property increased as coarticulated vowels were re-encoded during transmission.

An alternative explanation for typological asymmetries concerns the existence of an individual learning bias. In various phonological theories, typological asymmetries are reflected by phonetically motivated biases in the speaker's mind (Archangeli & Pulleyblank, 1994; Donegan & Stampe,

1979; B. P. Hayes & Steriade, 2004). These grammatical biases could induce learning biases, such that phonetically natural patterns are easier to learn and hence have an advantage in transmission across generations (Schane et al., 1974; C. Wilson, 2006). Note that this proposal is not inherently incompatible with the diachronic hypotheses already discussed; that is, a learning bias could be additive with phonetic grounding in leading to changes over time (for discussion, see Moreton, 2008).

Previous research, using artificial language learning paradigms, has provided evidence for the presence of a learning bias, both at the level of phonotactic learning and at the level of phonological rule learning. Examples of the latter are studies by Schane et al. (1974), who found that a typologically attested epenthesis rule (akin to the rule of French liaison) is learnt faster compared to a typologically unattested one) (Schane et al., 1974), and Peperkamp et al. (2006), who found that a rule applying to a typologically attested natural class is learnt better than one applying to an arbitrary group of sounds. Similarly, C. Wilson (2006) demonstrated asymmetrical generalisation of a newly learnt velar palatalization rule, with participants generalising the rule from mid to high vowels, but not vice versa. This follows the typological fact that languages that palatalise velar stops before mid vowels also do so before high vowels, whereas the inverse is not necessarily true.[1]

Concerning vowel harmony, mixed results have been observed. Vowel harmony is attested in various forms in a great number of languages (Rose & Walker, 2011; van der Hulst & van de Weijer, 1996), yet all rules share certain properties based in phonetic substance. A learning bias has been observed for at least some of these properties. For instance, like its phonetic precursor coarticulation, vowel harmony is always directional (i.e., vowel features spread either from left-to-right or from right-to-left), and there are no rules based on a majority count, whereby the feature that occurs most often in the word spreads to the vowels without that feature; accordingly, when exposed to input that is compatible with both a directionality-based and a majority-based harmony rule, participants overwhelmingly infer the former (Finley & Badecker, 2008). Furthermore, vowel harmony causes agreement of subsegmental features (e.g., [back]), and participants do indeed base generalisation on such features rather than on individual segments (Finley & Badecker, 2009). Finally, rounding harmony is more prevalent on mid than on high vowels, presumably because rounding is percep-

[1]It should be noted, though, that this study contained other experimental conditions that did not yield learning asymmetries that are predicted by typological facts.

tually more salient in the former than in the latter, and learners generalise a newly learnt rounding harmony rule from mid to high vowels, but not the other way around (Finley, 2012).

In the present article, we focus on a further cross-linguistic generalisation concerning vowel co-occurrences: while vowel harmony rules, which promote similarity between different vowels, are common, rules that promote dissimilarity (henceforth disharmony) are exceedingly rare (for a potential example, see Krämer, 1999). Again here, the attested rules (i.e., harmony) follow from patterns of coarticulation, whereas the rare ones (i.e., disharmony) do not. If there is a learning bias favouring natural rules, we should be able to observe better learnability for a phonetically natural harmony rule compared to an unnatural disharmony rule. Previous attempts to demonstrate such a bias, however, have not been fruitful. Both American English (Pycha et al., 2003) and French (Skoruppa & Peperkamp, 2011) listeners were found to be able to learn rules governing either harmonic or disharmonic co-occurrence restrictions in a novel language setting (i.e., when learning a mini-language or a new accent of their native language). In both of these studies, a harmony rule was pitted against a disharmony rule, with one group of learners being taught the former and another the latter. No differences between the groups were found. Furthermore, Rafferty, Griffiths, and Ettlinger (2013) used an iterative learning paradigm to examine whether a newly learnt harmony rule might be better transmitted than a newly learnt disharmony rule. Indeed, previous research using this paradigm has shown that weak learning biases, that cannot be reliably observed from the outset, may be amplified over time (Reali & Griffiths, 2009; Smith & Wonnacott, 2010). This is especially interesting when considering linguistic typology, as languages do indeed evolve according to pressures from the populations using them. If there is a small bias away from disharmony, this should be translated during transmission by the maintenance of harmony systems, and the disappearance of disharmony systems. Rafferty et al. (2013), however, found no evidence for better transmission of harmony compared to disharmony. They exposed participants to artificial languages in which the proportion of harmonic items could range from 0% (completely disharmonic) to 100% (completely harmonic), and presented the output of a given participant n as the input to participant $n + 1$, forming chains of transmission. Overall, the proportion of harmonic items in all such transmission chains approached chance level, regardless of the level of harmony at which the chain began.

To sum up, previous research has found no evidence for a bias favouring the learning of harmony

over disharmony, neither within nor across individuals. It has been argued that while learners have a bias favouring formally less complex patterns (which are often phonetically natural), they have no such bias favouring phonetic naturalness per se (Moreton & Pater, 2012b, 2012a). This fits well with the fact that both Pycha et al. (2003) and Skoruppa et al. (2011) found that participants were at chance in an additional condition where they were exposed to a more complex rule, involving harmony for some vowels and disharmony for others. Thus, the lack of a learning difference between harmony and disharmony would be due to the fact that they do not differ in complexity. However, there might be other reasons why no bias favouring harmony over disharmony has been observed. For instance, it has been suggested that the naturalness bias may be modality dependent (Skoruppa et al., 2011). All three of the aforementioned studies that have examined vowel harmony used perception tasks, but there is evidence that production tasks are more likely to reveal learning biases, possibly because they are cognitively more demanding (Peperkamp et al., 2006). Similarly, it has been shown that increasing task difficulty by presenting learners with exceptions to the rule can help reveal differences in learning (Baer-Henney, Kügler, & van de Vijver, 2014). It is therefore possible that, when faced with inconsistent input, it is more difficult to learn a disharmony than a harmony rule.

The aim of the present study is to re-assess whether there is a bias favouring the learning of natural vowel harmony compared to unnatural vowel disharmony. To this end, we first use an artificial language learning paradigm with a production task and manipulate the presence of exceptions during the exposure phase across groups of participants. Next, we use our experimental group results to design and implement a computational simulation of iterated learning.

3.2.2 Artificial language learning experiments

We focus on palatal vowel harmony, a rule whereby vowels within the domain of the word must share the same value along the front/back dimension. This long distance dependency between vowels is well attested in the typology, but its converse is not. That is, we know of only one language that has been reported to have a productive case of palatal *dis*harmony, i.e., Ainu (Krämer, 1999). As in previous work, we exploit an artificial language learning paradigm to test participants' learning of a constructed mini-language.

3.2.2.1 Experiment 1

3.2.2.1.1 Stimuli, procedure, and participants

Ninety-six CVCV items were created, each containing two different consonants and two different vowels. Half of the items contained two front vowels, drawn from the set /i, e, ɛ̃/, while the other half contained two back vowels, drawn from the set /u, o, ɔ̃/; consonants were drawn from the set /b, d, g, p, t, k, v, z, ʒ, f, s, ʃ, n, ʁ/. Each of the 12 possible vowel combinations (/i-e/, /e-i/, /i-ɛ̃/, /ɛ̃-i/, /e-ɛ̃/, /ɛ̃-e/, and likewise for the back vowels) occurred equally often (i.e., in eight items).

For each of these items, which were to be used as 'singulars', two 'plurals' were created, one containing a front vowel, and the other a back vowel: CVCV+/tɛl/ and CVCV+/tɔl/, respectively. Thus, half of the plural forms were harmonic (i.e., CVCV with front vowels + /tɛl/, for instance /pegitɛl/, or CVCV with back vowels + /tɔl/, for instance /gɔ̃dutɔl/), half of them disharmonic (i.e., CVCV with front vowels + /tɔl/, for instance /pegitɔl/, or CVCV with back vowels + /tɛl/, for instance /gɔ̃dutɛl/).

All items were recorded in a soundproof booth by a female native speaker of French using an M-Audio Micro Track II digital recorder and an M-Audio DMP3 pre-amplifier in 16-bit mono at a sampling rate of 44.1 kHz.

Participants were tested individually in a quiet room in front of a computer, with stimuli being presented binaurally through a headset. They were randomly attributed to one of five exposure conditions (four test conditions and one control condition), where they heard singular/plural pairs that were all harmonic (100% harmony condition), mostly harmonic (75% harmony condition), all disharmonic (0% harmony condition), mostly disharmonic (25% harmony condition), or half and half (control condition). They were told that they would hear words from a foreign language and that these words would be presented in their singular followed by their plural form.

For each participant, the 96 singulars were pseudo-randomly divided into two sets of 48 that each contained the same number of front and back items. The first set was used for the exposure and test phases, the second set for the test phase only. During exposure, a singular was played, followed after 500 ms by its plural form, e.g., /pegi/ - /pegitɛl/. The plural could be harmonic or disharmonic, depending on the condition the participant was in. For instance, if a participant was in the 75%

harmony condition, three quarters of the plural forms during exposure were harmonic, while one quarter were disharmonic. Participants heard each stem/plural pair only once during exposure. The exposure phase lasted about three minutes.

The test phase began immediately following the exposure phase. In each trial, a singular was presented auditorily, and participants had to orally provide the plural form that they believed to correspond to the stem. Participants were first tested on the exposure set ('old items'), i.e., the same 48 items that they had just heard, and then on the (complementary) set of 48 stimuli that they had not previously heard ('new items'). All responses were recorded, and then coded offline. Responses that contained errors in the initial consonant of the suffix (e.g., /pegizɛl/ instead of /pegitɛl/) were coded as if no error had been committed. However, responses that contained errors in either the suffix vowel or the root (e.g., /pegitil/ or /petɛl/ instead of /pegitɛl/) were excluded; responses with these types of errors and missing responses, including a few where the participant was too far away from the microphone, concerned less than 1.5% of the data. All other responses clearly contained either /ɛ/ or /ɔ/. At the end of the task, participants filled out a questionnaire concerning their strategies, foreign language experience, and knowledge of phonetics and phonology.

A total of 78 native speakers of French (55 women) participated, divided almost evenly across the 5 input conditions. They were aged from 18 to 35 years old (mean: 24±3.9). An additional 26 participants were excluded because they had studied phonetics or phonology (N=9), had knowledge of a language with vowel harmony (N=1), did not complete the task properly (e.g., did not follow recording instructions, or pronounced isolated suffixes; N=14), or because of experimenter error (N=2).

3.2.2.1.2 Results and discussion

Data were analysed using logistic mixed-effects models in R (Bates et al., 2014). Effects were included into all models using contrast coding, and significance was assessed through model comparison.

An initial analysis was performed on the control condition (50% harmonic input, 50% disharmonic input) to verify that participants had no baseline bias towards either harmony or disharmony. Participants in this condition cannot learn a rule from the input they are exposed to, as their input is

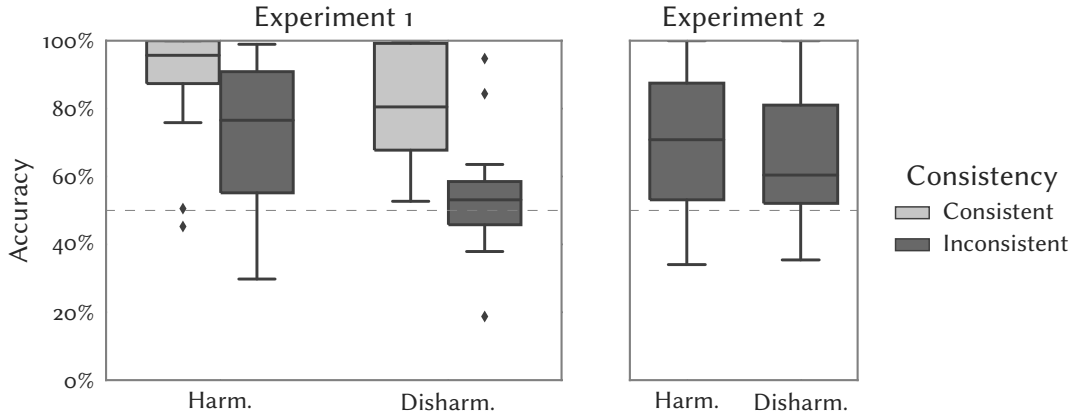

Figure 3.1: Accuracy scores for the test conditions averaged by participant, presented by Rule and Consistency in Experiment 1 (left side) and by Rule in Experiment 2 (right side).

mixed; we therefore compared their performance to chance level. We built a model with no fixed effects, including only an intercept and the random effects Participant and Stem. This model was not found to differ significantly from a null model which included only the random effects, i.e., with no intercept ($\beta = 0.02$, $SE = 0.27$, $\chi^2(1) < 1$), indicating that performance was not better than chance. Data from the control condition were not considered for further analysis.

For the four test conditions, we analysed the number of participants' correct responses. Recall that what is considered a correct response varies by group, so that for participants in the harmonic groups (100% and 75% harmony), a harmonic response is considered correct, while the opposite is true for participants in the disharmonic groups (0% and 25% harmony). An initial model was built with a fixed effect for Item Type (i.e., old items vs new items) and random effects for Participant and Stem. When compared to a model that excluded Item Type, no significant difference was observed ($\beta = 0.06$, $SE = 0.07$, $\chi^2(1) < 1$). For all subsequent analyses, we therefore collapsed data over old and new items to increase statistical power. Mean accuracy scores by Rule and Consistency condition (collapsed over old and new items) can be seen on the left side of Fig. 3.1.

We designed a full model, which included the following fixed effects: Rule (harmony or disharmony) and Consistency (consistent or inconsistent exposure), as well as the interaction between these two. It also included random effects for Participant and Stem. This model was compared to simpler models which each excluded one of the fixed effects or the interaction.

The full model was found to explain significantly more variance than a model which excluded Rule

(β = 0.98, SE = 0.43, $\chi^2(1)$ = 4.87, p = 0.03) and a model which excluded Consistency (β = 2.09, SE = 0.44, $\chi^2(1)$ = 20.6, $p < 0.0001$), showing that participants learnt the harmony rule better than the disharmony rule, and that they learnt better from consistent than from inconsistent input. However, the full model was not found to significantly differ from a model that excluded the interaction between Rule and Consistency ($\beta = -0.31$, SE = 0.878, $\chi^2(1) < 1$).

In addition to these models, we analysed subsets of the data in order to pinpoint the source of the observed effect of Rule. An effect of Rule was observed in the inconsistent input conditions (i.e., 75% versus 25% harmony; β = 1.10, SE = 0.41, $\chi^2(1)$ = 6.55, p = 0.01), but not in the consistent ones (i.e., 100% versus 0% harmony; β = 0.91, SE = 0.90, $\chi^2(1) < 1$). These results show, firstly, that a natural rule of vowel harmony is learnt better than an unnatural rule of vowel disharmony, and, secondly, that the asymmetry is due to the inconsistent input conditions. This is in line with previous research demonstrating that adding exceptions to the input in an artificial language learning paradigm can reveal differences in learning that are otherwise not observed (Baer-Henney et al., 2014).

We performed further subset analyses to test individual conditions against chance level. All of these analyses were carried out in the same manner as for the control condition described above. Performance was above chance in the 100% harmony (β = 3.63, SE = 0.68, $\chi^2(1)$ = 17.0, $p < 0.001$), 75% harmony (β = 1.41, SE = 0.38, $\chi^2(1)$ = 10.4, p = 0.001), and 0% harmony (β = 2.27, SE = 0.63, $\chi^2(1)$ = 13.3, $p < 0.001$) conditions, but not in the 25% harmony condition (β = 0.21, SE = 0.23, $\chi^2(1) < 1$).

Overall, the results of this experiment suggest that harmony is learnt better than disharmony, contrary to what was reported in previous studies (Pycha et al., 2003; Skoruppa et al., 2011). Note, though, that the true source for the learning difference was from a restricted comparison within our design, involving the inconsistent input conditions only. In the next experiment, we therefore use the same materials but focus on these crucial conditions. We predict to replicate the previous effects, that is, better performance in the 75% than in the 25% harmony condition, with above-chance performance for 75% harmony only.

3.2.2.2 Experiment 2

3.2.2.2.1 Procedure and participants

The procedure were the same as in the previous experiment, except that only the inconsistent conditions were tested (i.e., 75% and 25% harmony).

A total of 32 native speakers of French (22 women) participated, aged from 18 to 34 years of age (mean: 23.5). They were randomly assigned to one of the two experimental conditions. None of them had participated in the original experiment. Ten additional participants were excluded because they had studied phonetics or phonology (N=2), did not complete the task properly (e.g., did not follow recording instructions, or did not report the root with the suffix) (N=3), had knowledge of a vowel harmonic language (N=1), or because of experimenter error (N=4).

3.2.2.3 Results and discussion

As in Experiment 1, data were analysed using logistic mixed-effects models; effects were included into the models using contrast coding; and significance was assessed through model comparison. Mean accuracy scores by Rule can be seen on the right side of Fig. 3.1. Again, to increase statistical power, old and new items were collapsed.

We designed a full model, which included a fixed effect for Rule (harmony or disharmony) and random effects for Participant and Stem. It was not found to explain significantly more variance than a model which excluded the fixed effect of Rule (β = 0.10, SE = 0.57, $\chi^2(1) < 1$), indicating that participants who learnt the harmony rule did not perform better than those who learnt the disharmony rule. As in the original experiment, we also tested whether performance in both conditions was above chance. This was the case in both conditions (75% harmony: β = 1.24, SE = 0.38, $\chi^2(1) = 8.2$, $p < 0.01$; 25% harmony: β = 1.16, SE = 0.45, $\chi^2(1) = 5.9$, $p = 0.02$).

Thus, using the same number of participants per condition, we failed to replicate the asymmetry between harmony and disharmony observed in the two inconsistent input conditions of Experiment 1, finding only a numerical trend in the same direction (mean accuracy by condition: harmony =

69.0%, disharmony = 65.7%). Specifically, unlike in Experiment 1, participants in the present experiment were able to learn the inconsistent disharmony rule. This discrepancy highlights the issue of sample sizes. When we collapse data across the two experiments, and hence analyse twice as many data points in the inconsistent conditions, we observe a marginally significant effect of Rule (β = 0.70, SE = 0.36, $\chi^2(1)$ = 3.62, p = 0.057). It is therefore likely that Experiment 1 overestimated the size of the effect.

In addition to these analyses, we performed a post-hoc exploration of participants' performance on old items only, collapsing over Experiments 1 and 2, to test the extent to which they had memorized the individual plurals. In the consistent conditions, this is equivalent to testing their accuracy with regards to the rule they were to learn, so we focus here exclusively on the inconsistent conditions (i.e., 25% and 75% harmony). We designed mixed-effects models as described above, but with item-based correct response as the dependent variable. Thus, trials in which participants produced the same plural as the one they had heard during exposure (modulo errors in the suffix's consonant) were coded as correct. The full model included a fixed effect for Rule (harmony or disharmony) and random effects for Participant and Stem.

The full model was found to explain significantly more variance than a model which excluded Rule (β = 0.32, SE = 0.11, $\chi^2(1)$ = 8.10, $p < 0.01$), showing that plurals were memorized better in the harmony than in the disharmony condition. These results complement our full analysis targeting rule-adherence and including both old and new items. That is, participants not only learned the general rule they were exposed to better in the harmony than in the disharmony conditions, but they also learned individual harmonic plurals better than disharmonic ones.

3.2.2.4 Summary and discussion

To sum up, the overall experimental results are inconclusive: we observed a significant difference between harmony and disharmony in Experiment 1, no such difference in Experiment 2, and a marginal difference when analysing the data from the two experiments together. Recall that earlier studies using a perception task and completely consistent input found no difference between harmony and disharmony (Pycha et al., 2003; Skoruppa et al., 2011). The methodology in Skoruppa et al. (2011) was

very different from the present one: they exposed participants to 40-minute-long short stories spoken in a novel, systematically harmonic or disharmonic, accent of the participants' native language. Pycha et al. (2003), however, used an artificial language learning paradigm similar to ours. It should be noted that they observed a numerical difference (mean accuracy: harmony = 86%, disharmony = 75%) but tested only 10 participants per group (which was sufficient to show an advantage for learning harmony compared to learning a formally more complex rule mixing harmony and disharmony); hence, the study was likely underpowered.

All in all, it is impossible to draw firm conclusions concerning the presence or absence of a learning bias favouring harmony; still, no study so far has found an effect or even a numerical trend in the opposite direction, i.e. favouring *dis*harmony. Our working hypothesis, therefore, is that learners have a small bias that is difficult to observe experimentally. This is likely due to the presence of considerable individual variation, rendering even our study underpowered. We are encouraged by the fact that participants showed evidence of better memorization in the harmony than in the disharmony condition, and, furthermore, that they did not simply probability match indiscriminately. Indeed, were participants to probability match, we would not see the asymmetry between the 75% and 25% conditions (as performance would be distributed around 75% accuracy for both conditions). Recall in particular that in the original experiment, participants in the 25% condition did not learn the disharmony rule above chance level, indicating that they gave on average as many harmonic as disharmonic responses.

In the next section, we use computational modelling to examine whether a small difference, like that observed in our complete data set, can compound over time, such that it in the long run, it can influence typology.

3.2.3 Modelling transmission

Experimental work on iterated learning has shown that weak biases favouring certain linguistic patterns may be amplified by transmission within a population (Reali & Griffiths, 2009; Smith & Wonnacott, 2010). For instance, Reali and Griffiths (2009) demonstrated how a bias towards regularisation can quickly emerge during the transmission process. Participants in this study were ex-

posed to object-word mappings in an artificial language. Each object they saw was paired with two different words across trials, the frequency of which differed. The authors found that participants in the first generation generally produced words with the same frequency with which they were presented. Close examination of their results, however, reveals slight numerical deviations from the input. When one participant's output was used as the input for the next one, these deviations were quickly amplified, and after multiple iterations only the word that was most frequent at the beginning of the chain remained. In light of such results, it stands to reason that an explanation of typological facts may require investigation beyond experimentation at the individual level. Indeed, results from modelling work also point towards transmission strengthening weak biases (Griffiths & Kalish, 2007; S. Kirby, Dowman, & Griffiths, 2007; S. Kirby, 2001). These studies simulate behaviour of agents rather than testing transmission in the laboratory. Although they focus mostly on the emergence of compositionality or regular word-meaning mappings, their predictions are generally applicable to any linguistic phenomenon, including phonological rules.

Given that the data from our experiments provided inconclusive evidence of a learning bias favouring the learning of harmony, an iterated learning approach can be particularly insightful. Here, we examine whether the small asymmetry we observed in our experimental data will compound over time, or whether, as in the study by Rafferty et al. (2013), it will rather disappear. To this end, we construct a computational model of the transmission of the harmony and disharmony rules over time. Specifically, we build on S. Kirby (2001), who arranged simulated agents in linear chains of transmission, such that the output of a given agent is used as input for the next agent in the chain. We adopt a hybrid approach, though, building a computational model based off of a first generation of real learners' behaviour. That is, rather than testing a great many participants in transmission chains, we use our experimental data to initiate chains, whose evolution is simulated by probabilistically sampling from observed behaviour in the different experimental input conditions.

3.2.3.1 Methods

3.2.3.1.1 Test simulation

Data from the two experiments were pooled together and used as a base for the simulation; there were thus approximately twice as many data points for the inconsistent conditions as for the consistent and the control conditions. We simulate the propagation of harmony across generations of individuals, which we will refer to as *agents*. We directly model the average amount of harmony produced by an agent, rather than the specific items an agent might produce as harmonic or disharmonic. Specifically, given an input probability representing a proportion x_n of harmonic items, an agent n produces an output probability representing a proportion y_n of harmonic responses, which is presented as input x_{n+1} to agent $n + 1$, who then produces output y_{n+1} which is presented to agent $n + 2$ etc. Thus, the agents form a *transmission chain*, analogously to what we would obtain in our experimental paradigm if we provided the output of a given participant as input to the next one. We study simulated transmission chains with 50 agents per chain; the first agent of a chain is given one of the proportions of harmony x_0 that we used in our experiments (i.e., 0%, 25%, 50%, 75% or 100%). One hundred transmission chains are simulated for each of the five possible initial harmony conditions, for a total of 500 simulated transmission chains.

We assume that all agents are drawn uniformly at random from some underlying population. Then, to simulate the transmission chains, we estimate the probability distribution $P(x)$ of harmony proportions produced by randomly selected agents for any given amount x of harmony proportion in the input to these agents. Obtaining $P(x)$ if x is equal to one of the tested conditions (0%, 25%, 50%, 75% or 100%) is quite straightforward. For each of these conditions we have a sample y_1, ..., y_{15} or y_1, ..., y_{30} (one measure for each subject tested in the corresponding condition) from which we can estimate $P(x)$ parametrically using beta distributions. Beta distributions form a simple and classic family of distributions that are widely used to model the behavior of random variables whose values are bounded, as is the proportion of harmony responses in our experiments. Beta distributions are parameterised by two parameters, providing a good trade-off between simplicity and flexibility. We fitted these parameters with the method of moments (Hanson, 1991). As can be seen in the top row of Fig. 3.2, we obtain reasonably good fits to our experimental data.

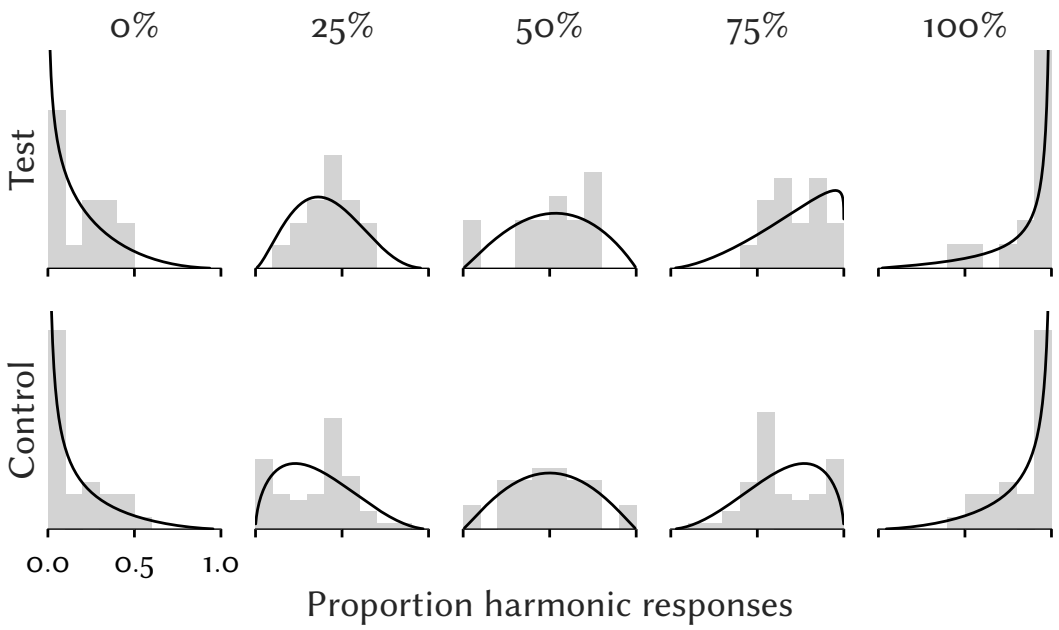

Figure 3.2: Beta distributions (probability density functions) estimated with the method of moments for the distribution of participants' responses from our experiments, shown as histograms. Each panel corresponds to one of the input conditions (levels of harmony listed above each column). In the top row, data from the original experiment and replication combined, in the bottom row, the same data combined with simulated 'flipped' data.

When x is not equal to one of the tested conditions (e.g., $x = 82\%$ harmony), we also modelled $P(x)$ using beta distributions. In this case, however, we did not have access to empirical samples from which to estimate the distribution parameters. Instead, we estimated the parameters from those obtained in the tested conditions. The estimation was performed using smoothing splines(Green & Silverman, 1993), a standard method for obtaining smooth fits. This method provides us with a y value for any given x level of harmony input by estimating a function based on the values estimated for the five tested conditions.

By way of example, let us develop how we obtain a transmission chain starting with $x = 25\%$ as the input for agent $n = 1$. First, we estimate the probability distribution $P(25\%)$ by fitting a beta distribution to the data produced by all participants exposed to 25% harmony. While the output values can lie anywhere between 0% and 100%, the second tile of the top row of Fig. 3.2 shows that the output values with the highest sampling probability are situated in the vicinity of 30–40% harmony (i.e., 0.3-0.4 on the horizontal axis). To start the simulation of the transmission process, we randomly sample a value from the estimated distribution $P(25\%)$, say $y = 0.435$. We interpret this as

meaning that agent n=1, the first agent in a simulated transmission chain starting with 25% harmony, produces an output of 43.5% harmony. Next, this output becomes the input given to the following agent in the chain, agent $n = 2$. In order to proceed in a manner analogous to agent $n = 1$, we now need to estimate the probability distribution $P(43.5\%)$. As this input harmony condition was not tested experimentally, we infer it by means of interpolation from the probability distributions that we estimated from the experimental data. We then sample a value from $P(43.5\%)$, say $y = 0.698$, and interpret this as meaning that the output of agent $n = 2$ is 69.8% harmony. This is then used as input for agent $n = 3$, and the simulation process is iterated until the chain reaches 50 agents. Transmission chains with other starting conditions are created in a similar fashion.

3.2.3.1.2 Control simulation

In order to ensure that any results obtained in our test simulation are due to biases in the data, rather than a bias inherent in the model, we also performed a simulation on a set of control data that does not contain any asymmetries in the patterns of responses for vowel harmony and vowel disharmony. This control data set consisted of a combination of the data used for the test simulation described above with a symmetrical, 'flipped' version. That is, for each participant, we constructed a new 'control' participant by switching both the input condition and the participant's output. For instance, a participant in the 75% harmony input condition who gave 88% harmonic responses would give rise to a new control participant in the 25% harmony condition with 12% harmonic responses. For the 50% condition, only the participant's output was flipped; thus, a participant in this condition who gave, say, 53% harmonic responses would give rise to a new participant in the 50% harmony condition, but with 47% harmonic responses. The overall data were thus symmetrical relative to the 50% harmony input condition. Values for the parameters of the beta distributions for inputs 0%, 25%, 50%, 75% and 100% were estimated in the same way as in the test simulation, and similarly, values for the unknown beta distributions between the tested conditions were interpolated as described above. The resulting beta distributions are shown in the bottom row of Fig. 3.2.

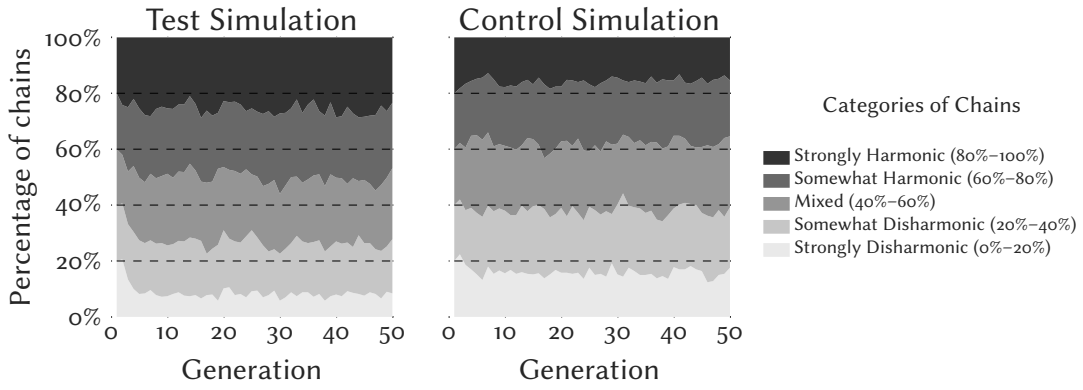

Figure 3.3: Results from the transmission simulation using experimental data (Experiments 1 and 2 combined, left side), and experimental and control 'flipped' data (right side). The shaded areas represent the percentage of chains that fall within a given interval over time.

3.2.3.2 Results and discussion

At each generation, we classified each chain as belonging to one of five categories: A chain was considered to be *strongly harmonic* if over 80% of its words were harmonic, *somewhat harmonic* if between 60% and 80% of its words were harmonic, *mixed* if between 40% and 60% of its words were harmonic, *somewhat disharmonic* if between 20% and 40% of its words were harmonic, and *strongly disharmonic* if less than 20% of its words were harmonic. The percentages of chains falling into each category at each generation are displayed in Fig. 3.3.

At generation 0, each category contains exactly 20% of the total set of chains (N=500). Over time, the category of strongly harmonic chains is seen to increase in the test simulation, while the category of strongly *dis*harmonic ones decreases. This pattern does not hold for the control simulation, where the strongly harmonic and strongly disharmonic categories contain approximately the same number of chains at each iteration.

Focusing on the strongly harmonic and strongly disharmonic chains, Fig. 3.4 shows the mean proportion of chains in each of these categories over the course of the simulation.

We performed a 2 × 2 ANOVA with the factors Simulation (test vs. control) and Harmony (strongly harmonic vs. strongly disharmonic) on these data. Main effects of both Simulation ($F(1, 96) = 7.5$, $p < 0.01$) and Harmony ($F(1, 96) = 383$, $p < 0.0001$) were observed, showing that both categories (strongly harmonic or disharmonic chains) contained higher proportions of the overall chains in the

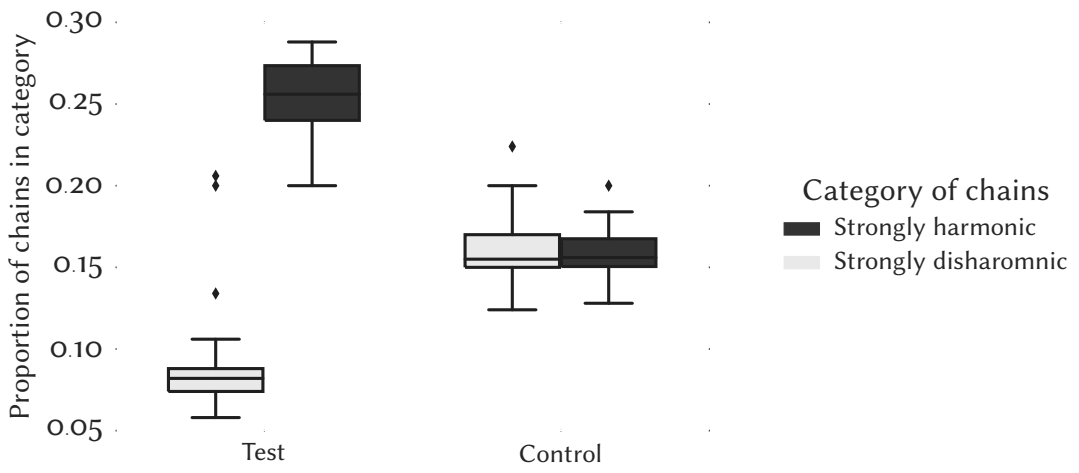

Figure 3.4: Distributions of the proportion of chains in the categories 'strongly harmonic' (80%–100% harmonic words) and 'strongly disharmonic' (0%–20% harmonic words) in the test (left) and control (right) simulations.

test compared to the control simulation, and that overall, more chains were strongly harmonic than strongly disharmonic. Crucially, there was a significant interaction of these two factors ($F(1, 96) = 407$, $p < 0.0001$), showing that in the test simulation, but not in the control simulation, there was a strong bias away from disharmony.

These results show that despite the initial bias being small, it compounded over time: after only a few generations, chains that were 'strongly harmonic' vastly outnumbered chains that were 'strongly disharmonic', and this pattern remained stable throughout the simulation. This is in line with previous results on iterated learning, using both experiments (Reali & Griffiths, 2009; Smith & Wonnacott, 2010) and modelling (Griffiths & Kalish, 2007; S. Kirby et al., 2007), in which even small biases were found to be amplified during transmission across individuals. However, our results are unexpected given Rafferty et al. (2013)'s experiment on the linear transmission of vowel harmony. Recall that they also used five initial input conditions, with 0%, 25%, 50%, 75% and 100% harmonic items, respectively, and that the proportion of harmonic items in all their transmission chains approached chance level, regardless of the level of harmony at which the chain began. We argue that their divergent result is due to the fact that they had at most four chains per input condition, such that individual variability had a strong impact on their results. For instance, two of the four chains that started with 100% harmonic items remained at 100% harmony in the first generation but then dropped to 50% in the second generation. Thus, in both chains, the first participant showed perfect accuracy

and the next one performed at chance. Unsurprisingly, these chains did not change with further participants. Indeed, a bias that renders the learning of harmony rules easier than the learning of disharmony rules is not the same thing as a bias for harmony to emerge. It is thus very unlikely that once there is no longer evidence for a harmony rule in the input, participants would reintroduce one (cf. our experimental control condition where the mean percentage of harmonic responses was not above chance). Hence, given that one participant's performance can be deterministic for the rest of the chain, and that this type of experiment typically shows quite a lot of individual variability, a considerably larger amount of chains per initial input condition would have been necessary to obtain reliable results. In our modelling, we simulated a great many chains (100 per starting condition), thus minimizing the risk of one participant biasing the results. Indeed, the probabilistic sampling from distributions of responses ensures that on average, chains represent the general behaviour of participants rather than the behaviour of one specific participant. The results from our simulation, then, reveal how a small underlying bias favouring the learning of harmony can compound over time and shape typology.

As our simulation depends directly on the experimental results we obtained, let us speculate on alternative outcomes that might have been observed had participants in our experiments behaved differently. It is likely that the asymmetry we observed in Experiment 1 between the 25% and 75% harmony conditions greatly influenced our model results. Indeed, in a simulation not shown here, based only on the results from Experiment 1, the bias towards harmonic languages and away from disharmonic ones is even greater than the one obtained above. Given the large inter-individual variability present in artificial language-learning experiments, one could wonder how accurate such a simulation is. We addressed this by collecting additional data in the critical conditions, and saw that in the simulation with the combined data, the bias - which was not much reduced - still compounded over time. Furthermore, in an additional simulation including only the data from Experiment 2, we saw that the small synchronic bias was enough for compounding to take hold. We predict therefore that even if the synchronic bias were nearly erased after collecting additional data in these critical conditions, the diachronic tendency would remain so long as even a small trend favouring the learning of harmony were present.

3.2.4 Conclusion

Using an artificial language-learning paradigm combined with a computational model simulating iterative learning, we showed that a weak, non-significant asymmetry favouring harmony over disharmony compounds over time, with transmission of the learnt rule across individuals. Our experimental data are inconclusive: while we found a bias favouring harmony in participants exposed to inconsistent input, i.e., input containing 25% exceptions to the rule they had to learn, we failed to observe more than a numerical trend in the same direction in a replication of these conditions. By contrast, our simulation of the transmission of these rules across generations in an iterated learning-style manner showed clear results: even the small difference that we observed in learning between harmony and disharmony compounded over time, yielding a great number of harmonic languages and, crucially, very few disharmonic languages. Thus, a learning bias, even small, can benefit from the iterative nature of language transmission (i.e., the learning process itself) and have a strong impact over time.

Our results are in line with several previous studies on linear transmission, which, using both computational modelling and experiments with iterated learning, show that even very weak biases may compound over simulated time (Griffiths & Kalish, 2007; S. Kirby et al., 2007; Reali & Griffiths, 2009; Smith & Wonnacott, 2010). For instance, Reali and Griffiths (2009)'s experiment with an iterated learning paradigm showed that although participants in the first generation tended to frequency-match their input, a bias for generalisation quickly set in. Recall that participants in their study were exposed to objects paired with two labels that appeared with different frequencies. During transmission, one of these two labels disappeared, yielding one-to-one mappings. We observe a similar pattern in our results, with participants generally frequency-matching, but deviating enough from input frequency for a small, numerical bias to compound over time in the simulation.

Iterated learning clearly provides a framework to study effects that may otherwise be difficult to observe. Our hybrid approach to iterated learning combining experimental data with modelling has a clear advantage over a purely experimental approach, in that the latter is rather costly in terms of the number of participants. Indeed, as mentioned earlier, many chains per initial input condition are necessary to obtain reliable results (unless the phenomenon to be considered is virtually

devoid of individual variation). This is because linear transmission occurs within agents arranged in Markov chains, such that any one participant's performance is directly influenced by the performance of only one other participant (namely, the previous one), whose performance may or may not be representative of the population mean. For example, if a given participant does not pay attention during exposure, their responses during test, although uninformative, will nonetheless determine the input for the next generation, and can therefore affect the trajectory of the chain. Thus, an accurate portrayal of transmission requires testing participants in tens or even hundreds of Markov chains. The strength of the stimulation method we propose is that it is based on probabilistic sampling from distributions of participant responses. That is, instead of testing, say, one hundred participants divided over ten linear transmission chains, each of which has a high chance of falling prey to inter-participant variability, we cluster roughly the same number of participants into groups that each represent a different level of input at the first generation, and then build a model of transmission. This hybrid method of sampling from distributions based on the responses from an initial pool of participants allows for the combination of observed behaviour in a restricted number of human learners with simulated behaviour in large numbers of agents (in Markov chains). Given the cost of running large-scale experiments, any step that can be simulated with a computer model seems beneficial, though it is important to understand the upper limits of interpolation. To this end, in future research we could expose listeners to levels of harmony that we did not include in our experiment, and compare their performance to our interpolations.

Another topic for future research concerns the social aspects of language learning. So far, almost all research on iterated learning, whether based on experiments or on modelling, has used simple Markov chains. Language acquisition, however, takes place in an interactive, multi-speaker environment, in which speakers have different social roles and levels of interaction with one another. It is possible to build a model in which learners are exposed to output from more than one teacher, and output from one teacher is given to multiple learners (e.g., J. P. Kirby & Sonderegger, 2015). Moreover, models incorporating more dynamic social structure have been shown to yield more realistic results than simple iterative learning (Niyogi & Berwick, 2009). Our hybrid approach of experimentally-based computational simulations can be adapted to include such interactive multi-agent networks.

Finally, let us get back to typology and the question of why vowel harmony is common while vowel disharmony is not. It should be noted that our simulation does not speak to the overall distribution of harmonic systems in the world's languages. This is because all five initial conditions contain two allomorphs, but in four of them (0, 25, 50 and 75% harmony) their conditioning is unattested in natural languages. Thus, our simulation only shows that all else being equal, vowel harmony has a higher probability of being transmitted across generations than vowel disharmony.[2] This is due to the fact that we observed a small numerical bias on the individual level that compounded over time; had the bias been observed in the opposite direction, we predict that disharmonic systems, rather than harmonic ones, would be more prevalent. For a modelling approach to the evolution of harmonic systems based on historical data, taking into account a multitude of phonological, lexical, and language-external parameters, see Harrison, Dras, and Kapicioglu (2002). Similarly, given the hypothesis that vowel harmony is due to a diachronic evolution of vowel-to-vowel coarticulation (e.g., Ohala, 1994; Blevins, 2004), our simulation does not speak to the emergence of vowel harmony either, which likely depends upon the strength of this phonetic precursor (Beddor, Krakow, & Lindemann, 2001).

To conclude, we have provided evidence for the existence of a learning bias, suggesting that listeners have knowledge of the phonetic naturalness of vowel harmony as part of their synchronic grammar, in accordance with substance-based theories of phonology (Archangeli & Pulleyblank, 1994; Donegan & Stampe, 1979; B. Hayes et al., 2004). More research is needed to investigate to what extent typology is influenced by phonetic pressure in speech perception and production on the one hand, and learning bias on the other hand.

3.3 Phonetic naturalness bias in the learning of vowel harmony

This section reports on two experiments that have not been included in any manuscript submitted for publication.

[2] We only analysed the difference between strongly harmonic and disharmonic chains, as this is the typological question underlying our research. Note, though, that in our test simulation, there is a numerically higher proportion of strongly harmonic (0.255), than somewhat harmonic (0.241), than mixed (0.23), than somewhat disharmonic (0.186), than strongly disharmonic (0.086) chains.

3.3.1 Introduction

Skoruppa and Peperkamp (2011) discuss the possibility that modality may influence learners in artificial language learning experiments. Indeed, as discussed in Section 1.2.2, vowel harmony has been proposed to emerge from a constraint on production: vowel harmony facilitates the articulation of words by promoting similarity between segments. Is it the case that the production-based task reported in Section 3.2 is the source of the asymmetry we observe between the learning of the natural versus the unnatural pattern? In this section, we report on two experiments that use the same stimuli and exposure as Section 3.2, but require participants to choose between two alternative plurals during the test phase, presented auditorily. This manipulation removes the influence of the production system by forcing participants to choose between alternative they perceive only.

3.3.2 Methods

3.3.2.1 Procedure

The procedure was identical to that reported in ?? 3.2.2.1.1, except that the test phase used a perception-only task.

The test phase began immediately following the exposure phase. Test trials consisted in the presentation of a stem followed by both possible plural forms (i.e., CVCV-tɛl and CVCV-tɔl) with an ISI of 500 ms. The order of presentation of the two plural forms was randomized across trials. After all three stimuli were played, two buttons appeared at random locations on the screen, labeled with the two plural suffixes. Participants were asked to select which of the two plural forms they thought was correct by clicking on the corresponding button. Participants were first tested on the exposure set (i.e., the same items that they had just heard) before being tested on another set of 48 stimuli that they had not previously heard, for a total of 96 trials.

At the end of the task, participants filled out a questionnaire concerning their response strategies, foreign language experience, and knowledge of phonetics and phonology.

3.3.2.2 Participants

A total of 92 native French-speaking participants were tested (61 women, 31 men), aged from 15 to 35 years old (mean: 23.5). The data of 10 participants were excluded because they had studied phonetics or phonology (N=9) or had knowledge of a language with vowel harmony (N=1). An additional seven participants were excluded because of an experimenter error. This yielded five groups of 15 participants.

3.3.3 Results

Data were analyzed using logistical mixed-effects models in R (Bates et al., 2014). Effects were included into all models using contrast coding, and significance was assessed through model comparison.

An initial analysis was performed on the control condition (50% harmonic input, 50% disharmonic input) to verify that participants had no baseline bias towards harmony or disharmony. Recall that participants in this condition cannot learn a rule from the input they are exposed to, as their input is mixed. We therefore compared their performance to chance level (50% since the task was 2AFC). We built a model with no fixed effects, including only an intercept and the random effects Participant and Stem. This model was not found to differ significantly from a null model which included only the random effects (i.e., with no intercept) ($\beta = 0.12$, $SE = 0.16$, $\chi^2(1) < 1$), indicating no preference for harmony or disharmony. Data from the control condition were not considered for further analysis.

For the four test conditions, we analyzed the number of participants' correct responses. Recall that what is considered a 'correct' response varies by group, so that for participants in the disharmonic group, a disharmonic response is considered correct, while the opposite is true for participants in the harmonic group.

An initial model was built with the fixed effect of Item Type (i.e., old items vs new items) and random effects of Participant and Stem. When compared to a model that excluded the fixed effect of Item Type, no significant difference was observed ($\beta = 0.06$, $SE = 0.05$, $\chi^2(1) = 1.36$, $p = 0.24$). For all subsequent analyses, we therefore collapsed data over old and new items to increase statistical

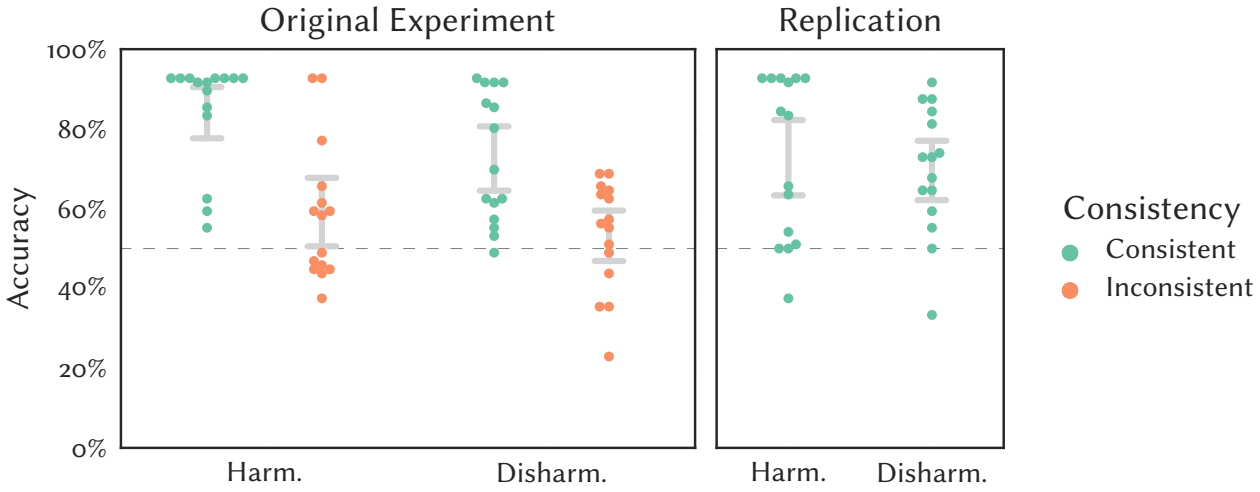

Figure 3.5: Accuracy scores for the test conditions averaged by participant, presented by Rule and Consistency in the original experiment (left side) and by Rule in the replication (right side). Grey bars in the middle represent one standard error above and below the mean.

power. Mean accuracy scores by Rule and Consistency condition (collapsed over old and new items) can be seen on the left side of Fig. 3.5.

We designed a full model, which included the following fixed effects: Rule (harmony or disharmony) and Consistency (consistent or inconsistent exposure), in addition to the interaction between these two. It also included random effects for Participant and Stem. This model was compared to simpler models which each excluded one of the fixed effects or the interaction.

The full model was found to explain significantly more variance than a model which excluded the fixed effect of Rule ($\beta = 0.57$, $SE = 0.22$, $\chi^2(1) = 6.22$, $p = 0.01$) and a model which excluded the fixed effect of Consistency ($\beta = 1.32$, $SE = 0.22$, $\chi^2(1) = 27.7$, $p < 0.0001$), showing that participants learnt the harmony rule better than the disharmony rule, and that they learnt better in consistent than in inconsistent conditions. However, the full model was not found to significantly differ from a model that excluded the interaction between Rule and Consistency ($\beta = 0.50$, $SE = 0.45$, $\chi^2(1) = 1.25$, $p = 0.26$).

In addition to these models, we analyzed subsets of the data in order to pinpoint the source of the observed effect of Rule. No significant effect of Rule was observed when considering only inconsistent input conditions (i.e., 75% harmony versus 25% harmony) ($\beta = 0.30$, $SE = 0.26$, $\chi^2(1) = 1.24$, $p = 0.26$), while a clear effect was observed in the consistent conditions (i.e., 100% harmony versus

0% harmony) (β = 0.99, SE = 0.44, $\chi^2(1)$ = 5.14, p = 0.02). This means that participants learned the harmony rule better, but only when there were no exceptions in the input.

We further verified that performance was above chance level in the inconsistent conditions. It is possible that no difference between harmony and disharmony was observed in these conditions simply because no learning was observed. When we compared a base model that included only an intercept and the same random effects as described above to a null model which did not include an intercept, we found that the base model did explain significantly more variance (β = 0.28, SE = 0.13, $\chi^2(1)$ = 4.15, p = 0.04), indicating that though performance was overall poor in the inconsistent conditions (mean: 56%), it was significantly above chance.

3.3.4 Replication

We additionally ran a replication of the principal result from our first experiment, which focused on the crucial conditions where a difference was observed (i.e., when input was consistent). Recall that no difference between harmony and disharmony was observed when participants were exposed to inconsistent input. However, when input was consistent, participants learned the harmony rule better than the disharmony rule. The following section reports on results from a replication of these critical consistent conditions.

3.3.4.1 Methods

3.3.4.1.1 Stimuli and procedure

The stimuli and procedure did not differ between the original experiment and the replication, except that only the consistent conditions were tested in the replication (i.e., 100% harmony and 0% harmony).

3.3.4.1.2 Participants

A total of 32 native French-speaking participants (none of whom had participated in the original experiment or the experiments reported in Section 3.2) were tested (21 women, 11 men), aged from

18 to 35 (mean: 23.5). One participant was excluded because they had knowledge of phonetics or phonology and another was excluded due to an experimenter error. This yielded two final test groups of 15 participants each.

3.3.4.1.3 Results

As in the original experiment, data were analyzed using logistical mixed-effects models; effects were included into the models using contrast coding, and significance was assessed through model comparison. Mean accuracy scores by Rule can be seen on the right side of Fig. 3.5. Again to increase statistical power, old and new items were collapsed.

We designed a full model, which included a fixed effect for Rule (harmony or disharmony) and random effects for Participant and Stem. It was not found to explain significantly more variance than a model which excluded the fixed effect of Rule (β = 0.56, SE = 0.44, $\chi^2(1)$ = 1.60, p = 0.21), indicating that participants who learnt the harmony rule did not perform significantly better than those who learnt the disharmony rule. We then collapsed the data between the original experiment and the replication. Recall that the replication tested only the consistent (0% and 100% harmony) conditions. The collapsed dataset therefore had twice as many datapoints in the consistent than in the inconsistent conditions. We then performed the same data analysis as for the original experiment (full model including the factors Rule and Consistency compared to reduced models). The full model was found to explain significantly more variance than models which excluded the factor Rule (β = 0.46, SE = 0.22, $\chi^2(1)$ = 4.38, $p < 0.05$) and the factor Consistency (β = 1.12, SE = 0.22, $\chi^2(1)$ = 23.1, $p < 0.0001$), but not the interaction of these two factors ($\chi^2(1) < 1$). This means that when considering both datasets together, participants generally performed better when learning harmony than when learning disharmony, and better in the consistent than in the inconsistent conditions, but no significant asymmetrical effects were observed. The main effects reflect what was found in the original experiment.

3.3.5 Discussion

When presented with two alternative plurals during the test phase, participants in our perception-only artificial language learning experiments show similar patterns of bias to those reported in the production-based experiments of Section 3.2. Overall, the natural rule of harmony is learnt better than the unnatural rule of disharmony (though as in the previous studies, our replication attempt highlights the sensitivity to sampling issues that are rampant in artificial language learning experiments). Interestingly, the bias is observed in different conditions than those of the previous studies. Recall that in the previous studies, the asymmetry between harmony and disharmony came out in the inconsistent conditions, whereas in the present experiments, it is in the consistent conditions that we observed the difference.

Overall, performance was lower in the perception tasks than the production tasks. This goes counter to what was suggested in Skoruppa and Peperkamp (2011), who claimed that a perception-only task was less cognitively demanding than a production task. We interpret the current results as having something to do with the precariousness of the effect to begin with. If, coming out of the exposure phase, participants are yet unsure of the rule they had to learn, being presented with both plurals at each trial may weaken their certainty with regards to what they heard during exposure (i.e., over-exposure during the test phase). This could explain why only those participants who received unambiguous input (i.e., those in the consistent conditions) demonstrated the asymmetry.

After having considered how a listener perceives a pattern (Section 3.3), then produces it (Section 3.2), and how this could influence the shaping of typology over several generations of learners (Section 3.2.3), in the next section, we explore a further possible factor that may influence the learning of phonological patterns: the role of sleep.

3.4 Sensitivity to phonetic naturalness in phonological rules: evidence from learning and consolidation with sleep

This section is an adaptation[3] of the following manuscript: Martin, A., & Peperkamp, S. (in revision). Sensitivity to phonetic naturalness in phonological rules: evidence from learning and consolidation with sleep. *Laboratory Phonology: Journal of the Association for Laboratory Phonology.*

3.4.1 Introduction

Phonetically natural rules (i.e., those that follow automatic phonological processes such as coarticulation) are common in the world's languages. By contrast, rules of the same level of complexity, but which do not follow phonetic processes (i.e., unnatural rules) are rarer, and often specific to one or a few languages at most. One hypothesis to explain this asymmetry is that of a learning bias, whereby phonetically natural rules are easier to learn and therefore more likely to be transmitted over time than their unnatural counterparts (Schane et al., 1974; C. Wilson, 2006). This hypothesis is in line with substance-based theories of phonology (Archangeli & Pulleyblank, 1994; Donegan & Stampe, 1979; B. Hayes et al., 2004), which posit phonetic knowledge on the part of the speaker in the synchronic grammar. Yet evidence supporting this hypothesis has so far been weak. Consider the typological asymmetry between vowel harmony, a common long-distance vowel co-occurrence restriction hypothesized to be born out of vowel-to-vowel coarticulation (Ohala, 1994), and vowel disharmony, a virtually unattested rule with no such basis in production (Moreton & Pater, 2012a, for a more complete review, including other rule types, see). Two previous studies have failed to show evidence for a learning bias in favor of vowel harmony. In an artificial language learning study, American English listeners exposed to singular/plural alternations that were either harmonic or disharmonic learned both rules equally well (Pycha et al., 2003). Similarly, French listeners who were exposed to short stories in a novel accent of their native language which followed a systematic rule of harmony or disharmony also showed equal learning of both rules (Skoruppa & Peperkamp,

[3]The stimuli section has been reduced to avoid repetitions and the discussion includes mention of the unpublished experiments in Section 3.3.

2011). However, a more recent artificial language-learning study with French speakers did provide some evidence for better learning of harmony compared to disharmony (Martin, Guevara-Rukoz, Schatz, & Peperkamp, submitted). This study used a production task, and its exposure phase included exceptions to the rule. A significant effect was obtained in a first experiment but not replicated in a second one. Interestingly, in a simulation of transmission across generations, the overall marginal asymmetry between harmony and disharmony, obtained by collapsing data from the original experiment and the replication, was shown to compound over time. Thus, a weak bias favoring the learning of harmony resulted in a higher probability for harmony to be transmitted across generations.

In the present study, we reexamine the question of a learning bias, and additionally investigate a novel factor that might influence the learning of phonological rules, namely sleep. Sleep is known to enhance the learning process by way of memory consolidation (M. P. Walker & Stickgold, 2004, for a review, see). Newly acquired knowledge consolidates overnight, yielding improved performance the following day. In the domain of language, sleep-dependent memory consolidation has been shown in adults for perceptual adaptation to synthetic speech (Fenn, Nusbaum, & Margoliash, 2003), lexical learning (Davis, Di Betta, Macdonald, & Gaskell, 2009; Dumay & Gaskell, 2007), morphosyntactic rule learning (Batterink, Oudiette, Reber, & Paller, 2014), and phonotactic learning (Gaskell et al., 2014). Here, we examine whether sleep enhances the learning of phonological rules, and if so, whether it does so differentially for phonetically natural versus unnatural rules. In particular, if sleep consolidates the former more than the latter, this would add to the evidence of a learning bias in favor of natural rules.

We focus on the case of vowel harmony versus disharmony, and use an artificial language learning experiment administered in two sessions (test and retest) separated by twelve hours, either with or without an intervening period of sleep. Like Pycha et al. (2003) and Martin et al. (submitted), our test case is palatal harmony, a rule whereby vowels within the domain of the word must share the same value along the front/back dimension. This long distance dependency between vowels is well attested in the typology, but its converse is not. That is, we know of only one language that has been reported to have a productive case of palatal disharmony, i.e., Ainu (Krämer, 1999).

Given the logistical difficulties of testing participants in the morning and the evening in the lab, we opted for an online experimental setup, allowing participants to take part in the study from home. Specifically, we recruited American English speaking participants on Mechanical Turk and tested them on the French stimuli used by Martin et al. (submitted). Steele, Denby, Chan, and Goldrick (2015) used this platform to test implicit phonotactic learning, comparing native stimuli to non-native stimuli, and showed that participants were able to learn implicit rules in both experiments with native and non-native (French) stimuli. This setup has two main advantages: First, as Steele et al. pointed out, online testing allows for large sample sizes to be recruited quickly and requires relatively few resources on the part of the experimenter. Second, the use of non-native stimuli reduces the likelihood that participants rely on a metalinguistic strategy to perform the task. As the sounds of one's native language have fixed mappings to orthographic symbols, participants in artificial language learning experiments might encode the stimuli in terms of orthographic rather than phonological categories. Non-native stimuli encourage more phonetic listening, as naïve listeners are less likely to have fixed mappings between non-native stimuli and graphemes.

3.4.2 Methods

3.4.2.1 Stimuli

The stimuli used were the same as those reported in Section 3.2.

3.4.2.2 Procedure

The experiment was run online from a server in our lab. Participants were therefore not present in the lab and all interfacing with them was done via email. Upon logging into our website for the first time, they were asked to provide their email address so that they could be contacted when it was time to come back for the second session. During the first session, participants received instructions regarding the exposure phase. They were told that they would hear words from an invented language, and that words would be presented in their singular and plural forms. The language was said to have two forms of the plural suffix: "tel" and "tol".

Participants were first exposed to two repetitions of one of the sets A, B, or C of 32 unique stems, chosen at random. During exposure, a stem was played, followed after 500 ms by its plural form (either harmonic or disharmonic, depending on the condition). Immediately following the auditory presentation of the stimuli, two boxes appeared on screen in random locations, each labeled with one of the plural suffixes (TEL or TOL). Participants were requested to click on the box corresponding to the plural form they had heard. If participants provided an incorrect response, the trial was repeated to ensure that they had correctly heard the singular and plural forms of all exposure stimuli. This task was added during the exposure phase to ensure that participants were paying attention to the stimuli and had their speakers turned on. The random button position manipulation was chosen so that participants were required to actively seek out their response and could not repeatedly press the same button. We recorded the number of errors participants committed and used this information to exclude those who were either not paying attention or did not have proper audio equipment.

Immediately after this exposure phase, the first test phase began. Test trials consisted in the presentation of a stem followed by both possible plural forms (i.e., CVCV-tɛl and CVCV-tɔl) with an ISI of 500 ms. The order of presentation of the two plural forms was randomized across trials. After all three stimuli were played, two buttons appeared at random locations on the screen (exactly as during the exposure phase) labeled with the two plural suffixes. Participants were asked to select which of the two plural forms they thought was correct. Participants were first tested on the exposure set (i.e., the same items that they had just heard) before being tested on another set of 32 stimuli that they had not previously heard. They were not told anything about the two types of items they were tested on.

Exactly twelve hours after having completed the exposure, participants received an email inviting them to log back into the website. They had exactly one hour to do so. If participants attempted to log in before the twelve hours had passed, they were instructed to return later. If they attempted to log in after the one-hour grace period, they were blocked from continuing and were excluded from the study. When they logged back in, they immediately began the retest. They were first tested on the exposure set again, before being tested on the third set of stimuli (i.e., items they had heard neither during exposure, nor during the first test). Participants were therefore tested on "old" items (which they had heard during exposure) and different sets of "new" items at both test (session 1) and

retest (session 2).

At the end of the retest, participants were asked to fill out a questionnaire concerning some basic personal information, as well as their sleep habits and strategies during the task.

3.4.2.3 Participants

Participants were American English speakers recruited on Amazon's Mechanical Turk platform. They were recruited in a total of 14 "batches" of around 50 "workers". That is to say, the task was made available 14 times to 50 workers at a time. "Wake" batches were launched at 16:00 CET (10:00 EST) while "sleep" batches were launched at 04:00 CET (22:00 EST). This means that half of the participants were recruited in the morning, to return 12 hours later at the end of their day. The other half of participants were recruited in the evening, completing the first session before going to bed, and then participating in the second session the following morning.

Once recruited, participants were redirected from the Mechanical Turk page to our website. Participants were randomly assigned to the harmony or disharmony condition upon logging into the website for the first time. After completing the first session, they received compensation for the "HIT", and if they completed the second session, they were awarded a bonus payment. A total of 538 participants were recruited, but 87 did not complete the first session, 240 did not return for the second session, and a further 23 completed either too few or too many trials (for instance, by doing the first session twice). Of the 188 participants who correctly completed both sessions, 69 were excluded from data analysis for the following reasons: made at least ten errors out of the 64 trials during the exposure phase (N=33), used only one response (i.e., either tel or tol) throughout an entire test phase (N=1), did not fill out questionnaire (N=16), napped during the day (N=11) or did not respond to question about napping (N=5), or took notes during exposure (N=13) or did not respond to the question about note taking (N=16).[4] Note that some excluded participants fall into multiple categories (i.e., took notes during exposure and did not answer question about napping).

A total of 119 participants (57 women and 62 men aged 20 to 61, with a mean of 37) were included

[4]We decided to be conservative by excluding participants who may have taken notes or napped, erring on the side of caution when no information was provided. Indeed note taking undermines our ability to observe learning, and even daytime napping has been shown to consolidate motor-related memories (Nishida & Walker, 2007).

	Harmony	Disharmony
Wake	$N = 26$	$N = 27$
Sleep	$N = 34$	$N = 32$

Table 3.1: Distribution of participants into experimental conditions.

	Wake	Sleep	Difference
Age	36.2	37.5	$t < 1$
Gender (ratio M:F)	1.12	1.06	$\chi^2 < 1$
Personality (ratio morning:evening)	0.77	0.61	$\chi^2 < 1$
Concentration at test (1–5)	4.70	4.53	$t = 1.2, p = 0.1$
Concentration at retest (1–5)	4.68	4.38	$t = 2.0, p = 0.05$
Fatigue before test (1–5)	1.55	1.85	$t = 2.0, p < 0.05$
Fatigue before retest (1–5)	2.22	1.56	$t = 3.5, p < 0.001$
Prior sleep quantity (hours)	6.91	6.78	$t < 1$
Prior sleep quality (1–5)	3.60	3.58	$t < 1$
Ideal amount of sleep (hours)	7.70	7.56	$t < 1$

Table 3.2: Average participant responses to questionnaire and test statistics comparing the wake and sleep groups.

in data analysis, distributed into the harmony/disharmony and wake/sleep conditions as shown in Table 3.1.

We compared the participants in the wake and sleep groups in a number of ways based on their responses to the questionnaire. For most measures, the two groups did not differ. Means of both groups and test statistics are reported in Table 3.2. Both groups reported similar, high concentration ratings at initial test, but the wake group reported being more concentrated at retest than the sleep group did. There was an expected asymmetry between the levels of fatigue at test and retest for the wake and sleep groups. The sleep group (initial test at the end of the day) reported an average higher level of fatigue at initial test than the wake group (initial test in the morning). Likewise, the wake group reported an average higher level of fatigue at their end of the day test (i.e., at retest) than the sleep group. Unsurprisingly, participants were thus more fatigued during the session they took part in at the end of the day.

3.4.3 Results

We calculated the accuracy of responses during the test phase for all participants. Data were analyzed using logistic mixed-effects models in R (Bates et al., 2014). Of course, what was considered a "correct" response depended on the rule participants were exposed to, such that disharmonic responses were considered correct in the disharmonic condition, and vice versa for the harmonic condition. All factors in the models were defined using contrast coding, and significance was assessed through model comparison. Results for old items and new items were analyzed together so as to increase statistical power.

We began by analyzing the initial test session only, so as to compare our results with those reported in the literature. Indeed, data from the initial test session are qualitatively comparable to results from Pycha et al. (2003) and Section 3.3, in that they represent performance just after exposure. Mean accuracy from the initial test are displayed in Fig. 3.6. We designed a model with fixed factors for Rule (harmony vs disharmony) and Group (wake vs sleep), and their interaction, and a random factor for Participant. We did not include a random factor for Stem, since we did not expect any effects to vary according to the individual stems. This model was compared to a simpler model that excluded the factor Rule. Unlike Pycha et al. (2003), but in line with Martin et al. (submitted) and Section 3.3, performance in the first test phase was better for harmony than for disharmony ($\beta = 0.46$, $SE = 0.15$, $\chi^2(1) = 8.74$, $p < 0.01$), indicating that just after exposure, participants had learned the phonetically natural rule better than the unnatural one. However, no significant difference was observed when comparing the full model to one that excluded the factor Group ($\chi^2(1) < 1$) or the interaction between Rule and Group ($\beta = -0.33$, $SE = 0.31$, $\chi^2(1) = 1.17$, $p = 0.28$).

We then analyzed the full dataset, including both sessions. Mean accuracy across the different manipulations can be seen in Fig. 3.7. A full model was designed that included the following fixed factors: Rule (harmony or disharmony), Group (wake or sleep), Session (test or retest), and all of their interactions. We also included a random factor for Participant, along with a random slope for Session. This model was compared to other simpler models that excluded one of the factors or interactions.

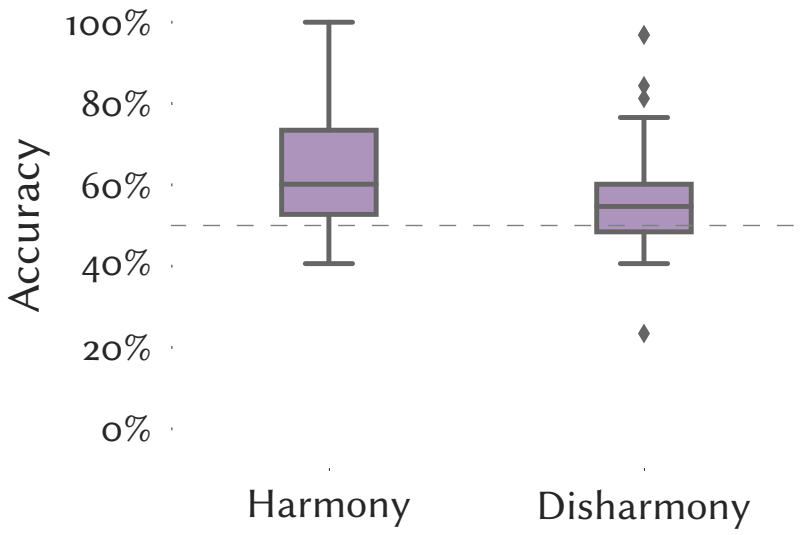

Figure 3.6: Boxplots showing mean accuracy scores at initial test as a function of rule.

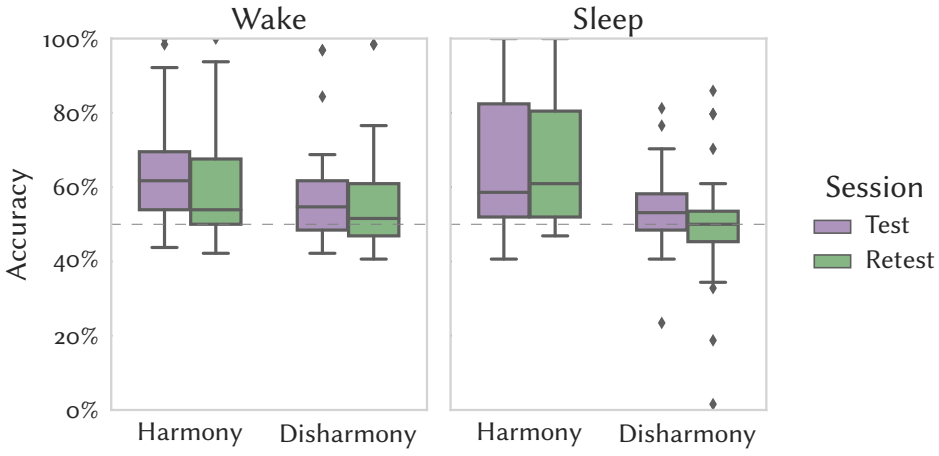

Figure 3.7: Boxplots showing mean accuracy at test and retest as a function of sleep and rule.

The full model was found to explain significantly more variance than a model which excluded Rule ($\beta = 0.57$, $SE = 0.18$, $\chi^2(1) = 9.36$, $p < 0.01$), but not than models which excluded Group ($\chi^2(1) < 1$) or Session ($\beta = -0.05$, $SE = 0.04$, $\chi^2(1) = 1.52$, $p = 0.22$). This indicates that harmony was generally learned better than disharmony, regardless of test session, and that no overall differences were found between the sleep and wake groups. A difference was found between the full model and a model excluding the interaction between Rule and Session ($\beta = 0.15$, $SE = 0.08$, $\chi^2(1) = 3.85$, $p < 0.05$), with performance decreasing more between test and retest for disharmony than for harmony. However, no difference was found between the full model and models excluding the interaction between Rule and Group ($\beta = -0.52$, $SE = 0.35$, $\chi^2(1) = 2.21$, $p = 0.14$), or the interaction between Session and Group ($\chi^2(1) < 1$). Finally, and crucially, the full model explained significantly more variance than a model which excluded the triple interaction amongst the three factors Rule, Group, and Session ($\beta = -0.32$, $SE = 0.15$, $\chi^2(1) = 4.49$, $p = 0.034$). This indicates that performance decreased between test and retest in all conditions, except for the participants in the sleep group who learnt the rule for harmony. That is, participants' performance increased after a night of sleep, but only when they had learned the phonetically natural rule.

3.4.4 Discussion

Our findings can be summarized in two points. First, a phonetically natural and typologically common rule of vowel harmony was learned better than a phonetically unnatural and exceedingly rare rule of vowel disharmony. Second, performance on the harmony but not on the disharmony rule improved after a night of sleep. We will discuss these findings in turn. Better learning of vowel harmony in the first test session, hence regardless of sleep, contrasts with some previous experiments comparing harmony to disharmony (Pycha et al., 2003; Skoruppa et al., 2011, i.e.,), but is in line with others (Martin et al., submitted, Section 3.3). Recall that in Martin et al. (submitted), participants were exposed to either a rule of harmony or a rule of disharmony, with or without exceptions, and had to produce the plural forms they thought best fit the stems they were presented with during the test phase. Taken together, the results of the two experiments reported in that paper showed a marginal effect in the conditions with exceptions, with harmony being learnt better than disharmony. A similar small effect was reported in Section 3.3 using a perception-only task. Here, using

exposure without exceptions and a perception task, we observed better performance for the harmony than for the disharmony rule during the initial test session. From a methodological point of view, these results are in line with those of Steele et al. (2015), showing that internet-based artificial language learning experiments can be run with non-native stimuli. More research is necessary to test whether such stimuli are actually more appropriate for exploring learning asymmetries, given the hypothesis that their processing is more phonetic and less likely to be influenced by orthographic knowledge.

As to the role of sleep, participants who learned the phonetically natural vowel harmony rule showed improved performance after a night of sleep, while those who learned the unnatural disharmony rule did not. Participants in the wake condition did not show improved performance at retest for either the natural or the unnatural rule. Recall that participants in both the wake and sleep groups reported being more fatigued during the test session that took place at the end of the day (retest for the wake group, initial test for the sleep group), and those in the sleep group further reported being less concentrated at retest than those in the wake group. However, as we did not observe an interaction between Group and Session (with increased performance during the test session that took place in the morning), these observations cannot explain our findings. Our results rather suggest that phonetically natural but not phonetically unnatural rules benefit from sleep-dependent consolidation, with one caveat: The wake and the sleep group differed not only with respect to the absence versus presence of sleep, but likely also with respect to the amount of language processing between test and retest. That is, participants in the sleep group spent on average seven hours without perceiving or producing speech (modulo any linguistic processing that occurs while dreaming), while those in the wake group did not necessarily have such a long language-free period. Thus, alternatively, it might be the absence of language processing (and not the presence of sleep per se) that gives a boost to the learning of phonetically natural compared to unnatural rules. The only clear way to address the causal role of sleep on consolidation is to record EEG (electroencephalography) during sleep and carry out individual correlation analyses. Previous work has shown this method to be an effective way of demonstrating the role of sleep (Batterink et al., 2014; Gaskell et al., 2014; Nishida & Walker, 2007; Tamminen, Payne, Stickgold, Wamsley, & Gaskell, 2010; M. P. Walker & Stickgold, 2004). Thus, if consolidation is sleep-dependent, we expect to see a positive correlation between the

amount of REM sleep and/or Slow-wave Sleep on the one hand and the improvement in performance on phonetically natural rules at retest on the other hand. Many studies have verified the role of sleep on consolidation by including a third test session, twenty-four hours (or more) after exposure and initial test (Dumay & Gaskell, 2007; Earle & Myers, 2014; Ellenbogen, Hu, Payne, Titone, & Walker, 2007; Fenn, Margoliash, & Nusbaum, 2013, 2003; Tamminen et al., 2010; M. P. Walker & Stickgold, 2004).

However, this is logistically very difficult with online testing. Indeed, nearly half of the participants who took part in the exposure and initial test phase did not return for the retest phase. Further work in the laboratory could combine EEG measures with a third testing session in order to provide direct evidence of sleep-related consolidation. We predict that the participants in the wake group would show improved performance after twenty-four, but not twelve hours, if they learned the harmony rule.

Under the assumption that there is indeed sleep-dependent consolidation of phonetically natural rules, we might speculate on the mechanism underlying this phenomenon. A hypothesis we would like to raise is related to the fact that phonetically natural rules typically reflect patterns of coarticulation (for the case of vowel harmony, see Ohala, 1994) and hence are rooted in the articulatory gestures associated with speech production. Specifically, activity in speech motor areas of the brain during sleep would benefit natural over unnatural rules. This could be the case despite the fact that we used a perception and not a production task. Indeed, speech perception activates areas in the motor cortex that are involved in speech articulation (Pulvermüller et al., 2006; S. M. Wilson, Saygin, Sereno, & Iacoboni, 2004). This activation, moreover, is increased for the perception of non-native compared to native speech sounds (C. Wilson, 2006); our use of non-native stimuli might thus have had a positive influence on our ability to find an effect of sleep. Some evidence in favor of the hypothesis linking activity in speech motor areas to consolidation of phonetically natural rules is the finding that the consolidation of motor skill learning is related to increased activation in the motor cortex during sleep (M. P. Walker, Stickgold, Alsop, Gaab, & Schlaug, 2005). More research is necessary, however, to directly test this hypothesis, using brain imaging techniques. In particular, we would predict to observe an increase of activation in the speech motor cortex during the retest in the sleep compared to the wake group, but only for those participants who were exposed to the

harmony rule.

To conclude, this study provides evidence that phonetically natural rules benefit from a learning bias, and suggests that sleep-dependent consolidation enhances this bias. While there is mounting evidence of sleep-dependent consolidation of learning in several linguistic domains (Batterink et al., 2014; Davis et al., 2009; Dumay & Gaskell, 2007; Fenn et al., 2003; Gaskell et al., 2014), our behavioral-only data do not allow us to draw firm conclusions concerning the role of sleep in the consolidation of phonological rule learning. The mechanism underlying a possible sleep-dependent bias favoring phonetically natural over unnatural rules similarly awaits further research. Overall, the present study paves the way for a novel research topic, concerning the role of sleep-dependent consolidation in shaping sound patterns in human language.

3.5 Discussion and conclusion

This chapter has presented three studies aimed at exploring a bias favoring the learning of a natural, typologically common phonological rule (vowel harmony) compared to an unnatural, typologically uncommon rule (vowel *dis*harmony). Section 3.2 considered the role that modality might play in learning bias for harmonic patterns. It has been suggested that vowel harmony has its roots in a production constraint (vowel-to-vowel coarticulation) (Ohala, 1994), and requiring participants to utilize their production system during the learning process might augment any observable bias favoring harmony to disharmony (which does not have clear phonetic motivation). We additionally manipulated consistency by presenting some groups of participants with exceptions in the input they were exposed to. What we observed was that only participants in the inconsistent disharmony condition (i.e., those who were exposed to the disharmony rule with exceptions) performed poorly, and indeed significantly less well than those in the inconsistent harmony condition. A replication of this experiment did not reveal a statistically significant bias, but a trend in the same direction.

Our difficulty in replicating the effects we observe points to just how subtle the learning bias may be. We argue, however, that even a subtle bias may have a lasting impact and be able to affect typology by compounding over time. This was explored in Section 3.2.3 by simulating the transmission of a

phonological rule over generations. We found that even a very small bias disfavoring disharmony could compound and yield significantly fewer languages with a productive disharmony rule than those with a productive harmony rule.

Section 3.3 reported on a perception-only artificial language learning experiment akin to that of Pycha et al. (2003). We again tested the manipulation of consistency, by presenting some groups of learners with exceptions to the rule they were to learn. Learners in an original experiment performed on average relatively poorly, with those in the inconsistent conditions close to, though significantly above, chance level performance (56%). Nonetheless, of the participants in the consistent conditions, those who were exposed to the harmony rule performed significantly better than those exposed to the disharmony rule. In a replication, we did not find as stark a difference between those two groups, with both groups performing less well than the participants in the original experiment, similar to what we found in Section 3.2.

Section 3.4 explored the same question from a different angle. Learning is a complex process that does not take place at one isolated point in time. Indeed, memories are known to consolidate during sleep, and newly-learnt linguistic rules have been shown to consolidate during naps or nighttime sleep (REFs). We therefore tested the learning of harmony versus disharmony and additionally manipulated whether or not participants slept in between two test periods separated by approximately twelve hours. For logistical reasons, we used Amazon's online participant recruiting system Mechanical Turk. This means that our participants were native speakers of English. Recall that the stimuli presented in Section 3.1 were recorded by a native French speaker, so many of the stimuli contained sounds unfamiliar to the participants we tested (e.g., /ɛ̃/, /ʁ/, /ɔ̃/). That, combined with the increased variability inherent in online testing and the differences in demographics between the Mechanical Turk population and the typical student population psycholinguistic experiments usually use yielded general lower performance than that observed for the French participants in Section 3.3. Like in the original experiment reported in Section 3.3, however, performance was on average higher for the group of participants exposed to the harmony rule than those exposed to the disharmony rule. Furthermore, we observed a small but significant amount of consolidation after twelve hours for participants who slept in between the test sessions, but not for those who did not, and crucially, only for those who had been exposed to the harmony rule. This suggests that not only

might there be a general learning bias favoring the harmony rule, but that sleep might differentially affect the consolidation of the harmony compared to the disharmony rule.

In all of the experiments reported in this chapter, we demonstrate a bias for the learning of vowel harmony patterns compared to vowel disharmony patterns. Our claim is that this represents a general preference for harmonic patterns (in line with typology) over disharmonic ones. But it is important to consider that the learners we tested (French speakers in Sections 3.2 and 3.3 and English speakers in Section 3.4) might be biased by patterns in their own language's lexicon. Indeed in Chapter 2, we showed how lexical patterns coalesce with bottom-up acoustic perception to influence word recognition. It seems reasonable therefore, to consider that language-specific lexical patterns might also influence the learning of vowel harmony. For example, although neither French nor English has a productive morphological system involving vowel harmony, it is possible that the languages happen to contain many harmonic words, and few disharmonic words. If this is the case, one would expect speakers of these languages to better learn the harmony pattern than the disharmony one. Below we detail an analysis we performed on the lexicons of French and English that we compared to the lexicon of Hungarian, a language with a productive vowel harmony system.

We began with French. We extracted all polysyllabic lemma forms from the Lexique corpus (New et al., 2001), collapsing the mid-vowels and /ɛ̃/~/œ̃/ contrasts.[5] Monosyllabic words are not informative with regards to vowel harmony, and all such words were therefore not included in the present analysis. For each word, a harmony score was calculated, where each vowel within the word was assigned a score, 0 or 1, depending on whether it was a back or front vowel, respectively. Schwas were not considered in the analysis, as it is unclear whether they are front or back (although they are realized phonetically as front in French). A word's harmony score was considered as the average palatality of its vowels, such that a word with two front vowels (1 + 1), would have a harmony score of 1, while a word with one front and one back vowel (1 + 0) would have a harmony score of 0.5. This gives a more or less continuous distribution of harmony scores bounded between 0 (fully harmonic back) and 1 (fully harmonic front), with a maximally disharmonic word (as many front as back vowels) having a score of 0.5. The distribution for words in French can be seen in the top left

[5]The mid vowel contrasts /e/~/ɛ/, and /o/~/ɔ/ are currently merging in Parisian French, and are furthermore inconsistently coded in the Lexique corpus. Especially given that this is not informative for the current research question, which focuses on palatal harmony, we merged these contrasts in our analysis.

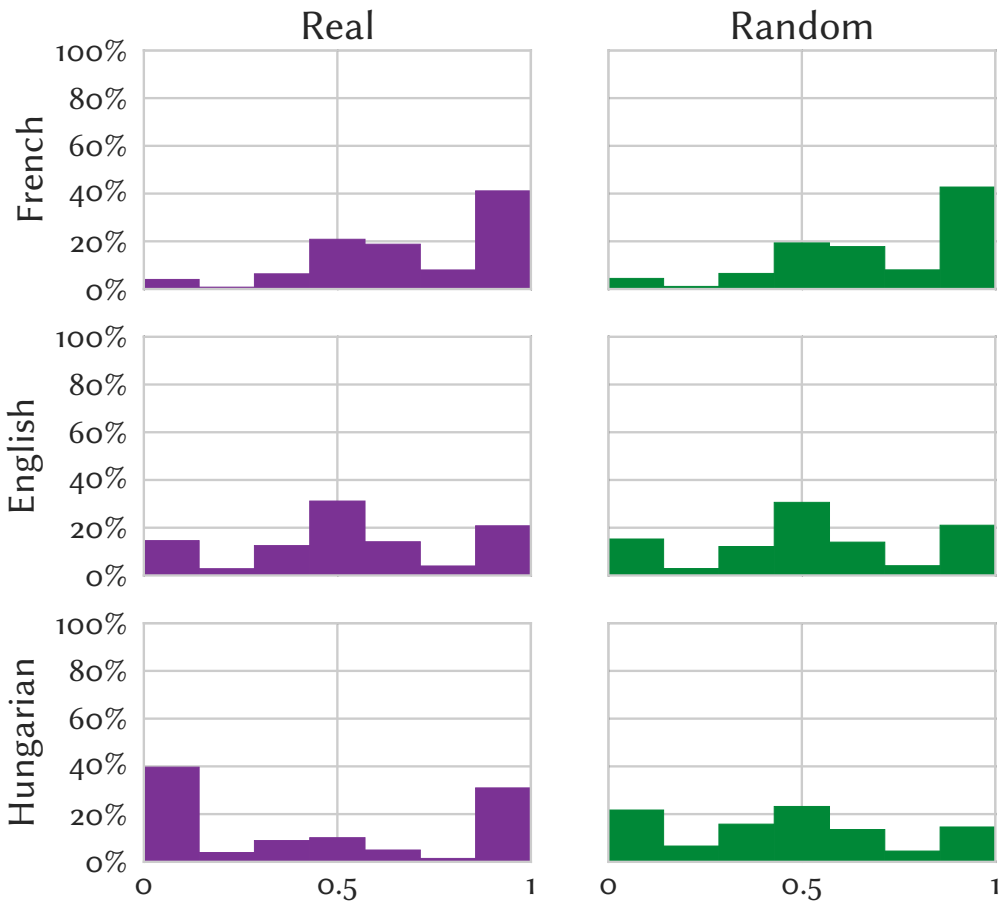

Figure 3.8: Distributions of harmony scores for real and random lexicons. Each bar represents the percentage of the lexicon that falls into that bin. A score of 0 means the words contain only back vowels, 1 only front vowels, and 0.5 as many front as back.

panel of Fig. 3.8.

There is a clear asymmetry, with a great deal of harmonic words containing front vowels (around 40% of the lexicon), and hardly any containing back vowels. This asymmetry is easily explained by the fact that French has many more front than back vowels: 7 (/i/, /y/, /e/, /ø/, /œ/, /ɛ̃/, /a/) vs 4 (/u/, /o/, /ɑ̃/, /õ/), respectively, and a word is thus simply randomly more likely to contain multiple front vowels. To demonstrate this, we generated a random version of French by extracting all vowels from all words, shuffling them, and reinserting them into the consonant frames. A word like éducation - /edykasjõ/ could therefore become /idokusja/ or /ydøkɛ̃sjɛ̃/, etc. We then recalculated the distribution of harmony scores for this random French lexicon and compared it to the real French lexicon. As can be seen in the top right panel of Fig. 3.8, the random distribution is almost completely identical to the real distribution; in fact only 2.55% of the real distribution is not overlapping

with the random one. This is a demonstration that while the French language shows a strong bias favoring words that contain mostly or only front vowels, this is due simply to the frequency of the very many front vowels compared to the back vowels. That is, a given word in French is likely to be harmonic, but by chance, rather than by design. Furthermore, since there is a large preference for front harmonic words, learners in the experiments reported in this chapter would need to learn the harmony pattern only for front vowels, and not for back vowels, in order for a language-specific bias to explain the learning we tested.

We performed a similar analysis on the lexicon of English ("CMU Pronouncing Dictionary," 2008). The results can be seen in the middle panels of Fig. 3.8. English, like French, demonstrates remarkable similarity between the real lexicon, and a random lexicon established by shuffling the vowels. Unlike French though, it does not show a preference for harmonic words in one direction or the other, but rather shows a large clump of mixed words, and additionally some front harmonic words, and some back harmonic words.

To ensure that our measure properly captures explicit harmony, we also performed the same analysis on the lexicon of a language with productive vowel harmony: Hungarian. The real Hungarian lexicon should contain more harmonic words than a random lexicon made by shuffling the vowels, because Hungarian words typically contain only front or back vowels. We used the Hungarian pronunciation dictionary (Grimes, 2006) to extract the phonological transcription of words. We followed the traditional divide of Hungarian vowels, with /y/, /y:/, /ɛ/, /e:/, /ø/, and /ø:/ considered front, /u/, /u:/, /o/, /o:/, /ɒ/, and /a:/ considered back, and /i/ and /i:/ considered to be neutral. As can be seen in the bottom left panel of Fig. 3.8, the Hungarian lexicon is symmetrically harmonic. That is, it contains a great number of harmonic words, both front and back. Note that unlike French, which has a great many more front vowels than back vowels, the categories in Hungarian are perfectly balanced, with six in each. We therefore verified that the harmonic organization of the lexicon is not due to this balance, but rather represents an active process in the language; we did this in the same way as described above for French.

As can be seen in the bottom right panel of Fig. 3.8, a random Hungarian lexicon shows a great deal more disharmonic words (the distribution around 0.5), and far fewer harmonic words. Indeed, 34.35%

of the real distribution is not overlapping with the random distribution (a great deal more than for the French data). Our method therefore seems able to distinguish between harmony by chance (due to the statistical distribution of vowels) and systematic harmony (due to an active process in the language).

A drawback to our method is that it ignores the order of vowels in a word. Compare the words *dégoûtant* - /degutɑ̃/ (disgusting) and *gouvernement* - /guvɛʁnəmɑ̃/ (government). Both contain two back vowels and a front vowel, and therefore have the same harmony score (namely, 0.33), but it may be reasonable to say that *dégoûtant* is more harmonic, as the two back vowels occur one after the other. Note though, that this only affects words that contain vowels from more than one category. Given the large amount of fully harmonic words present in both the French (over 40%) and Hungarian (over 70%) lexicons, it is unlikely that a measure that takes order into account would find results that contrast starkly with those reported here.

Of course a method is only as good as the corpora it uses. One issue with the Hungarian pronunciation dictionary compared to the Lexique database or the CMU pronunciation dictionary, is that we do not have morphological information, which means that we also analyzed a certain number of compound words, e.g., *olvasóközönség* - /olvɒʃoːkøzønʃeːg/ ("readership", composed of the words for "reader", *olvasó* and "audience", *közönség*, which are both themselves composed of two morphemes). Such words are not really informative for the question of harmony since they are composed of units which may themselves be harmonic (in this case both *olvasó* and *közönség* are harmonic individually, but one contains back vowels and the other front, such that the final compound is disharmonic). That means that our analysis probably underestimates the extent to which the Hungarian lexicon is harmonic. Nonetheless, we were able to observe a great difference between the real lexicon, and one randomly generated by shuffling vowels. Our method is therefore probably reliably able to measure the harmony in a lexicon, even when the corpora are not ideal.

All in all, it seems that the experimental results reported in this chapter cannot simply be explained by a language-specific bias, whereby learners show a preference for whatever pattern is present in their native lexicon, though more research is clearly necessary to assess how much L1 knowledge plays a role in artificial language learning experiments.

Chapter 4

General discussion

This thesis has presented two main angles of research. On the one hand, we explored how synchronic biases can affect speech processing, specifically during word recognition. On the other, we considered how synchronic bias in learning can affect typology by boosting certain phonological patterns over others. Here we will revisit the results reported in the previous chapters and discuss some of the remaining questions and angles of research that we have yet to explore. We will also discuss the implications of tying together these two lines of work and how they can inform our understanding of language change.

4.1 Bias in phonological processing

4.1.1 Summary of empirical work

The main question evoked in Chapter 2 was that of the relative role of phonological features during the process of word recognition. We began by testing listeners' ability to recognize mispronounced words when those mispronunciations concerned different phonological features (see Section 2.2), specifically voicing, manner, and place for the French obstruents. We showed that voicing mispronunciations allowed participants to more often recover the intended word than mispronunciations in manner or place. In Section 2.3, we then turned to the task of pinpointing the sources of this

asymmetry. We specifically considered two sources: bottom-up acoustic perception, and top-down lexical knowledge. In Section 2.3.2, we detailed an experiment using the ABX task that allowed us to assess the perceptual similarity of speech sounds without distorting their acoustics. We found that on average, manner contrasts were more reliably distinguished than voicing or place contrasts. This matches the fact that the manner contrasts in our stimuli were acoustically more distinct than the voicing or place contrasts. While word recognition will obviously be biased by the human auditory system, it is also likely to be influenced by listeners' experience. We explored this in Section 2.3.3 by testing the relative functional load of phonological features in the French lexicon. We found that French nouns were, all else being equal, more likely to be distinguished by the place feature than by voicing or manner. We then proposed that the initial finding reported in Section 2.2, that place and manner contrasts had a more important role in distinguishing words than the voicing feature, could be interpreted as the coalescence of the two sources of bias we tested. The human ear makes manner contrasts easier to perceive, while the lexicon of French renders place contrasts important in a lexical context. This means that in French, compared to voicing, manner and place are both important, but for different reasons.

4.1.2 Further questions

In Chapter 2, our conclusion was that during word recognition, listeners are biased by both top-down lexical knowledge, which, once the native language acquired, should be invariable, and bottom-up perceptual restrictions, for example that the human auditory system is more sensitive to certain contrasts and less to others. We proposed that it was necessary to provide a baseline of perceptual similarity of speech sounds in order to address our second point. We suggested that this could only be done in laboratory silence, without altering the sounds through the introduction of masking noise or manipulation of the acoustic signal, because noise affects features differentially. However, in our everyday lives, we are confronted with noisy conditions in which we are nonetheless able to recognize words and perceive speech. In fact, Wang and Bilger (1973) found that the relative importance of certain featural contrasts (e.g., voicing and nasality) was dependent on the listening conditions. That is, certain features were more confusable in quiet than in noise, and others vice versa. This highlights the importance of the question being asked. In a given situation, the specific

type of environmental noise may render certain features more or less important. While their lexical importance should not change[1], their relative acoustic salience clearly may. Therefore, in a given situation, French listeners might be even more biased to hear voicing contrasts as less important than place or manner contrasts, or on the contrary, hear voicing contrasts as more robust. For example, when hearing conversation from the room next door (similar to a low-pass filter), voicing contrasts may be relatively stable compared to manner contrasts which depend on high-frequency noise that would be filtered out by the wall. Word recognition is clearly a complex and *dynamic* process that requires a holistic research program if it is to be properly understood. Future studies could combine our ABX methodology with the classical introduction of different kinds of noise to measure to what extent certain features are distinguishable at different levels of masking noise. By comparing those results with ones obtained using a mispronunciation detection task with noise, we could have a better overall understanding of the relative influence of acoustics on word recognition. Indeed this combinatorial approach seems most necessary in order to understand how acoustic perception outside of lexical context interacts with lexical knowledge in the more realistic word recognition situation. Naturally, testing word recognition in controlled sentential context would be the fullest possible extension of this line of research, but we are still far from fully understanding the processes that underlie speech processing as a whole.

With regards to functional load (Section 2.3.3), we considered only a subset of the French consonant inventory and then performed our calculations on nouns only, considering that the effect we were trying to understand was present specifically for French nouns. We discussed how testing French verbs would be less straightforward than one might hope for (namely that what counts as the phonological form of a French verb is debatable). Future work could explore a language with clearer morphology both in the nominal and verbal domain, in order to see if our claim that lexical statistics (like functional load) are calculated and tracked within word classes holds up. We predict that if an opposite pattern were to be found between nouns and verbs, listeners should show bias in accordance with the word class they are being tested on, provided that word class is predictable. Given that word class tends to be predictable in real world situations (e.g., nouns and verbs tend to be preceded by different function words) and that even infants show sensitivity to word class in word

[1]Of course, in Section 2.3.4, we discussed how patterns may depend on lexical class. We tested only nouns and do not claim that the same pattern holds for verbs.

learning (Dautriche, Swingley, & Christophe, 2015), it seems reasonable that this hypothesis should hold true. Again, measuring statistics in the lexicon in addition to a combinatorial experimental approach clearly seems to add to our overall understanding of the phenomena we are exploring. We have presented methodologies in this thesis that can provide us with important data individually, but whose true impact comes from the more global view they provide together. This should be extended in future work to include other forms of bias that affect word recognition. Word-level effects such as frequency and length can easily be incorporated in future studies, for instance.

Of course the most interesting future work in this domain involves cross-linguistic comparisons. French presents a certain pattern of asymmetry in word recognition, specifically that place and manner contrasts are more important than voicing contrasts. Should another language be found that presents a different pattern, it would be the ideal testing ground for our hypotheses. We specifically predict that manner contrasts should be important in any language that phonetically contrasts stops from fricatives in a similar way to French, since the acoustic difference between the silence associated with a stop and the high frequency noise associated with a fricative is stark, but that place and voicing contrasts may be more or less important for the specific language in question based on its lexicon. Languages such as Icelandic or Faroese might be interesting places to start such a quest. Both languages distinguish voiced from voiceless sounds not only in the obstruents but also in the sonorants. This typological rarity, though not sufficient to say that voicing is more important in those languages than, say, French, indicates that there may be something special about these languages worth exploring. Indeed it is possible that although /r/ and /r̥/ are separate phonemes for example, they distinguish few words in the lexicon and therefore do not boost the voicing contrast as much as might be expected. This is why it is crucial to re-evaluate the meaning of contrastiveness. Having two sounds form a minimal pair does not necessarily mean that that contrast is as distinctive as any other. Overall, it is clear that contrast in phonological systems should not be viewed simply as 0 or 1, as gradient effects can be observed from statistical patterns in the lexicon to acoustic perception all the way to the level of word recognition. That is, some features, for one reason or another, simply are more contrastive than others.

4.2 Bias in phonological learning

4.2.1 Summary of empirical work

In Chapter 3, we turned to the question of learning bias and how it can affect linguistic typology. Section 3.2 considered the question of modality and its role in uncovering learning bias that has been hard to detect in previous work. We found in a first experiment that participants learned the typologically common rule of vowel harmony better than the practically unattested rule of vowel *dis*harmony, when we asked them to perform a production-based task, using the artificial language learning method. In an attempt to replicate this effect, we found only a numerical trend (non-significant) in the same direction, highlighting the power issues that are common in this type of experiment.

We then proposed a computational model of the transmission of (dis)harmony patterns over time using simulated iterated learning. We found that after only a few generations, even the small bias present in our experimental results was enough to lead to a sharp decrease in the amount of disharmonic chains that we simulated. We proposed that this small learning bias might indeed compound over time and explain part of the typological asymmetry. We then went on to see if using the same materials, but in a perception-only task, we could observe the same type of pattern. Indeed, we found that learners in our two-alternative forced-choice task also showed better learning of harmony over disharmony. Again, in an attempt to replicate, we failed to observe more than a numerical trend. While the learning bias may be small and difficult to detect, the fact that we can observe it in perception as well as production is important when considering the real world, as children who are learning vowel harmony systems acquire information regarding the rule long before they themselves are producing utterances.

Finally, we explored the role that sleep might play in learning bias. In addition to an immediate better learning of the natural pattern over the unnatural pattern, we hypothesized that the natural pattern might benefit from overnight memory consolidation more than the unnatural pattern and that the two effects combined might yield a greater asymmetry between harmony and disharmony.

We tested groups of participants in a perception-only task on the two rules and retested them twelve hours later, either with an intervening period of sleep or not. The group that learned the harmony rule and was able to sleep before retest showed better performance than any of the other groups.

Overall, there is clearly mounting evidence for synchronic learning bias favoring phonetically natural patterns over unnatural ones. Our simulation goes further by demonstrating how even a small bias can compound over time.

4.2.2 Further questions

Artificial language learning experiments simulate the transmission process using adult learners, but of course adult learners have resources available to them that child learners do not. Children acquire a large part of their language skills before being able to produce language themselves (thus only through perception), so any claimed learning bias for harmony over disharmony must be present on some level in perception for the claim to hold water. Our results are encouraging and suggest that even though vowel harmony has been suggested to be born out of a production constraint (Ohala, 1994), some perceptual advantage must also be present, and thus accessible to child learners in the real world.

An additional manipulation that would be of great interest to explore in order to better test the role of modality is that of articulatory suppression. Indeed, there is plenty of evidence that the production system is involved in speech perception (e.g., Pulvermüller et al., 2006; S. M. Wilson et al., 2004), so it is impossible to rule out the possibility that the bias we observed in our perception-only task was related to the rooting of vowel harmony in a production constraint. Articulatory suppression requires participants to perform a perception-based task while simultaneously repeating a nonce syllable (e.g., /ma/, /ma/, /ma/). The idea behind this manipulation is that if the articulatory system is being recruited to produce the nonce syllable, it cannot play a role in the perception task (for a discussion of this technique, see Baddeley & Lewis, 1981). It is possible to combine this technique with artificial language learning by having participants repeat a nonce syllable during the exposure phase when they are in the process of learning the (dis)harmony pattern. Should participants still show better learning of harmony than disharmony, we could more conclusively state that there must be a

perceptual benefit for the rule that influences the learning bias. Of course implementing such a task might prove rather difficult, as perception is also affected by articulatory suppression; participants hear their own productions at the same time that the stimulus is being presented auditorily. Recall that performance in the perception-only experiments was lower than in the production experiments, so making the learning situation even more difficult might require retuning of the exposure phase so that participants have enough input to be sure of the rule they have learnt.

A final way of exploring the learning bias hypothesis in a more ecological way would be to test preverbal infants. Saffran and Thiessen (2003) used the head-turn preference procedure to test infants' preferences of typological universals. It would be possible to use such a procedure to test infants' preference of harmony over disharmony to see if the former are learned better than the latter. If a bias in infants who are not yet producing speech sounds could be demonstrated, it would give credence to the learning bias hypothesis in a real world setting. Indeed, in order for learning bias to play a role in linguistic typology, it is not enough for it to facilitate the learning of phonological patterns by adults, given that most linguistic learning is done by children.

Section 3.4 included another manipulation that warrants discussion. Because of logistical constraints, we tested participants in our sleep-based memory consolidation study using the online platform Mechanical Turk. This meant that variability in our results was higher, but crucially, that we tested native English-speaking participants with stimuli produced by a native French speaker. The fact that the participants were therefore doing true non-native speech perception (note that most artificial language learning experiments use stimuli produced by a speaker of the same language as the participants) while attempting to learn the (dis)harmony pattern. This is important with regards to the role of modality, and should be further explored in future research. Indeed, it is known that non-native speech perception recruits the motor system more than native speech perception (S. M. Wilson & Iacoboni, 2006), so the effect we observed so robustly for English speakers may simply be a reflection of the more engaged role of the motor system due to non-native listening (again, perhaps tapping into the production basis of vowel harmony). Furthermore, the native French listeners in our other experiments could have used strategies less obviously available to the English speakers, which might also explain why the English speakers performed overall less well. For example, the French participants could recruit strategies involving orthography, since they were

invariably familiar with the mapping of sound to grapheme (not the case for the English speakers). Indeed, in some of the debriefing questionnaires, French participants referred to letters or patterns of letters when describing their strategies. It would be ideal to test French listeners with non-native stimuli and English listeners with native stimuli using similar paradigms, to see if the asymmetries we observed can be attributed to population-level differences, or, rather, to more revealing factors, such as non-native listening.

4.2.3 Combining sources of bias

Channel bias and learning bias (see Section 1.2.1 and Section 1.2.2) have often been the topic of research individually, but rarely have they been thought of as a combinatorial process. Moreton (2008) lays out this issue in an elegant way. On the one hand, what he refers to as channel bias (phonetically-based pressures from speaker to listener) leads certain patterns to emerge through transmission. On the other hand, what he refers to as "analytic bias" (what we have referred to as learning bias) is a pressure present in the individual learner. He refers to this as "selective learning", and it is this term, learning, that we have focused on.

Both channel bias and learning bias have been proposed as individual explanations for the shaping of typology. Moreton describes how learning bias has never been shown to be the sole possible source of explanation. This is not the case for channel bias, which is often claimed to be entirely sufficient to explain typology. He focuses his study on two phonetically motivated phonological patterns. On the one hand, height-height vowel harmony is a rule that dictates that co-occurring vowels should share the same value of the height feature. On the other, height-voicing harmony says that the height of a vowel should be determined by the voicing value (voice, aspiration, etc.) of the following consonant. Though both patterns are "phonetically natural" in that they facilitate articulation,[2] the former is overwhelmingly more common than the latter. In an artificial language learning experiment, Moreton showed that participants learned the height-height pattern (typologically attested) better than the height-voice pattern (typologically unattested), indicating a clear presence of learning bias. Given that channel bias is insufficient to explain this specific typological

[2]Though see Yu (2011) for a response to this claim.

asymmetry, Moreton suggests that learning bias too may shape the evolution of linguistic systems.

Thus while certain patterns may find their prevalence driven by phonetic pressures in the transmission from one generation to the next, others may indeed be best explained through a bias on the part of the learner that would favor certain patterns over others. In this case, it may be reduced to featural complexity. The height-height rule depends on only one feature (and even only one segment type), namely vowel height. The height-voicing rule, however, requires two features to explain. Thus, the simpler rule may just be easier to learn.

Returning to vowel harmony versus disharmony, we suggest that harmony might benefit from a learning bias, even though it also has clear phonetic grounding (that is, may be influenced by channel bias). Thus, it would be possible to explain the prevalence of vowel harmony compared to disharmony through phonetic pressures alone (see Mailhot, 2013). We showed, though, that there is likely *also* a learning bias that might influence the typology. How much each of these sources of influence matters is yet to be determined, and it will be interesting for future work to assess to what extent certain patterns are more likely to benefit from channel or learning bias.

4.3 Tying it all together

This thesis has largely been separated into two. On the one hand, we have considered biases in phonological processing that influence speech perception, specifically as it concerns the recognition of words. On the other hand, we have tested the hypothesis that certain, "natural", phonological patterns benefit from a learning bias and that this might explain, in part, their prevalence in the linguistic typology compared to their logically equivalent but "unnatural" counterparts. Both of these questions are in fact a small part of a bigger picture regarding the factors that coalesce to influence both humans and therefore linguistic systems.

Certain typological universals are easy to explain simply by the fact that languages are used by human beings, and are thus restricted by what humans are capable of. The fact that CV syllables are the most prevalent in the typology is a natural consequence of the human auditory and articulatory systems. For example, acoustic events associated with consonants are more readily identified in that

prevocalic position than, say between two other consonants. We considered how word recognition will be biased by the human auditory system in Section 2.3, but we discussed very little about the consequences this might have on the typology. If manner contrasts are more reliably identified than place or voicing contrasts due to their acoustic salience, we should expect to find that manner contrasts are more present in the typology than place or voicing contrasts.

If linguistic systems are indeed shaped by the fact that they are used by humans, then we should expect to see other usage factors influence certain languages but not others. A recent example of this is the correlation between altitude and the presence of ejectives in phonological systems (Everett, 2013). The author claims that environmental context could play a role in shaping phonological systems over time. In the case of ejectives, that higher altitude is associated with lower ambiant air pressure, thus facilitating the articulation of these otherwise articulatorily complex consonants.

Another line of research that has considered this question has focused on Australian Aboriginal languages. Butcher (2006) looked at patterns in Australian languages that showed a preference for VC syllables over CV syllables, a typological rarity. Australian languages are unusual in the great number of place of articulation contrasts they have compared to the number of manner contrasts they have (Butcher, 2012). Butcher suggested that this may have to do with a particularity present in these populations. The World Health Organization reported that Australian Aboriginals have the highest proportion of chronic otitis media (a middle ear infection) in the world. This infection can damage hearing both at the very low end of the spectrum (where voicing contrasts are distinguished), and at the very high end of the spectrum (where fricative noise is most prevalent). The ensuing hypothesis is that over time, the languages spoken by these populations developed to take advantage of the auditory systems of the learners. That means avoiding fricative noise and low-frequency voicing contrasts in favor of a great number of place contrasts. The idea is interesting and can of course be extended to all human populations, the crucial difference being that most humans have nearly identical auditory systems, so languages may independently be shaped by human usage, and thus wind up resembling one another, in a case of "parallel evolution".

These hypotheses remain to be confirmed and thoroughly tested, but they show the possible relationship between usage and the evolution of the linguistic system. While some approaches to

diachrony specifically rebut teleology (e.g., Blevins, 2004), it might be more reasonable to assume that certain aspects, but not all, of languages might be goal-directed. One prime argument against this is that any facilitation in one dimension (e.g., ease of articulation) is likely to make a pattern or phenomenon less ideal in another dimension (e.g., difficulty in perception). Each language finds its balance amongst the various factors influencing it through the transmission process. If we can imagine general, universal, constraints that are shared by all members of a community (even all humans), however, surely these factors could be considered simultaneously goal-oriented, and ideal. Consider again auditory perception. If manner contrasts are universally easier to perceive than voicing contrasts, we can expect that even if they are articulatorily more complicated (though they need not be), they will be preferred, because their benefit in perception outweighs their tax in production. This would then be a case of auditory perception shaping languages over time, as manner contrasts become preferred to voicing ones.

If on the one hand auditory perception can shape the way languages evolve, and on the other hand learnability has been shown to be able to play a role, the idea that learning bias should find its root in phonetic knowledge on the part of the learner seems more reasonable. Listeners know what they can hear (either through their experience with others, or with themselves), and this might bias them in the learning process to assume certain patterns over others. Thus, the synchronic bias we demonstrated in Section 2.3 for certain contrasts to be perceived better than others might combine with learning bias to explain why certain contrasts are more prevalent and others rarer. As mentioned in Section 2.3, Johnson and Babel (2010) showed that the contrast [f]~[θ] was difficult for both Dutch and English speakers in their AX task compared to [f]~[s]; this could explain why the former is typologically rarer than the latter. If the difference between [f] and [θ] is difficult to perceive, and if there is not enough evidence for the existence of the sound [θ] in the input, learners may fail to acquire the contrast, and this could lead to change over time. Indeed this process is currently happening in certain dialects of English, where [θ] is fronted to [f]. This is a case of perceptual bias combining with a bias in the usage of the sound to influence change. Indeed, the transmission process, like speech perception itself, must be considered from angles other than simple phonetic pressures.

Chapter 5

Conclusion

This thesis has focused on two main questions: 1) What are some synchronic sources of bias in the processing and learning of phonological patterns? 2) How can these sources combine to, over time, influence linguistic typology?

In Chapter 2, we focused on the question of word recognition and considered what synchonic sources of bias influence that process. We specifically looked at two sources of bias in phonological processing: human auditory perception and language-specific lexical knowledge, and suggested that these two influences coalesce during the word recognition process and can explain asymmetries we observed when testing French participants' sensitivity to mispronunciations along different featural dimensions. We found that both the manner and place contrasts were important in word recognition, the former due to acoustics, and the latter due to its status in the French lexicon.

In Chapter 3, we turned our attention to bias in phonological learning. We explored the hypothesis that certain "natural" patters (e.g., vowel harmony) are easier to learn than their logically equivalent but unattested counterparts (e.g., vowel *dis*harmony). We tested the effects of modality (production versus perception) and sleep-based memory consolidation on the learning of these kinds of patterns and found that the natural patterns did indeed tend to be learnt more easily than the unnatural ones. Though the effects we observed were small, we used a computational simulation to demonstrate how they can compound over time, yielding strong asymmetries in the typology.

Combining all of these sources of bias on a synchronic level will mean that certain patterns will be

more likely than others to succeed in interaction. In turn, this means that during the transmission process, certain patterns are more likely than others to survive, yielding shifts in typology over time. Some of these sources of bias are language-specific, and should yield different effects cross-linguistically. Most of what we have focused on in this thesis, however, concerns biases that should be present in all human learners, and thus can explain why even genetically very distinct languages often resemble each other so strongly.

We have only just begun to pick at the surface of how bias can influence typology over time. Future simulations should attempt to weigh the different sources of bias that we have discussed. Is the auditory system a more important constraint than learnability? Does this vary with regards to the type of pattern being learnt? It is imperative that future research consider these types of questions in order to study the shaping of linguistic typology from a holistic point of view.

Appendix A

Sound symbolism

This chapter details two studies that were conducted exploring the classic bouba-kiki phenomenon, whereby non-words containing certain sounds are more strongly associated with round shapes, and non-words containing other sounds are more strongly associated with spiky shapes. This bias to associate sounds with shapes is present cross-linguistically, but we are far from understanding why or how this phenomenon is so omni-present. The following two sections are reprints of articles that look at phonological effects in the bouba-kiki phenomenon. First, we examine what specific sounds and sound combinations influence adults' decisions in bouba-kiki two-alternative forced-choice tasks. Then, we attempt to test the developmental roots of the phenomenon by exploring the effect in preverbal infants.

A.1 Consonants are more important than vowels in the bouba-kiki effect

This section is a reprint of the following article: Fort, M., Martin, A., & Peperkamp, S. (2015). Consonants are more important than vowels in the bouba-kiki effect. *Language and Speech*, *58*(2), 247–266. Appendices from the original article are not reproduced here.

A.1.1 Introduction

In natural languages, the vast majority of words show an arbitrary link between form and meaning (de Saussure, 1959). However, languages typically also contain words that are sound-symbolic (e.g., in French: Chastaing, 1958; in English: Bloomfield, 1933; in Japanese: Imai, Kita, Nagumo, and Okada, 2008). For instance, in the English lexicon, the onset /gl/ is often used for words with meanings that are related to "vision" and "light" (e.g., "glimmer", "glisten", "glitter", "gleam", "glow", "glint"), whereas the consonant cluster /kr/ is often associated with "noisy impact" meanings in verbs (e.g., "crash", "crack", "crunch"). While the lexicons of individual languages may contain sound-symbolic items such as the English ones above, listeners across languages are sensitive to certain universal sound-symbolic associations that might or might not be exploited in their native lexicon. In particular, in sound-shape matching tasks, participants systematically map certain pseudowords, such as "bouba" and "maluma" onto round shapes, and others, such as "kiki" and "takete", onto spiky ones, a phenomenon known as the bouba-kiki, or alternatively, the maluma-takete effect (adults: Köhler, 1929, Köhler, 1947; Ramachandran and Hubbard, 2001; toddlers: Maurer, Pathman, and Mondloch, 2006). But why is a "bouba" more likely to be round whereas a "kiki" more probably refers to something spiky? To date, most studies either investigated these and other sound-symbolic associations across languages (Bremner et al., 2013; Imai et al., 2008; Kantartzis, Imai, & Kita, 2011; Nygaard, Cook, & Namy, 2009), or focused on the emergence of such associations in the course of human ontological development (Fort, Weiss, Martin, & Peperkamp, 2013; Maurer et al., 2006; Oztürk, Krehm, & Vouloumanos, 2013; Peña, Mehler, & Nespor, 2011; Spector & Maurer, 2013). However, the nature of the information in the speech signal that is actually matched with the visual shape is still unclear. In other words, little is known about which specific speech sounds in pseudowords like "bouba" and "kiki" make them more likely to be associated with a round and a spiky shape, respectively, rather than the reverse. In the present study, we investigate the respective influence of consonants and vowels in the bouba-kiki effect.

Audiovisual cross-modal correspondences for non-speech stimuli have been found in different populations, showing links between simple physical stimulus dimensions such as loudness and brightness (e.g., adults and children: Bond and Stevens, 1969; 20- to 30-day-old infants: Lewkowicz and

Turkewitz), pitch and brightness or visual lightness (e.g., human adults: Marks, 1987; chimpanzees: Ludwig, Adachi, and Matsuzawa, 2011; toddlers: Mondloch and Maurer, 2004), pitch and size (e.g., adults: Gallace and Spence, 2006), pitch and visual elevation (11-month-olds: Wagner, Winner, Cicchetti, and Gardner, 1981; three- to four-month-olds: M. P. Walker and Stickgold), and so on (for reviews see Marks, 2004; Spence, 2011; P. Walker, 2012). For instance, regarding sound-shape associations, Marks (1987, Experiment 4) used a speeded classification task where participants had to determine the spikiness of visually presented stimuli. Results showed slower reaction times and higher error rates when the object was accompanied by an incongruent auditory tone (i.e., high-pitched tone + round shape, low-pitched tone + spiky shape) rather than by a congruent one (i.e., high-pitched tone + spiky shape, low-pitched tone + round shape). Several studies have also documented cross-modal correspondences between auditory speech stimuli on the one hand and different visual properties of objects on the other hand, such as size (Parise & Spence, 2012; Peña et al., 2011; Sapir, 1929), brightness (Parise & Pavani, 2011), and shape (Köhler, 1929, 1947; Kovic, Plunkett, & Westermann, 2010; Maurer et al., 2006; Monaghan, Mattock, & Walker, 2012; Nielsen & Rendall, 2011; Oztürk et al., 2013; Parise & Pavani, 2011; Parise & Spence, 2012; Ramachandran & Hubbard, 2001; Sweeny, Guzman-Martinez, Ortega, Grabowecky, & Suzuki, 2012; Westbury, 2005).

With regards to speech-shape associations, while many studies have investigated the bouba-kiki effect, only a few of them have explicitly manipulated specific auditory components in the speech material (Monaghan et al., 2012; Nielsen & Rendall, 2011; Oztürk et al., 2013; Parise & Pavani, 2011; Parise & Spence, 2012; Sweeny et al., 2012). For instance, in two different experiments, Monaghan et al. (2012) compared different types of consonants (i.e., stops vs. continuants) and vowels (i.e., close front vs. open back vowels), respectively. Using a cross-situational learning paradigm, they found that participants better learned congruent sound-symbolic pairings between pseudowords and shapes (i.e., spiky shapes paired with stop consonants or close front vowels, and round shapes paired with continuant consonants or open back vowels) rather than incongruent ones. This study indicates that both consonantal and vocalic features play a role in speech-shape correspondences. However, little is known about the *respective* roles of consonants and vowels in the bouba-kiki effect.

This sound-symbolic phenomenon has often been claimed to depend mostly on the nature of the vowels in the speech stimuli (Maurer et al., 2006; Ramachandran & Hubbard, 2001; Tarte, 1974,

1982), possibly because perceivers match the visual shape with the shape of the lips when producing the vowels of the speech stimuli (e.g., presence of lip rounding in /u/, as in "bouba" or "maluma", vs. absence of lip rounding in /i/ and /e/ as in "kiki" and "takete"). However, two recent studies found that consonants seem to play a more important role than vowels in the bouba-kiki effect. Nielsen and Rendall (2011) used a forced-choice task, where in each trial participants had to match an auditory CVCV[1] or CVCVCV pseudoword with one of two visually presented shapes, one round, one spiky. Results indicated that, regardless of the vowels, auditory pseudowords containing certain consonants, including /b/, /m/, /l/ and /g/, are mapped more often onto round shapes, while pseudowords containing other consonants, including /p/, /t/ and /k/, are mapped more often onto spiky shapes. These data suggest the existence of a bias to rely more on consonants than on vowels when performing such sound-symbolic pairings. Using the same design, Oztürk et al. (2013) directly examined which type of segments (consonants, vowels, or both) are used to match the pseudowords /bubu/, /kiki/, /bibi/ and /kuku/ with round and spiky shapes. For instance, when the pseudoword /bibi/ was paired with a round shape, the choice was considered to be based on the consonant (/b/ $\longrightarrow$ round), whereas when it was paired with a spiky shape, the choice was considered to be based on the vowel (/i/ $\longrightarrow$ spiky). In accordance with Nielsen & Rendall (2011), the authors found more consonant-based than vowel-based responses.

These results mesh well with findings on the differential roles of vowels and consonants in speech processing. In particular, consonants have been shown to be more important for lexical access (Bonatti, Peña, Nespor, & Mehler, 2005; Caramazza, Chialant, Capasso, & Miceli, 2000; Cutler, Sebastián-Gallés, Soler-vilageliu, & Ooijen, 2000; Mehler, Peña, Nespor, & Bonatti, 2006; Nespor, Peña, & Mehler, 2003; New & Nazzi, 2014; Toro, Nespor, Mehler, & Bonatti, 2008; Toro, Shukla, Nespor, & Endress, 2008) whereas vowels have been argued to carry more information for the purpose of syntactic processing (Nespor et al., 2003; Toro, Nespor, et al., 2008; Toro, Shukla, et al., 2008). To the extent that mapping pseudowords onto visual shapes triggers processing at a more lexical rather than a syntactic level, the greater influence of consonants than vowels reported by Nielsen & Rendall (2011) and Oztürk et al. (2013) is thus unsurprising. However, more research is needed before drawing the conclusion that consonants play a more important role than vowels in the bouba-kiki effect,

[1] C=Consonant, V=Vowel.

for two reasons. First, both of the studies only tested a small number of stimuli: Nielsen & Rendall (2011) used five pairs of pseudowords while Oztürk et al. (2013) used only two, including a limited subset of the English phoneme inventory. Second, they used only consonant-initial pseudowords, making it impossible to rule out that the consonantal bias reflects an onset effect.

The goal of the present research is to provide further evidence regarding the respective roles of consonants and vowels in the bouba-kiki-effect. First, we test a larger set of segments (nine vowels and 15 consonants) to explore the robustness of the consonant-vowel asymmetry. Second we use both consonant- and vowel-initial stimuli to test for a possible onset bias. We report on three experiments with French adult participants, using a forced-choice association task as in Nielsen & Rendall (2011) and Oztürk et al. (2013). In Experiments 1 and 2, we explore the role of vowels on the bouba-kiki effect, while in Experiment 3 we focus on the influence of consonants.

A.1.2 Experiment 1

The goal of Experiment 1 is to explore whether changing the identity of the vowels in the stimuli used in Köhler (1947)'s original study influences the sound-symbolic matching process. We constructed pseudowords with varying vowels and two fixed consonant pairs based on those of "maluma" (i.e., /l,m/) and "takete" (i.e., /t,k/). We used disyllabic instead of trisyllabic stimuli because disyllables constitute the most prevalent word form in the French lexicon. Participants were asked to match these pseudowords with one of two visually presented shapes, one round, one spiky. If they rely more on consonants than on vowels to perform the sound-shape association, they should match pseudowords containing /m/ and /l/ more often with round shapes and those containing /t/ and /k/ with spiky shapes, regardless of the vowels.

A.1.2.1 Methods

A.1.2.1.1 Participants

Twenty-four native French-speakers (five men and 19 women, mean age: 26 years, range: 18-58) participated in the experiment. None of them had a known history of hearing or language impairment.

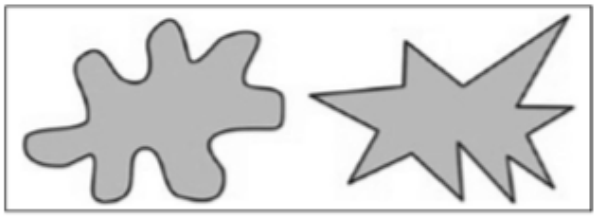

Figure A.1: Example of a round and a spiky shape used in experiments 1, 2 and 3

A.1.2.1.2 Materials

We constructed 32 CV_iCV_i pseudowords in which the vowel varied among 9 Standard French vowels (/i/, /y/, /e/, /ø/, /a/, /u/, /o/, /ɛ̃/, /ɑ̃/) and the consonants were either /l/ and /m/ (irrespective of order) or /k/ and /t/ (again, irrespective of order) (e.g., /lumu/, /mili/, /koto/, /teke/[2]). The pseudowords were recorded in a sound-proof booth by a female native French speaker. As the stimuli were also used for an infant study not reported on here, they were recorded in infant-directed speech. Stimuli were balanced across two different experimental lists (16 per list), such that for each combination of consonant pair and vowel, one pseudoword was part of List 1 and the other of List 2 (e.g., List 1 contained /lumu/ and /mɛ̃lɛ̃/, while List 2 contained /mulu/ and /lɛ̃mɛ̃/).

To construct the visual stimuli, we created a set of 33 spiky and round black-outlined shapes with Adobe Photoshop CS6, and filled each of them with two different colors (yellow and orange)[3]. We then asked 50 participants in a Mechanical Turk online study to judge the form of each shape on a scale from 1 (very round) to 7 (very spiky), and selected eight round and eight spiky shapes that - regardless of color - were considered overall to be roundest (scored 1-2) and spikiest (scored 6-7), respectively. The yellow and the orange versions of these shapes constituted the final set of visual stimuli, for a total of 32 shapes. Examples of a round and a spiky shape type are shown in Fig. A.1.

A.1.2.1.3 Procedure

Participants were randomly assigned to List 1 or List 2. They were seated at 40 cm from a computer monitor in a soundproof booth, and listened to the pseudowords through headphones. For each

[2]Out of the 36 possible items (9 vowels x 2 consonant pairs x 2 orders) 4 were excluded because they corresponded to French words

[3]These similar colors were chosen such as to minimize the possibility of color having an effect.

trial, one round and one spiky shape (both either yellow- or orange-filled) appeared side-by-side on the screen, against a white background. Then, after 500 ms, participants heard a pseudoword and had to press one of two labelled keys (one was on the left side and the other on the right side of the keyboard) to indicate whether they felt that the pseudoword referred to the shape on the left or on the right, respectively. There was no time pressure on their response and they could replay the auditory stimulus as often as they wanted by pressing the "R" key. The next trial started immediately after the participant had answered.

Participants were tested on 16 trials, one for each pseudoword. For each participant, each pseudoword was randomly matched with one spiky and one round shape of the same color. Round and spiky shapes appeared an equal number of times on the left and right sides of the screen. Each colored shape was used exactly once. The order of the stimuli was randomized. The E-Prime 2.0 software (Psychological Software Tools, Pittsburgh, PA, USA) was used to generate the stimuli and to collect participants' responses.

A.1.2.2 Results and discussion

For each trial, we coded each participant's choice as "1" or "0", depending on whether it corresponded to the round or the spiky shape, respectively. These scores were then averaged between subjects for each condition. The level equivalent to chance (i.e., 50%) was subtracted from these means, showing either an overall preference for round or spiky shapes. Results for each condition are reported in Fig. A.2. Positive scores indicate a preference for round shapes, negative scores indicate a preference for spiky shapes. For all the analyses reported in this study, we used a logistic mixed-effects model to analyze the data with R (R Core Team, 2012) and the lme4 package (Bates et al., 2014). Participants were a random factor, while Vowel Identity (/i/, /y/, /e/, /ø/, /a/, /u/, /o/, /ɛ̃/, /ɑ̃/) and Consonant Pair (/l,m/, /t,k/) were within-participant fixed factors. Initially, List (1, 2) was also declared as a random factor, but it did not significantly increase the variance accounted for and was thus excluded from the final model. The vowel /ɑ̃/ in the condition /t,k/ was used as the intercept because it yielded a mean score that was not significantly different from the 50% chance level (M = 42%, SD = 0.5, $t(23) < 1$).

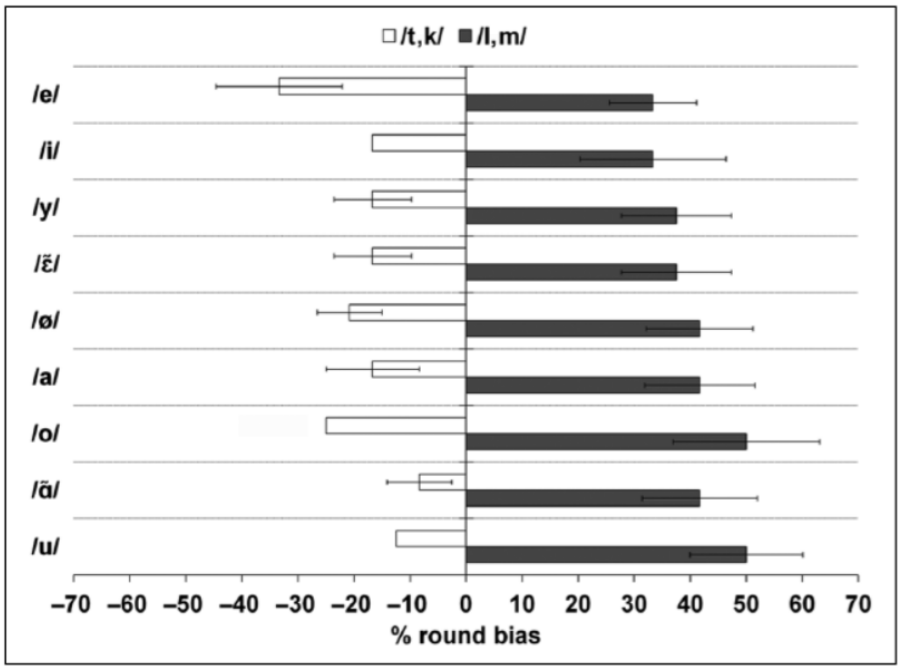

Figure A.2: Percentage of bias to choose round shapes over spiky shapes (round bias) as a function of vowel identity (/u/, /ɑ̃ /, /o/, /a/, /ø/, /ɛ̃ /, /y/, /i/, /e/) and consonant pair (/l,m/ vs. /t,k/). The error bars represent standard error.

Results revealed a significant influence of Consonant Pair (β = 2.84, SE = 0.88, z = 3.28, p = .001), indicating that pseudowords containing /t/ and /k/ were more often mapped onto spiky shapes whereas pseudowords containing /l/ and /m/ were more often mapped onto round shapes. For Vowel Identity, however, the analysis yielded no effect except a marginally significant one for /e/ (β = 1.33, SE = 0.69, z = 1.91, p = .059), showing that Vowel Identity did not influence the sound-shape mapping process. There was no interaction between Consonant Pair and Vowel Identity.

A.1.2.3 Discussion

The goal of this experiment was to explore whether varying the vowels influences the cross-modal correspondences observed in the classic bouba-kiki effect. The results indicate that this is not the case: regardless of the vowel, participants consistently mapped pseudowords containing /l/ and /m/ to round shapes and ones containing /t/ and /k/ to spiky shapes. In other words, only the consonants influenced the sound shape mappings. By selecting a larger set of stimuli than previously used (Nielsen & Rendall, 2011; Oztürk, et al., 2013) and by systematically varying the vowel and consonant pairings, we thus demonstrated a robust consonant-vowel asymmetry. Our results suggest that, in

line with the findings of Nielsen & Rendall (2011) and Oztürk et al. (2013), consonants have a greater influence than vowels in the bouba-kiki effect. However, this experiment cannot rule out that an onset bias rather than a consonant bias is responsible for the pattern of results, as all the stimuli were consonant-initial. To disentangle these two possible explanations, we ran a second experiment with vowel-initial stimuli.

A.1.3 Experiment 2

In this experiment, we used the same design as in Experiment 1, except that the pseudowords were of the form V_iCV_i (e.g., /ulu/, /imi/, /yky/, /ata/) instead of CV_iCV_i. Thus, our stimuli were vowel-initial, thereby encouraging participants to rely on vowels to perform sound-shape associations. If the effect observed in Experiment 1 is due to an onset bias, we should observe a consistent influence of vowel identity regardless of the consonant pair.

A.1.3.1 Methods

A.1.3.1.1 Participants

Twenty-four native French-speakers (eight men and 16 women, mean age: 21.9 years, range: 19-28) participated in the experiment. None of them had a known history of hearing or language impairment and none had participated in Experiment 1.

A.1.3.1.2 Materials

As in Experiment 1, we constructed 32 V_iCV_i pseudowords in which the vowel varied among 9 Standard French vowels (/i/, /y/, /e/, /ø/, /a/, /u/, /o/, /ɛ̃/, /ɑ̃/) and the consonants were either /l/ and /m/ (irrespective of order) or /k/ and /t/ (again, irrespective of order) (e.g., /lumu/, /mili/, /koto/, /teke/. The stimulus preparation and recording procedure were the same as in Experiment 1, except that the speaker was told to produce the stimuli in adult-directed speech For the visual stimuli, we created one round and one spiky shape and added them to the eight original round and spiky shapes used in Experiment 1. As in Experiment 1, we filled them with two different colors (yellow

and orange) and then mirrored them either horizontally or vertically to increase the number of each shape type to 36 (e.g., 9 pairs of shapes x 2 colors x 2 mirror-image versions).

A.1.3.1.3 Procedure

The procedure was the same as in Experiment 1 except that each participant was presented with all the pseudowords.

A.1.3.2 Results

he coding procedure for the participants' responses was the same as in Experiment 1. Results for each condition are reported in Fig. A.3. As in Experiment 1, positive scores indicate a preference for round shapes and negative scores a preference for spiky shapes. As in Experiment 1, we analyzed the data using a logistic mixed-effects model with participants as a random factor and Vowel Identity (/i/, /y/, /e/, /ø/, /a/, /u/, /o/, /ɛ̃/, /ɑ̃/) and Consonant Pair (/l,m/, /t,k/) as within-participants fixed factors. The vowel /ɑ̃/ in the condition /t,k/ was used as intercept because it yielded a mean score that was not significantly different from the 50% chance level (M = 40%, SD = 0.5, $t(47)$ = 1.46, p = .15).

The results revealed a significant effect of Consonant Pair (β = 2.88, SE = 0.61, z = 4.74, $p < .001$). A main effect of Vowel Identity was significant, indicating the pseudowords containing /i/ were significantly more often associated with a spiky shape than those containing /ɑ̃/ (β = 1.37, SE = 0.51, z = 2.69, $p < .01$). A similar, marginally significant, tendency was observed for /e/ (β = 0.81, SE = 0.46, z = 1.77, p = .08). There was no interaction between Consonant Pair and Vowel Identity.

A.1.3.3 Discussion

In this experiment, we tested whether the consonant-vowel asymmetry observed in Experiment 1 could be due to an onset bias, by using V_iCV_i instead of CV_iCV_i stimuli. The results showed that overall, pseudowords with /k/ and /t/ were more often associated with spiky shapes than those with /l/ and /m/. The magnitude of this effect was similar to the one observed in Experiment 1

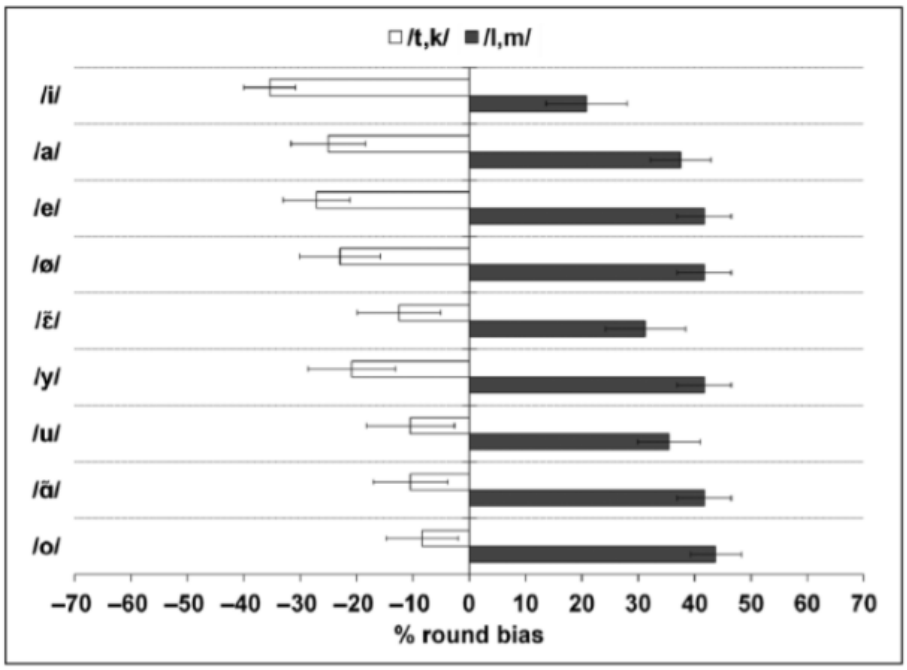

Figure A.3: Percentage of bias to choose round shapes over spiky shapes (round bias) as a function of vowel identity (/u/, /ɑ̃ /, /o/, /a/, /ø/, /ɛ̃ /, /y/, /i/, /e/) and consonant pair (/l,m/ vs. /t,k/). The error bars represent standard error.

(Experiment 1: $\beta = 2.84$; Experiment 2: $\beta = 2.88$). Moreover, there was only a small influence of the vowels: /i/ and /e/ were more often associated with spiky shapes than the other vowels (/y/, /ø/, /a/, /u/, /o/, /ɛ̃/ and /ɑ̃/). These effects were small (/i/: $\beta = 1.37$, /e/: $\beta = 0.81$) and significant only for /i/. Thus, despite the fact that the pseudowords were vowel-initial, participants relied mostly on the consonantal information. This, then, shows that the consonant-vowel asymmetry observed in Experiment 1 cannot be due to the fact that the pseudowords in that experiment were consonant-initial.

One caveat is still in order, though. As participants had to make a two-alternative forced choice in both Experiment 1 and 2, they might have developed a response strategy to focus on the consonants rather than on the vowels, because there were likewise two pairs of consonants but many different vowels. If the consonant-vowel asymmetry is due to such a response strategy, we should observe the reverse asymmetry when using two vowel pairs and multiple consonants. Conversely, if this strategy is not the main factor that induced this pattern of results, we should still observe a consonant-vowel asymmetry, at least for the consonants used in Experiments 1 and 2 (i.e., /l/, /m/, /t/ and /k/). In the next experiment, we thus use the same paradigm as in Experiments 1 and 2, but with stimuli

that exhibit a wide range of consonants and only two pairs of vowels. This experiment will also allow us to gain insight into which consonants besides /l/ and /m/ tend to be associated with round shapes, and which ones beside /t/ and /k/ are more often associated with spiky shapes. For instance, Monaghan et al. (2012) found that the association of individual consonants with round or spiky shapes by English listeners depends on the feature continuancy, with sonorants and fricatives being associated with the former and stops with the latter. If continuancy is likewise the determining feature for French listeners, we expect to observe the same response pattern.

A.1.4 Experiment 3

In this experiment, we constructed C_iVC_iV stimuli by crossing 15 different consonants with two pairs of vowels (/o,u/ and /i,e/) that are typically mapped to round and spiky shapes, respectively (Maurer, et al., 2006; Monaghan & Mattock, 2012; Ramachandran & Hubbard, 2001; Tarte, 1974, 1982).

A.1.4.1 Methods

A.1.4.2 Participants

Twenty-three native French-speakers (nine men and 14 women, mean age: 26 years, range: 21-40) participated in the experiment. None of them had a known history of hearing or language impairment, and none had participated in Experiment 1 or 2.

A.1.4.2.1 Material

We created 56 C_iVC_iV pseudowords in which the consonant varied among 15 Standard French consonants (6 stops: /p/, /b/, /t/, /d/, /k/, /g/ and 9 continuants: /f/, /v/, /s/, /z/, /ʃ/, /ʒ/, /l/, /m/, /n/) and the vowels were either /o/ and /u/ (irrespective or order) or /i/ and /e/ (again, irrespective of order) (e.g., /pupo/, /popu/, /kike/, /keki/[4]). The pseudowords were recorded by a female native French speaker. As the stimuli were also used for an infant study not reported on here, they were recorded in infant-directed speech. As in Experiment 1, stimuli were balanced across two different experimental lists

[4]Out of the 60 possible items (15 consonants x 2 vowel pairs x 2 orders) 4 were excluded because they correspond to French words.

(28 per list), such that for each combination of vowel pair and consonant, one pseudoword was part of List 1 and the other of List 2 (e.g., List 1 contained /popu/ and /fufo/, while List 2 contained /pupo/ and /fofu/). To construct the visual stimuli we randomly selected 14 round and spiky shapes from Experiment 1. As in Experiment 1, we filled them with 2 different colors, to increase their number to 28 for each shape type.

A.1.4.2.2 Procedure

The procedure was the same as in Experiment 1.

A.1.4.3 Results

Scoring of participants' responses was done in the same way as in the previous two experiments. Results for each condition are reported in Fig. A.4. Positive scores indicate a preference for round shapes and negative scores indicate a preference for spiky shapes. As in Experiments 1 and 2, we analyzed the data using a logistic mixed-effects model with participants as a random factor and Consonant Identity (/p/, /b/, /t/, /d/, /k/, /g/, /f/, /v/, /s/, /z/, /ʃ/, /ʒ/, /l/, /m/, /n/) and Vowel Pair (/o,u/, /i,e/) as within-participants fixed factors. Initially, List (1, 2) was also declared as random factor, but it did not significantly increase the variance accounted for and was hence excluded from the final model. The vowel /ɑ̃/ in the condition /t,k/ was used as intercept because it yielded a mean score that was not significantly different from the 50% chance level (M = 40%, SD = 0.5, $t(47)$ = 1.46, p = .15). The consonant /f/ in the condition /o,u/ was used as intercept because they yielded mean scores that were not significantly different from the 50% chance level (M = 50%, SD = 0.51, $t(23) < 1$).

Results revealed a main effect of Consonant Identity, such that pseudowords containing the consonants /b/, /ʃ/, /d/, /l/, /m/, /n/ (all β = 2.43, all SE = 0.85, all z = 2.85, all $p < .005$), /s/, /p/, /ʒ/ (all β = 1.36, all SE = 0.65, all z = 2.09, all $p < .05$), and /g/ (β = 1.98, SE = 0.745, z = 2.65, $p < .01$) were significantly more often associated with round shapes than those containing /f/; Pseudowords containing the remaining consonants (/k/ and /t/: both $z < 1$; /v/: β = 0.7, SE = 0.58, z = 1.18, p = .23; /z/: β = 0.9, SE = 0.6, z = 1.5, p = .14) did not significantly differ from those containing /f/. The main effect of Vowel Pair was not significant ($z < 1$). However, there was an interaction between

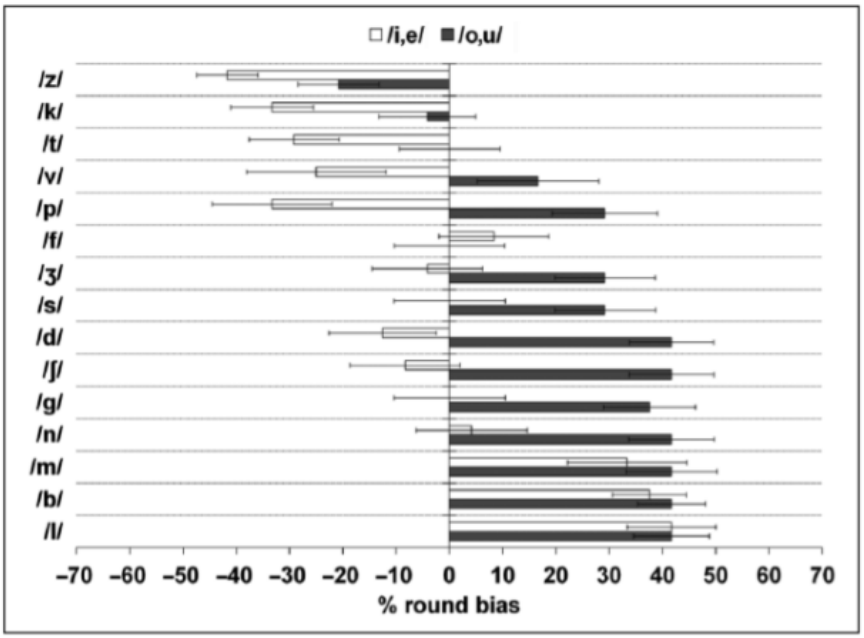

Figure A.4: Percentage of bias to choose round shapes over spiky shapes (round bias) as a function of vowel identity (/u/, /ɑ̃ /, /o/, /a/, /ø/, /ɛ̃ /, /y/, /i/, /e/) and consonant pair (/l,m/ vs. /t,k/). The error bars represent standard error.

Consonant Identity and Vowel Pair: the effect of Vowel Pair was significant for some consonants (/ʃ/: $\beta = 3.12$, $SE = 1.03$, $z = 3.014$, $p < .005$; /d/: $\beta = 3.3$, $SE = 1.04$, $z = 3.18$, $p < .005$; /g/: $\beta = 2.32$, $SE = 0.95$, $z = 2.45$, $p < .05$; /ʒ/: $\beta = 1.87$, $SE = 0.88$, $z = 2.14$, $p < .05$; /k/: $\beta = 1.81$, $SE = 0.90$, $z = 2.01$, $p < .05$; /n/: $\beta = 2.6$, $SE = 0.1$, $z = 2.52$, $p < .05$; /p/: $\beta = 3.4$, $SE = 1.1$, $z = 3.08$, $p < .005$). For the remaining consonants, the effect of Vowel Pair was either marginally significant (/s/ and /t/: both $\beta = 1.17$, both $SE = 0.87$, both $z = 1.95$, both $p = .052$; /z/: $\beta = 1.87$, $SE = 1.05$, $z = 1.78$, $p = .08$), or not significant (/m/, /l/ and /b/: all $z < 1$).

A.1.4.4 Discussion

In this experiment we further investigated the consonant bias observed in Experiments 1 and 2 by testing a wider range of consonants, combined with two vowel pairs traditionally thought to be associated with round and spiky shapes, respectively. First, we did not observe any significant main effect of vowel pair. Regarding the main effect of consonant identity, we found that /b/, /ʃ/, /d/, /l/, /m/, /n/, /s/, /p/, /ʒ/ and /g/ were more often associated to round shapes than /f/, /k/, /t/, /v/ and /z/. Finally, interactions between consonant identity and vowel pair indicated that vowel pair had

a significant influence in the context of most of the consonants, within the same range as that of consonant identity (1.8β3.5), indicating that /i/ and /e/ were more often associated with spiky shapes than /o/ and /u/, except in the context of /t/, /s/ and /z/ (where the effect was marginally significant) and /b/, /m/ and /l/ (where the effect was not significant).

These results indicate, first of all, that the consonant-vowel asymmetry observed in Experiments 1 and 2 is likely not due to a response strategy according to which listeners favored consonants because they came in two pairs, thus matching the number of proposed alternatives in the forced-choice task. Indeed, according to this hypothesis we should have observed a reversed vowel-consonant asymmetry in the present experiment, contrary to fact. Second, this data shows that consonants are not associated with round and spiky shapes according to their continuancy feature. Recall that Monaghan et al. (2012) found that English listeners tend to associate sonorants and fricatives with round shapes and stops with spiky shapes. The present experiment shows a more subtle pattern of results, by unpacking the coarser distinction made in Monaghan et al. (2012): Even though most continuant consonants were more often associated with round shapes, two of them (i.e., /v/ and /z/) were more often associated with spiky shapes. Moreover, among the stop consonants, only two (i.e., /t/ and /k/) were more often associated with spiky shapes, the remaining four (i.e., /b/, /p/, /g/ and /d/) were more often associated with round shapes. Finally, the fact that we observed a significant influence of the vowel pair in items containing nine out of the 15 consonants shows that vowels do play a role, albeit a minor one, in the bouba-kiki effect.

A.1.5 General discussion

Since Köhler et al.'s (1947) original study, the bouba-kiki effect has been extensively investigated, but the exact mechanism underlying this specific sound-symbolic phenomenon remains unclear. The present study examined the respective roles of consonants and vowels in this effect. We conducted three experiments in French monolingual adults, using a forced-choice association task in which pseudowords were to be mapped to one of two shapes. In Experiment 1, participants consistently mapped CVCV pseudowords containing /l/ and /m/ to round shapes and those containing /t/ and /k/ to spiky shapes, regardless of the nine different vowels with which they could be paired. Experiment

2 yielded basically the same pattern of results with VCV stimuli, ruling out the possibility that the consonant-vowel asymmetry observed in Experiment 1 reflected an onset bias. In Experiment 3, we used CVCV pseudowords by crossing 15 different consonants with two pairs of vowels and found that vowels nonetheless do have an influence, when paired with consonants other than /m/, /l/ and /b/: participants more often mapped pseudowords containing /o/ and /u/ to round shapes and those containing /i/ and /e/ to spiky shapes. In the following, we first discuss separately the influence of vowels and consonants, respectively, in the bouba-kiki effect. Then, we focus on the respective roles of consonants and vowels and discuss the possible implications of our findings for theories that describe different functions of consonants and vowels in speech perception.

Regarding the influence of vowels in the bouba-kiki effect, this study indicates that, as previously shown (Maurer, et al., 2006; Monaghan, et al., 2012; Ramachandran & Hubbard, 2001; Tarte, 1974, 1982), front unrounded vowels (e.g., /i/, /e/) are more often associated with spiky shapes, while back rounded vowels (e.g., /u/, /o/) are more often associated with round shapes. As the two types of vowels differ both in rounding and in backness, either one of these features might be responsible for the effect. On the one hand, vowel rounding is directly visible as lip protrusion and it has been argued that this is what leads to the association of round/rounded and unrounded vowels with round and spiky shapes, respectively (Maurer et al., 2006, Ramachandran & Hubbard, 2001). On the other hand, front vowels have a higher F_2 than back vowels (the two types of vowels do not greatly differ in F_1), and previous studies have shown sound-symbolic associations between high-pitched and low-pitched auditory stimuli with spiky and round shapes, respectively, in both infants and adults (Marks, 1987; Parise & Spence, 2012; Walker, 2012; Walker, et al., 2010). Most languages have only front unrounded and back rounded vowels. French, however, also has front rounded vowels, and although the present study was not designed to examine this question, we did include two front rounded vowels, /y/ and /ø/ in Experiments 1 and 2 (cf. Fig. A.2 and Fig. A.3). Interestingly, the mean round bias (averaged across Consonant Type) observed for both of these vowels in both Experiments 1 and 2 is comprised between the mean bias observed for /i/ and /e/ on the one hand that for /o/ and /u/ on the other hand. Thus, both rounding and backness might contribute to the bouba-kiki effect found in vowels. More research combining the three types of vowels (front rounded, front unrounded, back rounded) with consonants that yield no strong association with either round or

spiky shapes (e.g., /p/ and /v/), cf. Fig. A.4) is needed to further investigate this issue.

With regards to the influence of consonants, the overall pattern of data observed in this study is more complex. One thing is clear: the consonants /m/ and /l/ stand out as "round" insofar as they were systematically associated with round shapes across all experiments. As both of these consonants are sonorants, one could hypothesize that it is this feature (or more basically their status as continuants) that leads to such consistent sound-symbolic associations (as claimed by Monaghan et al. 2012; Westbury 2005; see also Parise & Spence, 2012, for similar distinctions). However, as previously stated, this hypothesis cannot entirely explain the pattern observed in Experiment 3, in that certain fricatives (which are also continuants) were more often associated with spiky shapes and certain plosives (which are not continuants) were more often associated with round shapes. Another noticeable difference between /m/ and /l/ vs. /t/ and /k/ is that the former are voiced whereas the latter are voiceless. Interestingly, Monaghan et al. (2012) found that in the English lexicon, words referring to the concept of angularity or spikiness are more likely to contain voiceless consonants, whereas words referring to roundness are more likely to contain voiced consonants. Thus, one might raise the hypothesis that voiced consonants are associated with round shapes and voiceless ones with spiky shapes. Voiced consonants are produced with vocal fold vibration and contain more intensity in lower frequency bands; either of these characteristics might lead to the association with round shapes. Yet again, our data do not entirely support this hypothesis. The most striking counterexample is /s/, which was more often associated with round shapes: not only is this consonant voiceless, it also contains the largest amount of energy in high-frequency bands among all French consonants.

Considering the respective influence of vowels and consonants in the bouba-kiki effect, this study indicates, first of all, that at least for some consonants there is little or no influence of the vowel. This result holds even with VCV pseudowords and hence cannot be attributed to an onset effect. Note also that it is independent of the type of speech in which the pseudowords had been recorded, i.e. infant-directed as in Experiments 1 and 3 or adult-directed as in Experiment 2. As infant-directed speech is characterized, among other things, by lengthened vowels, one might expect that this type of speech encourages participants to rely on vowels, but this does not appear to be the case in the present study.

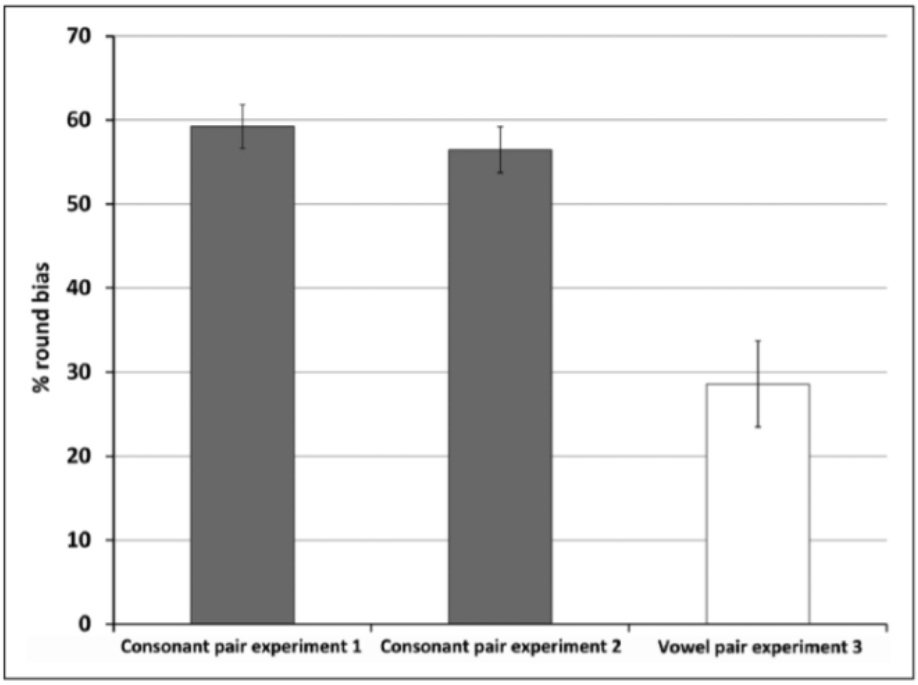

Figure A.5: Effect of consonant pair in experiment 1 and experiment 2 (mean round bias for /l,m/ minus mean round bias for /t,k/) and the effect of vowel pair in experiment 3 (mean round bias for /o,u/ minus mean round bias for /i,e/). The error bars represent standard error.

To further compare the respective influence of consonants and vowels across our three experiments, we compared the effect of consonant pair obtained in Experiment 1 and 2 to the effect of vowel pair observed in Experiment 3. To do so, we estimated the effect of consonant pair in Experiment 1 and 2 by subtracting the mean round bias score obtained across the nine modalities of vowel identity in the /t,k/ condition from the one obtained in the /l,m/ condition. Similarly, we estimated the effect of vowel pair by subtracting the mean round bias score obtained across the 15 modalities of consonant identity in the /o,u/ condition from the one obtained in the /i,e/ condition. We then conducted a one-way ANOVA on these measures declaring Effect Type (Consonant Exp.1, Consonant Exp.2, Vowel Exp.3) as a between-subject factor. Results are shown in Fig. A.5.

The analysis yielded a significant main effect of Effect Type ($F(2, 30) = 15.5$, $p < .001$, $\eta_p^2 = .50$). Planned comparisons revealed that it was due to the fact that the difference observed for Vowel Pair in Experiment 3 was significantly smaller than the difference observed for Consonant Pair in Experiments 1 and 2 ($F(1, 30) = 29.96$, $p < .001$, $\eta_p^2 = .97$), while the effect of Consonant Pair in Experiments 1 and 2 did not statistically differ from one another ($F < 1$). Overall, this data indicates that the effect of consonant pair was greater than the effect of vowel pair across the three experiments. The

fact that the effect of consonant pair between Experiment 1 was not significantly different from the one observed in Experiment 2, moreover, strengthens our claim that the consonant-vowel asymmetry in the bouba-kiki effect is not due to an onset effect. This result indeed shows that initial consonants do not contribute more to the bouba-kiki effect than non-initial ones. It also provides a statistical underpinning of the claim that infant- and adult-directed speech stimuli are not treated any differently.

Finally, the present study used a wider range of consonants and vowels than previous ones that likewise showed a consonant-vowel asymmetry (Nielsen & Rendall, 2011; Oztürk, et al., 2013). This allowed us to show that the contribution of each phoneme to this asymmetry is not the same. Specifically, we showed that even for vowels that have previously been found to be consistently associated to round and spiky shapes (e.g., /o,u/ vs. /i,e/), the influence varies as a function of the consonantal context: when combined with certain consonants, the effect was weak (/l/, /m/, /t/, /z/ in Experiment 2) or even absent (i.e., /b/ in Experiment 3). By contrast, the influence of consonants is more stable in that it depends less on the vocalic context.

All in all, our data clearly demonstrates that listeners are influenced more by consonants than by vowels to associate pseudowords with visual shapes. This is in line with the literature pertaining to the differential role that consonants and vowels may play in speech perception (Bonatti, et al., 2005; Caramazza, et al., 2000; Cutler, et al., 2000; Mehler, et al., 2006; New & Nazzi, 2013; Toro, Nespor, et al., 2008; Toro, Shukla, et al., 2008). Cutler et al. (2000) showed that both Dutch and Spanish participants would rather alter a vowel than a consonant to turn a pseudoword into a real word. For instance, they turned "kebra" into "cobra" rather than into "zebra". In the same vein, Bonatti et al. (2005) found that adults can track statistical regularities among consonants but not among vowels when segmenting words from a continuous auditory stream of artificial speech. Findings like these demonstrate that consonants are more important than vowels in lexical access. Assuming that the bouba-kiki effect involves some processing at a lexical level, this hypothesis could explain why this sound-shape mapping process depends more upon consonants than upon vowels. Whether the fact that consonants are more informative than vowels for lexical access is a result of an innate language module (Bonatti et al., 2005) or the result of exposure to the lexical regularities of one's native language (Keidel, Jenison, Kluender, & Seidenberg, 2007) remains an open question and should be

explored in future research.

To conclude, among all the cross-modal or cross-sensory correspondences observed so far (see Marks, 2004 and Spence, 2011 for reviews), the bouba-kiki effect seems to be the most complex one. Indeed, since its first discovery in 1947, no phonological or acoustic feature or set of features has been identified as being responsible for this effect. This is not surprising, as unlike most other cross-modal correspondences, the bouba-kiki effect involves auditory speech events rather than simple stimuli that vary across a single physical dimension, such as pitch or loudness. Thus, the bouba-kiki effect could be influenced by a host of factors, including acoustic, articulatory, and phonological properties of the speech stimuli[5] (see Spence, 2011 and Walker, 2012 for similar arguments). Further research is needed to systematically test the possible influences of these different factors on the bouba-kiki effect.

A.2 Looking for the bouba-kiki effect in prelexical infants

This section is a reprint of the following article: Fort, M., Weiss, A., Martin, A., & Peperkamp, S. (2013). Looking for the bouba-kiki effect in prelexical infants. In: S. Ouni, F. Berthommier & A. Jesse (eds.) Proceedings of the 12th International Conference on Auditory-Visual Speech Processing, INRIA, 71–76. Appendices and tables from the original publication are not reproduced here.

A.2.1 Introduction

The link between a speech sound and its meaning is supposed to be arbitrary (de Saussure, 1959). However, most languages contain sound-symbolic words (e.g., English: Bloomfield, 1933, Spence, 2011; Japanese: Imai et al., 2008, Kantartzis et al., 2011). For instance, the English lexicon has several sets of verbs in which a shared initial consonant cluster seems to reflect a common part of the verbs‟ meanings (e.g., /kr/ is associated with “noisy impact” in verbs like crash, crack, and crunch). Moreover, adults and toddlers are sensitive to sound symbolism: They more easily learn novel sound-meaning mappings when the words are sound-symbolic compared to when they are not (Imai et al.,

[5]Note that other factors, such as orthography, also may play a role (cf. Westbury, 2005).

2008; Kantartzis et al., 2011; Nygaard et al., 2009). In various sound-shape matching tasks, they also consistently associate certain pseudowords, such as „bouba" or „maluma", with round shapes, and others, such as „kiki" or „takete", with spiky ones (Maurer et al., 2006; Ramachandran & Hubbard, 2001; Köhler, 1929). This so-called bouba-kiki effect holds across different cultures and languages (Spence, 2011; Bremner et al., 2013), and hence it may well be universal.

One question concerning these spontaneous sound-symbolic associations concerns its ontological origin. Indeed, whether the bouba-kiki effect is an unlearned aspect of perception or rather emerges with language exposure is largely unknown. Its presence could depend upon the acquisition of sound-symbolic words in the native language, or upon direct experience with one's own vocal tract gestures when producing speech sounds. Still other possibilities are that it arises with passive exposure to speech sounds, or even that it is present at birth. The goal of the present research is to explore whether prelexical infants who have neither lexical knowledge nor experience with babbling already show a bouba-kiki effect.

To our knowledge, only one study has reported a bouba-kiki effect in infants (Oztürk et al., 2013). In a preferential listening procedure, four- month-olds looked longer at a shape when the accompanying speech sound is judged as incongruent by adults (round shape + /kiki/ or spiky shape + /bubu/) than when it is judged as congruent (round shape + /buba/ or spiky shape + /kiki/). This study used only two stimulus pairs, and the effect was limited in scope: contrary to adult control subjects, the infants failed to show a preference in two additional experiments in which either the consonants (/kiki/ vs. /kuku/) or the vowels (/bubu/ vs. /kuku/) were held constant.

In the present research, we further explore the presence of a bouba-kiki effect in prelexical infants. Moreover, we investigate whether consonants and vowels might play differential roles. The bouba-kiki effect has often been claimed to be mostly driven by the influence of vowels (Maurer et al., 2006; Ramachandran & Hubbard, 2001), possibly because perceivers match the visual shape with the shape of the lips when producing the vowels within the speech stimuli (e.g. presence of lip rounding in /u/, as in /maluma/ and /buba/, vs. absence of lip rounding in /i/ and /e/, as in /takete/ and /kiki/). Prelexical infants are sensitive to this cue: as early as two months of age they can match auditory /i/ and /u/ onto silent videos of a talking face showing the corresponding articulatory

gestures (Patterson & Werker, 2003). However, consonants have also been shown to play a role in the bouba-kiki effect (Westbury, 2005; Monaghan, Christiansen, & Fitneva, 2011), and recent data even provide evidence for a stronger influence of consonants than of vowels (Oztürk et al., 2013; Nielsen & Rendall, 2011; Fort, Martin, & Peperkamp, 2015). For instance, French adults map CVCV pseudowords systematically onto round shapes when the consonants are /m/ and /l/ and onto spiky shapes when they are /k/ and /t/, regardless of the vowels (e.g., both /lumu/ and /limi/ are mapped onto round shapes and both /koto/ and /kiti/ onto spiky shapes). By contrast, their mapping of CVCV pseudowords onto round shapes when the vowels are /o/ and /u/ and onto spiky shapes when the vowels are /e/ and /i/ is less systematic: depending on the consonants, they sometimes prefer the reverse mappings (e.g., while both /pipe/ and /dedi/ are consistently mapped onto a spiky shape, /bibe/ and /memi/ are mapped onto a round shape) (Fort et al., 2015). Thus, in addition to exploring the presence of a bouba-kiki effect in prelexical infants, we examine whether, like adults, infants are more sensitive to consonants than to vowels in this type of sound symbolic matching.

In the following, we report on three experiments with five- and six-month-old infants. In order to test the robustness and the possible generalization of infants' sound-symbolic matching between speech sounds and visual shapes, we use at least six different pairs of auditory and visual stimuli in each of them.

A.2.2 Experiment 1

We use an intermodal preferential looking procedure with five-month-old infants. Infants hear both homogeneous stimuli, where both consonants and vowels are consistently matched to round (e.g., /lomo/) or spiky shapes (e.g., /tiki/) by French adults (Fort et al., 2015), and heterogeneous stimuli that combine either round consonants with spiky vowels (e.g., /limi/) or spiky consonants with round vowels (e.g., /toko/). The latter type of stimuli allows us to investigate whether infants are more sensitive to consonants or to vowels when performing sound-shape mappings.

A.2.2.1 Methods

A.2.2.1.1 Participants

Twenty-four five-month-old infants (range: 4;24-5;28; mean: 5;14; 16 girls) participated. The data from ten more infants were excluded from the analyses due to fussiness (N=8) or strong side bias (N=2).

A.2.2.1.2 Stimuli

The auditory stimuli consisted of 28 CVCV disyllabic sequences. They were constructed by combining two consonants that are both consistently associated with either round or spiky shapes by French adults (Fort et al., 2015) ("round": /l,m/; "spiky": /k,t/) and two vowels that are likewise both associated with either round or spiky shapes ("round": /o,u/; "spiky": /i,e/). Thus, half of the stimuli contained vowels and consonants that are all judged as either round (e.g., /lumu/) or spiky (e.g., /kiti/). The other half contained either round consonants and spiky vowels (e.g., /mili/) or round vowels and spiky consonants (e.g., /tuku/). The stimuli were recorded by a French female speaker in infant-directed-speech. Their mean duration, minimum Fo and maximum Fo - averaged for the items within each condition - were not significantly different among the four conditions (all $p > .05$).

To construct the visual stimuli we created two pairs of black outlined shapes (one round, one spiky for each pair); within each pair, the shapes were matched in surface but not in number of curves and spikes. We filled both pairs with 7 different colors (yellow, red, brown, blue, pink, green, purple), and finally mirrored each pair of shapes to increase the total number of pairs to 28 (2 shapes x 7 colors x 2 mirror-image versions). Color-filled examples of each pair are shown in Fig. A.6.

A.2.2.1.3 Procedure

Infants were held in the lap of a parent and tested in a quiet room. Their eye gazes were monitored and recorded on video. The parents wore headphones with masking music that prevented them from hearing the auditory stimuli. Each trial started with the presentation of an attention getter, a colorful

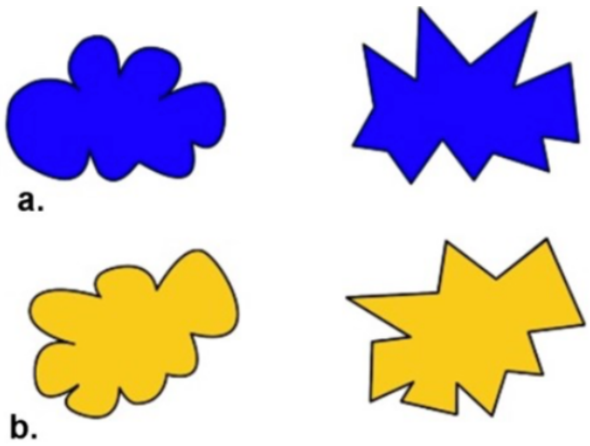

Figure A.6: Example of the first (a) and the second pair (b) of round and spiky shapes used in Experiment 1.

moving image, on a television screen. As soon as the infant had fixated it for 1 s., it was replaced by a pair of still shapes (one round, one spiky) presented side-by-side on a white background, followed after 500 ms by the presentation of five repetitions of a single token of one auditory stimulus. Each trial lasted 10.5 s. The order of the stimuli and the sound-shape pairing was pseudo-randomized across four different lists. Each infant was presented with only one list, consisting of 28 different trials divided into two blocks. Within each block, the round shape appeared at the left side half of the times. Trials with the same type of auditory stimulus or with the same color of the visual stimulus were presented not more than twice in a row.

A.2.2.2 Results and discussion

Infants' eye gazes were coded off-line frame-by-frame. Frames in which the infant looked towards the round shape were coded "1" and those in which they looked towards the spiky shape were coded "-1". Figure A.7 shows these scores averaged across all infants in intervals of 1 second from the beginning of the auditory stimulus until its end, and separated for the homogeneous and the heterogeneous auditory stimuli. A positive mean score indicates a preference for the round shape, a negative one a preference for the spiky shape.

Note that while infants showed an overall preference for the round shapes, they did not seem to look differentially to the round and the spiky shapes as a function of auditory stimulus, whether these were homogeneous or heterogeneous. We thus conducted our analyses over the total trial duration.

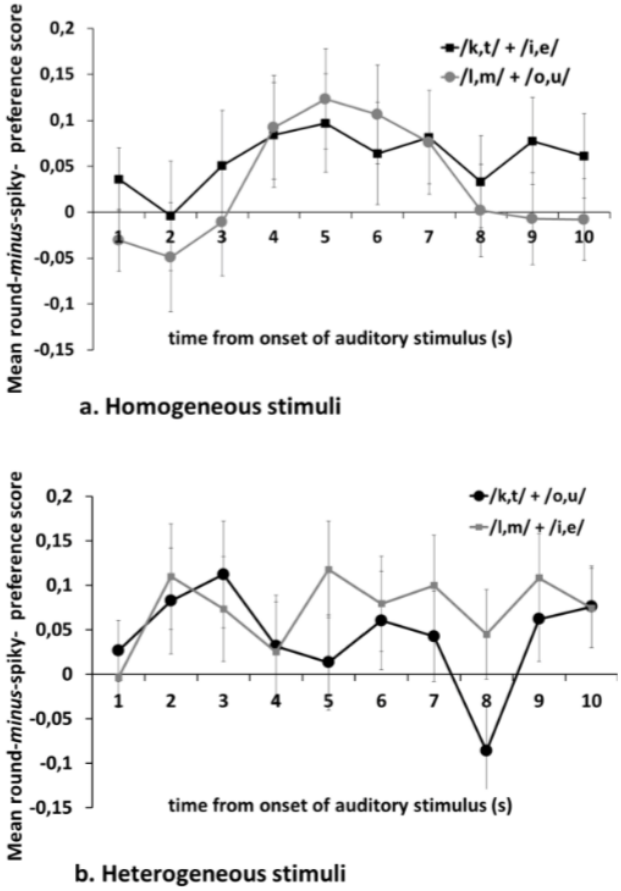

Figure A.7: Mean preference scores for the homogeneous (a) and the heterogeneous (b) auditory stimuli over time. Positive scores indicate a preference for the round shape. Error bars represent standard errors from the mean.

These difference scores were submitted to a 2x2 repeated measures ANOVA with the factors Consonant Type (/l,m/ vs. /t,k/) and Vowel Type (/o,u/ vs. /i,e/). The analysis revealed neither a main effect nor an interaction (Consonant Type: $F < 1$; Vowel Type: $F(1, 23) = 2.25$, $p = .15$; Consonant Type x Vowel Type: $F < 1$). Furthermore, planned restricted analyses of the two conditions in which both the vowels and the consonants were either round or spiky, showed no significant differences either (both $F < 1$). Overall, infants only showed positive difference scores, indicating a general preference for the round shapes ($t(23) = 3.02$, $p < .01$), in accordance with what has been reported in (Jadva, Hines, & Golombok, 2010). The same pattern of results was obtained analyzing the first and the second block separately: no significant effects in the ANOVA, (all $p > .10$), and a preference for the round shapes (first block: $t(23) = 2.33$, $p < .05$; second block: $t(23) = 2.05$, $p = .05$).

Thus, we found no sign of a bouba-kiki effect in this experiment, not even when we considered only the two conditions in which according to French adults the pseudowords were composed of consonants and vowels that are either all round or all spiky (Fort et al., 2015). One possible explanation for the absence of an effect is that half of the stimuli contained either spiky consonants with round vowels or round consonants with spiky vowels; this might have prevented infants to perform sound-symbolic associations. Moreover, the preference for the round shapes overall might have masked the expected bouba-kiki effect. In the next experiment, we address these points as follows. First, we only use auditory stimuli in which both the consonants and the vowels are either round or spiky, hence leaving aside the question as to the relative roles of consonants and vowels in sound-shape matching. Second, we use a preferential looking procedure as in (Oztürk et al., 2013), where in each trial only one shape is displayed.

A.2.3 Experiment 2

We use a preferential looking paradigm with six-month-old infants. Infants only hear stimuli where both consonants and vowels are consistently matched onto round (e.g., /buba/) or spiky shapes (e.g., /kike/) by French adults in a pre-test. As in (Oztürk et al., 2013), we manipulate the type of the shape displayed in each trial. That is, we use sound-symbolic pairings that are either congruent (e.g., round shape + /buba/; spiky shape + /kike/) or incongruent (e.g., round shape + /kike/; spiky shape

+ /buba/). If infants are sensitive to this type of sound symbolism, they should look differentially in the two types of trials.

A.2.3.1 Methods

A.2.3.1.1 Participants

Twenty-four six-month-old infants (range: 5;24–6;17; mean: 6.;26; 12 girls) participated. The data from six more infants were excluded from the analyses due to parental interference (N=4), equipment failure (N=1), or failure to look away on more than half of the trials (N=1).

A.2.3.1.2 Stimuli

The auditory stimuli consisted of 12 CVCV pseudo-words. Half of them were of the bouba-type: their vowels had been chosen from among /o,u,a/ and their consonants from among /b,d,g,v,ʒ/. The other half were of the kiki-type: their vowels had been chosen from among /i,e, / and their consonants from among /p,t,k,f,s/. There were no significant differences in duration, in minimum and maximum Fo, or in maximum Fo difference between the bouba-type items and the kiki-type items. A female native speaker of French, different from the one used in Experiment 1, recorded all stimuli eight times in infant-directed speech.

The visual stimuli were similar to the ones used in Experiment 1 and consisted of six pairs of color pictures with black contours, one of a round and one of a spiky shape. Within pairs, the pictures had the same number of curves or spikes and they had the same color, but they were not matched for their surface. The colors were red, pink, yellow, green, light blue, and dark blue.

A.2.3.1.3 Procedure

Infants were held in the lap of a parent and tested in a quiet room. Their eye gazes were monitored and recorded on video. The parents wore headphones with masking music that prevented them from hearing the auditory stimuli.

Each infant was tested on 12 trials, six congruent and six incongruent ones. There were two counterbalancing groups, such that sounds that were presented in the congruent condition for one group were presented in the incongruent condition for the other group. For instance, /ʒuvo/ was paired with the pink round shape (congruent) and /daʒo/ with the yellow spiky shape (incongruent) for one group, while the reverse pairings were used for the other group.

Each trial started with the presentation of an attention getter, a colorful moving image, on a television screen. As soon as the infant had fixated it for 1 s., it was replaced by the visual stimulus, shown on a white background, followed after 300 ms by the presentation of the auditory stimulus. The trial ended when the infant had looked away for a consecutive period of more than 2 s. or when all 8 tokens of the item had been played three times. The ISI between tokens was 1 s., and the maximum trial duration around 40 s.

Trials were presented semi-randomly, with no more than three congruent or incongruent trials in a row and with the two shapes of the same color never being presented one after the other.

A.2.3.2 Results and discussion

Infants‟ eye gazes were coded off-line frame-by-frame. A 2x2 repeated measures ANOVA with the factors Shape Type (round vs. spiky) and Sound (bouba-type vs. kiki-type) revealed a main effect of Shape Type ($F(1, 23) = 9.16$, $p < .01$), indicating a preference for the round shapes. However, there was neither a main effect of Sound ($F < 1$) nor an interaction ($F(1, 23) = 1.0$, $p = .27$). Separate analyses of the first and the second half of the experiment revealed no significant main effects or interactions at all (all $F < 1$).

As in the previous experiment, we found no sign of a bouba- kiki effect. Moreover, despite the fact that infants no longer had to choose between a round and a spiky shape in each trial, we still obtained an overall preference for the round shapes, in that infants looked longer during trials with a round shape.

One possible reason for the absence of a bouba-kiki effect is that infants might have failed to make the association between the auditory and the visual stimuli. In the next experiment, we use the

same experimental design but make the visual stimuli move in synchrony with the auditory ones, thus facilitating the sound-shape association (Cogate & Bahrick, 1998).

A.2.4 Experiment 3

As in Experiment 2, we use a preferential looking paradigm with six-month-old infants.

A.2.4.1 Methods

A.2.4.1.1 Participants

Twenty-three six-month-old infants (range: 5;29–6;16; mean: 6.26 month-old; 10 girls) participated in the study. Seven more infants were tested but excluded from the analyses due to fussiness (N=3), experimenter error (N=2), or failure to look away on more than half of the trials (N=2).

A.2.4.1.2 Stimuli

The auditory and visual stimuli were the same as those in Experiment 2.

A.2.4.1.3 Procedure

The procedure was the same as that in Experiment 2, with one exception: The visual stimuli moved in synchrony with the auditory ones, decreasing and increasing in size (20% of size variation); they reached their maximum size during the second, stressed, syllable of the pseudoword, and remained immobile until just before the presentation of the next token.

A.2.4.2 Results and discussion

Infants' eye gazes were coded off-line frame-by-frame. In a 2x2 repeated measures ANOVA with the factors Shape Type (Round vs. Spiky) and Sound (Bouba-type vs. Kiki-type) neither the main effects nor the interaction were significant (Shape Type: $F(1, 23) = 2.16$, $p = .16$; Sound: $F < 1$; shape type

x sound: $F(1, 22) = 1.94$, $p = .18$). An ANOVA restricted to the first half of the experiment revealed no main effects or interaction either (all $F < 1$). For the second half, there was an effect of Sound $F(1, 22) = 5.17$, $p = .03$), indicating that infants looked longer at the shapes when kiki-type stimuli were displayed, but no effect of Shape Type ($F(1, 22) = 1.74$, $p = .20$) and no interaction ($F < 1$).

Thus, despite the fact that we facilitated the sound-shape association by making the shape move in synchrony with the auditory stimuli, we again failed to find a bouba-kiki effect.

A.2.5 General discussion

In three experiments, we failed to find a hint of a bouba-kiki effect in five- and six-month old infants. It is unlikely that methodological issues can explain this absence of evidence for sound-symbolic speech-shape associations in prelexical infants: We tested more than 20 infants in each experiment, using different paradigms that have been shown to be suited for testing infants' capacities to link auditory and visual stimuli at this age (Peña et al., 2011; M. P. Walker & Stickgold, 2010). We may also discard the possibility that infants did not discriminate between the round and the spiky shapes and therefore failed to show a bouba-kiki effect. Indeed, in two out of three experiments, infants looked longer at the round than at the spiky shapes, showing that they had no difficulty discriminating them.

Of course, our null results remain difficult to interpret, especially in light of the fact that a bouba-kiki effect was reported earlier in four-month-old infants (Oztürk et al., 2013). Recall that in that study, only one pair of stimuli was used, whereas in our experiments each infant was tested on at least six different pairs. We tentatively argue that the increased complexity of our design might have masked infants‟ emerging sound-symbolic matching abilities. In other words, the bouba-kiki effect in infants might be weak and detectable only with the simplest designs. The fact that Oztürk et al. (2013) failed to find the effect when the pair of sounds only differed by either its consonants or its vowels might be another sign of the lack of robustness of the effect in prelexical infants; specifically, they would only show it when all the phonemes in the auditory stimuli provide congruent sound-symbolic information.

Possibly, for the bouba-kiki effect to become more robust infants need to experience their own vocal

tract gestures when producing speech sounds. Thus, sound-symbolic associations should become stronger as infants gain experience with the production of speech sounds through babbling and as they recruit sensorimotor areas to decode spoken language. In line with this idea, infants' cross-modal binding has been shown to be initially broad and to become tuned with perceptual experience (Pons, Lewkowicz, Soto-Faraco, & Sebastián-Gallés, 2009; Lewkowicz & Ghazanfar, 2009; Imada et al., 2006). More specifically, speech motor areas in the left inferior frontal gyrus are activated by the perception of speech in 6.5 and 12- month-old infants but not in neonates, suggesting that experience producing speech sounds is required to perform sensorimotor bindings (Imada et al., 2006).

Two recent studies have examined related forms of audio- visual sound-symbolic associations in prelexical infants (Peña et al., 2011; M. P. Walker & Stickgold, 2010). One of them showed that three- to four-month-old infants look longer at a spiky shape when it is accompanied by a high-pitch rather than a low-pitch sound, and that conversely, they look longer at a round shape when it is accompanied by a low-pitch rather than a high-pitch sound (Peña et al., 2011). The other study reported that four-month-old infants prefer to look at a large version of a shape rather than a small one when they hear /o/ or /a/, while they prefer to look at a small version when they hear /i/ or /e/ (Cogate & Bahrick, 1998). Thus, like adults (P. Walker, 2012; Sapir, 1929), infants are able to perform systematic cross-modal sound-shape and sound-size mappings only a few months after birth.

Note that both of these studies used low-complexity auditory stimuli (pure tones and isolated vowels, respectively). It would be interesting to investigate the emergence of the bouba-kiki effect using isolated vowels. Prelexical infants might have less difficulty matching isolated vowels rather than entire CVCV sequences onto visual shapes, especially in light of their spontaneous matching of vowel sounds with silent videos of a talking face showing the corresponding articulatory gestures (Patterson & Werker, 2003; Bristow et al., 2009). In particular, we would expect them to match rounded vowels such as /u/ and /o/ onto round shapes and unrounded ones, such as /i/ and /e/, onto spiky shapes. Such findings would be in line with the idea that at least part of the correspondences between speech sounds and shapes is due to the mappings of the shape of the lips to produce a vowel onto the visual properties of a shape. More complex cross-modal mappings between consonants and visual shapes might emerge later. Of course, if this is the case, then we have to account for the qualitative shift that takes place during development, with consonants at some point in time

becoming more important than vowels for the bouba-kiki effect (Oztürk et al., 2013; Nielsen & Rendall, 2011; Fort et al., 2015). The preponderance of the role of consonants in the bouba-kiki effect in adults meshes well with studies showing that adults rely more on consonants than on vowels for the purposes of lexical processing (Cutler et al., 2000; Toro, Nespor, et al., 2008). Interestingly, 12-month-old infants likewise rely more on consonants than on vowels when distinguishing among words (Hochmann, Benavides-Varela, Nespor, & Mehler, 2011). It would thus be particularly interesting to compare younger and older infants in a simple design as the one used by Oztürk et al. (2013), to see whether for the purposes of sound-shape matching they initially pay more attention to vowels and come to rely more on consonants by the time they start to learn words.

To conclude, infants' sensitivity to sound symbolism has received little attention in research on the ontological development of multimodal speech perception. The evidence for a bouba-kiki effect in prelexical infants so far is weak, and we argue that null results like the present ones should not be kept in a drawer. More research is necessary to investigate the role of experience in the emergence of the bouba-kiki effect, as well as in the differential role that consonants and vowels may play in infants' cross-modal correspondences.

References

Archangeli, D. & Pulleyblank, D. (1994). *Grounded Phonology*. Cambridge, Massachusetts: MIT Press.

Baddeley, A. & Lewis, V. J. (1981). Inner active processes in reading: the inner voice, the inner ear and the inner eye. In A. M. Lesgold & C. A. Perfetti (Eds.), *Interactive processes in reading* (pp. 107–129). Hillsdale, NJ: Lawrence Erlbaum.

Baer-Henney, D., Kügler, F., & van de Vijver, R. (2014). The interaction of language-specific and universal factors during the acquisition of morphophonemic alternations with exceptions. *Cognitive Science*, 1–33.

Bailey, T. M. & Hahn, U. (2005, April). Phoneme similarity and confusability. *Journal of Memory and Language*, *52*(3), 339–362.

Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4.

Batterink, L. J., Oudiette, D., Reber, P. J., & Paller, K. A. (2014, October). Sleep facilitates learning a new linguistic rule. *Neuropsychologia*, *65*, 169–179.

Beddor, P. S., Harnsberger, J. D., & Lindemann, S. (2002). Language-specific patterns of vowel-to-vowel coarticulation: acoustic structures and their perceptual correlates. *Journal of Phonetics*, *30*, 591–627.

Beddor, P. S., Krakow, R. A., & Lindemann, S. (2001). Patters of perceptual compensation and their phonological consequences. In E. Hume & K. Johnson (Eds.), *The role of speech perception in phonology* (pp. 55–78). London: Academic Press.

Bell, T., Dirks, D. D., & Carterette, E. C. (1989). Interactive factors in consonant confusion patterns. *The Journal of the Acoustical Society of America*, *85*(1), 339–346.

Blevins, J. (2004). *Evolutionary Phonology: the Emergence of Sound Patterns*. Cambridge University Press.

Blevins, J. (2006). A theoretical synopsis of Evolutionary Phonology. *Theoretical Linguistics*, *32*(2), 117–166.

Bloomfield, L. (1933). *Language*. New York: Henry Holt.

Bonatti, L. L., Peña, M., Nespor, M., & Mehler, J. (2005). Linguistic Constraints on Statistical Computations. *Psychological Science*, *16*(6), 451–460.

Bond, B. & Stevens, S. (1969). Cross-modality matching of brightness to loudness by 5-year-olds. *Perception & Psychophysics*, *6*, 337–339.

Bremner, A. J., Caparos, S., Davidoff, J., de Fockert, J., Linnell, K. J., & Spence, C. (2013). "Bouba" and "Kiki" in Namibia? A remote culture make similar shape-sound matches, but different shape-taste matches to Westerners. *Cognition*, *126*(2), 165–172.

Bristow, D., Dehaene-Lambertz, G., Mattout, J., Soares, C., Gliga, T., Baillet, S., & Mangin, J. F. (2009). Hearing faces: how the infant brain matches the face it sees with the speech it hears. *Journal of Cognitive Neuroscience*, *21*(5), 905–921.

Butcher, A. (2006). Australian Aboriginal Languages: Consonant-Salient Phonologies and the "Place-of-Articulation Imperative". In J. Harrington & M. Tabain (Eds.), *Speech production: models, phonetic processes, and techniques* (pp. 187–210). New York: Psychology Press.

Butcher, A. (2012). On the phonetics of long, thin phonologies. In C. Donohue, S. Ishihara, & W. Steed (Eds.), *Quantitative approaches to problems in linguistics: studies in honour of phil rose* (pp. 133–154).

Caramazza, A., Chialant, D., Capasso, R., & Miceli, G. (2000). Separable processing of consonants and vowels. *Nature*, *403*(6768), 428–430.

Chastaing, M. (1958). Le symbolisme des voyelles: Signification des "i". *Journal de Psychologie*, *55*, 461–481.

Cogate, L. J. & Bahrick, L. E. (1998). Intersensory Redundancy Facilitates Learning of Arbitrary Relations between Vowel Sounds and Objects in Seven-Month-Old Infants. *Journal of Experimental Child Psychology*, *69*, 133–149.

Cole, R. A., Jakimik, J., & Cooper, W. E. (1978, July). Perceptibility of phonetic features in fluent speech. *The Journal of the Acoustical Society of America*, *64*(1), 44–56.

Connine, C., Blasko, D., & Titone, D. (1993). Do the beginnings of spoken words have a special status in auditory word recognition? *Journal of Memory and Language*, *32*, 193–210.

Côté, M.-H. (2011). The role of the syllable in the organization and realization of sound systems. In A. Cohn, C. Fougeron, & M. Huffman (Eds.), *The oxford handbook of laboratory phonology* (pp. 232–242). Oxford: Oxford University Press.

Culbertson, J., Smolensky, P., & Legendre, G. (2012). Learning biases predict a word order universal. *Cognition*, *122*(3), 306–329.

Cutler, A., Sebastián-Gallés, N., Soler-vilageliu, O., & Ooijen, B. v. (2000). Constraints of vowels and consonants on lexical selection: Cross-linguistic comparisons. *Memory & Cognition*, *28*(5), 746–755.

Cutler, A., Weber, A., Smits, R., & Cooper, N. (2004). Patterns of English phoneme confusions by native and non-native listeners. *The Journal of the Acoustical Society of America*, *116*(6), 3668–3678.

Dautriche, I., Swingley, D., & Christophe, A. (2015). Learning novel phonological neighbors: Syntactic category matters. *Cognition*, *143*, 77–86.

Davis, M. H., Di Betta, A. M., Macdonald, M. J., & Gaskell, M. G. (2009). Learning and consolidation of novel spoken words. *Journal of Cognitive Neuroscience*, *21*(4), 803–820.

Davis, M. H. & Johnsrude, I. S. (2007). Hearing speech sounds: Top-down influences on the interface between audition and speech perception. *Hearing Research*, *229*(1-2), 132–147.

de Saussure, F. D. (1959). *Course in general linguistics*. New York: Philosophical Library.

Dell, F. (1995, March). Consonant clusters and phonological syllables in French. *Lingua*, *95*(1-3), 5–26.

Donegan, P. J. & Stampe, D. (1979). The Study of Natural Phonology. In D. A. Dinnsen (Ed.), *Current approaches to phonological theory* (pp. 126–173). Bloomington & London: Indiana University Press.

Dumay, N. & Gaskell, M. G. (2007, January). Sleep-associated changes in the mental representation of spoken words. *Psychological science*, *18*(1), 35–9.

Earle, F. S. & Myers, E. B. (2014, October). Building phonetic categories: an argument for the role of sleep. *Frontiers in Psychology*, *5*(October), 1–12.

Ellenbogen, J. M., Hu, P. T., Payne, J. D., Titone, D., & Walker, M. P. (2007). Human relational memory requires time and sleep. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(18), 7723–8.

Ernestus, M. & Mak, W. M. (2004). Distinctive phonological features differ in relevance for both spoken and written word recognition. *Brain and language*, *90*(1-3), 378–392.

Everett, C. (2013). Evidence for Direct Geographic Influences on Linguistic Sounds: The Case of Ejectives. *PLoS ONE*, *8*(6).

Farmer, T. A., Christiansen, M. H., & Monaghan, P. (2006, August). Phonological typicality influences on-line sentence comprehension. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(32), 12203–12208.

Farmer, T. A., Monaghan, P., Misyak, J. B., & Christiansen, M. H. (2011). Phonological typicality influences sentence processing in predictive contexts: Reply to Staub, Grant, Clifton, and Rayner (2009). *Journal of experimental psychology. Learning, memory, and cognition*, *37*(5), 1318–1325.

Fenn, K. M., Margoliash, D., & Nusbaum, H. C. (2013, September). Sleep restores loss of generalized but not rote learning of synthetic speech. *Cognition*, *128*(3), 280–6.

Fenn, K. M., Nusbaum, H. C., & Margoliash, D. (2003). Consolidation during sleep of perceptual learning of spoken language. *Nature*, *425*(October), 614–616.

Finley, S. (2012, December). Typological asymmetries in round vowel harmony: Support from artificial grammar learning. *Language and Cognitive Processes*, *27*(10), 1550–1562.

Finley, S. & Badecker, W. (2008). Analytic biases for vowel harmony languages. In N. Abner & J. Bishop (Eds.), *Proceedings of the 27th west coast conference on formal linguistics* (pp. 168–176). Somerville, MA: Cascadilla Proceedings Project.

Finley, S. & Badecker, W. (2009). Artificial language learning and feature-based generalization. *Journal of Memory and Language*, *61*(3), 423–437.

Fort, M., Martin, A., & Peperkamp, S. (2015). Consonants are More Important than Vowels in the Bouba-kiki Effect. *Language and Speech*, *58*(2), 247–266.

Fort, M., Weiss, A., Martin, A., & Peperkamp, S. (2013). Looking for the bouba-kiki effect in prelexical infants. In S. Ouni, F. Berthommier, & A. Jesse (Eds.), *Proceedings of the international conference on auditory-visual speech processing* (pp. 71–76). Annecy, France.

Frisch, S. A., Large, N. R., & Pisoni, D. B. (2000). Perception of wordlikeness: Effects of segment probability and length on the processing of nonwords. *Journal of Memory and Language*, *42*(4), 481–496.

Fromkin, V. (1971). The non-anomalous nature of anomalous utterances. *Language*, *47*(1), 27–52.

Gallace, A. & Spence, C. (2006). Multisensory synesthetic interactions in the speeded classification of visual size. *Perception & Psychophysics*, *68*(7), 1191–1203.

Gaskell, M. G., Warker, J. A., Lindsay, S., Frost, R., Guest, J., Snowdon, R., & Stackhouse, A. (2014). Sleep Underpins the Plasticity of Language Production. *Psychological science*, *25*, 1457–1465.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "L" and "R". *Neuropsychologia*, *9*(3), 317–323.

Green, P. J. & Silverman, B. W. (1993). *Nonparametric regression and generalized linear models: a roughness penalty approach*. Chapman & Hall.

Griffiths, T. L. & Kalish, M. L. (2007). Language evolution by iterated learning with bayesian agents. *Cognitive science*, *31*(3), 441–480.

Grimes, S. M. (2006). On the creation of a pronunciation dictionary for Hungarian. In *Proceedings of the 6th midwest computational linguistics colloquium*. Indiana University.

Hahn, U. & Bailey, T. M. (2005). What makes words sound similar? *Cognition*, *97*(3), 227–267.

Hall, K. C. (2009). *A Probabilistic Model of Phonological Relationships from Contrast to Allophony* (Doctoral dissertation, The Ohio State University).

Hall, K. C. (2013). A typology of intermediate phonological relationships. *The Linguistic Review*, *30*(2), 215–276.

Hall, K. C., Allen, B., Fry, M., Mackie, S., & McAuliffe, M. (2015a). Calculating functional load with pronunciation variants. Stuttgart, Germany.

Hall, K. C., Allen, B., Fry, M., Mackie, S., & McAuliffe, M. (2015b). Phonological CorpusTools.

Hall, K. C. & Hume, E. V. (2013). Perceptual confusability of French vowels. In *Proceedings of meetings on acoustics* (Vol. 19). Montreal, Canada.

Hanson, B. A. (1991). *Method of Moments Estimates for the Four-Parameter Beta Compound Binomial Model and the Calculation of Classification Consistency Indexes.*

Hansson, G. Ó. (2008). Diachronic Explanations of Sound Patterns. *Language and Linguistics Compass*, *2*(5), 859–893.

Harrison, K. D., Dras, M., & Kapicioglu, B. (2002). Agent-Based Modeling of the Evolution of Vowel Harmony. In *Proceedings of nels* (Vol. 32, pp. 217–236).

Hay, J., Pierrehumbert, J. B., & Beckman, M. (2004). Speech perception, well-formedness, and the statistics of the lexicon. In *Papers in laboratory phonology vi* (pp. 58–74). Cambridge, UK: Cambridge University Press.

Hayes, B. P. & Steriade, D. (2004). The phonetic basis of phonological Markedness. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically based phonology* (pp. 1–33). Cambridge: Cambridge University Press.

Hayes, B., Kirchner, R., & Steriade, D. (Eds.). (2004). *Phonetically Based Phonology*. West Nyack: Cambridge University Press.

Heller, J. R. & Goldrick, M. (2014). Grammatical constraints on phonological encoding in speech production. *Psychonomic Bulletin & Review*, *21*(6), 1576–1582.

Hochmann, J.-R., Benavides-Varela, S., Nespor, M., & Mehler, J. (2011). Consonants and vowels: different roles in early language acquisition. *Developmental Science*, *14*(6), 1445–1458.

Hockett, C. (1955). A Manuel of Phonology. *International Journal of American Linguistics*, *21*(4).

Hockett, C. (1967). *The Quantification of Functional Load: A Linguistic Problem.*

Hooper, J. B. (1976). *An Introduction to Natural Generative Phonology*. New York: Academic Press.

Hume, E. V. & Mailhot, F. (2013). The role of entropy and surprisal in phonologization and language change. In A. Yu (Ed.), *Origins of sound change: approaches to phonologization* (pp. 29–50).

Imada, T., Zhang, Y., Cheour, M., Taulu, S., Ahonen, A., & Kuhl, P. K. (2006). Infant speech perception activates Broca's area: a developmental magnetoencephalography study. *Neuroreport*, *17*(10), 957–962.

Imai, M., Kita, S., Nagumo, M., & Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition*, *109*(1), 54–65.

Jadva, V., Hines, M., & Golombok, S. (2010). Infants' preferences for toys, colors, and shapes: sex differences and similarities. *Archives of Sexual Behavior*, *39*(6), 1261–1273.

Johnson, K. & Babel, M. E. (2010). On the perceptual basis of distinctive features: Evidence from the perception of fricatives by Dutch and English speakers. *Journal of Phonetics*, *38*(1), 127–136.

Kantartzis, K., Imai, M., & Kita, S. (2011). Japanese sound-symbolism facilitates word learning in English- Speaking children. *Cognitive science*, *35*, 575–586.

Keidel, J. L., Jenison, R. L., Kluender, K. R., & Seidenberg, M. S. (2007). Commentary on Bonatti, Peña, Nespor, and Mehler's "Does Grammar Constrain Statistical Learning?" *Psychological Science*, *18*(10), 922–923.

Kirby, J. P. & Sonderegger, M. (2015). *Bias and population structure in the actuation of sound change.*

Kirby, S. (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, S., Dowman, M., & Griffiths, T. L. (2007). Innateness and culture in the evolution of language. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(12), 5241–5245.

Köhler, W. (1929). *Gestalt Psychology*. New York: Liveright.

Köhler, W. (1947). *Gestalt psychology: An introduction to new concepts in modern psychology*. New York: Liveright.

Kovic, V., Plunkett, K., & Westermann, G. (2010). The shape of words in the brain. *Cognition*, *114*(1), 19–28.

Krämer, M. (1999). A Correspondence Approach to Vowel Harmony and Disharmony.

Lewkowicz, D. J. & Ghazanfar, A. A. (2009). The emergence of multisensory systems through perceptual narrowing. *Trends in cognitive sciences*, *13*(11), 470–478.

Lewkowicz, D. J. & Turkewitz, G. (1980). Cross-modal equivalence in early infancy: Auditory-visual intensity matching. *Developmental Psychology*, *16*, 597–607.

Luce, P. A. (1986). *Neighborhoods of Words in the Mental Lexicon* (Doctoral dissertation, Indiana University).

Ludwig, V. U., Adachi, I., & Matsuzawa, T. (2011). Visuoauditory mappings between high luminance and high pitch are shared by chimpanzees (Pan troglodytes) and humans. *Proceedings of the National Academy of Sciences*, *108*(51), 20661–20665.

Luke, S. G. (2016). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, (2000), 1–9.

Mailhot, F. (2013). Modeling the emergence of vowel harmony through iterated learning. In A. Yu (Ed.), *Origins of sound change: approaches to phonologization* (pp. 247–261). Oxford: Oxford University Press.

Marks, L. E. (1987). On cross-modal similarity: Auditory–visual interactions in speeded discrimination. *Journal of Experimental Psychology: Human Perception and Performance*, *13*(384-394).

Marks, L. E. (2004). Cross-modal interactions in speeded classification. In I. G. A. Calvert, C. Spence, & B. E. Stein (Eds.), *Handbook of multisensory processes* (pp. 85–105). Cambridge, Massachusetts: MIT Press.

Martin, A. & Peperkamp, S. (2015). Asymmetries in the exploitation of phonetic features for word recognition. *The Journal of the Acoustical Society of America*, *137*(4), EL307–EL313.

Martinet, A. (1955). *Économie des changements phonétiques* (Francke). Bern.

Maurer, D., Pathman, T., & Mondloch, C. J. (2006). The shape of boubas: Sound-shape correspondences in toddlers and adults. *Developmental Science*, *9*(3), 316–322.

Mehler, J., Peña, M., Nespor, M., & Bonatti, L. L. (2006). The "soul" of language does not use statistics: Reflections on vowels and consonants. *Cortex*, *42*, 846–854.

Milberg, W., Blumstein, S. E., & Dworetzky, B. (1988). Phonological factors in lexical access: Evidence from an auditory lexical decision task. *Bulletin of the Psychonomic Society*, *26*(4), 305–308.

Miller, G. & Nicely, P. (1955). An analysis of perceptual confusions among some English consonants. *The Journal of the Acoustical Society of America*, *27*(2), 338–352.

Monaghan, P., Christiansen, M. H., & Fitneva, S. A. (2011). The arbitrariness of the sign: learning advantages from the structure of the vocabulary. *Journal of Experimental Psychology: General*, *140*(3), 325–347.

Monaghan, P., Mattock, K., & Walker, P. (2012). The role of sound symbolism in language learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(5), 1152–1164.

Mondloch, C. J. & Maurer, D. (2004). Do small white balls squeak? Pitch–object correspondences in young children. *Cognitive, Effective, and Behavioral Neuroscience*, *4*(2), 133–136.

Moreton, E. (2008). Analytic bias and phonological typology. *Phonology*, *25*(1), 83–127.

Moreton, E. & Pater, J. (2012a). Structure and Substance in Artificial-Phonology Learning , Part I: Structure. *Language and Linguistics Compass*, *6*(11), 686–701.

Moreton, E. & Pater, J. (2012b). Structure and Substance in Artificial-Phonology Learning, Part II: Substance. *Language and Linguistics Compass*, *6*(11), 702–718.

Nespor, M., Peña, M., & Mehler, J. (2003). On the different roles of vowels and consonants in speech processing and language acquisition. *Lingue e Linguaggio*, *2*, 203–229.

New, B. & Nazzi, T. (2014). The time course of consonant and vowel processing during word recognition. *Language and Cognitive Processes*, *29*(2), 147–157.

New, B., Pallier, C., Ferrand, L., & Matos, R. (2001). Une base de données lexicales du français contemportain sur internet: LEXIQUE. *L'année psychologique*, *101*, 447–462.

Nielsen, A. K. & Rendall, D. (2011). The sound of round: Evaluating the sound-symbolic role of consonants in the classic Takete-Maluma phenomenon. *Canadian Journal of Experimental Psychology*, *65*(2), 115–124.

Nishida, M. & Walker, M. P. (2007). Daytime naps, motor memory consolidation and regionally specific sleep spindles. *PLoS ONE*, *2*(4).

Niyogi, P. & Berwick, R. C. (2009). The proper treatment of language acquisition and change in a population setting. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(25), 10124–10129.

Nygaard, L. C., Cook, A. E., & Namy, L. L. (2009). Sound to meaning correspondences facilitate word learning. *Cognition*, *112*(1), 181–186.

Oh, Y. M., Coupé, C., Marsico, E., & Pellegrino, F. (2015). Bridging phonological system and lexicon: Insights from a corpus study of functional load. *Journal of Phonetics*, *53*, 153–176.

Ohala, J. J. (1993a). The phonetics of sound change. In C. Jones (Ed.), *Historical linguistics: problems and perspectives* (pp. 237–278). London: Longman.

Ohala, J. J. (1993b). The phonetics of sound change. In C. Jones (Ed.), *Historical linguistics: problems and perspectives* (pp. 235–278). London: Longman.

Ohala, J. J. (1994). Towards a universal, phonetically-based, theory of vowel harmony. In *Proceedings of the 3rd interational conference on spoken language processing* (pp. 491–494). Yokohama, Japan.

CMU Pronouncing Dictionary. (2008). Carnegie Mellon University.

Oztürk, O., Krehm, M., & Vouloumanos, A. (2013). Sound symbolism in infancy: Evidence for sound-shape cross-modal correspondences in 4-month-olds. *Journal of Experimental Child Psychology*, *114*(2), 173–186.

Parise, C. V. & Pavani, F. (2011). Evidence of sound symbolism in simple vocalizations. *Experimental Brain Research*, *214*(3), 373–380.

Parise, C. V. & Spence, C. (2012). Audiovisual crossmodal correspondences and sound symbolism: A study using the implicit association test. *Experimental Brain Research*, *220*(3-4), 319–333.

Patterson, M. L. & Werker, J. F. (2003). Two-month-old infants match phonetic information in lips and voice. *Developmental Science*, *6*, 191–196.

Peña, M., Mehler, J., & Nespor, M. (2011). The role of audiovisual processing in early conceptual development. *Psychological Science*, *22*(11), 1419–1421.

Peperkamp, S., Skoruppa, K., & Dupoux, E. (2006). The role of phonetic naturalness in phonological rule acquisition. In D. Bamman, T. Magnitsakaia, & C. Zaller (Eds.), *Proceedings of the 30th annual boston university conference on language development* (pp. 464–475). Somerville, MA: Cascadilla Press.

Plauché, M. C., Delogu, C., & Ohala, J. J. (1997). Asymmetries in consonant confusion. In *Proceedings of the 5th european conference on speech communication and technology* (pp. 2187–2190).

Pons, F., Lewkowicz, D. J., Soto-Faraco, S., & Sebastián-Gallés, N. (2009). Narrowing of intersensory speech perception in infancy. *Proceedings of the National Academy of Sciences*, *106*(26), 10598–10602.

Pulvermüller, F., Huss, M., Kherif, F., Moscoso del Prado Martin, F., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(20), 7865–7870.

Pycha, A., Nowak, P., Shin, E., & Shosted, R. (2003). Phonological rule-learning and its implications for a theory of vowel harmony. In M. Tsujimura & G. Garding (Eds.), *Wccfl 22 proceedings* (Vol. 22, pp. 101–114). Somerville, MA.

Rabiner, L. & Juang, B.-H. (1993). *Fundamentals of Speech Recognition*. Prentice-Hall, Inc.

Rafferty, A. N., Griffiths, T. L., & Ettlinger, M. (2013, October). Greater learnability is not sufficient to produce cultural universals. *Cognition*, *129*(1), 70–87.

Ramachandran, V. S. & Hubbard, E. M. (2001). Synaesthesia – A window into perception, thought and lan- guage. *Journal of Consciousness Studies*, *8*(12), 3–34.

Reali, F. & Griffiths, T. L. (2009). The evolution of frequency distributions: Relating regularization to inductive biases through iterated learning. *Cognition*, *111*(3), 317–328.

Rose, S. & Walker, R. (2011). Harmony Systems. In J. Goldsmith, J. Riggle, & A. Yu (Eds.), *The handbook of phonological theory* (Second Edi, pp. 240–290). Blackwell Publishing.

Saffran, J. R. & Thiessen, E. D. (2003). Pattern induction by infant language learners. *Developmental Psychology*, *39*(3), 484–494.

Samuel, A. G. & Lieblich, J. (2014, August). Visual speech acts differently than lexical context in supporting speech perception. *Journal of experimental psychology. Human perception and performance*, *40*(4), 1479–90.

Sapir, E. (1929). A study in phonetic symbolism. *Journal of Experimental Psychology: Human Perception and Performance*, *12*, 225–239.

Schane, S. a., Tranel, B., & Lane, H. (1974, January). On the psychological reality of a natural rule of syllable structure. *Cognition*, *3*(4), 351–358.

Schatz, T. (2016). *ABX-Discriminability Measures and Applications* (Doctoral dissertation, École Normale Supérieure).

Schatz, T., Peddinti, V., Bach, F., Jansen, A., Hermansky, H., & Dupoux, E. (2013). Evaluating speech features with the Minimal-Pair ABX task: Analysis of the classical MFC/PLP pipeline. In *Proceedings of interspeech*.

Severen, L. v., Gillis, J. J. M., Molemans, I., Berg, R. v. d., De Maeyer, S., & Gillis, S. (2013, September). The relation between order of acquisition, segmental frequency and function: the case of word-initial consonants in Dutch. *Journal of Child Language*, *40*(4), 703–740.

Shannon, C. E. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal*, *27*, 379–423.

Skoruppa, K., Lambrechts, A., & Peperkamp, S. (2011). The Role of Phonetic Distance in the Acquisition of Phonological Alternations. In S. Lima, K. Mullin, & B. Smith (Eds.), *Proceedings of the 39th annual meeting of the north east linguistic society* (Vol. 2, *2008*, pp. 717–729). Amherst, MA.

Skoruppa, K. & Peperkamp, S. (2011, March). Adaptation to novel accents: feature-based learning of context-sensitive phonological regularities. *Cognitive science*, *35*(2), 348–366.

Smith, K. & Wonnacott, E. (2010). Eliminating unpredictable variation through iterated learning. *Cognition*, *116*(3), 444–449.

Spector, F. & Maurer, D. (2013). Early sound symbolism for vowel sounds. *i-Perception*, *4*(4), 239–241.

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention Perception & Psychophysics*, *73*(4), 971–995.

Steele, A., Denby, T., Chan, C., & Goldrick, M. (2015). Learning Non-Native Phonotactic Constraints Over the Web. In *Proceedings of the 18th international conference of the phonetic sciences* (pp. 1–5).

Stemberger, J. P. (1991). Apparent Anti-frequency Effects in Language Production : The Addition Bias and Phonological Underspecification. *Journal of Memory and Language*, *30*, 161–185.

Stevenson, S. (2015). *The Strength of Segmental Contrasts: A Study on Laurentian French* (Doctoral dissertation, University of Ottawa).

Stokes, S. F., Klee, T., Carson, C. P., & Carson, D. (2005). A phonemic implicational feature hierarchy of phonological contrasts for English-speaking children. *Journal of Speech, Language and Hearing Research*, *48*(4), 817.

Strand, J., Simenstad, A., Cooperman, A., & Rowe, J. (2014). Grammatical context constrains lexical competition in spoken word recognition. *Memory & Cognition*, *42*(4), 676–87.

Studdert-Kennedy, M., Shankweiler, D., & Pisoni, D. B. (1972). Auditory and Phonetic Proceses in Speech Perception: Evidencefrom a Dichotic Study. *Cognitive Psychology*, *3*(3), 455–466.

Sumby, W. H. & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The Journal of the Acoustical Society of America*, *26*(2), 212–215.

Surendran, D. & Niyogi, P. (2003). *Measuring the Usefulness (Functional Load) of Phonological Contrasts*. Department of Comuter Science, University of Chicago.

Surendran, D. & Niyogi, P. (2006). Quanitifying the Functional Load of Phonemic Oppositions, Distinctive Features, and Suprasegmentals. In O. Nedergaard Thomsen (Ed.), *Competing models of linguistic change: evolution and beyond* (pp. 49–64).

Sweeny, T. D., Guzman-Martinez, E., Ortega, L., Grabowecky, M., & Suzuki, S. (2012). Sounds exaggerate visual shape. *Cognition*, *124*(2), 194–200.

Tamminen, J., Payne, J. D., Stickgold, R., Wamsley, E. J., & Gaskell, M. G. (2010). Sleep spindle activity is associated with the integration of new memories and existing knowledge. *The Journal of Neuroscience*, *30*(43), 14356–14360.

Tarte, R. D. (1974). Phonetic symbolism in adult native speakers of Czech. *Language and Speech*, *17*, 87–94.

Tarte, R. D. (1982). The relationship between monosyllables and pure tones: An investigation of phonetic symbolism. *Journal of Verbal Learning and Verbal Behavior*, *21*, 352–360.

Toro, J. M., Nespor, M., Mehler, J., & Bonatti, L. L. (2008). Finding words and rules in a speech stream: Functional differences between vowels and consonants. *Psychological Science*, *19*(2), 137–144.

Toro, J. M., Shukla, M., Nespor, M., & Endress, A. D. (2008). The quest for generalizations over consonants: Asymmetries between consonants and vowels are not the by-product of acoustic differences. *Perception & Psychophysics*, *70*(8), 1515–1525.

van der Hulst, H. & van de Weijer, J. (1996). Vowel Harmony. In J. A. Goldsmith (Ed.), *The handbook of phonological theory* (pp. 495–534). Blackwell Publishing.

Versteegh, M. (2015). Spectral.

Wagner, S., Winner, E., Cicchetti, D., & Gardner, H. (1981). "Metaphorical" mapping in human infants. *Child development*, *52*, 728–731.

Walker, M. P. & Stickgold, R. (2004, September). Sleep-dependent learning and memory consolidation. *Neuron*, *44*(1), 121–33.

Walker, M. P. & Stickgold, R. (2010). Overnight alchemy: sleep-dependent memory evolution. *Nature Reviews Neuroscience*, *11*(3), c1–c2.

Walker, M. P., Stickgold, R., Alsop, D., Gaab, N., & Schlaug, G. (2005). Sleep-dependent motor memory plasticity in the human brain. *Neuroscience*, *133*(4), 911–917.

Walker, P. (2012). Cross-sensory correspondences and cross talk between dimensions of connotative meaning: Visual angularity is hard, high-pitched, and bright. *Attention Perception & Psychophysics*, *74*(8), 1792–1809.

Wang, M. D. & Bilger, R. C. (1973). Consonant confusions in noise: a study of perceptual features. *The Journal of the Acoustical Society of America*, *54*(5), 1248–1266.

Weber, A. & Smits, R. (2003). Consonant And Vowel Confusion Patterns By American English Listeners. In M. J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th international congress of phonetic sciences* (pp. 1427–1440).

Wedel, A. (2015). Biased variation shapes sound system change: Integrating data from modeling, experiments and corpora. Stuttgart, Germany.

Wedel, A., Kaplan, A., & Jackson, S. R. (2013, August). High functional load inhibits phonological contrast loss: a corpus study. *Cognition*, *128*(2), 179–186.

Westbury, C. (2005). Implicit sound symbolism in lexical access: Evidence from an interference task. *Brain and Language*, *93*(1), 10–19.

Wickelgren, W. (1969). Auditory or Articulatory Coding in Verbal Short-term Memory. *Psychological Review*, *76*(2), 232–235.

Wiener, S. & Turnbull, R. (2015). Constraints of Tones, Vowels and Consonants on Lexical Selection in Mandarin Chinese. *Language and Speech*.

Wilson, C. (2006). Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science*, *30*(5), 945–982.

Wilson, S. M. & Iacoboni, M. (2006). Neural responses to non-native phonemes varying in producibility: Evidence for the sensorimotor nature of speech perception. *NeuroImage*, *33*(1), 316–325.

Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature neuroscience*, *7*(7), 701–2.

Yu, A. (2011). On measuring phonetic precursor robustness: a response to Moreton. *Phonology*, *28*(3), 491–518.

Résumé

Pendant la perception de la parole, les locuteurs sont biaisés par un grand nombre de facteurs. Par exemple, il existe des limitations cognitives comme la mémoire ou l'attention, mais aussi des limitations linguistiques comme leur langue maternelle. Cette thèse se concentre sur deux de ces facteurs : les biais de traitement pendant la reconnaissance des mots, et les biais d'apprentissage pendant le processus de transmission. Ces facteurs peuvent se combiner et, au cours du temps, influencer l'évolution des langues.

Dans la première partie de cette thèse, nous nous concentrons sur le processus de la reconnaissance des mots. Des recherches antérieures ont établi l'importance des traits phonologiques (p. ex. le voisement ou le lieu d'articulation) pendant le traitement de la parole. Cependant, nous en savons peu sur leur poids relatif les uns par rapport aux autres, et comment cela peut influencer la capacité des locuteurs à reconnaître les mots. Nous avons testé des locuteurs français sur leur capacité à reconnaître des mots mal prononcés et avons trouvé que les traits de mode et de lieu sont plus importants que le trait de voisement. Nous avons ensuite considéré deux sources de cette asymétrie et avons trouvé que les locuteurs sont biaisés et par la perception acoustique ascendante (les contrastes de mode sont plus facile à percevoir à cause de leur distance acoustique importante) et par la connaissance lexicale descendante (le trait de lieu est plus exploité dans le lexique français que les autres traits). Nous suggérons que ces deux sources de biais se combinent pour influencer les locuteurs lors de la reconnaissance des mots.

Dans la seconde partie de cette thèse, nous nous concentrons sur la question d'un biais d'apprentissage. Il a été suggéré que les apprenants peuvent être biaisés vers l'apprentissage de certains patrons phonologiques grâce à leurs connaissances phonétiques. Cela peut alors expliquer pourquoi certains patrons sont récurrents dans la typologie, tandis que d'autres restent rares ou non-attestés. Plus spécifiquement, nous avons exploré le rôle d'un biais d'apprentissage sur l'acquisition de la règle typologiquement commune de l'harmonie vocalique comparée à celle de la règle non-attestée (mais logiquement équivalente) de la disharmonie vocalique. Nous avons trouvé des preuves d'un biais d'apprentissage aussi bien en perception qu'en production. En utilisant un modèle d'apprentissage itéré simulé, nous avons ensuite montré comment un biais, même petit, favorisant l'un des patrons, peut influencer la typologie linguistique au cours du temps et donc expliquer (en partie) la prépondérance de systèmes harmoniques. De plus, nous avons exploré le rôle du sommeil sur la consolidation mnésique. Nous avons montré que seul le patron commun bénéficie d'une consolidation et que cela est un facteur supplémentaire pouvant contribuer à l'asymétrie typologique.

Dans l'ensemble, cette thèse considère certaines des sources de biais possibles chez l'individu et discute de comment ces influences peuvent, au cours du temps, faire évoluer les systèmes linguistiques. Nous avons démontré la nature dynamique et complexe du traitement de la parole, à la fois en perception et dans l'apprentissage. De futurs travaux devront explorer plus en détail comment ces différentes sources de biais sont pondérées les unes relativement aux autres.

Abstract

During speech perception, listeners are biased by a great number of factors, including cognitive limitations such as memory and attention and linguistic limitations such as their native language. This thesis focuses on two of these factors: processing bias during word recognition, and learning bias during the transmission process. These factors are combinatorial and can, over time, affect the way languages evolve.

In the first part of this thesis, we focus on the process of word recognition. Previous research has established the importance of phonological features (e.g., voicing or place of articulation) during speech processing, but little is known about their weight relative to one another, and how this influences listeners' ability to recognize words. We tested French participants on their ability to recognize mispronounced words and found that the manner and place features were more important than the voicing feature. We then explored two sources of this asymmetry and found that listeners were biased both by bottom-up acoustic perception (manner contrasts are easier to perceive because of their acoustic distance compared to the other features) and top-down lexical knowledge (the place feature is used more in the French lexicon than the other two features). We suggest that these two sources of bias coalesce during the word recognition process to influence listeners.

In the second part of this thesis, we turn to the question of bias during the learning process. It has been suggested that language learners may be biased towards the learning of certain phonological patterns because of phonetic knowledge they have. This in turn can explain why certain patterns are recurrent in the typology while others remain rare or unattested. Specifically, we explored the role of learning bias on the acquisition of the typologically common rule of vowel harmony compared to the unattested (but logically equivalent) rule of vowel disharmony. We found that in both perception and production, there was evidence of a learning bias, and using a simulated iterated learning model, showed how even a small bias favoring one pattern over the other could influence the linguistic typology over time, thus explaining (in part) the prevalence of harmonic systems. We additionally explored the role of sleep on memory consolidation and showed evidence that the common pattern benefits from consolidation that the unattested pattern does not, a factor that may also contribute to the typological asymmetry.

Overall, this thesis considers a few of the wide-ranging sources of bias in the individual and discusses how these influences can over time shape linguistic systems. We demonstrate the dynamic and complicated nature of speech processing (both in perception and learning) and open the door for future research to explore in finer detail just how these different sources of bias are weighted relative to one another.