FUNCTIONALISM, LEXICAL CONTRAST AND SOUND CHANGE

Andrea Ceolin

A DISSERTATION

in

Linguistics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy
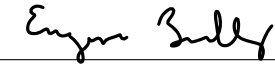
2020

Co-Supervisor of Dissertation

_____

Donald A. Ringe, Professor of Linguistics

Co-Supervisor of Dissertation

_____

Charles Yang, Professor of Linguistics

Graduate Group Chairperson

_____

Eugene Buckley, Associate Professor of Linguistics

Dissertation Committee:

Robin L. Clark, Professor of Linguistics

FUNCTIONALISM, LEXICAL CONTRAST AND SOUND CHANGE

COPYRIGHT

2020

Andrea Ceolin

Dedicated to the Italian taxpayers, who funded my education before my PhD studies.

ACKNOWLEDGMENT

The first two people that I want to thank are my advisors, Don Ringe and Charles Yang.

Words of encouragement are not very common in academia, but Don has always had plenty of them for me, from when I was admitted to the program to when I wrote up my dissertation. His patience and his kindness are remarkable. Every time I felt lost, meeting with Don was enough to immediately get back on track.

Charles has always been excellent at making sure I was asking the right questions, and that I was answering them convincingly. He has always been inspiring when we talked about why and how we do science, and he has been a role model to me. I think he always gave me much more credit than I deserved.

I also want to thank the two members of my dissertation proposal committee, Robin Clark and Rolf Noyer, and Tony Kroch, who encouraged me to think about information theory since the days in which 'logistic' was just one of the many things to take care of during event planning. Then, thanks to all the faculty members of the department who found the time to meet with me to discuss my research ideas (Mark Liberman, Meredith Tamminga, Gareth Roberts, Beatrice Santorini, Kathryn Schuler, Gene Buckley, Florian Schwarz, and Julie Legate), to Amy Forsyth, and to those outside of the department who gave me extensive feedback on my dissertation research (Geoffrey Sampson, Mary Beckman, David Ingram, Bert Vaux, and Marilyn Vihman).

The graduate students that I met during my journey at Penn will always deserve my gratitude. In particular, thanks to my cohort (Jordan Kodner, Spencer Caplan, Caitlin Richter, Ava Creemers, Hong Zhang, Wei Lai, Lacey Arnold Wade, and Jia Tian) and to Ollie Sayeed, Ryan Budnick, Hassan Munshi, Aini Li, Ugurcan Vurgun, Lefteris Paparounas, Faruk Akkuş, Nari Rhee, Nikita Bezrukov, Aletheia Cui, Milena Šereikaitė, Ava Irani, Luke Adamson, Kajsa Djärv, Nattanun 'Pleng' Chanchaochai, Sabriya Fisher, Rob Wilder, Betsy Sneller, Aaron Freeman, Akiva Bacovcin, David Wilson, Duna Gylfadóttir, Einar Freyr Sigursson, Samuel Andersson, and Naomi Lee.

I also want to thank the people in the Computer and Information Science department

ABSTRACT

FUNCTIONALISM, LEXICAL CONTRAST AND SOUND CHANGE

Andrea Ceolin

Donald A. Ringe, Charles Yang

One of the most discussed questions in the literature on sound change is whether functional factors play a role in shaping the lexicon over time, for instance by blocking the occurrence of a sound change (Gilliéron, 1918), or whether sound change is entirely mechanical and phonetically determined (Labov, 1987). Interestingly, the functional factors that have been proposed in the literature, like the principle of least effort (Zipf, 1939) and the minimization of entropy loss (Hockett, 1967), have not been related to the literature on language acquisition, with few exceptions.

In this work, I address this question by formalizing a model of sound change over time, in order to compare its predictions with the findings of historical and contemporary investigations on sound change (Chapter 2), by revisiting the hypothesis that sound change is blocked when it leads to homonymy (Chapter 3), and by identifying the factors that predict phonological development in children (Chapter 4).

While in the first part of the dissertation I argue that lexical change over time can be modeled without reference to functional considerations (contrary to the models in Martin, 2007, Graff, 2012 and Dautriche et al., 2017), in the second part I focus on the proposal that sound change seems not to occur when it would lead to homonymy (the 'Functional Load Hypothesis' in Martinet, 1955), and I argue that while this might be true for sound changes that result in the loss of a sound contrast in the speaker's grammar, it is not true for other common types of sound change. I motivate this hypothesis by showing that lexical contrast is a factor that influences early phonological development.

Table of Contents

List of Tables

List of Figures

Preface

This dissertation has been inspired by the work of Dinoj Surendran and Partha Niyogi. Dinoj and Partha both passed away exactly 10 years ago, in 2010. I didn't have the opportunity to meet either of them, and I only know about them through the voices of their friends and colleagues and through their papers.

The main goals that I was hoping to achieve with this dissertation were to improve on the work they did, and to keep their voice alive. One of the most powerful aspects of writing is that it allows us to hear from those that were here before us, and to talk to those that will come after, and this has been my main motivation in the writing of this dissertation.

I hope that this work will succeed in showing that we still have a long way to go on the road they paved.

Chapter 1

Introduction

1.1 Research framework

A question that is often discussed in linguistics is whether adaptation has a role in explaining language variation and language change, in the same way in which it explains biological evolution.

In a series of papers, Pinker & Bloom (1990), Jackendoff & Pinker (2005), and Pinker & Jackendoff (2005) defend the hypothesis that language itself has emerged in humans as a complex adaptation driven by communication needs, while Hauser et al. (2002), Fitch et al. (2005), and Chomsky & Berwick (2016) argue that the hypothesis is ill-posed, since human language exhibits fundamental differences with other communication systems identified in living organisms. Other works, instead of considering the origin of language, focus instead on the dynamics of language change, and show that some of the mechanisms that determine biological evolution can be used to explain language change over time (Yang, 2000, Niyogi & Berwick, 2009, Newberry et al., 2017).

One domain in which adaptation has been ruled out empirically is the study of sound change over time. In Labov (1987, 1994), one can find several case studies in which functionalist explanations of sound change are subjected to evaluation, and then discarded on the basis of empirical investigations of sound changes in progress. Labov concludes that sound change is essentially a mechanical and phonetically determined process, where functional considerations play no role in determining whether a particular sound change will occur or not.

While Labov's conclusions have essentially remained undisputed for two decades, the trend has changed in recent years. This change has been documented in an article that appeared in the *Journal of Sociolinguistics* in 2016:

> "Labov [...] comes to the conclusion that function plays no role in sound change and variation [...] In recent years the question has been reopened with new sophisticated statistical techniques" (Kiparsky, 2016, 14)

Indeed, the last decade has seen the emergence of a new 'functionalist' wave, inspired by information theory and statistical modeling. The recent works by Piantadosi et al. (2011), Kaplan (2011), Cohen Priva (2012), Graff (2012), Wedel et al. (2013), Bouchard-Côté et al. (2013), Dautriche et al. (2017), and Mahowald et al. (2018) provide both direct and indirect arguments in favor of functional pressures on sound change, which in some cases (Sóskuthy, 2013, Cohen Priva, 2017) are presented as solutions to the actuation problem in Weinreich et al. (1968). The conclusion reached by many of these works is that languages are essentially optimized for communication (Piantadosi et al., 2012, Fedzechkina et al., 2012, Bentz, 2018).

This dissertation is an attempt to revisit this question and to reflect on the dynamics of sound change and acquisition, using the same quantitative tools that have been used to argue in favor of functional pressures on sound change, and trying to go beyond, by identifying the cognitive mechanisms that underlie sound change and acquisition. In particular, I will first address the question of how languages would change over time if sound change were a stochastic process not influenced by functional factors, and then the question of whether such factors emerge when looking at sound change over time and sound acquisition in children.

The dissertation is organized in the following chapters:

- In Chapter 2, I propose an artificial model of lexical change, to study the effect of sound change on a lexicon over time. Models of lexical change have been proposed as a way to study the effect of functional biases on the lexicon (Martin, 2007, Graff,

2012). The model proposed here shows that once we incorporate the notion of sound change, languages drift into certain specific directions over time regardless of functional considerations. These results challenge the common assumptions that reference to functional biases is necessary to explain the fact that lexical change exhibits drifts over time, and that such drifts need to be countered by functional forces, in order for languages to preserve their communicative functions.

- In Chapter 3, I address the hypothesis that sound change over time is biased against the introduction of homonymy, namely the Functional Load Hypothesis (King, 1967a, Surendran & Niyogi, 2006). I show that contrary to some recent proposals (Silverman, 2010, Wedel et al., 2013, Babinski & Bowern, 2018), the hypothesis has only limited explanatory role.

- In Chapter 4, I show that there is an interesting parallel between factors that can be related to sound change, and factors that predict phonological development in children. I propose that in both cases it is not acoustic misperception the main predictor of sound change over time and production errors in child language, but lexical contrast, operationalized as contrast among word-initial syllables. I conclude that the Functional Load Hypothesis is not an independent factor that influences the acquisition of sound contrasts, but it is the outcome of a simple acquisition model.

The findings of this dissertation challenge the idea that language communicative functions are predictors of language change over time, and they show that once the mechanisms that underlie sound change and acquisition are formalized, the role of language communicative functions is explained away.

1.2 Theoretical background

This dissertation will use some assumptions and some notions that are grounded in theoretical and empirical research. These will be clarified in the next paragraphs.

1.2.1 Phonemes

One of the most important contributions of linguistic theory is the notion of 'phoneme'. Even though the same word might be produced in different ways, from the acoustic viewpoint, by different speakers, or by the same speaker in different instances, speakers only pay attention to a restricted subset of sound contrasts in perception and production. While the first consonant of the word *tie* might be produced slightly differently given the linguistic context, the accent of the speaker, or the situation, the listener will correctly perceive the target word as long as it does not sound too close to the first consonant in *die*, an instance in which a misunderstanding might arise. For this reason, in English, the sounds [t] and [d] are not typically described as two different sounds, but they are two different categories of sounds. Such categories are characterized by the property that a meaningful contrast is perceived across the two categories, but not within them. It is for this reason that languages can be represented in writing systems using a limited set of language-specific symbols, despite the fact that humans are able to produce and perceive many more sounds than those that they rely on for communicating in their own native tongue. Languages select only a subset of categories to distinguish among words, and as a consequence, the sound system of each language can be described with reference to a limited amount of broad categories (consonants, vowels, stress, and in some instances tones) that are useful to contrast among different words in that language. These categories are usually referred to as 'phonemes', and the set of categories which is relevant for the speakers of a language is defined as the 'phonemic inventory' of that language. For a history of the notion of phoneme, see Dresher (2011). When referring to phonemes, slashes ('//') will be used instead of brackets ('[ ]').

1.2.2 Minimal Pairs

When the distinction between two words relies on a single phonemic contrast, the two words form a 'word minimal pair', or simply 'minimal pair'. The two words mentioned above, *tie* and *die*, form a minimal pair, because their disambiguation relies on the distinction between the two phonemes /t/ and /d/. While most of the times a minimal pair can be identified

when two words are distinguished by a single letter in orthography, there are cases in which the orthography can be misleading (e.g., *knight* and *right* form a minimal pair, but just because the sequence 'kn' maps to a single phoneme /n/) and other cases where the same contrast is repeated in multiple parts of the word (e.g., *papa* and *mama*). In these cases, we can still have minimal pairs.

The notion of minimal pair is relevant when studying the phonemic inventory of a language, because it is a good diagnostic for phonological contrast: if two words form a minimal pair, then we have to postulate that the two sounds used to distinguish them belong to two different phonological categories. As we will see, the literature on language acquisition has investigated the hypothesis that minimal pairs are the cues that children use to learn the phonemic inventory of their target language.

1.2.3 Regular sound change

Languages change over time, because words run out of use, and new words appear. Words are also affected by change in pronunciation: the way most of the words are pronounced in North American English is different from the way they are pronounced in British English.

During the 19th century, scholars interested in the study of languages made a crucial discovery: change in pronunciation typically affects sounds, not words. When a certain sound changes, the change is regular across the lexicon (Grimm, 1848, Osthoff & Brugmann, 1880, Paul, 1891). As a consequence, when a sound change occurs, regular correspondences emerge in the lexicon. Many of these correspondences can survive for millennia: by going through a Spanish-English vocabulary, anyone can notice that many of the Spanish words starting with the sound [p] (*padre*, *para*, *pie*, *pez*, *primero*) have a starting [f] in English (*father*, *for*, *foot*, *fish*, *first*). This is evidence for the fact that English and Spanish are descendant of an ancestral language, and they both contain reflexes of its original lexicon. Note that this fact alone would allow us to postulate a historical relationship between English and Spanish even if we had no historical record of its occurrence.

In the 19th century, regular sound correspondences allowed this group of scholars, who

became known as the Neogrammarians, to relate most of the languages spoken in Europe to several languages spoken in Asia, and to postulate the existence of an ancestral language, Proto-Indo-European, which could be reconstructed by identifying sound correspondences among languages for which historical records were available, and by using chronological evidence to determine the order of the sound changes. This discovery allowed philologists to reconstruct a part of the phylogenetic history of human populations in Europe and Asia 100 years before DNA could be used for the same goal (Cavalli-Sforza & Edwards, 1967).

Exceptions to regular sound change are mostly due to word borrowing and to analogical change, namely a change in pronunciation which is motivated by similarities in sound or meaning among different words.

Another exception which has been proposed in the literature is lexical diffusion (Wang, 1969, Bybee, 2002, Phillips, 2006), namely the case in which a sound change is not regular across the lexicon, but it appears to be conditioned on external factors like word frequency and co-occurrence with other words in speech (Bybee, 2017), and it can gradually affect the entire lexicon. While there is evidence that some sound changes proceed this way, lexical diffusion has found limited application in language reconstruction and in modeling sound change over time. While some reference to lexical diffusion will appear during the dissertation, this type of change will not be discussed in detail.

For a more complete account on regular sound change, see Labov (1994), Campbell (2013) and Ringe & Eska (2013).

1.2.4 Mergers and Splits

Among different types of sound change, one can distinguish changes that involve the pronunciation of specific sounds, and changes that instead affect the size of the phonemic inventory, for instance by collapsing together two phonemes or by adding new contrasts in the inventory. The latter are the objects of investigation of this work, because they are those that have received the larger amount of attention in historical linguistics, since their interaction is what causes languages to become mutually unintelligible over time. Following

Hoenigswald (1960), we can divide them into three categories.

The first category is *unconditioned merger*. According to this type of sound change, two different phonemes in a language collapse into one, therefore causing homonymy among the words which could be distinguished only by that contrast. One widespread type of unconditioned merger is that between the vowel in *lot*, /ɑ/, and the vowel in *thought*, /ɔ/, which affected some English varieties spoken in North America (West American English, New England English, and Canadian English) and in the British Isles. As a consequence of this merger, words like *caught* and *cot* become indistinguishable in pronunciation.

The second category is *primary split*, or *conditioned merger*, for which a distinction between two phonemes is lost only in a specific phonological environment. In this case, the contrast is still present outside of the environment, and therefore the phonemic inventory is not altered, because speakers can still distinguish the two sounds given the proper context. A case which can be ascribed to North American English is the neutralization of the contrast between /t/ and /d/ after a stressed syllable: in words like *butter* and *wedding*, the two phonemes are perceived as an unique sound, described as an alveolar flap, [ɾ]. In other varieties of English, the two words still mantain the distinct [t] and [d] in pronunciation. Since the two phonemes contrast outside of this specific context (*tie* and *die* are still well distinguished), the phonemic inventory is not affected.

The third category is *secondary split*, or *phonemic split*, or simply *split*, and it refers to a change in which two sounds that are simply variations of a unique phoneme, and therefore not contrastive, are reanalyzed as independent phonemes. For instance, in Old English, [f] and [v] were not contrastive: the letter 'f' was used to represent a sound which was more *f*-like when it was used in word-initial or word-final position, and more *v*-like when it occurred between vowels, but there was no contrast between the two sounds in the lexicon (Eng. *love* < OE. *lufu)*. However, when Old French words entered the English vocabulary, many words with a word initial *v*-like sound suddenly appeared (like *very, visit, voice* and *vegetable*), and this led to a reanalysis of /f/ into two different phonemes, /f/ and /v/, to account for word contrasts like *ferry* and *very*.

A crucial property of splits is that they are always dependent on other changes in the system: one of the most common sources of phonemic split is the loss of a conditioning environment due to other sound changes, but another one can be lexical borrowing, like in the case just mentioned.

Chapter 2

Are languages drifting?

According to many of the Neogrammarians, sound change is 'essentially random drift in phonetic space' (Newmeyer, 2003, 26). Sound change is unpredictable, because it is the result of physical or ethnographic pressures. For instance, this quote from Grimm, via Lightfoot (2006), is self-explanatory:

> "connected with the German's mighty progress and struggle for freedom...the invincible German race was becoming ever more vividly aware of the unstoppability of its advance into all parts of Europe...How could such a forceful mobilization of the race have failed to stir up its language at the same time, jolting it out of its traditional rut and exalting it? Does there not lie a certain courage and pride in the strenghtening of voiced stop into voiceless stop and voiceless stop into fricative?" (Grimm, 1848)

Other variants of this explanation are, for instance, that spirantization results from mountain-climbing (Meyer-Benfey, 1901), rounding is triggered by low temperatures (Sweet, 1990, 32) and lenition follows from starvation (Straka, 1979, 285).[1] Essentially, 'languages were seen as external objects floating smoothly through time and space' (Lightfoot, 2006, 37).

As the investigation of contemporary languages widened in scope, certain sound changes were repeatedly found in many language families, while others remained unattested. In particular, many common sound changes could be explained with reference to ease of articulation:

---

[1] References from Newmeyer (2003) and Scheer & Ségéral (2019).

"Hermann Paul suggested that the more common changes tended to be 'in some respect more convenient...where greater or lesser degree of convenience is...purely physiological' (Paul, 1891, 43)" from Newmeyer (2003)

The question of whether there is a global pressure on languages to change in particular directions over time has been investigated since Sapir (1921). In particular, Sapir first popularized the notion of linguistic 'drift': the observation that language change is not a random walk, but 'language moves down time in a current of its own making'. The search and the explanation for linguistic drifts have become popular since then, especially in the domain of morpho-syntax (Greenberg, 1963, Lakoff, 1972, Vennemann, 1973).

The existence of natural sound changes is received without much controversy today, to the extent that reading quotes like the one by Grimm appears, to the least, 'a bit wild and farfetched' (Lightfoot, 2006, 34). However, the idea of natural and unnatural sound changes presents some practical problems. If sound laws were universal, one might expect all languages to be converging toward the same state, making it unlikely to find languages where the initial input of a natural sound change is still preserved (Lass, 1981, 16-17).

One common reaction to this problem is to postulate that ease of articulation is just one of many forces acting on sound change, and that other forces, like phonetic (Martin, 2007) and learnability biases (Graff, 2012), or pressures to avoid ambiguity (Dautriche et al., 2017), might actually counter any drift. If this were true, then one should be able to identify some types of diachronic change that have the opposite effect of natural sound changes, and determine if and how frequently they occur typologically.

Dautriche et al. (2017, 144) propose that forces acting against a drift in the lexicon might actually be identified:[2]

"For example, 'flour'/'flower' were originally two senses of a single word, and so pronounced identically. Now that English speakers perceive them as entirely different words, it is plausible that processes of dispersion could act to bring

---

[2]I found it strange that no references are reported in this passage. A footnote in the original article suggests that this passage was inspired by the comments of one of the reviewers of the paper, and so it can be that it was not present in the authors' original draft.

their pronunciations apart, as happened with 'one'/'a(n)' and 'thyme'/'time' "

Dautriche et al. (2017, 144)

The passage refers to the possibility of an unconditioned split driven by ambiguity considerations as a diachronic change that favors phonological dispersion in the lexicon. However, this is at odds with the historical records: splits are, in fact, always conditioned on a phonetic environment, and in fact none of these cases is an unconditioned split. In the case of 'one'/'an', the Online Etymological Dictionary[3] reports that the two words were homophones in Old English, but then 'one' changed its pronunciation in South West and West England in the 14th century, and it started to spread in the 18th century. The change was not idiosyncratic, but applied to other words with the same word-initial vowel, and therefore is better described as a case of dialectal borrowing.[4] As for the second case, the authors are probably referring to Gahl (2008), in which 'thyme/time' and other homonyms are indeed shown to be pronounced with a different vowel length due to their difference in frequency. However, this work does not prove that this effect is specific to homophones: for instance, it does not show that speakers can disambiguate the vowels in 'time/thyme' better than the vowels in 'time/rime'. Therefore, it is not clear how one can make the prediction that the pronunciations of 'flour' and 'flower' are under any pressure to change.[5]

Moreover, even if some types of diachronic change had the effect of countering any drift, one would have to make sure that they are frequent enough to have an effect on the system. For instance, Labov posed a similar problem in 1994, when reasoning on the frequency of mergers and splits:

"most reports of phonemic change involve mergers [...] [this fact] would lead to the odd conclusion that most languages are steadily reducing their vowel inventory [...] it stands to reason that just as many phonemic splits must take place as merger" Labov (1994, 331)

---

[3]https://www.etymonline.com

[4]Cf. 'woak' as an alternative spelling of 'oak' (Upward & Davidson, 2011).

[5]According to the Online Etymological Dictionary, the two words have been homophones since the 13th century, and actually they were also spelled in the same way until the beginning of the 19th century.

In this passage, Labov says that pointing to different types of sound change that push languages in different directions, in this case in terms of number of segments in the phonemic inventory, is not sufficient: one needs to show that their frequencies of occurrence are balanced, in order for a drift over time to be prevented.

In this chapter, I will address these questions with reference to language drift along two dimensions: the size of phonemic inventories, as mentioned in Labov (1994), and the amount of phonological dispersion in the lexicon (Martin, 2007, Graff, 2012, Dautriche et al., 2017). In particular, I will show that language drifts in these two domains can result from a stochastic model of sound change of the Neogrammarian type, and they can be easily prevented through a simple mechanism of lexical replacement, without the necessity of postulating any functional pressure on lexical change over time.

2.1   Two cases of drift

In this section, I summarize the debate around two cases of drifts discussed in the literature: the drift toward smaller phonemic inventories, and the drift toward a more 'compact' lexicon.

2.1.1   The size of phonemic inventories

As previously mentioned, Labov (1994, 331) states that even though most of the reported sound changes in progress are mergers, an equivalent number of splits must occur in order to keep the size of phonemic inventories stable over time. Labov's conjecture can be true: for instance, Tse (2016) proposes that splits are more likely to emerge in intense contact settings, and the reason why they are underreported in the literature is that sociolinguists have not focused on contact areas. Still, more than two decades after Labov (1994), the majority of linguists would agree that mergers are still being reported at a higher rate than splits.

There is clearly something odd about this conclusion. It does not sound realistic that one day we will just communicate using sequences of *ba* and *pa*. However, there is an ongoing debate about whether the hypothesis of a drift toward smaller phonemic inventories is more

plausible than one might initially imagine. A *Language* paper published by Hay & Bauer (2007) noticed that reports of phonemic inventories for languages with large populations contain on average a larger number of phonemes compared to those associated with smaller populations. Even though the claim has been rejected by several studies on statistical grounds (for instance in Donohue & Nichols, 2011 and Moran et al., 2012), the hypothesis was interesting enough to lead to a study that modeled the size of phonemic inventories using a serial founder effect model, borrowed from population genetics (Atkinson, 2011). According to this model, when a group of people leaves a population to migrate, they carry only a subset of the genes which characterize the population: for this reason, the genetic diversity in the new population is reduced, compared to the original one. The analogy in linguistics would be that when a group of people migrates, they might develop sound changes that have the effect of reducing their phonemic inventory. The model was corroborated by showing that the languages that contain the highest amount of phonemes are all located in Africa, a finding that matches the genetic traces of the out-of-Africa migration of Homo Sapiens about 60,000-70,000 years ago (Wallace et al., 1999).

Even though replications of this experiment did not confirm the results (cf. Creanza et al., 2015), this scenario is surprisingly compatible with the fact that mergers, which are the main source of phoneme loss, are the most common type of sound change.[6]

2.1.2 Phonological dispersion in the lexicon

There are two kinds of functional pressures that are often mentioned in the literature as potential motivations for lexical change over time. One is the need of minimizing effort on the part of the speaker (the 'principle of least effort' in Zipf, 1939 and Wright et al., 2004), and the second one is the need of minimizing ambiguity on the part of the listener (Liljencrants & Lindblom, 1972, Flemming, 2004). These two factors are complementary:

---

[6]This finding is also compatible with the assumption that mergers only spread if a certain threshold of the population exhibits them (Yang, 2009), because significant thresholds are easier to reach in smaller populations, where individual grammars have a greater weight. Even though, ultimately, the only thing that matters is the exposure of children to adult speech, a fact which is more influenced by social norms than by the overall size of the population.

languages with many phonemes distributed in a uniform way would be very efficient from the viewpoint of clarity, but they would require a great cognitive effort for learners and a great physical effort for speakers when it comes to producing sounds. On the other hand, languages where phonemes are distributed unevenly, with sequences of sounds that are easy to produce being more frequent than sequences requiring more articulatory effort, would be easier to articulate, but more likely to produce misunderstanding.[7]

Previous works studied whether either of the two pressures has any quantifiable impact on the shape of the lexicon of real languages, by counting word minimal pairs (Graff, 2012, Dautriche et al., 2017). Similar measures that are used to quantify phonological dispersion include the average Levenshtein distance among words (namely, the number of insertions, deletions and substitutions needed to make two strings identical) and phonological neighborhood density (Munson & Solomon, 2004), but the use of minimal pairs is the most popular one. The argument is that low phonological dispersion correlates with a high number of minimal pairs, while high phonological dispersion correlates with a low number of minimal pairs.[8]

Establishing a baseline for minimal pairs is not intuitive, because they are a function of the average word-length and the size of the vocabulary, which might include the presence of inflectional and derivational morphemes. Dautriche et al. (2017) manage to do so by taking data from the three languages in CELEX (Dutch, English and German, Baayen et al., 1996) plus French, from Lexique (New et al., 2004), and excluding all the words that contain derivational morphemes. Then, they train various language models using n-grams. This strategy is adopted to 'teach' phonotactics to the model, since the model will learn

---

[7]The problem of ambiguity is overstated in this simplistic formulation: in the real world, speakers can rely on several strategies to disambiguate among alternative readings and retrieve the intended message. This point is usually not stressed enough in the functional literature, probably because it is very hard to quantify: speakers of English know that *cat* and *cap* are two different words, but it would be very difficult to isolate the contexts in which this phonetic distinction is useful at all to retrieve the meaning of an utterance. The only work that I'm aware of that tried to address this question is the study of natural misunderstandings presented in Labov (2010).

[8]One problem with using minimal pairs as a measure of phonological dispersion is that this measure does not handle homonymy well: a lexicon in which all the words have the same phonemic shape would have no minimal pairs at all, in spite of being clumpy, which is a paradox, since a low number of minimal pairs is usually correlated with high dispersion. When dealing with homonymy, a Levenshtein distance would yield a more interpretable result. For this reason, this is the measure that I will employ later in the chapter.

that certain sound sequences are common, and certain others never occur. These models are then used to generate pseudo-words, using a window of n-1 phonemes as a probabilistic conditional environment. After the model is trained for each language, they select a subset of real words of a certain length, and they compare it with a subset of artificial words of the same length generated by the model, to control for word-length. The results are interesting: they find that the number of minimal pairs in real languages (among other measures) is much higher than that found in the pseudo-vocabulary they generated, and they derive from this result that the pressure for phonemic 'clumpiness' must be stronger than the pressure for phonological dispersion in real languages.[9]

The explanation they propose for their result is that diachronic change is responsible for this pattern, and to avoid the problem of having to postulate a language drift, they hypothesize that diachronic change can push languages in different directions:

> "[...] it is likely that we may have a lot to learn from diachronic data to observe how clumpiness evolves in the lexicon as new words appear in the language. While we discussed the possibility that there are pressures for clumpiness exerting on the lexicon, another possibility is that there are only pressures for dispersion and not clumpiness, but that word coining leads to clumpy initial states" Dautriche et al. (2017, 143-144)

The hypothesis that they present can be tested, and it is clearly related to the problem of determining whether the size of phonemic inventories tends to be stable over time. This is another case of potential drift that we will investigate in the next sections.

2.2 Models of lexical change

Scholars who are interested in testing the hypothesis that languages are optimized for communication typically focus on analyzing synchronic data from modern languages, while there

---

[9]'The space of wordforms for Dutch, English, German and French is clumpier than what would be expected by the best chance model, across a wide variety of measures: minimal pairs, average Levenshtein distance and several network properties. The strongest evidence comes from minimal pairs, for which the effect size was quite large.' Dautriche et al. (2017, 144).

are only few attempts to study the relationship between diachronic change and synchronic variation.[10] In particular, a question to ask is whether sound change has any role in shaping the lexicon over time.

Linguists have been able to formalize sound change explicitly since the 19th century. While the details of specific sound changes might appear intricate, the phenomenon of regular sound change is explicit from a formal viewpoint. For this reason, I argue that linguists are in a good position to build formal models of sound change, and to use them to observe the consequences of sound change over time (see, among others, Sóskuthy, 2013, Morley, 2015, Beguš, 2018 and Kodner & Cerezo Falco, 2018). In particular, several works in the literature stressed the importance of formalizing null models of change in order to evaluate claims about factors influencing language change over time (cf. Baxter et al., 2009, Kauhanen, 2017, Newberry et al., 2017), and therefore in this chapter I will attempt to implement a null model of sound change. Before proceeding, I will briefly summarize some previous attempts of modeling lexical change over time.

2.2.1 Previous models of lexical change

The most explicit attempt to formalize a model of lexical change over time is found in Martin (2007). The question that Martin wants to address is the origin of 'gradient phonotactics' in the lexicon:

> "[...] a growing body of research has focused on *gradient phonotactics*, statistical
> rather than categorical sound patterns that hold over the lexicon. Although this
> work has been a welcome corrective to the categorical bias in the field, it has
> been largely restricted to answering the learning problem [...] Much less work
> has been devoted to the typology problem - understanding why language exhibit
> the pattern they do, or indeed why they have gradient phonotactics at all. This
> is the question I will attempt to answer [...]" Martin (2007, 2-3)

---

[10]This is also true for syntactic change: apart from classical works like Kroch (1989), Lightfoot (2006), Roberts (2007) and the LANGELIN project (Longobardi et al., 2013, Ceolin et al., 2020) one does not find many works trying to link diachronic change with a theory of synchronic variation.

In his work, Martin does not attempt to model sound change directly, but mostly focuses on lexical replacement. He proposes a model in which lexical change is driven by synonymy resolution: given that speakers are often faced with several options to refer to objects and concepts, there might be a pressure for selecting some wordforms over others. His model is inspired by the spreading activation model in Dell (1986), according to which lexical access involves the activation of the phonemic representation of a word. According to this model, retrieving the meaning of words that contain more frequent phonemes is easier than retrieving the meaning of words containing rare phoneme sequences. For this reason, speakers are inclined to resolve synonymy by using words whose phonemic sequences are more common. This model resembles a classic rich-get-richer model (Simon, 1955), a model through which frequent elements tend to increase their frequency over time at the expense of less frequent elements, which slowly lose probability mass.

Such a model clearly has the problem of dealing with drift over time: unless the lexicon is infinite, frequent phonemes become as frequent as to occupy the entire lexicon. To solve this problem, Martin addresses the possibility of some constraint on lexical change. One potential source of 'counter drift' that he proposes is an arbitrary lexical weight, which causes some words to be more resistant to change in spite of their phonological representation, a technical implementation of the notion of basic vocabulary. A second source of stability is introduced through a phonetic bias: since some sounds might be 'attractive' because of phonetic or articulatory reasons, i.e. they are easy to perceive or produce, speakers will favor them whenever they have a chance to do so. The most striking example in support for this view is the fact that Latin borrowed a considerable number of 'b-' words from Greek (35) with respect to, for instance, the number of 'd-' words (15), even though in Greek 'd' was twice as frequent as 'b' as a word-initial consonant. The trend continued in the Romance languages: in particular, French retained a large part of Latin's 'b-' words compared to words starting in 'd-', and neologisms in French derived from 'b-' words were more frequent than those derived from 'd-' words. Martin's motivation for this statistical trend is a bias toward 'b' as a phoneme in borrowing, word-retention and word-formation. The presence

17

of a phonetic bias is also shown to have affected the retention of onset consonant clusters from Middle English and neologisms in Modern English.

Martin does not present evidence for the existence of a lexical weight, even though the idea has been popular in the lexicostatistics tradition (Swadesh, 1955, Oswalt, 1971, Dyen et al., 1992, Kessler, 2007, Holman et al., 2008, Tadmor et al., 2010), and measures like word frequency have been proposed as a proxy for lexical weight (Lieberman et al., 2007, Pagel et al., 2013). However, the assumption of a phonetic bias behind word borrowing, word retention and word formation is supported only by a limited number of cases that he refers to.

The model proposed by Martin has been revised in Graff (2012). Graff proposes to abandon the notion of lexical weight by re-implementing the model introducing instead a 'lateral inhibition' factor (Bard, 1991, Berg & Schade, 1992). According to the model, the introduction of a new word in a lexicon should be penalized if it is very similar to other existing words.[11] This factor has the consequence of making the model converge to a more uniform distribution of phoneme frequencies and to a decreased amount of lexical similarity over time, rather than an increased one.

2.2.2   A null model of sound change

A limitation of the previous approaches to lexical change is that they make the assumption that words are monolithic entities: they enter a lexicon, and they stay as they are until they are replaced. Historical linguists documented that even though words can survive in a lexicon for millennia, they can completely change their form because of regular sound change. Among the pressures on the lexicon to change over time, regular sound change is probably the most attested one. For this reason, while the models described in the previous sections studied lexical change using a mechanism of word replacement, in the next sections I will try to develop a model which includes regular sound change.

In order to determine how sound change shapes the lexicon over time, I propose the

---

[11]The fact that (near-)homonymy can inhibit word learning has also been corroborated in acquisition studies (Peters & Zaidel, 1980, Mazzocco, 1997, Doherty, 2004, Swingley, 2016).

strategy of devising a null model of sound change, studying its implications for the lexicon over time, and then gradually adding additional factors to see if they change the predictions of the null model.

A naive implementation of a null model of sound change would be a direct application of the Wright-Fisher model (Fisher, 1930, Wright, 1931) on a list of phonemes, which can be obtained by concatenating the phonemes associated with the words of a written text. The analogy with population genetics would be that each phoneme represents a haploid individual which reproduces asexually (like in Saunders et al., 1984 and Möhle & Sagitov, 2001). In such a model, change over time is represented through a random sampling of the entire phoneme population, which would lead to change in phoneme frequencies over time (cf. Reali & Griffiths, 2009, Sindi & Dale, 2016, Sayeed & Ceolin, 2019 and Ceolin & Sayeed, 2019 for similar models).

This is a toy example of how such a model can be implemented. Suppose that we start with a text that contains three words: {'dog', 'cat', 'pig'}. Our population of phonemes can then be represented by the sequence 'dogcatpig', if we make the idealization that each letter maps onto a different phoneme (which is true in this case, but is not the norm in English). A resampling of the sequence might first yield to 'dotcatpig', if the word 'dog' changes into 'dot', and then into 'dotcatpit', if also the word 'pig' changes into 'pit' in a second stage. By plotting the phoneme counts in a graph, we can represent the phonemic drift over time (see Figure 2.1). This is, essentially, genetic drift applied to language.

The main problem with such a model is that it is completely unrealistic when applied to sound change: the target of a regular sound change is not a change of a specific phoneme in a specific word, but a change affecting classes of sounds in specific phonological environments. In order to simulate sound change more realistically, we need a lexicon, a formal notion of sound change, and a notion of phonological feature. These three components will be defined in the next subsections.
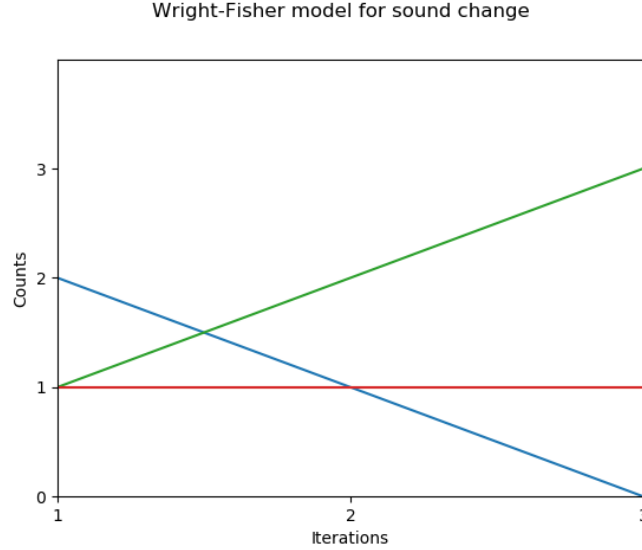
Figure 2.1: Changes in phoneme counts over time in the toy example {'dog', 'cat', 'pig'}. The blue line represents changes in the counts of 'g', the green line represents changes in the counts of 't', while the red line represents all the other phonemes.

2.2.2.1 Lexicon

Defining a lexicon is the first step toward simulating lexical change over time. Martin (2007)'s initial lexicon has 100 words, and 5 segments, and all the words have the same form 'abcde'. While it makes sense to use a language with segments being equally distributed, it would not represent a distribution that we expect to find in real languages.[12] Moreover, having all the words sharing the same phonological form is not ideal for the study of regular sound change, because there would be no diversity of phonological environments.

In theory, one could use a real-language lexicon as a starting point, for instance the English lexicon. Limiting the model to simple CVC roots will make it simple to implement and easy to interpret. Notice that this would be sufficient to implement syllable structure considerations: with a CVC lexicon, we can distinguish between consonants and vowels, and among onsets, nuclei and codas.

Since the English lexicon is full of simple V, CV, VC and CVC roots, I decided to select

---

[12]Phonemes in real languages follow a Yule/polylogarithmic distribution (Tambovtsev & Martindale, 2007, Nichols & Kauhanen, 2019).

the 150 most frequent ones in the COCA corpus (Davies, 2009) as the initial lexicon. This is the list of words that were extracted:

- Lexicon: {a, air, as, at, each, eat, if, in, it, i, of, on, or, oil, our, out, up, us, bad, bar, be, bed, big, bit, book, both, boy, box, but, buy, can, car, cup, cut, dad, day, dead, death, deal, deep, do, dog, door, fail, far, fear, feel, few, four, food, foot, for, get, go, goal, god, gun, guy, head, hear, hair, him, his, hit, he, her, hour, hot, how, join, job, keep, kid, law, lay, lead, leg, let, look, lot, low, man, main, may, mean, meet, me, mom, need, new, near, no, not, now, pain, pay, per, poor, put, reach, read, real, room, red, run, she, show, shoot, say, see, sea, seat, seem, seek, set, sex, sir, sit, six, so, son, soon, such, than, that, the, them, then, they, their, this, thus, two, tax, team, ten, too, to, top, year, yes, yet, you, your, what, when, which, who, wait, wear, war, way, we, week, win, with}

The next step requires coding the words in a format that represents a phonological transcription. For the purpose of the simulation, anything will do: bits, arbitrary symbols, or an IPA transcription. However, using these representations would just introduce another layer of abstraction, and make the words and their changes less identifiable when we look at how they change in the course of the simulations. Especially in the case of an IPA transcription, it would also be misleading: until we define a phonemic feature representation, our inventory is just made of arbitrary symbols with no real phonetic content. Therefore, I decided to use the words in their original form, and to define an arbitrary inventory represented by the following English letters and digraphs, divided among consonant and vowel symbols:

- C: {b, c, ch, d, f, g, h, j, k, l, m, n, p, r, s, sh, t, th, tw, v, w, wh, x, y, z _ (space)}

- V: {a, ai, e, ea, ee, ei, i, io, ie, o, oa, oi, oo, ou, u}

This inventory does not map directly onto English phonemes: for instance, there is not a 'c' phoneme in English, and the symbol 'w' is phonemic only in onset position (while in

coda is part of a vowel diphthong). Some of the words selected do not even map onto a CVC phonological representation (for instance, 'six' [sɪks] is CVCC). However, since the simulations are not meant to say anything about the English language, any symbolic representation would do, as long as it is unambiguous, and each symbol can be interpreted as representing a phoneme in the language.

2.2.2.2 Sound change

The next step is to apply changes to this lexicon in order to simulate sound change. Sound change can be described using the classic formalism in Chomsky & Halle (1968):

$$X \rightarrow Y \ / \ \_\_ \ [\pm\text{features}]$$

Where we have a segment X becoming Y in an environment defined by a set of features.[13] If we want to apply sound change to a lexicon, we can just create a function, and apply it over the lexicon, as follows:

a. Select one position in the syllable (Onset, Nucleus, Coda).

b. Select two segments in the inventory among those available for that syllable position (X and Y)

c. Determine a conditioning environment.

d. X becomes Y in the conditioning environment.

The function can have five possible outcomes depending on the distribution of the segments in the lexicon:

a. X and Y are expressed in the lexicon in the position targeted by the change, and all instances of X become Y (Unconditioned Merger)

---

[13]Since phonemes are expressed in terms of features, the formalism can also be used to describe changes involving classes of segments which have some features in common. This is something that I will not focus on in this work, because of some implementation difficulties that I address later.

b. X and Y are expressed in the lexicon in the position targeted by the change, and only some instances of X become Y (Conditioned Merger)

c. X is expressed in the lexicon in the position targeted by the change, Y is not part of the lexicon, and only some instances of X become Y (Split)

d. Only X is expressed in the lexicon in the position targeted by the change, and all instances of X become Y (Phonetic change)

e. None of the instances of X become Y (No change)

There are at least two delicate points in the design of the simulation.

First, the outcome of a sound change depends on several factors: small vocabularies, or vocabularies in which the distribution of the symbols is restricted, will have (3d) and (3e) as frequent outcomes, while large vocabularies, or vocabularies in which the symbols are distributed over all possible environments, will have (3b) and (3c) as common outcomes instead; (3a) can be seen as an extreme case of (3b), when the change has the consequence of removing X from the language; moreover, the size of the inventory has an impact on whether mergers are more or less common than splits. These are all factors that need to be controlled for.

Second, implementing the notion of feature in the system is not straightforward. While generative phonology has provided feature vectors that can be implemented directly to represent phonemes (Hayes, 2011), the fact that defining the class of possible sound changes is not trivial, and that some of them have been described as more 'marked' than others, in ways which have not been clearly formalized (Dresher, 2009, ch. 5.2), makes it practically impossible to pre-determine 'plausible' conditioning environments in a way that does not bias the results of the experiments toward certain obvious outcomes. For instance, if all consonants are allowed to become [-voice] in word-final position, but the opposite cannot happen, one should not be surprised to end up with languages where [+voice] consonants disappear from word-final position.[14]

---

[14]Moreover, if following Dresher (2009) our model of phonology includes feature hierarchies which can change over time, then the conditioning environments can also change over time, adding a further layer of complexity.

For the purpose of defining a null model, I decided to run a first experiment in which conditioning environments are randomly selected by sampling through the possible segments that can constitute a conditioning environment. The choice is justified by the fact that the goal of this chapter is not to make claims about specific phonemes or features, but to see how a lexicon changes over time when we implement the technical notions of regular sound change and conditioning environment. For an attempt of implementing the notion of phonological features, see Appendix A.

In the case of a change affecting the onset or coda position, the conditioning environment is represented by the nucleus, while in the case of changes targeting the nucleus, the conditioning environment would be a subset of the consonants that can appear either on the left or on the right. From this last point, it follows naturally that a nucleus should have more opportunities to be targeted for sound change than an onset or a coda, because it is exposed to two possible different environments.

To summarize, the four possible changes are:

a. Change in the Onset (conditioned on the Nucleus)

b. Change in the Nucleus (conditioned on the Onset)

c. Change in the Nucleus (conditioned on the Coda)

d. Change in the Coda (conditioned on the Nucleus)

These four types of changes are selected by the algorithm with equiprobability (0.25 each). This is also an arbitrary factor that will be controlled for.

One could object that the beginning and the end of the word are also conditioning environments (in the case of sound changes like final devoicing, for instance), and the representation of the word should include word boundaries (i.e., #CVC#), but by adding word boundaries to the model there would still be no diversity in terms of environment for word-initial or word-final segments, beyond the nucleus. A sound change conditioned on word boundaries could then generate an unconditioned merger, but no other change, and so it would just introduce a bias toward mergers in the model.

On a final note, while this model reproduces sound change mechanically, it is by no means capturing the fact that mergers and splits are not simply the result of 'errors' in perception or production. There is a large tradition of studies that investigated factors that might favor or constrain the occurrence of a merger (Martinet, 1955, King, 1967b, Labov, 1994, Yang, 2009), and even though less studies focused on splits (cf. Gylfadóttir, 2018), we mentioned in the previous chapter that splits are not independent events, but they result from the loss of a conditioning environment that follows from other sound changes, or from the appearance of loanwords from a foreign language. The purpose of the model is simply to study how the lexicon changes as a function of regular sound change, regardless of the way in which the change was initially introduced in the lexicon.

Of course, one can object that it is impossible to study sound change without explicit assumptions about the constraints and the factors that (might) influence possible sequences of sound change. On the one hand, this brings us back to the problem of determinism, presented at the beginning of the chapter (cf. Lass, 1981). On the other hand, since further assumptions would require increasing the complexity of the model, I think a plausible strategy is that of starting from a model which is as unassuming as possible, and then moving to more complex models, to see if there are patterns of change that are consistent, and patterns that instead vary depending on the initial assumptions.

2.3   Results

In this section, we consider the results produced by the null model, and we examine their implications for the two main questions of the chapter.

2.3.1   The size of the phonemic inventory and phonological dispersion in the lexicon

In order to study the change in the size of phonemic inventories over time, all we need to do is keeping track of the number of segments in our lexicon as we apply sound change on it. On the side, we can also keep track of the average Levenshtein distance among words,
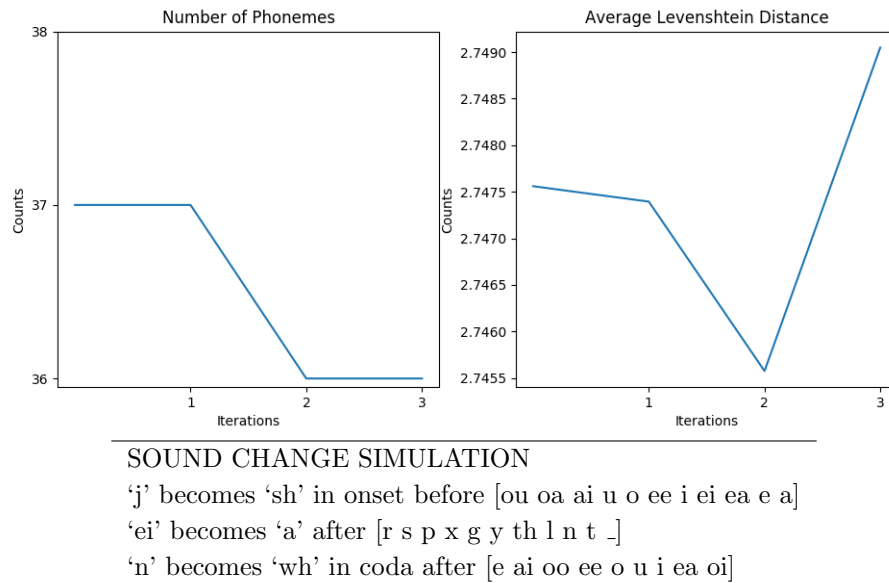
Figure 2.2: The output of a toy sound change simulation. Three sound changes are applied to the initial lexicon.

as a measure of phonological dispersion in the lexicon. Since the words have a fixed length, each word pair will have a distance in the range [0-3] depending on how many segments are shared in the same position, with '0' corresponding to homonymy and '3' corresponding to two words that do not share any segment.

A first example can be seen in Figure 2.2, in which three sound changes are applied to the lexicon. The y-axis displays the size of the inventory and the average Levenshtein distance, while the x-axis tracks the sound changes as they occur in the simulation.

The first change has 'j' becoming 'sh' in onset position before a subset of nuclei. This subset does not include 'oi', so the word 'join' stays the same, but it includes 'o', so that 'job' becomes 'shob'. This is a conditioned merger which does not have any effect on the number of phonemes ('j' and 'sh' are both present before and after the merger), but it slightly decreases the average distance.

The second change has 'ei' becoming 'a' after a subset of onsets. Since 'ei' is only present in the word 'their', which contains the conditioning environment and becomes 'thar', the merger is *de facto* unconditioned, and therefore it leads to a decrease of the size of the phonemic inventory, that now has lost 'ei'. The average distance decreases, as expected.

26
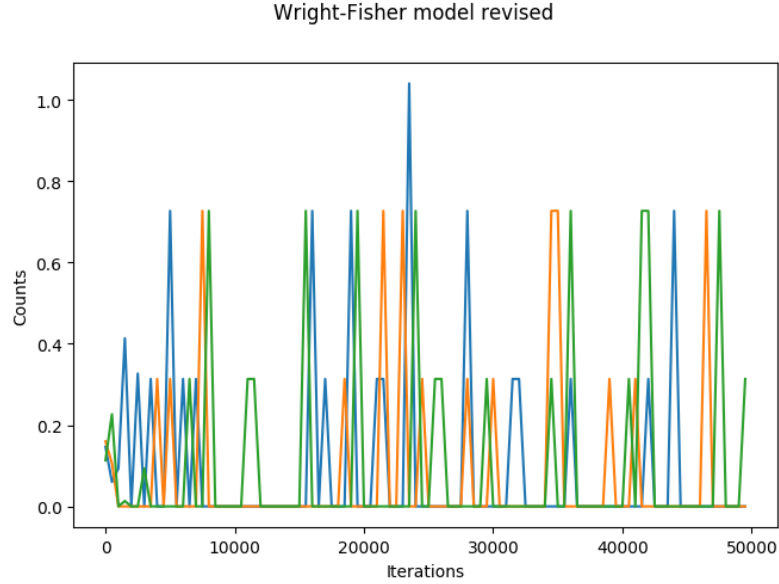
Figure 2.3: Relative frequencies of the vowels 'a', 'e' and 'i' after a sound change simulation, where 50000 sound changes are applied to the initial lexicon.

The third change has 'n' becoming 'wh' in coda position. Since 'n' is a frequent coda, some words do change, but others keep their form, for instance all the '-an' words, since 'a' is not in the conditioning environment. The number of segments stays the same, because 'wh' is a segment available in onset position, but the fact that this segment now also appears in coda position makes the average distance increase, because it creates more diversification in codas.

These changes appear structurally like regular sound changes, the only difference being that they involve no feature content, because they are applied over arbitrary symbols rather than real phonemes. Now, the interesting thing is to see what happens if instead of having three sound changes applied to a lexicon, we have many of them occurring in sequence.

To begin with, in Figure 2.3 I replicated the Wright-Fisher analysis of segment frequencies that was presented in Figure 2.1, to show how segment frequencies change over time under a null model in which, contrary to the one in Figure 2.1, regular sound change is implemented. For illustrative purposes, I kept track of the frequencies of the segments 'a', 'e', and 'i' over the course of a single simulation where 50000 sound changes are applied to the

27
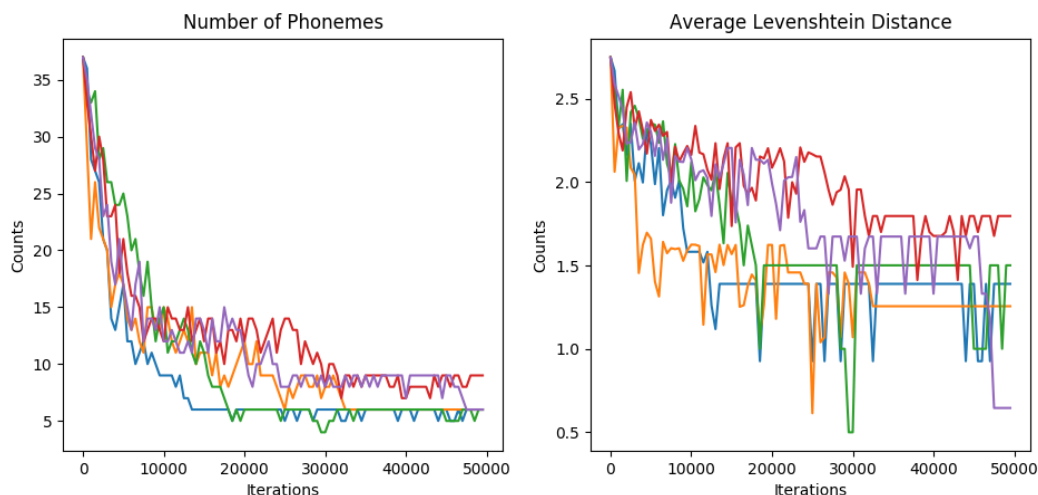
Figure 2.4: The output of five sound change simulations, where 50000 sound changes are applied to the 150-word English list. Number of possible segments: 41.

lexicon. It is clear that the output is very different from any model of genetic drift, because it is very common for the segments to abruptly drop to frequency '0' or rise to high frequencies. While in real languages we do not have cases in which only a single vowel is reported for a language, or at least they are extremely rare (cf. Kuipers, 1960 and Halle, 1970), the fact that a common phoneme can disappear as a consequence of sound change, or that some mergers can create very frequent phonemes, is compatible with regular sound change. This shows that once we model regular sound change, language drift follows dynamics that are completely different from those associated with genetic drift, for which patterns like the one in Figure 2.3 would be implausible, given the independence of reproductive events.

Now, we can see what happens to the phonemic inventory and to phonological dispersion over time, by running the simulation for five parallel runs, in Figure 2.4. The results are surprising and, to some extent, counterintuitive. In a null model, what we expect to see is lines branching more or less at random in the space of the possible outcomes, while in this case the change seems to follow a particular direction: the number of phonemes decreases, as does the average Levenshtein distance. An inspection of the output reveals that at the end of the run these phonemes are distributed around four or five different wordforms that

cover the whole vocabulary: this means that the lexicon has moved toward states of massive homonymy. The outcome is similar to that found in Martin (2007), with the difference that in the model by Martin high-frequency segments are more likely to reappear in the course of the iterations, while here there is no such bias: as Figure 2.3 shows, high-frequency segments can completely disappear from the lexicon after a merger, and low-frequency segments can significantly increase their probability mass by merging with high-frequency ones.

This is an outcome that must be explained.

2.3.2 The probability of the sound change functions

One of the arbitrary decisions that was made in the design of the simulation is the equal probability of applying the four sound change functions to the lexicon. For this reason, I replicated the experiment by randomly weighting the functions with four different weights {1, 2, 3, 4}, so that one of the four functions is four times as likely as the least likely, one is three times as likely, and the last one is twice as likely. This means that in some simulations onsets will be changing faster than codas, or vice versa.

The simulations, summarized in Figure 2.5, still show drift over time. The factors that cause a drift do not appear related to the probabilities of the sound change functions.

2.3.3 The set of possible segments

One of the possible explanations for this pattern is that the model favors mergers over splits just because in the 150-word English list that we used, the great majority of the symbols we used to represent phonemes was present (37 out of 41). It is clear that if X and Y are selected at random in the set of possible segments, the outcome of the change X > Y is much more likely to be a segment already present in the language rather than a segment outside of it. Notice, though, that this would predict a fixation point at around half of the possible segments ($\approx$20), while the number of final segments appears to be between 5 and 10, and therefore lower, in the simulations run so far.

In a second simulation, I expanded the number of consonants and vowels from 15 vowels
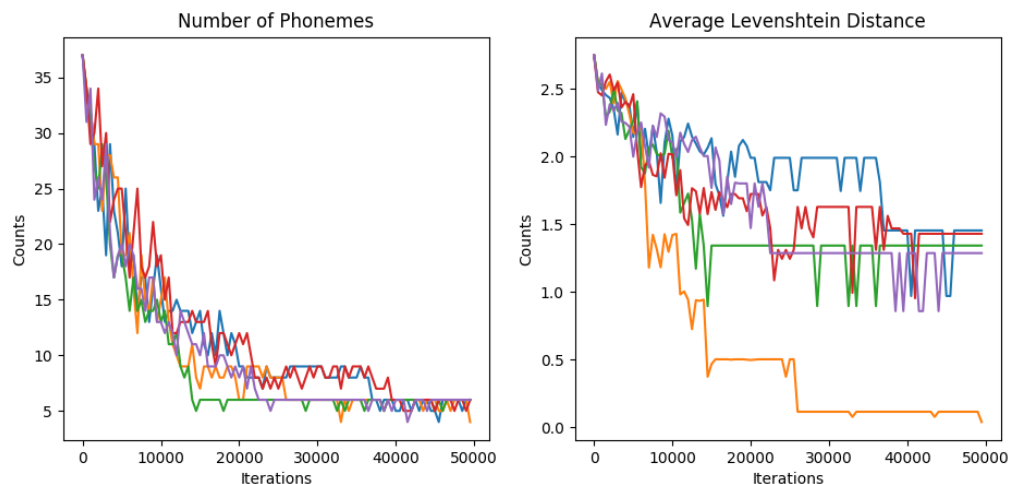
Figure 2.5: The output of five sound change simulations, where 50000 sound changes are applied to the 150-word English list. Number of possible segments: 41. The probability of the four sound change functions is arbitrarily weighted.

and 26 consonants to 30 vowels and 52 consonants, while the initial lexicon is the same as the previous experiment. As a consequence of this change, there is a pressure toward more splits.

Figure 2.6 shows that this change has the effect of slowing down the trend, but it does not revert it. In particular, we see that the number of phonemes increases from 37 to about 65 in the first iterations, but then it starts to decrease, and by the end of the simulation it goes down to ≈20. A similar pattern can be observed for the average Levenshtein distance, where none of the simulations end up in a state comparable to the starting point.

This simulation shows that it is not the number of possible segments that is responsible for this trend. It also shows that the reduction in the number of segments also occurs when we control for the relative frequency of mergers and splits: by making the size of the inventory bigger, we are in fact introducing a bias toward splits, because a segment outside of the inventory should have a higher probability of being the outcome of a sound change than a segment within the inventory, when the number of segments represented in the lexicon is less than 41. This fact makes Labov (1994)'s conjecture even more intriguing, because it shows that one can obtain a drift toward smaller phonemic inventories even when

30
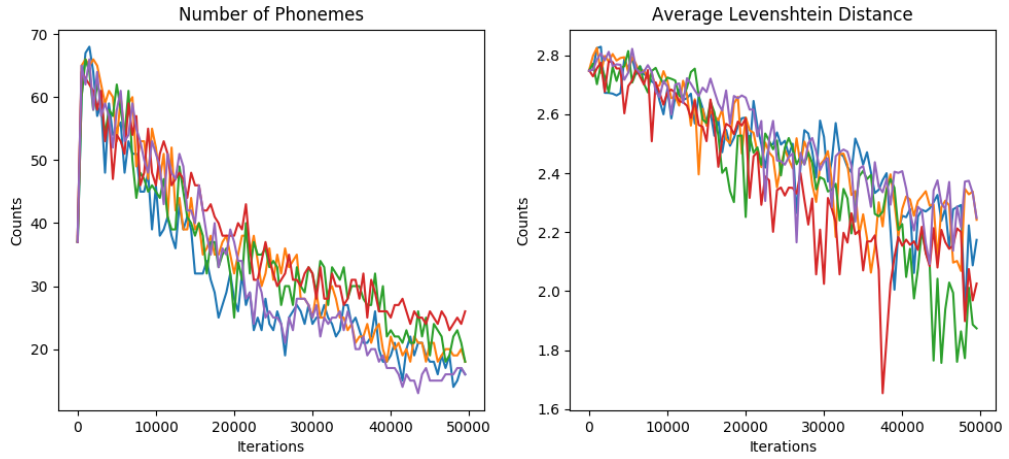
Figure 2.6: The output of five sound change simulations, where 50000 sound changes are applied to the 150-word English list. Number of possible segments: 82.

splits are not less frequent than mergers.

In the next section, we test whether the drift depends on the initial distribution of the segments.

2.3.4   The initial distribution of segments

Since the wordlist employed so far is derived from the English lexicon, the distribution of its segments is not uniform. In order to evaluate whether a skewed initial distribution had an impact on the outcome of the simulations, I replicated the experiment by creating a wordlist of length 150, sampling uniformly through the space of consonants and vowels, and I re-ran the experiments of the previous sections. This strategy guarantees a uniform distribution in the initial lexicon (like in Martin, 2007 and Sayeed & Ceolin, 2019). As we see in Figure 2.7, in the conditions of the first simulation, we obtain exactly the same results.

Similarly, if we generate a wordlist according to the segment inventory we predetermined, and then we double the inventory of possible segments, like in 2.3.3, the results are comparable to what we obtained in the experiment on the English wordlist (Figure 2.8).

31
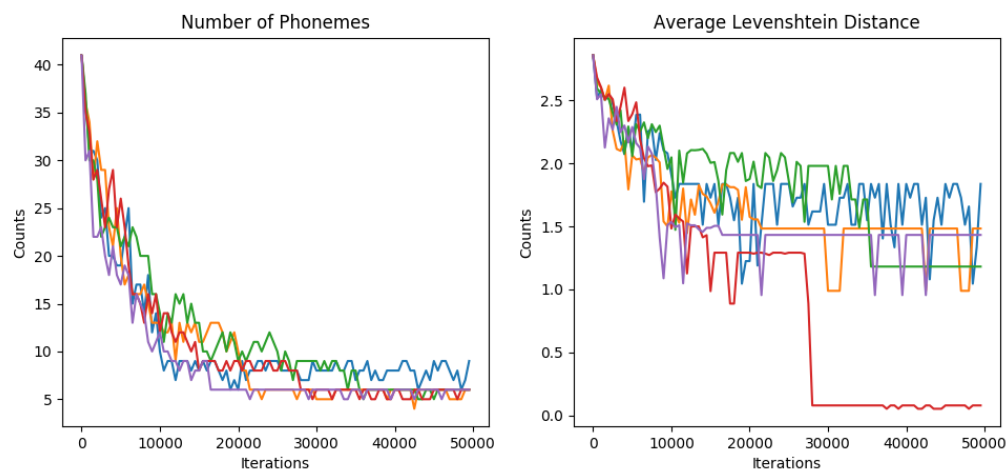
Figure 2.7: The output of five sound change simulations, where 50000 sound changes are applied to a 150-word artificial vocabulary generated using the same segments represented in the English vocabulary, through uniform sampling. Number of possible segments: 41.
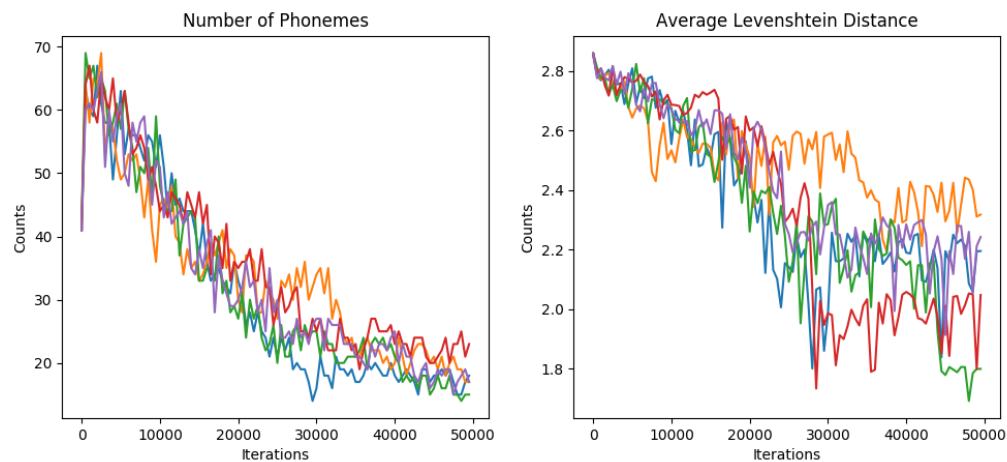
Figure 2.8: The output of five sound change simulations, where 50000 sound changes are applied to a 150-word artificial vocabulary generated using the same segments represented in the English vocabulary, through uniform sampling. Number of possible segments: 82.
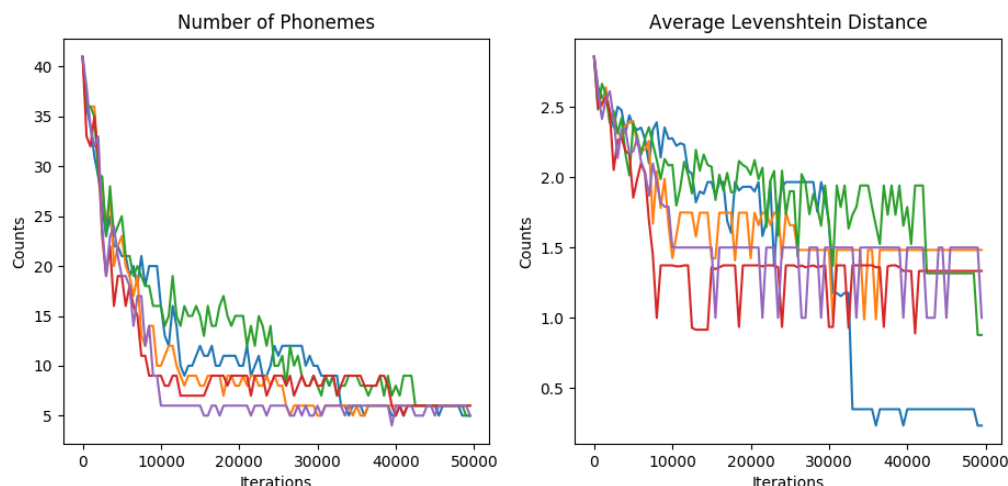
Figure 2.9: The output of five sound change simulations, where 50000 sound changes are applied to a 500-word artificial vocabulary generated using the same segments represented in the English vocabulary. Number of possible segments: 41.

2.3.5 The length of the wordlist

I also decided to replicate the simulations using a longer artificial list of 500 words instead of 150.[15] The results are practically identical to the previous ones (Figure 2.9 and Figure 2.10). The length of the wordlist does not appear to be a relevant factor.

2.3.6 Reliability of mergers and splits as the main actors in sound change

Another question that needs to be addressed is whether modeling sound change just in terms of mergers and splits is reasonable. In particular, are there other sources of phoneme creation that can contribute, like splits, to an increase of the number of segments, and therefore an increase of phonological dispersion?

This question has already been addressed in Ringe (2011). In response to Atkinson (2011), Ringe (2011) investigates the changes that affected the phonemic inventory of four Indo-European families (Indo-Iranian, Germanic, Greek and Latin) starting from Proto-Indo-European, in order to determine the frequency of mergers with respect to other kinds

---

[15]Evaluating lists of higher length would be more difficult, because computing the average Levenshtein distance requires $O(n^2)$ time, and therefore the time complexity grows exponentially.
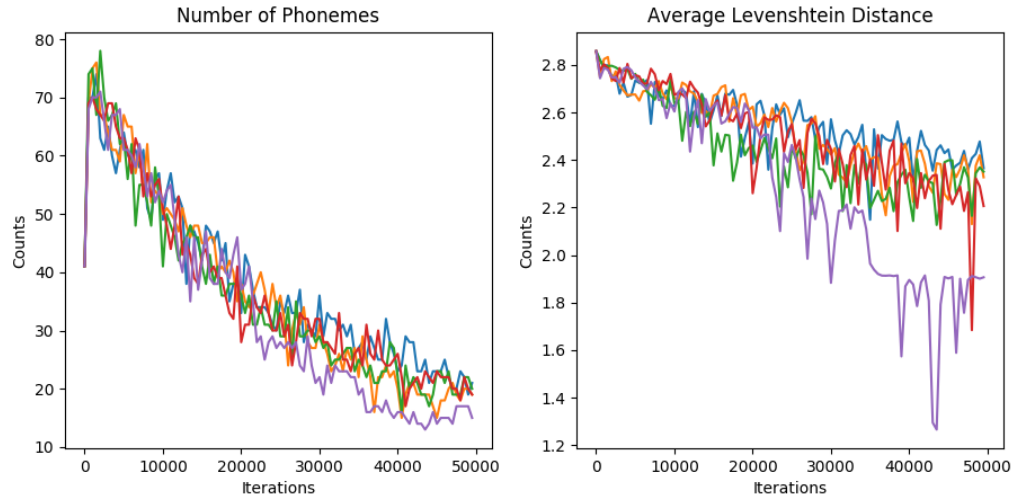
Figure 2.10: The output of five sound change simulations, where 50000 sound changes are applied to a 500-word artificial vocabulary generated using the same segments represented in the English vocabulary. Number of possible segments: 82.

of sound change.

Interestingly, there is one process which turned out to have a prominent role in removing and creating segments: a phoneme with multiple places of articulation can be reanalyzed as a sequence of two phonemes (resolution), while a sequence of two phonemes can be reanalyzed as a single phoneme (contraction). Examples of these types are the loss of labiovelar consonants in Latin (/kʷ/, /gʷ/ > /kw/, /gw/) and the creation of vowel diphthongs from vowel sequences in Latin and Ancient Greek, respectively. In Ringe's pilot study, these phenomena were quite widespread. In particular, contractions, namely the reanalysis of particular sequences as an independent phoneme, might constitute the force that balances the loss of phonemes caused by the frequency of mergers.

In order to corroborate this hypothesis, we can look at other language families for which historical data are available. A great source that can be consulted is Sammallahti (1988), which contains accurate reconstructions of the phonological systems of several Uralic proto-languages. Another source is Robbeets (2003), which reconstructs the phonemic inventories of five language families traditionally ascribed to a single Altaic macro-family (Japanese, Korean, Tungusic, Mongolian and Turkic). A summary of the survey can be found in

34

Lineage	Contractions	Borrowing	Splits	Resolutions	Mergers
Proto-Indo-European to Proto-Indo-Iranian	5		3		11
Proto-Indo-European to Proto-Germanic	3		2	4	8
Proto-Indo-European to Greek	14		1	2	12
Proto-Indo-European to Latin	6			6	8
Proto-Indo-Iranian to Sanskrit	1	5	3	2	4
Proto-Germanic to Old English	4		14		8
Proto-Uralic to Proto-Samoyed			2		3
Proto-Uralic to Proto-Finno-Ugric	5				
Proto-Samoyed to Proto-South-Samoyed					1
Proto-Samoyed to Proto-North-Samoyed					1
Proto-Finno-Ugric to Proto-Ugric			1		5
Proto-Finno-Ugric to Proto-Finno-Permic					1
Proto-Japanese to Old Japanese	8				
Proto-Korean to Middle Korean	4				
Proto-Tungusic to Manchu		1	1		2
Proto-Mongolic to Mongolian	20	1	1		1
Proto-Turkic to Turkish	1	2			8
TOTAL	**71**	**9**	**28**	**14**	**73**

Table 2.1: A summary of processes regulating phoneme creation and deletion.

Appendix B.

The counts for Indo-European (which are taken from Ringe, 2011), along with those for Uralic and Altaic, are summarized in Table 2.1. The patterns are confirmed: if anything, contractions are even more widespread in the changes that affected families different from Indo-European. This also provides a possible solution to Labov's conjecture: if mergers are more common than splits, but contractions are as common as mergers, then contractions provide stability over time. Also, loanwords constitute an additional source of phoneme creation: even though the phenomenon is uncommon, new phonemes entered the lexicon via loanwords from other languages in Sanskrit, Manchu, Mongolian and Turkish. Resolutions, instead, are not reported outside of Indo-European.

As a consequence, it might be appropriate to introduce the possibility of contractions in the null model.

2.3.7 Pseudo-Contractions

Modeling contractions artificially is problematic, because contractions affect the length of the word, a factor that is usually kept constant in models of sound changes: this is true in Nowak et al. (1999), Martin (2007) and Graff (2012).

For this reason, I decided to implement 'pseudo-contractions': a sound change reduces two adjacent segments to a new segment, but instead of reducing the length of the word, an extra segment which is already represented in the lexicon is added. The algorithm works as follows:

a. Pick one C-place in the syllable (onset, coda).

b. Select a segment X_C for that position and a segment X_V among the vowels. Both X_C and X_V are part of the lexicon.

c. Select a C-segment Y_C which is NOT part of the lexicon.

d. X_C and X_V contract to Y_C.

e. X_V is substituted with another nucleus (Y_V) present in the lexicon in the current state.

Point 5e is what makes the contraction function a 'pseudo-contraction'. The fact that the vowel is chosen at random is necessary to avoid any bias toward specific segments.

This function will introduce in the model sound changes of the kind:

pen > (pm) > pam

Where a CV combination ('en') contracts into a new symbol ('m') and then a vowel among those available in the language is added ('a'). Like mergers and splits, contractions can apply vacuously if the particular CV or VC combination is not represented in the lexicon. However, since their outcome is guaranteed to introduce a new symbol in the inventory, they serve the purpose of modeling a realistic change that increases phonological dispersion.

With contractions, the updated list of functions that are applied to the lexicon is the following:

a. Change in the Onset (conditioned on the Nucleus)

b. Change in the Nucleus (conditioned on the Onset)
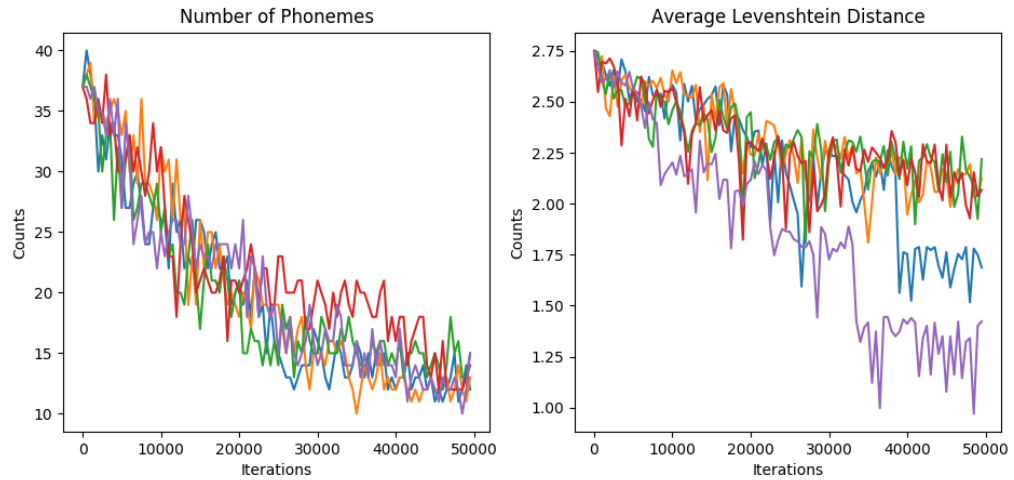
Figure 2.11: The output of five sound change simulations, where 50000 sound changes are applied to a 150-word artificial vocabulary generated using the same segments represented in the English vocabulary. Number of possible segments: 41. Contractions included.

c. Change in the Nucleus (conditioned on the Coda)

d. Change in the Coda (conditioned on the Nucleus)

e. Contractions of a CV sequence to a new segment.

f. Contractions of a VC sequence to a new segment.

With each of the functions occurring with equal probability.

The first simulation in the original conditions (150-word English list, and 41 phonemes), in Figure 2.11, differs from the first one (in Figure 2.4) for the fact that we see more cases of phoneme creation in the course of the simulations, and that after 50000 changes are applied, the lexicon converges to a state in which the range of segments is [10,15] instead of [5,10], but the trend toward a reduction of both the phonemic inventory and phonological dispersion is still evident.

The fact that contractions are not sufficient to counter the pressure toward dispersion is even more evident when we replicate the simulation on the condition with a wider space of possible segments (82 instead of 41), in Figure 2.12.

In this condition, the lexicon is initially pushed toward a state in which there is a high
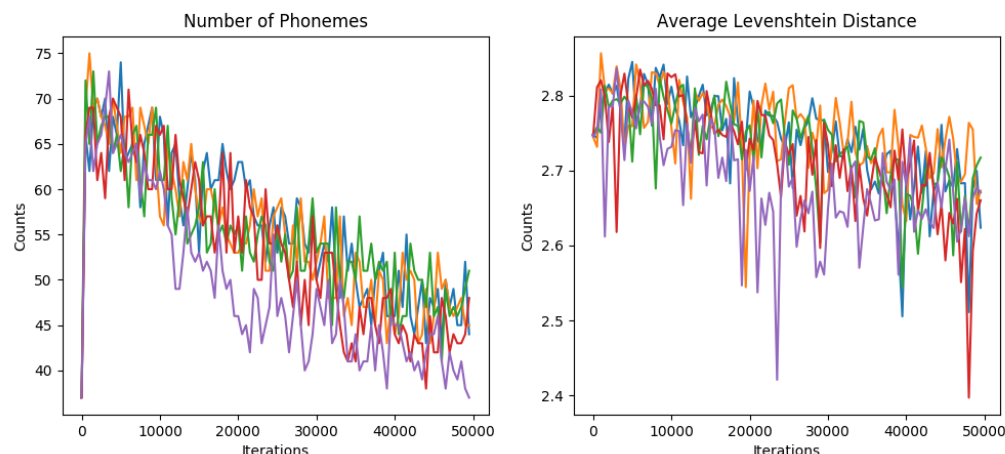
Figure 2.12: The output of five sound change simulations, where 50000 sound changes are applied to a 150-word artificial vocabulary generated using the same segments represented in the English vocabulary. Number of possible segments: 82. Contractions included.

number of phonemes and there is high dispersion, and then it gradually retracts toward states with less segments and lower dispersion. Even in this condition, therefore, we see a clear trend toward a reduction of the phonemic inventory and of phonological dispersion in the lexicon over time.

Finally, I re-ran the main experiment by increasing the probability through which contractions apply to the lexicon. In a first simulation, contractions affecting the onset or the coda position apply with a probability which is two times higher than the probability of a regular sound change targeting that position (Figure 2.13), while in a second simulation the probability of a contraction is three times higher than that of a regular sound change (Figure 2.14). We can see that increasing the likelihood of contractions slows the drift down, but does not prevent it.

2.4   Discussion

The results of the previous section show that both the number of segments and the average distance among words decrease over time when a model of regular sound change is implemented. This result is independent from:
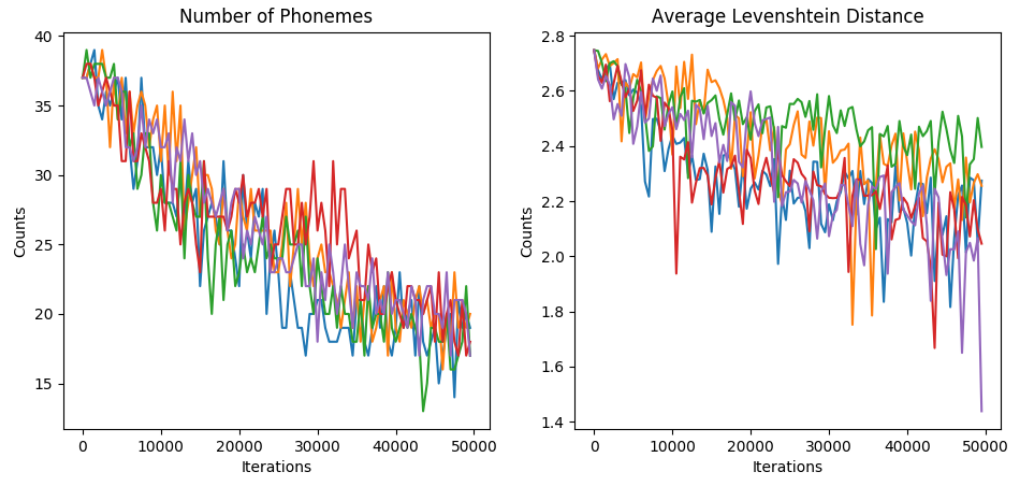
Figure 2.13: The output of five sound change simulations, where 50000 sound changes are applied to a 150-word artificial vocabulary generated using the same segments represented in the English vocabulary. Number of possible segments: 41. Contractions included with probability of occurrence doubled.
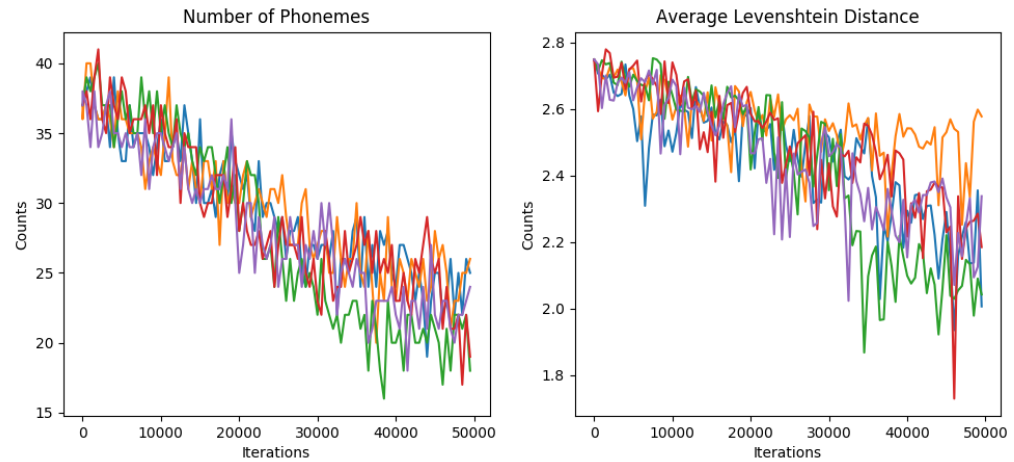
Figure 2.14: The output of five sound change simulations, where 50000 sound changes are applied to a 150-word artificial vocabulary generated using the same segments represented in the English vocabulary. Number of possible segments: 41. Contractions included with probability of occurrence tripled.

(A)	⟶	(B)	/e/ > /a/ +2 Minimal Pairs (MP)
sad		sad	
set		sat	
far		far	
her		har	
(A)	⟵	(B)	-2 MP complicated in a single step

Table 2.2: Minimal pair creation is not always reversible.

a. the probability of a sound change targeting one position of the word more frequently than other positions

b. the size of the segment inventories (and, indirectly, the likelihood of mergers over splits)

c. the initial distribution of the segments in the lexicon

d. the length of the wordlist

e. the presence of contractions to introduce new segments in the lexicon and their likelihood of occurrence

A first possible conclusion is that the model is not capturing sound change in the appropriate way. However, the next subsection shows that this behavior is the result of well acknowledged properties of the sound changes described so far.

2.4.1 The irreversibility of mergers

By taking a look at the output words of the artificial model, one can identify the patterns which are causing the results of the simulations. For explanatory purposes, instead of keeping track of the average distance among words, I will just report minimal pair counts, since the two measures are both proxies for phonological dispersion, but the latter is more intuitive.

The pattern in Table 2.2 is frequent in the simulations. At a certain stage (A) there are no minimal pairs within a subset of the words, but each pair shows a contrast of at

least two segments. Then, a stage (B) is reached, in which a merger of two nuclei causes two minimal pairs to appear in the subset. A property of this stage is that this result cannot be reverted in a single step: this is a context in which mergers are irreversible. A split of the nucleus conditioned on the onset, for instance, could potentially revert back the minimal pair 'far'-'har' to its previous stage, but it would keep the minimal pair 'sad'-'sat' in the language. The opposite is true for a split of the nucleus conditioned on codas. In other words, we have pushed the lexicon toward a state where two bigrams ('sa' and 'ar') are attested in more than one word, at the expense of other bigrams, and reverting the change is costly, because no single change can decrease the frequencies of the two bigrams simultaneously. This problem would also arise for other sound changes not included in this model, like deletion or epenthesis. Reverting the merger would require at least two splits or contractions.

Crucially, the same would not be true in case of phonemic splits: since splits are always conditioned on an environment, a conditioned merger can always revert them back to the previous stage. This means that for each change that pushes a language toward a state characterized by high dispersion, there is always the possibility that a change will pull the language back to its previous stage. The same is true for contractions: contractions can be reverted through their complementary sound change, resolutions.[16]

The irreversibility of mergers has been carefully investigated in the sound change literature (Labov, 1994, 331). While there are historically attested cases of apparently reversed mergers (cf. Chomsky & Halle (1968) and Labov et al. (1991) on the *mate/meat* merger in Early Modern English), scholars agree that these are cases that do not simply involve a regular sound change, but require the preservation of the contrast at the phonetic level (and at the level of the underlying representation), even though speakers might fail to perceive it; a situation which Labov describes as a 'near-merger' case. For the purposes of our problem,

---

[16]Even though I did not implement resolutions, conceptually they could be implemented in a way similar to contractions: one segment is deleted, two segments available in the language appear at its place, and then the previous nucleus is deleted to force the new word to have length 3 instead of 4. The reason why I did not implement this sound change is that it would constitute an additional pressure against dispersion, which at this point would not be interesting, given that we acknowledged that regular sound change already leads to low dispersion states.

the irreversibility of mergers creates an asymmetry in the way sound change interacts with minimal pairs: in fact, mergers are expected to create minimal pairs at a rate which is faster than the rate of minimal pair deletion associated with splits and contractions. This has an indirect effect on the global number of segments: less variation in possible conditioning environments means less chances of a split or a contraction that increases dispersion by introducing a new segment, rather than applying vacuously or completely replacing an existing segment with another one, with no consequences in terms of phonemic dispersion.

Moreover, minimal pairs can be lost after a merger if the merger results in homonymy, as in Table 2.3. Notice that the irreversibility of mergers causes some additional problems in this second case. When a pair of words reaches stage (C), going back to the previous stage is impossible. As a consequence, the words are bound to the same outcomes if their phonemes are affected by sound change. This is another outcome that contributes to reducing the number of phonemes and phonological dispersion in the lexicon, and ultimately can lead to a saturation of the whole vocabulary.

This scenario is, of course, unrealistic. Languages have many ways of disambiguating in case of massive homonymy, like word compounding, word borrowing, and the use of synonyms. However, for the purposes of our investigation, it is important to note that sound change is not a neutral force when it comes to the shape of the lexicon: expecting the size of phonemic inventories and phonological dispersion in the lexicon to be stable over time is unjustified, if we acknowledge the fact that mergers are irreversible. This fact is sufficient to predict low phonological dispersion over time, because while it is possible to create homonymy via regular sound change, retreating from homonymy via sound change is impossible.

Note that there would be a way to solve this problem by proving the existence of unconditioned phonemic splits: if unconditioned mergers, or mergers that lead to homophony, could be reverted, than the problem would be solved. However, as we said in the introduction of this chapter, unconditioned splits are not attested as a type of sound change.[17]

---

[17]There is one potential case of surface unconditioned split: if, through some mechanism of lexical diffusion, a word, or a class of words, shows the effect of a regular sound change, while the rest of the lexicon is

(B)	→	(C)
sad		sat
sat		sat
(B)	←	(C)
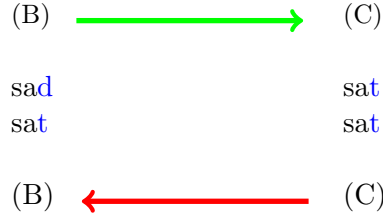
Table 2.3: Mergers and homonymy

Another point worth mentioning is that these considerations are independent from the phonemic representation of the segments: implementing phonological features in the system would not solve the problem of the irreversibility of mergers.

In the next subsection, we address the question of how lexical replacement might interact with this system.

2.5 Lexical replacement

From the discussion in the previous section, it is clear that regular sound change has the effect of increasing phonological dispersion in the lexicon over time, if we do not have a formal mechanism to revert mergers. While this has the potential of endangering communication, we know that there are at least two strategies that speakers can adopt to circumvent the problem, and they both involve lexical replacement.

First, as previously mentioned, it is well known that word compounding is one of the main strategies that can be employed to avoid ambiguity, and this has been proven to be the optimal strategy to increase the amount of information transmitted by languages (Nowak et al., 1999, Plotkin & Nowak, 2000). For instance, one of the most studied cases of this kind is the fact that after several mergers had reduced the phonemic inventory of Middle Chinese, word compounding was extensively employed as a way to avoid ambiguity (cf. Sampson, 2015 and Behr, 2015). Moreover, in North American regions affected by the

---

unaffected. While technically this does not revert a merger, because it would require speakers to guess which were the words that had form X and which were those that had form Y before X and Y merged, it can have the effect of breaking homonymy. Since this phenomenon is analogous to what happens in cases of lexical replacement through borrowing, the results of the next subsection can be interpreted as the outcome of a model that allows lexical diffusion.

pin/pen merger, speakers can refer to the second word as *ink pen* in order to disambiguate (Finegan, 2014), using a process of word compounding which is parallel to that identified in Middle Chinese.

A second strategy is borrowing from other languages or dialects. Borrowing is not conceptually different from population migration in biology, and it is well known that a low amount of migration can prevent genetic drift between two populations (Bodmer & Cavalli-Sforza, 1968, Felsenstein, 1976). Borrowing can also lead to an increase in the phonemic inventory, even though we saw in Table 2.1 that it is less common than splits and contractions as a form of phoneme creation (but see Gylfadóttir (2018) for a detailed case study on phonemic borrowing).

In theory, a third option for lexical replacement is word coining. However, while both word compounding and borrowing are well attested in languages, it is less clear how often speakers invent new words from scratch without reference to pre-existing material, which can be considered a form of word compounding. This would be the equivalent of entries in etymological dictionaries that are completely empty, and of onomatopoeia words. In the case of languages for which we have a long tradition of historical records, this seems to be a rare phenomenon, and therefore I will not take it into account.

Word compounding and borrowing can be both integrated in a sound change model. Modeling word compounding would be difficult, because it requires abandoning the idealization of a lexicon which is composed of (C)V(C) syllables and allowing words to be longer. Moreover, it would also be uninteresting: the fact that allowing longer sequences of words will increase dispersion is straightforward. On the other hand, a model of borrowing would be easier to implement, and would also be more interesting from a theoretical viewpoint.

One of the motivations that previous studies had for studying phenomena like the presence of phonetic biases behind borrowing (Martin, 2007) or lexical inhibition as a factor in lexical change (Graff, 2012) was the need to counteract the presence of phonemic drift in models of word-coining which are inspired by rich-get-richer models (Simon, 1955): in any model in which frequent segments tend to be more active in lexical replacement, since the

lexicon is finite, the prediction is lexical saturation over time.

While both Martin's and Graff's models of lexical competition are, in a sense, 'functional', because they make reference to communication, a question that one might ask is whether a neutral model of borrowing would be sufficient to address the problem of lexical saturation. Historical records show us that borrowing is overwhelming in natural languages. While languages do not easily accomodate foreign sounds in their lexicon, and they adapt the loanwords to their native phonotactics by replacing foreign sounds with native sounds (Calabrese & Wetzels, 2009), the loanwords that are compatible with the native language phonotactics can become part of the native linguistic input for the next generation of learners, and then be added to the language vocabulary. For this reason, borrowing qualifies as another component of a model of lexical change.

2.5.1   Modeling borrowing

Implementing borrowing in an artificial model is straightforward. A strategy that would be close to how borrowing works in real languages is that of memorizing one word at each iteration of the sound change model, and adding it to a list of 'potential loanwords'. When borrowing occurs, we replace one word of the vocabulary with a word randomly sampled from the 'potential loanwords' list. Since this list of 'potential loanwords' increases at each iteration, it will be more unlikely over time for the language to borrow a word that was present at the very first iterations. The analogy would be that in a modern language like English, it would not be inconceivable to see words which resemble Germanic forms, which could have just survived in other lineages and entered the language later as loanwords; similarly, if some phonemes were particularly frequent in Proto-Indo-European but they have been lost in English, their presence in neighbor languages would make it more likely for them to be borrowed, rather than phonemes that were never present at earlier stages. Note that this model would sample from a limited domain, in the sense that it would never borrow a word or a phoneme that was never present in the history of the language. This constraint is needed to avoid a scenario in which, by sampling from the entire space of

combinations, we have an explosion of phoneme combinations which are completely different from everything the language has ever seen.

An additional filter that we need, in order to make borrowing more realistic, is excluding the possibility of loanwords that are not 'compatible' with the language: in this case, we know that loanwords are usually adapted to the phonotactics of the original language (Calabrese & Wetzels, 2009), and this would require us to implement the notion of adaptation. However, it is not entirely clear to me how 'compatible' should be defined: in the history of English, as we said in the previous chapter, Old French words starting with /v-/ were borrowed into the English lexicon without being adapted to the language phonotactics, that only allowed /f-/ in word-initial position. It could be that the borrowing was possible because even though the sound was not allowed in word-initial position, it was an allophone of /f/ that could occur intervocalically (Labov, 1994). One possible definition of 'compatible', then, could be that the loanwords contain sounds which are familiar to the speakers of the language. Since our model does not include underlying representations and phonological rules, the constraint of forcing loanwords to have minimal pairs with at least one word in the vocabulary is one of the simplest ways to enforce that they be close enough to the original lexicon to be accepted without adaptation. The presence of minimal pairs also has psychological implications, because it is treated as a requirement for learning phonological contrasts in acquisition models (Yang, 2009, Cui, 2020).

To summarize, a borrowing function should:

a.   store one word in a list at each iteration

b.   when the function is called, replace one word in the lexicon with one word drawn from the list of memorized words, if it creates a minimal pair with words which are already in the native vocabulary

With the addition of this function, the null model now has seven functions which are applied to the lexicon.

a.   Change in the Onset (conditioned on the Nucleus)

b. Change in the Nucleus (conditioned on the Onset)

c. Change in the Nucleus (conditioned on the Coda)

d. Change in the Coda (conditioned on the Nucleus)

e. Contractions of a CV sequence to a new phoneme.

f. Contractions of a VC sequence to a new phoneme.

g. Lexical replacement

In the next subsection, we test the impact of lexical replacement on the English wordlist.

2.5.2 Results

In a first simulation, I ran the model on the English lexicon and with lexical replacement applied with the same probability of the other functions; which means that, on average, every six sound changes, one word of the lexicon is replaced.

The results, in Figure 2.15, show that this mechanism of lexical replacement is sufficient to stabilize the system. The number of phonemes in the late iterations is comparable to the number of phonemes at the beginning, and the same is true for the average distance.

By replicating the analysis expanding the number of phonemes (Figure 2.16), we see that lexical replacement has the opposite effect of sound change: it tends to increase lexical dispersion over time, by expanding the phonemic inventory, and therefore increasing phonological dispersion in the lexicon.

One problem with this factor is that its counterbalancing effect crucially depends on the rate at which borrowing applies. In Table 2.4, we see that by decreasing the probability of the function, in the same conditions of the simulation in Figure 2.15, the counterbalancing effect of borrowing becomes weaker.

Interestingly, even with lower values of p, the lexicon does not saturate in the same way that it did in the previous simulations: when $p=1/31$, we still end up with an average of at least ≈ 31 segments, while in the previous simulations the average number was always

47
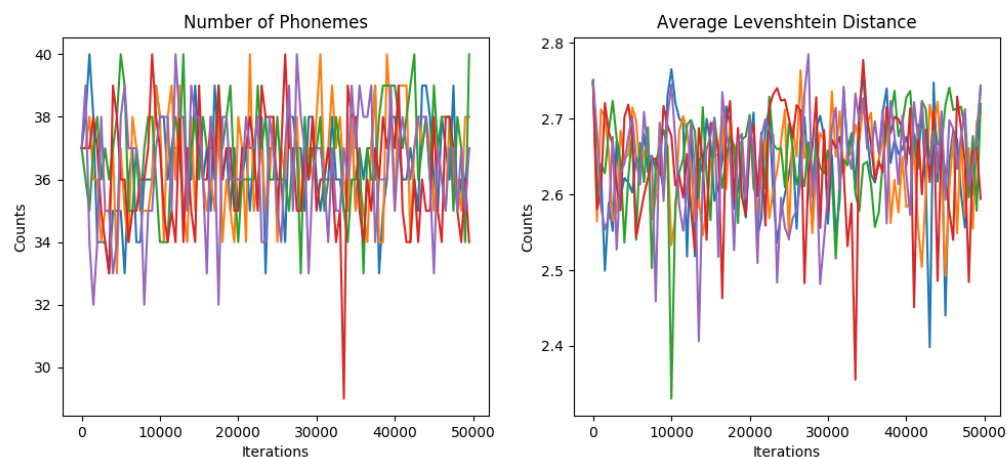
Figure 2.15: The output of five sound change simulations, where 50000 sound changes are applied to the 150-word English list. Number of possible segments: 41. Contractions and lexical replacement are included.
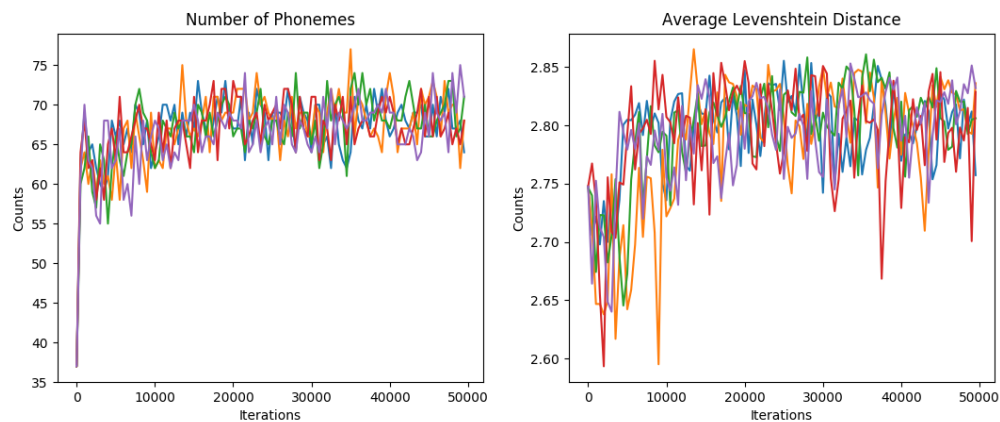
Figure 2.16: The output of five sound change simulations, where 50000 sound changes are applied to the 150-word English list. Number of possible segments: 82. Contractions and lexical replacement are included.

$p.$ of borrowing	Phonemes (range)	Average Levenshtein Distance (range)
1/7	[34, 40]	[2.589, 2.719]
1/13	[30, 38]	[2.320, 2.565]
1/19	[28, 37]	[2.235, 2.578]
1/25	[31, 36]	[2.067, 2.567]
1/31	[28, 34]	[2.169, 2.440]

Table 2.4: Expected range of phonemes and average distance as a function of the frequency of borrowing.

below 25, starting from the 41-segment inventory, even when the frequency of contractions was increased.

2.5.3 Conclusions

In the last subsection, we have seen that in the presence of lexical replacement, the drift detected in the previous sections is prevented. This fact has an analog in population genetics: as previously mentioned, low amounts of migration can have a tangible effect on preventing the drift of a population and maintaining genetic diversity over time (Bodmer & Cavalli-Sforza, 1968, Felsenstein, 1976).

It also has some direct implications for rich-get-richer models of word formation, in which frequent phonemic sequences tend to increase over time: as long as there is a mechanism of lexical replacement for which previous words can survive in the lexicon, the lexicon will not saturate. This means that we do not need to postulate any cognitive bias regulating borrowing or any hard constraint on sound change in order to avoid lexical saturation.

Of course, this hypothesis can raise further speculations. If what I'm arguing here is true, namely that the presence of borrowing might be beneficial for the phonemic inventory, then it follows that languages which are surrounded by different neighbors, and add many loanwords to their native vocabulary, should show less word compounding than languages which are isolated and less exposed to other language varieties. I leave this claim here as a speculation, since it is not clear to me how this hypothesis can be tested, but I will report that Trudgill (2002) independently suggested that isolation can be a factor affecting phonological structures.

2.6 Summary

In this chapter, we addressed the question of how sound change relates to the size of phonemic inventories and phonological dispersion in the lexicon.

When we modeled sound change, all the simulations were compatible with a scenario in which, all things being equal, languages have easier strategies to decrease both the size phonemic inventories and phonological dispersion in the lexicon rather than increasing them. If this is true, it would raise some doubts about the possibility of using low dispersion to make claims about functional pressures that optimize languages for learnability and production (Dautriche et al., 2017).

Moreover, by including lexical replacement in our artificial model, we showed that a simple model of sound change that includes lexical replacement is more stable over time from the viewpoint of the phonemic inventory and phonological dispersion. This suggests that we do not need functional considerations, at least when it comes to borrowing and word formation, to explain why languages do not exhibit high amounts of homonymy, in spite of mergers.

While this chapter shows that functional pressures are not *necessary* to model lexical change over time, it does not rule out their existence. In particular, both Martin (2007) and Graff (2012) present historical and typological data that suggest some functional bias behind the presence of certain lexical patterns, with reference to lexical retention and lexical replacement. This chapter makes the argument that such biases are not needed, but further research is required to establish if concepts like markedness or lateral inhibition can be emergent from a model of sound change (cf. Sayeed & Ceolin, 2019), or if they need to be modeled as independent factors.[18]

Finally, there is one last point that has not been addressed in this chapter. Since we have seen that sound change causes a drift mainly because of the irreversibility of mergers, one could wonder whether mergers that would cause a high amount of homonymy in the

---

[18]Appendix C shows that, like in Sayeed & Ceolin (2019), we can use the model to derive Yule's distribution of phoneme frequencies.

lexicon are typically prevented. In the next chapter, I will consider this possibility, which has been formalized in the literature as the Functional Load Hypothesis.

Chapter 3

The Functional Load Hypothesis

The Functional Load Hypothesis is one of the most discussed factors in the literature on sound change. The hypothesis states that the likelihood of a particular sound change does not entirely depend on the articulatory and acoustic properties of two sounds, but also on their contrastive function: if a language relies on a particular contrast to distinguish several words in the lexicon, then we might expect a pressure toward preserving that contrast. This hypothesis was formulated in the context of the Prague School in the first half of the 20th century, in particular in the works by Martinet, and it is consistent with the fact that phonemes are defined as contrastive units. King (1967a), who was the first one to attempt an empirical test of the hypothesis, stressed the importance of studying functional load with the aim of explaining sound change:

> 'Martinet has cogently and persistently argued that functional load has rich yet unexplored possibilities for the linguist who attempts to plumb the causality of sound change' (King, 1967a, 833).

The first attempts to test functional load found no relationship between phonemic contrast and sound change (King, 1967a, Surendran & Niyogi, 2006). However, recent works which looked at a wider sample of languages found a functional load effect in various cases of sound change (Gurevich, 2004, Silverman, 2010, Bouchard-Côté et al., 2013, Wedel et al., 2013, Eychenne & Jang, 2015, Babinski & Bowern, 2018), in contrast with previous findings. On the one hand, linguists have been skeptical toward functional load, either backing King's and Surendran and Niyogi's findings, or pointing to cases where the hypothesis makes the

wrong predictions (Weinreich et al., 1968, 133-137, Sampson, 2013, 2015). On the other hand, the methodological advances and the availability of large electronic corpora have contributed to the suggestion that previous failures of validating the hypothesis were mostly due to the lack of statistical power (Wedel et al., 2013, Kaplan, 2015). I add to this point that a positive test of the Functional Load Hypothesis would be a welcome outcome for many scholars from different theoretical backgrounds: from those traditionally associated with 'functional' views on language, to formal phonologists (cf. Newmeyer, 2003). For this reason, it is of no surprise that the recent positive tests have been welcomed without critical engagement (Sampson, 2015).

In this chapter, I address this longstanding question by looking at many cases of sound changes in progress, and by trying to determine when functional load makes the right predictions and when it fails. Then, I draw some generalizations that will hopefully explain the inconsistency of the literature findings.

3.1 Functional Load in the '60s

The term 'functional load' can be traced back to Martinet (1955), who refers to the number of minimal word pairs whose difference relies on a single phonemic contrast as the *rendement fonctionnel* of such contrast. While most of the first works referring to it translated the concept with 'functional load' (Hockett, 1955, Kučera, 1963, Wang, 1967, King, 1967a), the translation 'functional yield' is also used (Greenberg, 1959). In Martinet's work, we find the first explicit suggestion that this measure can be a factor in sound change, though similar ideas had been around in the first half of the 20th century (Gilliéron, 1918, Mathesius, 1929, Jakobson, 1931, Trnka, 1931, Trubetzkoy, 1939).

Several proposals for implementing functional load appeared in the literature in the following years. The most influential one is by Hockett (1955, 1967), who proposes to use Shannon's formula for estimating the entropy of a text of written English (Shannon, 1951). Hockett's idea is that the loss of a phonemic contrast through a merger of two sounds x-y in a language yields to a potential increase of the ambiguity of its lexicon,

since less phonemic sequences are possible. In information-theoretic terms, the entropy of the language decreases. The difference in entropy between the pre-merger and the post-merger lexicon can be thought as the functional load carried by the contrast. This can be summarized by the following formula:

$$FL(x, y) = \frac{H(L) - H(L_{xy})}{H(L)}$$

Where $H(L)$ represents the entropy of the language before the merger between x and y, and $H(L_{xy})$ represents the entropy of the language after the merger. The calculation of $H(L)$ depends on assumptions about the level of representation, and it will vary depending on whether we want to study phonemic contrasts globally, phonemic contrasts in context, or featural contrasts. I will expand on this point later.

Several papers proposed alternative measures or improved on Hockett's formula (Greenberg, 1959, Rischel, 1962, Kučera, 1963, Wang, 1967), but the most influential work of the period was King's attempt to empirically validate functional load by looking at some cases of historically attested mergers in Germanic (King, 1967a). King derives a functional load measure through a formula that multiplies the frequencies of two phonemes in a text and a measure of their contrastiveness in their possible phonological environments. For King, a phonological environment is a sequence of the type X__Y, in which X is the preceding phoneme and Y is the following one. Word boundaries are coded as independent phonemes. The formula has been presented in King (1967b), and it ultimately reduces to:

$$L(i, j) = \frac{1}{N^2} \left(\sum_{k=1}^{m} f_{ik} * f_{jk} \right)$$

Where N is the length of the texts (in phonemes), k is one of m possible phonological environments, and f_{ik} and f_{jk} are the number of occurrences of two phonemes i and j in the environment k. The result is a functional load measure $L(i,j)$ that represents the contrastiveness of the two phonemes. The final measure can be thought as the covariance of two vectors of length m representing the distribution of the two phonemes over all possible

environments. A complementary distribution would yield the value 0, since one of the two factors inside the summation will always be 0, while if the two phonemes co-occur in the same environment, the measure can increase up to a number determined by the length of the text.

Although this measure differs from Hockett's information-theoretic formula, King argues that the two are highly correlated, because essentially they are different ways of capturing the overlap between two different distributions. King (1967a) applies the formula to some phonemic contrasts at different diachronic stages in four different Germanic languages (Modern Standard German, Old Icelandic, Old Saxon, and Middle High German) before the mergers took place, in order to compare the measures calculated for the phoneme pairs that ended up merging with those calculated for the pairs that did not merge.

King tests three different versions of the Functional Load Hypothesis:

- the 'weak point' version states that mergers affect phonemic pairs which are the least contrastive in the system, compared to all possible phonemic pairs

- the 'least resistance' version states that assuming a pressure toward a phoneme x that leads to a merger, it will merge with the phoneme y which is the least contrastive with respect to x in the system

- the 'frequency' version states that given a pressure for a merger between x and y, the merger will go in the direction of the higher-frequency phoneme

His results, however, did not support any version of the hypothesis, and therefore he concluded that 'functional load, if it is a factor in sound change at all, is one of the least important of those we know anything about' (King, 1967a, 831).

King's work has been influential among linguists, and his results are probably the reason why the debate around functional load faded away by the end of the '60s.[1]

---

[1]Interestingly, even though Hockett did not test functional load empirically, he had an opinion on King's work, which was expressed as a note at the end of Hockett (1967): 'Robert D. King, "Functional Load and Sound Change" Language XLIII (1967) 831-852, appeared after the present paper was in proof. It is a sketchy report of an empirical study leading King to conclude that "functional load, if it is a factor in sound change at all, is one of the least important of those we know anything about...". This puts the matter

Another possible reason is that functional load was not part of the variationist agenda proposed by Weinreich et al. (1968). In reference to functional load, they stated:

"There are few quantitative studies bearing on it, and they suffer from a rather narrow conception of the frame in which contrasts important for communication must be maintained. They take a rather simplified approach to language by calculating the yield of oppositions among minimal pairs uttered as isolated lexical items. Other studies of functional yield have also erred by setting too narrow an environmental frame (following and preceding element), making it impossible to deal with such phenomena as 'breaking', vowel harmony, umlaut, or 'preconsonantal r' " (Weinreich et al., 1968, 134).

In that passage, they also mention that while Martinet's hypothesis of a relationship between structural considerations and sound change was corroborated by empirical studies, in particular the study of New York City English (Labov, 1966) and Moulton's study on Swiss German dialects (Moulton, 1962), it is also easy to identify cases in which mergers that lead to homonymy are not blocked. A few examples cited are the vowel mergers in the spontaneous speech of lower-middle-class English speakers in New York City (in rhotic environments, Labov, 1966) and high vowels mergers in Yiddish (Herzog, 1965). Moreover, even if the hypothesis were true, one would be left with the problem of explaining why certain languages which meet the structural conditions for a merger develop it, while others do not. These observations led Weinreich, Labov and Herzog to conclude that "the homonymy-prevention theory contributes little to the solution of the 'actuation riddle' " (Weinreich et al., 1968, 137).

---

awkwardly, since as a matter of fact we know nothing about any other factor in sound change either. King used a mere 20,000 running phonemes of text for each stage of each language, which is not nearly enough, and the general tenor of his exposition does not inspire confidence. Of course, it may well be that functional load has no significant bearing on sound change. But King has by no means proved his thesis'. At least for Hockett, the evidence provided by King was not sufficient to settle the debate.

3.2 Functional Load in the 2000s

3.2.1 Surendran and Niyogi (2006)

After the '60s, for several decades there is a lack of interest in functional load, even though references to it appear in Lass (1981), Pye et al. (1987), Ingram (1989) and Lass (1997), among others.[2] However, starting from the early 2000s, many publications begin to readdress the topic. This renewed interested is mostly due the work of Dinoj Surendran, who at the time was a master student in Computer Science at the University of Chicago (Surendran, 2003).

The reason why functional load was discussed in Computer Science departments was that the problem of estimating the contrastiveness of phonological features was relevant for automatic speech recognition (Huttenlocher, 1985, Carter, 1987), and the implementations based on information-theory were surprisingly similar to Hockett's first proposal in the '50s.

Surendran & Niyogi (2006) propose a straightforward way to estimate the entropy of a language, *H(L)*. First, they obtain a distribution over *n*-grams from a corpus *S*. Then, they use Shannon's formula to calculate the entropy of the distribution. The formula they propose is the following:

$$H_{kS}(L) = \frac{1}{k+1} \left(-\sum_{x \in X} p(x) log_2 p(x) \right)$$

In this formula, *k* represents the order of the Markov process that generated the language, which means that *k+1* is equivalent to the *n* of the corresponding *n*-gram representation. In linguistic terms, it corresponds to the amount of contextual information available before generating the next symbol (x), which can be a phoneme, a syllable, a word, or any meaningful linguistic unit of interest that is defined *a priori*. If k=1 and the level of representation is phonemic, the entropy will be calculated over bigrams, or more specifically,

---

[2]I excluded from this list works like Campbell (1996), in which it is argued that languages avoid homonymy not by preventing mergers across the board, but by exceptionally preserving the contrast in cases where homonymy would result. A similar argument is made in Blevins & Wedel (2009) and Baerman (2011). See Mondon (2009) for a critical review of this proposal.

Bigram	Frequency
at	6
tu	4
ua	4
ta	3
tt	2
aa	1
au	1
ut	1
uu	1
Total:	23

Table 3.1: Bigrams and frequencies for the toy example in Surendran & Niyogi (2006).

over the probability of observing the phoneme x after each possible phoneme. This is a generalization of King's concept of phonological environment.

An example: if $k=1$ and $S=$'atuattatuatatautuattuua', we have nine possible bigrams (see Table 3.1). We can then estimate the entropy of the language from their frequency: $H_{1S}(L) = \frac{1}{2}[-(\frac{6}{23}log_2\frac{6}{23} + \frac{4}{23}log_2\frac{4}{23} + ...\frac{1}{23}log_2\frac{1}{23})] = 1.43$. The estimate is influenced by both k and S: in the limit, as both numbers approach infinity, $H_{kS}(L) \approx H(L)$.

With this formula, we can estimate the entropy of a language in a state in which it has a contrast, and the entropy of the language after such a contrast is lost, by just replacing the two symbols that are involved in the merger with a third symbol (an archiphoneme) and calculating the entropy of this second state. This will lead to a decrease in entropy. The information-theoretic interpretation of functional load is the difference in entropy between the two states of the language, and the estimation of such difference is shown to be robust to different values of k and S. This is consistent with King's observation that functional load estimations are well correlated even when the environments are simplified.

Surendran & Niyogi (2006) use the formula to quantify the functional load of the contrast between word-initial /n/ and /l/ in Hong Kong Cantonese, for which an ongoing merger is reported (Zee, 1999, To et al., 2015). However, they show that this merger is not associated with low functional load when compared to other contrasts among coronal segments, basically confirming the skepticism by King (1967a).

Several studies in the following years, instead, begin to report cases in which functional

load seems to be correlated with the absence of mergers. Silverman (2010) presents a convincing case study on Korean, showing that among the possible synchronic assimilation patterns that could affect Korean consonant clusters, only those which do not lead to a high amount of homophones are attested.[3] Bouchard-Côté et al. (2013) apply King's formula to a set of (automatically) reconstructed sound changes in a large dataset of Austronesian languages, and they find a functional load effect. Eychenne & Jang (2015) use Surendran & Niyogi (2006)'s entropy formula, calculated using phoneme trigrams, to show that the /e/-/ɛ/ merger in Seoul Korean carries low functional load. Cohen Priva (2017) notes that by applying Surendran & Niyogi (2006)'s formula to North American English final-/t/ deletion, one would not expect final-/t/ do be deleted rather than other consonants, which carry a lower functional load, but he suggests an alternative formula, based on average rather than local informativity, to explain why /t/ is deleted instead of other segments. Babinski & Bowern (2018) investigate several mergers in nine Australian languages and find that they are associated with a low amount of minimal pairs, suggesting that functional load might explain their occurrence.

All these works uniformly conclude that at least some version of the Functional Load Hypothesis must be true.

3.2.2   Wedel et al. (2013)

A work which deserves a detailed discussion is Wedel et al. (2013). In their study, Wedel et al. collect a large sample of ongoing mergers in different contemporary languages, and use a Mixed Effects Logistic Regression model to test the correlation between factors associated with pairs of phonemes (like relative type and token frequency) and the presence of mergers involving those pairs in some language varieties. In their model, a dependent variable is coded for many possible phonemic pairs in several languages, and the variable distinguishes phonemic pairs for which mergers have been reported in the literature ('1') versus phonemic

---

[3]Note that Silverman uses the number of minimal pairs as a proxy for functional load, focusing on the traditional suggestion by Martinet that functional load is directly proportional to the amount of homophones that a merger would introduce in the language.

pairs for which no merger has been reported in the varieties of a language ('0').

The model has two independent variables. The first independent variable is the number of minimal pairs associated with each pair of phonemes, which is a proxy for the functional load of the phonemic contrast, following Martinet's initial proposal. In order to control for phonetic similarity, the phonemic pairs that are compared do not differ by more than one feature. For instance, if a language undergoes the /f/-/θ/ merger, it might be informative to ask whether /f/-/v/ and /θ/-/ð/ are also reported as mergers, but it would not be informative to extend the same question to contrasts like /f/-/k/, because since they differ by two features (continuant and place), we might already expect a merger between the two sounds to be unlikely. The second independent variable is the frequency of the most common phoneme of the pair, which is shown to be the measure that yields the best model fitting, among those that could be derived for each phonemic pair (like the frequency of the least common pair, or the sum of the frequencies of the two phonemes).

Their model shows that the number of minimal pairs is inversely correlated with mergers: phonemic contrasts that are represented by a high amount of minimal pairs tend not to be lost, while contrasts for which minimal pairs are not present can be lost. Phoneme frequency is marginally significant, and only in the absence of minimal pairs. Crucially, Wedel et al. also try to substitute the raw number of minimal pairs with the entropy measure proposed by Surendran & Niyogi (2006), and they show that the measure does not improve the model, and in fact it is a worse predictor.

This result is the most convincing evidence provided toward the Functional Load Hypothesis, and it had great influence on subsequent work (Sóskuthy, 2013, Sampson, 2015, Kiparsky, 2016, Babinski & Bowern, 2018, Eychenne & Jang, 2015). Moreover, since its conclusions are based on a large amount of empirical data, it provides an excellent starting point for an empirical reassessment of the Functional Load Hypothesis.

In the rest of this chapter, I will evaluate the data and the ongoing mergers reported by Wedel et al. (2013), and I will try to classify them in the following three groups:

- 'functional' mergers: mergers that carry the lowest amount of functional load among

the possible mergers

- 'unclear' mergers: mergers that carry a functional load that is in the range of the other possible mergers

- 'anti-functional' mergers: mergers that carry the highest amount of functional load among the possible mergers

The aim of the chapter is to explain why the literature findings have not been consistent, and to draw some generalization about cases in which functional load makes correct predictions, and cases in which it does not.

3.3   Methodological issues

There are some methodological issues that arise when testing functional load: the choice of an appropriate corpus to derive functional load measures, the choice of an appropriate linguistic unit, the choice of the phonemic pairs whose comparison is meaningful, and the formalization of the hypothesis to test. These issues will be addressed in the next subsections.

3.3.1   Selecting a corpus

Written corpora are required to estimate word and phoneme frequencies. Databases like CELEX (Baayen et al., 1996) are popular choices to derive language statistics. Such databases can be used to obtain more accurate estimates of the entropy of a language, but their use to derive minimal pair counts is not uncontroversial.

One of the reasons why Martinet's proposal of using minimal pair counts as a proxy for functional load is appealing, is that it suggests a hypothesis about how phonemic contrasts are acquired by language learners. Recent investigations have shown that phonemic contrasts in English can be learned using a simple model which is sensitive to lexical contrast, and therefore minimal pairs (Cui, 2020), even if counts are based on a corpus of child di-

rected speech. The results suggest that paying attention to minimal pairs is sufficient for English learners to build a phonological inventory.[4]

With this hypothesis in mind, if we want to imply that functional load is a psychological factor that influences language acquisition, when counting minimal pairs we should rely on conversational corpora of child directed speech, like CHILDES (MacWhinney, 2014), rather than CELEX, because it is unlikely that children have access to words like *abjuration*, *kilohertz* and *phlegmatic* in the construction of their phonemic inventories.

A second reason to use child directed speech is that because of Zipf's law, most of the words that one can find in a dictionary are rare, and therefore they cannot be used to represent the vocabulary of an ideal speaker. This is particularly true for CELEX, in which one can find many words which are never used outside of specialized domains, and that are unknown to the majority of the speakers. While filtering out rare words has only a modest effect when calculating functional load using word token measures, like the functional load formula by Surendran and Niyogi, the effect is substantial when using word type measures, like the number of minimal pairs. Without controlling for word frequency, we would end up counting a great amount of minimal pairs that contain words which are extremely rare in conversational speech. In this case, a frequency threshold is necessary to remove the 'noise' contained in the tail of the Zipfian distribution, otherwise the amount of noise will be one of the main factors determining the number of minimal pairs.

For this reason, I will count the number of minimal pairs and estimate functional load using a restricted list of frequent words derived from CHILDES corpora. I decided to include the most frequent 5000 words in each corpus investigated. This is in the order of magnitude of the vocabulary of a 5-year-old child, an age period in which phonological development should be close to completion (Jakobson, 1931, Ingram, 1989, Hoff, 2009, McLeod & Crowe, 2018).

---

[4]There are two caveats here: first, every language has phonemes which are not very contrastive, for instance /ð/ or /ʒ/ in English, but are still learned; second, for languages whose average word length is greater than English, minimal pairs can be less common in general. For this reason, it is likely that minimal pairs are not the only cues for phonemic contrasts in the lexicon. This will be discussed in the next chapter.

3.3.2 The type of linguistic unit

Surendran & Niyogi (2006) base their calculations for Cantonese on a word unigram model. This means that the level of representation is the word, and their conditioning environment is null (k=0), because they use a unigram model (n=1). The reasoning for this model is the following: after a merger, some wordforms that were contrastive based uniquely on the contrast that was lost, namely all the minimal pairs associated with the contrast, become homophones. Therefore, if one calculates entropy over wordforms, since some of them become homophones, we have a reduction in entropy.

Deciding to calculate entropy over whole wordforms rather than a subpart of them (like bigrams or trigrams of phonemes) seems plausible when dealing with Cantonese: the average wordform is short (typically one syllable), and the presence of suprasegmental features like tones makes it more complicated to use bigrams or trigrams calculated over phonemes to represent the phonological environment. Since most of the words are monosyllabic, asking the question of how many of these words will become homophones after a merger is sufficient to measure the contrast associated with two different phonemes.

However, for languages in which wordforms can be long, and where tones are not present, focusing on minimal pairs might create a sparsity problem: the longer the average wordform, the harder it is to find minimal pairs in support of a phonemic contrast. Even frequent phonemes might end up yielding just a few minimal pairs. In these languages, one might expect that learners are not only sensitive to contrasts among entire words (i.e., minimal pairs), but to contrasts among subparts of words (i.e., near-minimal pairs), or even syllables. In these cases, a phoneme n-gram implementation can be used to capture most of the properties of the phonological environment, like the preceding and following segments (as in King, 1967b) and, with some tweaks, also contexts like 'before intervocalic /r/'. It will also capture distributional asymmetries: for instance, if a phoneme x is mostly used word-initially, while a phoneme y is mostly used word-finally, we might expect a merger between them to be more likely that a merger between x and a phoneme z which also occurs word-initially more than it occurs word-finally. In this second case, the evidence for this contrast

63

would be higher, even though the contrast might not yield minimal pairs in the lexicon.

The decision on the linguistic unit is arbitrary, and it probably depends on one's assumptions about how lexical processing and phonological acquisition work. However, given that all previous attempts of using models based on word unigrams failed to find a functional load effect, and that the only positive result has been achieved using a model based on phoneme rather than word n-grams (Eychenne & Jang, 2015), I think it is worth exploring a more generalizable measure, and therefore I will adopt a trigram model over phoneme sequences (k=2, n=3) as a default for all the languages investigated in this chapter.[5]

Another delicate decision is whether to base the counts for entropy calculations on word types or word tokens. Most of the works I investigated base their counts on tokens (with the exception of Cychosz, 2017), and the motivation for using tokens is that, contrary to the number minimal pairs, the calculation is sensitive to word frequency. If a contrast is only visible in two words, but they are both very high-frequency, we might expect their contrast to be more important compared to a contrast which is only visible in two low-frequency words. Since our estimation of minimal pairs is already at the type level, having a token-level measure is a good way to capture a different dimension of contrast, one that includes word frequency. For this reason, I will estimate functional load using both minimal pairs and entropy loss over word tokens. An alternative measure that would be interesting to study is also entropy loss calculated over word types, but I will postpone the discussion of this measure to the next chapter.

3.3.3 Phonetic control

Another question is which pairs of phonemes should be compared to those for which a merger has been attested, and this depends on whether we want to test the 'weak point' hypothesis or the 'least resistance' hypothesis, in terms of King (1967a). When testing the former hypothesis, King compares all the possible pairs of phonemes that are one or two features apart (King, 1967a, fn. 10), while when testing the latter he compares pairs of

---

[5]This is the same model used by Eychenne & Jang (2015).

phonemes that merged and other pairs that did not merge, involving: i) the phoneme which is lost after the merger and ii) other phonemes which are one or two features far from it. One might be surprised by the fact that the choice was not narrowed down to phonemes which are only one feature apart, but thinking about vowel mergers, there are indeed many attested cases of vowels that merge even if they are two features apart (e.g., [high] and [ATR], in the case of /ɪ/-/e/ and /ʊ/-/o/, in several Romance languages), because they are close in acoustic space.

Surendran & Niyogi (2006), when studying the merger between /n/ and /l/, suggest to limit the investigation to other coronal pairs. Wedel et al. (2013) instead control for phonetic similarity by limiting the investigation of potential mergers to phonemes which differ by one feature only.

Surendran and Niyogi do not explain the motivation behind limiting part of their investigation to coronal segments, but the choice of constraining the place of articulation is reasonable when working in the consonantal domain: for instance, while /θ/ and /f/ are articulated in the front part of the mouth, /χ/ is articulated in the far back, and therefore the assumption that /θ/-/f/ and /θ/-/χ/ should be treated equally because they both involve a change in one single feature (place) is unmotivated. Changes that affect place typically involve segments whose places of articulation are close to each other. This might constitute an issue for unconstrained tests of the weak point hypothesis: for instance, in Wedel et al. (2013)'s attempt to compare their mergers with other possible mergers that could have occurred in the phonological inventory, the set of tested pairs also contains pairs like /p/-/k/. Technically, /p/ and /k/ differ by only one feature (place of articulation), but mergers between them are typologically rare.[6] It might well be that regardless of functional load considerations, it would be implausible to obtain this merger through misperception or misproduction, and therefore it does not qualify as a 'potential merger' for our purposes.

---

[6]A merger /k/>/p/ is attested in Romanian, in which we have Lat. OCTŌ, FRŪCTUS > Rm. *opt, frupt*. This merger affects all velar consonants (Lat. SIGNUM > Rm. *semnum*). While in theory this merger could have gone through an assimilation stage (/kt/>/tt/), which we know occurred in other Romance branches (see Italian *otto, frutto*), and then through a dissimilation stage, there is no evidence for dissimilatory processes in Romanian when we look at the reflexes of Latin geminates (Lat. ANNUS, QUATTŬŎR > Rm. *an, patru*). I thank Don Ringe for pointing this out to me.

One way to control for phonetic plausibility would be using confusion matrices derived from lab experiments (like in Hualde & Prieto, 2014 and Eychenne & Jang, 2015), but there are at least two problems with this strategy: first, confusion matrices are not usually available for many languages, and second, they usually do not capture confusion in context, which is crucial for sound change.

In this situation, King (1967a)'s approach is the most conservative one, and it can be simply amended with the two following points : i) when studying consonants, limit the possible mergers to phonemes which are one feature apart, and ii) in the case of place of articulation, limit the possible mergers to those which are close enough that no other phoneme could serve as a 'bridge' between them.

It appears reasonable to allow the comparison with phonemes which are two features apart in case of vowels, because we know that there are vowels which are close in the acoustic space despite being two features apart. This should be sufficient to avoid biasing the results by including pairs of phonemes whose merger would be implausible for independent reasons.

3.3.4   Which Functional Load Hypothesis?

Another issue is which hypotheses among those in King (1967a) should be tested. In the previous paragraph, we excluded any reference to the 'frequency' hypothesis because it does not involve any functional load measurement, but just an empirical examination of ongoing mergers and their relative frequencies. The hypothesis that the low-frequency phoneme is always the one being lost after a merger is plausible, but it is orthogonal to the problem of studying the relationship between lexical contrast and mergers, because it just has to do with the phonetic outcome of a sound change.

The other two hypotheses are more interesting. The 'weak point' hypothesis simply states that mergers are always associated with low contrast. The 'least resistance' hypothesis is, in a sense, more conservative, because it assumes that there are independent reasons that put phoneme x under pressure, and therefore limits its predictive power to determining the target of the merger (namely, given that x is under a pressure to merge, predict the phoneme

y with which it will merge).

In practice, the least resistance hypothesis is the easiest one to test, because the domain is restricted and the predictions are clear. The weak point hypothesis suffers from the problem that one would need to define how weak is 'weak', and the interpretation would be very different depending on whether one wants to test stronger versions (namely, no minimal pairs at all, or the lowest entropy loss of the system) or weaker versions (namely, minimal pairs and entropy loss which are beyond one standard deviation from the overall mean). For this reason, the discussions in the chapter will mostly refer to the least resistance hypothesis.[7]

There are two specific instances for which the weak point version makes more sense as the hypothesis to test. First, when testing cases of multiple sound changes attested in a single environment, it makes little sense to ask specifically why a particular sound change happened, but it seems more plausible to ask whether the sound changes attested carry a lower functional load with respect to other changes involving pairs that could have merged, but did not. Second, when testing cases of several mergers which are not independent, because they involve the loss of the same feature (like for instance, the loss of a voicing contrast), the scenario is similar: we do not want to subject to test the single changes, but we want to check whether other contrasts involving the same feature (or the same phonemes) were associated with similar functional load values. In this case, since we have several mergers occurring in the same environment, determining a single optimal outcome (as required by the least resistance version) is not the best strategy, because it will rule out the possibility that all the mergers that occurred were optimal in that environment. Therefore, in this case, we need to relax the criteria as follows:

- 'functional' mergers: mergers that carry a functional load which is the lowest among those associated with the pairs of phonemes investigated OR that is more than one standard deviation below the mean of the unmerged pairs

---

[7]There will be some cases in which the phonetic outcome of a merger between x and y is unclear, because sometimes the outcome is x, and in some other cases it is y, or an intermediate sound. In this case, the comparison will involve all the pairs in which *either* phoneme is present.

- 'unclear' mergers: mergers that carry a functional load that is within one standard deviation of the mean of the unmerged pairs

- 'anti-functional' mergers: mergers that carry a functional load which is the highest among those associated with the pairs of phonemes investigated OR that is more than one standard deviation above the mean of the unmerged pairs

With this background in mind, we can proceed to examine some cases of mergers in progress.

3.4 Reported mergers

In this section, we investigate the mergers used in Wedel et al. (2013), and we add to the sample other reported mergers for the same languages.

3.4.1 English (UK varieties)

For the study of English varieties in the United Kingdom, we can rely on many electronic resources. The language is well represented in CHILDES, with 18 corpora, and it is one of the three languages for which phonological transcriptions are available in CELEX. I collected all the lemmas that appear in the speech of mothers and fathers in CHILDES and that occurr at least twice in the corpora used in CELEX, obtaining 7776 lemmas. Then, I used the most 5000 frequent ones, according to CHILDES frequencies, to calculate minimal pairs and entropy measures. After the filter, we are left with ≈4M tokens. Proper nouns were excluded. The choice of lemma is motivated by the fact that we want to exclude redundancy due to morphology.

2.4.1.1 Consonants The consonant mergers reported for English are all taken from the classic Wells (1982), and they involve the two interdental fricatives /θ/ and /ð/. The merger of the two phonemes with /f/ and /v/, also known as th-fronting, is a characteristic of the Cockney dialect (Wells, 1982, 328). The merger with /t/ and /d/, also known as

th-stopping, is reported for Irish English, specifically in the urban dialects of Cork and Dublin (Wells, 1982, 129, 430). Finally, a merger between /θ/ and /s/ is reported for Gaelic English, while the voiced interdental fricative appears as /ts/ (Wells, 1982, 413).[8]

When it comes to these specific mergers, there are two problems that need to be addressed, and that are discussed at length in Honeybone (2016). In general, a common problem in dealing with mergers is that we need to distinguish cases in which two sounds merge from cases in which there is a variable rule that optionally neutralizes the contrast. Then, we need to distinguish sound changes which are, to some extent, plausibly language-internal, from changes which are clearly the result of language contact. Note that in both cases the Functional Load Hypothesis cannot be properly tested: in the first case, there would be no structural change from the viewpoint of the grammar, and functional load in Martinet's sense makes a prediction for sound change, not for variable rules; second, in case of obvious language contact, there is an external pressure that would be difficult to account for, and that could represent a confound.[9] For most of the cases mentioned in this work, the literature can provide the information necessary to evaluate whether the mergers are 'regular and endogenous' or 'irregular and exogenous' (using Honeybone's terms). This is necessary when one wants to isolate the factors that influence a sound change.

In this case, Sampson (2020) correctly points out that Wells describes th-fronting as a variable rule, not as a merger, and this is how it has also been described in North American English (Labov, 1966). Adult speakers of the Cockney dialect clearly have the two fricatives in their inventory: they are able to distinguish between them, and they never hyper-correct. As for th-stopping, Wells reports the merger as occurring in Irish English, in particular in the cities of Dublin and Cork (Wells, 1982, 428-430). Honeybone (2016) argues that many instances of th-stopping are both irregular and exogenous, namely attested only in bilinguals or language contact areas. Moreover, Wells (1982, 413) is clear about the fact that the merger of the interdental fricatives with /s/ and /ts/ is a substratum effect from

---

[8]Wedel et al. report a merger of /ð/ with /z/ instead of /ts/, but I could not locate it in Wells (1982), and so I assume it was a mistake.

[9]The same point is made in Blevins (2006): 'if we are interested in discovering the [...] origins of a particular sound change, we must filter out contact-induced change' (Blevins, 2006, 9).

Phonemes	MinPairs	Entropy loss
θ-f	5	.0010
ð-v	1	.0011
θ-t	14	.0012
θ-s	14	.0010
θ-ð	0	.0009
ð-z	2	.0004
ð-d	6	.0018
	μ=7.2 σ=6.6	μ=.0011 σ=.0005

Table 3.2: Functional Load for interdental fricative mergers in English (UK varieties).

Gaelic. For these reasons, none of these cases qualifies as a real merger in progress.

Even though these specific cases will not be considered as an argument in favor or against the Functional Load Hypothesis, because of their unclear status as regular sound changes, we can still check the functional load values associated with these mergers, as an example of the methodology that will be employed in this section.

Table 3.2 compares some mergers involving the interdental fricatives and phonemes which are one-feature far from them. In this case, neither of the mergers resulted the one carrying the lowest amount of functional load. In terms of minimal pairs, the voicing contrast among the two interdental fricatives is the weakest one, while in terms of entropy loss, the contrast between /ð/ and /z/ has the lowest functional load.

The voicing contrast is the one for which learners have the least amount of evidence, but we know that voicing is associated with many other phonemic contrasts, and therefore there might be a structural constraint against the merger. We can see it in Table 3.3, in which many of the voicing contrasts are associated with dozens of minimal pairs, even if we just look at the 5000 most frequent words in child-directed speech.

If we ignore this potential merger, then the merger with the labial would be the one associated with the least amount of minimal pairs if we look at /θ/ (5: *three-free, thin-fin, wreath-reef, half-hearth, deaf-death*), and this is in fact a merger that according to Honeybone (2016) could qualify as a regular sound change. The entropy loss associated with the contrast between /θ/ and /f/ is also quite low, but not beyond one standard deviation from the mean.

Phonemes	MinPairs	Entropy loss
p-b	66	.0034
f-v	14	.0013
t-d	87	.0098
s-z	12	.0017
ʃ-ʒ	0	.0001
tʃ-dʒ	8	.0004
k-g	30	.0023

Table 3.3: Functional Load for voicing contrasts in English (UK varieties).

As for /ð/, the phoneme has one minimal pair with its labial counterpart (*than-van*), but entropy loss points toward other contrasts as weaker, in particular the contrast with /z/. There is an explanation for this pattern: since /z/ is relatively uncommon in word-initial position (it appears in 8 words in our sample), while /v/ is more frequent (45 words), this makes the contrast between /ð/ and /z/ the least important, since most of the occurrences of /ð/ are in word-initial position, in function words. For this reason, a child sensitive to distributional cues should be more inclined to merge /ð/ with /z/.

The low functional load associated with the contrast between /ð/ and /θ/ and that between /ð/ and /z/, and the interdependence between the two attested mergers, make it difficult to argue in favor of the Functional Load Hypothesis in this case, according to our criterion, but it could explain while th-fronting is more widespread as a type of sound change compared to other alternatives.[10]

Another possibility is that the merger occurred just because the two phonemes are highly confusable from the acoustic viewpoint. Table 3.4 summarizes the results from Wang & Bilger (1973)'s main experiment on CV sequences that involve the phoneme contrasts mentioned, by deriving probabilities of misperception from the responses of the participants of a laboratory experiment, in which they were exposed to recordings of syllables in the presence of noise. In the first column, we see that the two probabilities of misperceiving

---

[10]An alternative test would be applying functional load at the feature level, and ask for instance if distinguishing labiodental from interdental fricatives is more important than distinguishing interdental from alveolar fricatives in terms of minimal pairs or entropy loss. This would be a valid test. However, defining the domain of the possible featural contrast to test appears to be a more complicated task. In particular, there are instances where a feature is lost only for a subset of the contrasts that are associated with it: for instance, the voicing contrast between /t/ and /d/ is neutralized intervocalically in North American English, but the same is not true for the voicing contrasts /p,b/ and /k,g/.

Phonemes	Confusability(1)	Confusability(2)
θ-f	0.39	0.10
ð-v	0.31	0.11
θ-t	0.04	0.02
θ-s	0.12	0.04
θ-ð	0.02	0.04
ð-d	0.05	0.05
ð-z	0.11	0.07
p-b	0.01	0.08
f-v	0.03	0.05
t-d	0.04	0.03
s-z	0.09	0.02
ʃ-ʒ	0.02	0.01
tʃ-dʒ	0.02	0.01
k-g	0.03	0.01
p-f	0.06	0.12
s-ʃ	0.03	0.05
z-ʒ	0.04	0.08

Table 3.4: Reciprocal confusability indeces calculated from the responses to the CV stimuli in Wang (1967)'s main experiment 1.

the interdental fricatives for their labiodental counterpart are the highest of the table (0.39 and 0.31). The reverse pattern, namely the probability of misperceiving /f/ and /v/ for the interdental fricatives, is also high (0.10 and 0.11), but comparable to other contrasts.

Interestingly, the contrast between /p/ and /f/ is also weak acoustically. In particular, the probability of misperceiving /f/ as /p/ is noticeable (0.12). This is corroborated by the fact that an empirical study of speech errors, in Labov (1994, 2010), found that the confusion between *copy* as *coffee* is one of the most widespread cases of misunderstanding in conversations. In our sample, this contrast is supported by many minimal pairs (53) and is associated with relatively high entropy loss (.0034). The fact that a merger between the two sounds is not attested in the English speaking world is compatible with functional load.

To summarize, in this case the evidence is mixed, because there might be an interaction among functional load considerations, structural constraints, and phonetic confusability, that explain the outcomes of the mergers involving the interdental fricatives in the English world. This example shows that determining a single criterion to evaluate functional load is challenging, because there are many factors which might be interacting with it, and a careful

evaluation of the mergers is necessary to determine whether lexical contrast is playing a significant role.

In practice, this is not a case which is incompatible with the Functional Load Hypothesis, but it fails to provide clear evidence for it. As previously mentioned, this case will not be considered in the final statistics, because it is unclear whether we are dealing with regular sound changes to begin with.

Fortunately, the cases discussed in the next subsections will be easier to evaluate.

2.4.1.2 Vowels There are four vowel mergers reported for English in the UK, and they are all taken from the classic Wells (1982).

The first merger is /aɪ/-/ɔɪ/ (lexical sets PRICE-CHOICE), and it is attested in the west country (Wells, 1982, 347) and in the north (Wells, 1982, 425), even though in this second case it might be a substratum effect from Irish, which lacks the second diphthong. In this case, it seems that /aɪ/ enters the acoustic space usually associated with /ɔɪ/ (cf. Wells, 1982, 308, 347).

The second merger is /ʊə/-/ɔː/ (lexical sets CURE-THOUGHT), and it is well attested in different parts of England (Wells, 1982, 237, 287, 374). The merger is a recent development, and it occurred after the loss of historical /r/ (/ʊər/>/ʊə/>/ɔː/, cf. Wells, 1982, 213-216).

The third merger is /ɪə/-/ɛə/ (lexical sets NEAR-SQUARE). This merger is attested in East Anglia (Wells, 1982, 338), in New Zealand, and in the US, specifically in Charleston (Baranowski, 2006) and New York (Labov, 1994, 343), with various outcomes included in the range of the two original diphthongs. In this case, we do not know if its origin is anterior, posterior, or related to the loss of historical /r/, but Lass (1992) reports instances of this merger that pre-date the Great Vowel Shift, and therefore the loss of historical /r/.[11]

The fourth merger is /ɜː/-/ɛə/ (lexical sets NURSE-SQUARE), which is reported in

---

[11]I also note that Wikipedia reports the merger as /ɪər/-/ɛər/, but it does not report any source apart from Wells, who never explicitly says whether the merger is conditioned on the following /r/ or not, and some literature on New Zealand English, where instead the merger is apparently unconditioned because historical /r/ was lost (see Gordon et al., 2004 and Bauer et al., 2007).

Phonemes	MinPairs	Entropy loss	Phonemes	MinPairs	Entropy loss
aɪ-ɔɪ	**7**	**.0005**	ʊə-ɔː	2	.0018
aɪ-eɪ	63	.0035	ʊə-ɔɪ	0	.0000
aɪ-aʊ	19	.0026	ʊə-uː	0	.0010
aɪ-iː	76	.0047	ʊə-əʊ	0	.0000

Phonemes	MinPairs	Entropy loss
ɪə-ɛə	**22**	.0015
ɜː-ɛə	14	.0010
ɪə-iː	7	.0012
ɪə-ɜː	11	.0012
ɪə-ʊə	1	.0006
ɜː-iː	21	.0018
ɜː-a:	13	.0011
ɜː-ɔː	26	.0018
ɜː-uː	9	.0010
	μ=12.6 σ=8.48	μ=.0012 σ=.0004

Table 3.5: Functional Load for vowel mergers in English (UK varieties).

Liverpool, Merseyside, adjacent parts of Lancashire, and some parts of the Manchester area (Wells, 1982, 361). The outcome of the merger is usual toward the vowel of NURSE, but sometimes it can appear as [ɛə] or [ɛ:]. Since these are also areas where historical /r/ tends to be preserved, it is not clear if this merger is conditioned on the presence of /r/ or not (Sampson, 2020).

The first two mergers can be studied in isolation, because they are both unconditioned. This means that we will be testing the least resistance hypothesis. The last two mergers, however, involve the same vowel, and occur in the same environment (historical /r/). Therefore, we will have to test the weak point hypothesis.

The results in Table 3.5 show that both minimal pairs and entropy loss are quite low for the PRICE-CHOICE merger. This is a sound change that is compatible with the Functional Load Hypothesis.

For the CURE-THOUGHT merger the evidence is unclear: the dictionary has two minimal pairs, *you're-your* and *cruel-crawl*, but this second one is a mistake, because the word *cruel* does not belong to the CURE set. The dictionary also has *poor* transcribed as /pɔ:/, with the merger already completed. For these reasons, CELEX is not a reliable

source to study this case. This merger is thus excluded from further investigation.

As for the other two cases, the NEAR-SQUARE and the NURSE-SQUARE mergers are both associated with a large amount of minimal pairs:

- ɪə-ɛə, (NEAR-SQUARE): ear-air, ear-heir, weir-where, we're-where, here-hair, hear-hair, here-hare, hear-hare, cheer-chair, beer-bear, beer-bare, peer-pair, sheer-share, fear-fair, peer-pear, dear-dare, deer-dare, weird-where'd, steer-stare, steer-stair, we're-where, really-rarely, deary-dairy

- ɜː-ɛə, (NURSE-SQUARE): err-air, err-heir, per-pair, per-pear, fur-fair, fur-fare, purr-pair, purr-pear, furry-fairy, her-hair, her-hare, word-where'd, steer-stair, steer-stare

Interestingly, the NEAR-SQUARE merger yields the second-highest amount of minimal pairs, and is beyond one standard deviation from the mean. For this reason, it is an anti-functional merger, according to our criteria. However, its entropy loss is close to the average.

Apart for some obvious mistakes like *where'd* and *we're*, which are considered independent lemmas in CELEX just because orthographically they are not separated by a space (but notice that CELEX excludes *were*, since it is an inflected form), and *err*, which is a filler, but it ends up in the list because it can be used as a verb, it seems that English has a high amount of homophones in the historical rhotic environment to begin with. In other terms, the reason why entropy loss is limited is because the entropy was already low before the mergers.

This sound change represents a problem for the Functional Load Hypothesis: the hypothesis is at odds with the fact that the pre-rhotic environment exhibits a great amount of homonymy, because this should have been prevented to begin with. For the NURSE-SQUARE merger, the amount of functional load is within the expectations according to both measures.

A potential reaction to these high-homonymy cases is that the number of minimal pairs would be much lower if we only consider minimal pairs that share the same syntactic category, since in the other cases the context can disambiguate the meaning. However,

there are several problems with this argument: first, this is also true for the number of minimal pairs calculated for the *unmerged* pairs, which represent our baseline, and therefore repeating the analysis and counting the number of minimal pairs only among words of the same syntactic category would likely confirm the results; second, experimental studies have showed that homophones can influence lexical access and create confusion or delay even if the ambiguity is between two words of different syntactic categories (Boland & Blodgett, 2001); third, the argument of contextual disambiguation is so general that can be used against the idea of functional load almost *a priori*: if we consider a large context, information-theoretic models and functional load accounts simply do not work, because ambiguity is almost non-existent when the speaker has access to contextual information.

3.4.2 North American English

2.4.2.1 Consonants There are two consonant mergers reported in the Atlas of North American English (Labov et al., 2008): /w/-/ʍ/ and /t,d/-'flapping'. In both cases, since we have a single sound change, we can check whether the functional load of these specific mergers is the lowest among the alternatives that the language had, according to the least resistance hypothesis.

The merger between /w/ and /ʍ/ (*witch-which*) is widespread across North America. Even though in the CMU dictionary /ʍ/ is typically not present as an independent phoneme, and all its instances appear as /w/, the dictionary lists varieties with /ʍ/ as a secondary pronunciation, and therefore we can measure functional load by selecting all the alternative pronunciations offered by the dictionary.

There are seven minimal pairs associated with the merger:

- w-ʍ (WITCH-WHICH): why-'y', where-wear, whether-weather, which-witch, whine-wine, wise-whys, win-when (only in varieties with the *pin/pen* merger)

Following the least resistance hypothesis, we might ask why /ʍ/ merged with /w/ rather than with the other phonemes in its phonological neighborhood. In terms of minimal pairs, the merger between /w/ and /v/ would be preferred, as the merger with /w/ is only the

Phonemes	MinPairs	Entropy loss
w-ʍ	7	**.0026**
v-ʍ	1	.0008
f-ʍ	10	.0011
h-ʍ	15	.0018
b-ʍ	15	.0018
p-ʍ	12	.0010

Phonemes	MinPairs	Entropy loss
t-d	**4**	**.0346**
p-b	2	.0235
f-v	3	.0133
s-z	2	.0086
ʃ-ʒ	0	.0016
tʃ-dʒ	0	.0014
k-g	0	.0198

Table 3.6: Functional Load for consonant mergers in North American English.

second best option. On the other hand, entropy loss calculated over tokens suggests that merging with /w/ is the worst option. The reason for this outcome is that /w/ (like /h/) occurs in the same environment in which /ʍ/ is found: it usually appears in simple onsets, and never in codas (even though it can occur in consonant clusters). Therefore, from the viewpoint of entropy loss, /v/ would be a better candidate for the merger, since its distribution in the lexicon is different (e.g., it occurs at the end of a word more than at the beginning of a word, 74 vs 50, in the sample employed here). We conclude that according to entropy loss, the attested merger is anti-functional.

The neutralization of the voicing contrast between /t/ and /d/ after a stressed vowel (also known as 'flapping') is typical of North American English varieties. Lenition of consonants when they occur between vowels is a natural sound change, well attested typologically (Gurevich, 2004) and well motivated acoustically (Kaplan, 2010). For this reason, a meaningful question to ask in this case is why the voicing contrast is neutralized for these two phonemes, but not for others.

In terms of minimal pairs, as we see from Table 3.6, neutralizing the contrast is not optimal, because the voicing contrast between /t/ and /d/ is the one that yields the highest amount of minimal pairs (*writer-rider, writing-riding, petal-pedal, metal-medal)*, even though one of them is redundant because CELEX codes *writer-writing* and *rider-riding* as independent lemmas.

One could make the argument that while according to the least resistance hypothesis this merger is not optimal, the number of minimal pairs involved is small, but then we would

be left with explaining why a sound change as general as the neutralization of voicing does not apply across the board, but only between the two segments for which the contrast stands out from a distributional viewpoint. This merger, then, appears anti-functional.

As for entropy loss, estimating entropy within conditioning environments requires a tweak to the procedure described in the preceding section. Surendran & Niyogi (2006) propose to substitute the phonemes in the conditioning environment with a third symbol (in this case, a flap consonant) and leave them intact outside of the environment. Then, the usual formula can be applied. Note that by introducing an additional symbol, a functional load formula applied at the level of phonemes can yield a negative value, because we are introducing an extra distinction in the lexicon, instead of removing one.[12]

Alternatively, there is a simpler procedure that can be adopted. Instead of creating a distribution of n-grams, we can create a unigram distribution that only contains the phonemes that can appear intervocalically, and we keep track of their frequency in that specific context. Then, we can calculate functional load over a unigram distribution of phonemes that is just based on their frequency. This will capture the idea that losing a distinction between phonemes which often appear intervocalically will be more costly than losing a distinction between phonemes which are uncommon in intervocalic position.

By calculating functional load in this way, the entropy values also point toward an anti-functional merger.

2.4.2.2 Vowels   There are two unconditioned vowel mergers for North American English reported in Wedel et al. (2013). The first is the /ɑ/-/ɔ/ merger (LOT-THOUGHT, or *cot-caught*), which has been well documented in the literature (Labov et al., 2008). This merger led to the the loss of /ɔ/ in most of the western varieties of American English, and is now spreading within eastern varieties. The second merger is /ɔɪ/-/ɜr/ (CHOICE-NURSE). Even though I could not find this merger reported in Labov et al. (2008), it is discussed in Labov (1966), and it was found in some speakers of New York City English that were born

---

[12]I suspect that this possibility was not discussed in Surendran & Niyogi (2006) only because if we use an entropy model at the word level, like they did, this does not happen.

Phonemes	MinPairs	Entropy loss
ɑ-ɔ	**5**	.0014
ʊ-ɔ	6	.0010
ʌ-ɔ	11	.0036

Phonemes	MinPairs	Entropy loss
ɔɪ-ɝ	**2**	**.0005**
aɪ-ɝ	23	.0024
eɪ-ɝ	25	.0028
aʊ-ɝ	9	.0017
oʊ-ɝ	21	.0024
i-ɝ	61	.0045
u-ɝ	18	.0013

Table 3.7: Functional Load for vowel mergers in North American English.

at the beginning of the 20th century (it is also reported in Labov, 1994). This merger causes homonymy among minimal pairs like *coil-curl*, and words like *girl* to appear as [gɔɪl]. Since these two mergers involve two different classes of vowels (short vowels for the first merger, and long vowels for the second merger), we can investigate these two mergers independently according to the least resistance hypothesis.

Using the CMU dictionary for estimating functional load in these two cases must be done with some caveat. In North American English, rhotic environments can preserve unconditioned mergers or trigger conditioned mergers, and this is a problem for functional load calculations. According to the CMU dictionary, *star-store* is a minimal pair which is associated with the LOT-THOUGHT merger, but this is a mistake, because the rhotic environment preserves the historical short-/o/, which instead merges with /ɔ/ in all the other environments in the majority of the North American English varieties. The CMU dictionary does not encode /o/ among its phonemes. For this reason, a proper study of the LOT-THOUGHT merger requires coding rhotic environments as independent symbols.

Table 3.7 contains the functional load calculations for the mergers. Most of the minimal pairs for LOT-THOUGHT are resulting from noise. Apart from *stock-stalk*, we have the interjections *ha-ho*, *pa-po*, *don-dawn* (the first word appears in CELEX and CMU, but it's a misspelling for *don't* in CHILDES), and *la-law*, in which the first word is an onomatopoeia. The contrast is indeed the weakest one according to minimal pairs, but entropy loss points toward the contrast between /ʊ/ and /ɔ/ as the least informative one.

The confusion matrix reported in Hillenbrand et al. (1995), who derive the probability

Phonemes	Confusability(1)	Confusability(2)
ɑ-ɔ	0.035	0.138
ʊ-ɔ	0	0.001
ʌ-ɔ	0.018	0.038

Table 3.8: Reciprocal confusability indeces from Hillenbrand et al. (1995).

of misperceiving one vowel for another (and viceversa) through a laboratory experiment on a group of speakers, shows that even though from the featural viewpoint the two contrasts are comparable, it is the case that when it comes to confusability they are very different, with the contrast between /ɑ/ and /ɔ/ definitely being the weakest one (Table 3.8). In this case, the merger is associated with high confusability and a low number of minimal pairs, and is compatible with functional load, while the evidence from entropy loss is unclear.

The second table in Table 3.7 shows measurements for /ɝr/ and other diphthongs and long vowels, and it shows that this is a case in which functional load seems to be fully compatible with the merger, since it only distinguishes two minimal pairs (*oil-earl, soy-sir*) and it causes the least amount of entropy loss.

2.4.2.3 Vowels before /r/ Many vowel mergers in North American English are limited to the pre-rhotic environment. Wedel et al. (2013) selected one of them, the merger between /ɑ/ and /ɔ/ (START-NORTH, or *far-for*). Some varieties of English merge instead /ɔ/ with /o/ in pre-rhotic environment (NORTH-FORCE, or *for-four*). As we said in the previous section, the CMU phonological dictionary has a unique symbol in this case, because it neutralizes the distinction between /ɔ/ and /o/, and therefore it represents a variety of English in which the second merger occurred. However, this second merger and the first one are in complementary distribution: one does not find varieties in which START-NORTH-FORCE share the same vowel. This fact makes it impossible to count minimal pairs for the START-NORTH merger, because the NORTH class and the FORCE class are overlapping.

There are several other mergers reported in Labov et al. (2008) before an intervocalic /r/: we have the merger between /ʌ/ and /ɝr/ (the *hurry-furry* merger, the merger among /eɪ/, /æ/ and /ɛ/ (the *Mary-marry-merry* merger), the merger between /ɛ/ and /ʌ/ (the

Phonemes	MinPairs	Entropy loss
ʌ-ɝ	0	.0027
ɛ-æ	0	.0457
ɛ-eɪ	0	.0023
ɛ-ʌ	0	.0026
ɪ-i	0	.0026
ɛ-ɝ	4	.2138
æ-ɝ	0	.0462
eɪ-ɝ	0	.0023
ɪ-ɝ	0	.0654
ɛ-ɪ	0	.0646
eɪ-i	0	.0017
æ-ʌ	1	.0017
æ-ɑ	0	.0316
ɔ-ɑ	1	.0672
ɔ-ʊ	0	.0121
ʌ-ɑ	0	.0022
ʌ-ɔ	0	.0022
eɪ-ɪ	0	.0017
ɝ-ɑ	1	.1104
ɝ-ɔ	0	.1145
	$\mu=0.35$ $\sigma=0.93$	$\mu=.0397$ $\sigma=.0552$

Table 3.9: Functional Load for some possible vowel mergers before intervocalic /r/ in North American English.

merry-Murray merger) and the merger between /ɪ/ and /i/ (the *mirror-nearer* merger). This is a case in which the weak point hypothesis should be tested, since we have multiple mergers reported for the same environment, and therefore studying them individually would not be informative.

Table 3.9 shows that in this case we have a sparsity problem: minimal pairs in this environment are rare in general. Technically, most of the contrasts in this environment are compatible with the functional load hypothesis, because they yield no minimal pairs.

Entropy loss is calculated following the strategy adopted for flapping: namely, the counts of the vowels that appear before an intervocalic /r/ are first extracted, and then entropy is calculated over vowel frequencies using a phoneme unigram model. In this case, entropy loss does not show mergers which are clearly non-contrastive, because one can find several potential mergers which are equally or less contrastive than the pairs which ended

Phonemes	MinPairs	Entropy loss
ɪ-ɛ	8	.0603
æ-ɛ	9	.0418
ʌ-ɛ	7	.0683

Table 3.10: Functional Load for some possible vowel mergers before nasals in North American English.

up merging. Since in this case the two measures cannot be properly compared, I will exclude these cases from the final summary table.

2.4.2.4 Vowels before nasals   Another well-studied conditioned merger is the merger between /ɪ/ and /ɛ/ before nasal consonants, the *pin-pen* merger. Wedel et al. (2013) describe the merger as occurring before /n/, but actually the merger occurs before all nasals (/m/, /n/ and /ŋ/).

The outcome of the merger is clear (/ɛ/ > /ɪ/). This is a case that qualifies for the least resistance hypothesis. Table 3.10 contains the calculations for some other potential vowel mergers involving /ɛ/, and we can see that in this case functional load seems to play little role. In terms of minimal pairs, we find many of them associated with this merger:

- ɪ-ɛ, (PIN-PEN): win-when, tin-ten, pin-pen, din-den, mini-many, in-'n', since-sense, gym-gem

This is more or less the same amount of minimal pairs that we find in other contrasts before nasal phonemes, and merging the mid vowel to /ʌ/ would actually yield less minimal pairs.

For entropy loss, we can use the phoneme unigram model of the previous subsections to calculate functional load in the nasal environment. In this case, the merger with /æ/ would be preferred instead. In this case, it is unclear how functional load is related to the merger.

2.4.2.5 Vowels before /l/   In North American English we find many mergers before /l/. Wedel et al. (2013) report, from Labov et al. (2008), the following mergers: /ɪ/-/i/ (*hill-heel*), /ʊ/-/u/ (*pool-pull*), /ʊ/-/oʊ/ (*bull-bowl*), /ʌ/-/ɔ/ (*hull-hall*) and /ʌ/-/ʊ/ (*hull-*

Phonemes	MinPairs	Entropy loss
ɪ-i	**13**	.0402
ʊ-u	2	.0094
ʊ-oʊ	3	.0124
ʌ-ɔ	**0**	.1116
ʌ-ʊ	**0**	.0197
ɛ-eɪ	**12**	.0190
ʌ-oʊ	**0**	.0544
ɪ-ɛ	10	.0738
ɪ-ʌ	1	.0854
ʊ-ɔ	3	.0179
ʊ-aʊ	3	.0020
oʊ-aʊ	3	.0029
ʌ-ɛ	1	.1143
ʌ-ɑ	2	.0206
ʌ-eɪ	1	.0207
ʌ-aʊ	0	.0040
ɪ-eɪ	16	.0164
i-eɪ	17	.0131
ɑ-ɔ	0	.0186
æ-ɛ	1	.0174
ɑ-æ	2	.0068
ɔ-oʊ	10	.0474
ɔ-aʊ	0	.0037
u-oʊ	9	.0191
u-aʊ	0	.0025
	μ=4.39 σ=5.52	μ=.0270 σ=.0322

Table 3.11: Functional Load for some possible vowel mergers before /l/ in North American English.

bull). However, Labov et al. (2008) also mention areas where the mergers involve /ɛ/-/eɪ/ (*fell-fail*) and /ʌ/-/oʊ/ (*hull-hole*), for a total of seven vowel mergers. This is another case that qualifies for the weak point hypothesis test.

Table 3.11 lists all phonemic pairs which could have potentially merged. From the viewpoint of minimal pairs, two of these mergers appear anti-functional, because they lead to a great amount of homonymy:

- ɪ-i, (HILL-HEEL): will-wheel, fill-feel, hill-heel, sill-seal, pill-peal, mill-meal, filling-feeling, will-we'll, hill-heal, still-steal, still-steel, willing-wheeling

- ɛ-eɪ, (FELL-FAIL): well-whale, tell-tale, tell-tail, bell-bale, sell-sale, sell-sail, belle-bale, hell-hail, cell-sale, cell-sail, 'l'-ale, cellar-sailor

In terms of minimal pairs, these mergers are beyond one standard deviation from the mean, and therefore they are both anti-functional. These cases are similar to the vowels mergers in English (UK), for which the main problem is that the language already had homonyms that resulted from historical mergers. Like previously mentioned, limiting the investigation to the minimal pairs that share the same syntactic category would not resolve the issue, because it would also reduce the number of minimal pairs for the other contrasts, and still leave us with the question of why they have not merged instead.

Some minimal pairs are also found for /ʊ/-/u/ (*pull-pool, full-fool*) and /ʊ/-/oʊ/ (*pull-pole, full-foal- bull-bowl*).

Considering minimal pairs, we have two mergers which are clearly anti-functional, two mergers which are neutral, and three mergers which can be considered functional (namely, they yield no minimal pairs at all).

As for entropy loss, still calculated using the phoneme unigram formula of the preceding subsections, the mergers do not carry a functional load which is significantly higher or lower than the mean, but mostly because the standard deviation of the sample is quite high.

3.4.3 German

For the study of German, we can rely on 11 corpora from CHILDES. I extracted the lemmas that appeared in CHILDES associated with the speech of mothers, fathers and some of the investigators, and that appeared at least twice in CELEX. Since all German nouns are inflected for gender, the lemmas contain the dictionary-entry morpheme. Proper nouns were included, because since in the German orthography both proper and common nouns have capital letters, it is not straightforward to tell them apart. In this way, I obtained about 4500 lemmas. The total amount of filtered wordforms is ≈600K.

2.4.3.1 Vowels   In German there is a well studied merger between /ɛ:/ and /e:/, also referred to as the *Bären-Beeren* merger (Wiese, 2000), after which /ɛ:/ disappears from the phonemic inventory. /ɛ:/ is extremely rare in this restricted list (only one occurrence in the word *Trainingsanzug*), and therefore we find no minimal pairs with all the vowels that would be plausible candidates for a merger, and virtually no variation in entropy for all cases. For this reason, I excluded this case from further investigation.

3.4.4   Dutch

For Dutch, we have 14 corpora on CHILDES. Putting together the speech of mothers, fathers, siblings and investigators, and filtering the words through CELEX, we have about 7700 lemmas that in CELEX have frequency of at least two, and I selected the top 5000. After the filter, we are left with ≈832K tokens. Proper nouns were excluded.

2.4.4.1 Consonants   Kissine et al. (2003) report that in the northern part of the Netherlands, voiced fricatives are disappearing, and they merge with their voiceless counterpart.

This is another case in which we need to test the weak point hypothesis, by comparing the mergers with other possible mergers involving the same feature or the same segments. In Table 3.12, we see that the contrast between /f/ and /v/ has only one minimal pair (*fee-'v'*) while the contrast between /s/ and /z/ has three (*zee-'c', set-'z', set-zet*). The merger between /χ/ and /ɣ/ has zero. All of them are more than one standard deviation lower than the mean. For entropy loss, the patterns are similar.

For Dutch, we also have a detailed study on phonetic confusability among consonants, which contains data from a laboratory experiment on a large group of native speakers (Smits et al., 2003). Table 3.13 summarizes the mutual phonetic confusability for the voicing contrasts represented in the study, and it shows some interesting asymmetry.

In the first column, we have the likelihood of hearing a voiced consonant, while instead a voiceless consonant is the target. When speakers hear voiceless consonants, they are more confused when they hear /t/ (0.34) than when they hear other phonemes. This is an

Phonemes	MinPairs	Entropy loss
f-v	**1**	**.0012**
s-z	**3**	.0029
χ-ɣ	**0**	**.0006**
t-d	23	.0085
p-b	33	.0032
v-b	36	.0042
v-z	16	.0051
v-m	26	.0050
v-w	18	.0056
z-d	24	.0093
z-l	22	.0050
z-n	8	.0062
z-ɣ	4	.0014
h-ɣ	0	.0003
j-ɣ	4	.0017
ŋ-ɣ	0	.0005
	μ=16.5 σ=12.3	μ=.0043 σ=.0028

Table 3.12: Functional Load for voicing contrasts in Dutch.

Phonemes	Confusability(1)	Confusability(2)
f-v	0.21	0.17
s-z	0.15	0.12
t-d	0.34	0.02
p-b	0.20	0.06

Table 3.13: Reciprocal confusability indeces derived from Smits et al. (2003).

interesting pattern, because /t/ is among the phonemes for which the voicing distinction is maintained.

In the second column, we have the likelihood of hearing a voiceless consonant, while instead a voiced consonant is the target. In this case, confusability for /d/ (0.02) and /b/ (0.06) are low compared to /v/ (0.17) and /z/ (0.12). This can be explained with the fact that since some of the speakers might have the merger, they would not have the voiced version of the phoneme, and therefore they should perform worse on the task. Interestingly, all these values are lower than those we see in the first column, and therefore one cannot simply say that voiced segments are more difficult to perceive than voiceless segments from the acoustic viewpoint.

These observations imply that reference to a misunderstanding driven by the acoustic properties of the sounds is not sufficient to explain the merger pattern, but there is something beyond acoustic properties which determines which contrasts are maintained and which contrasts are lost. In this case, functional load appears to be a plausible factor: the voicing contrasts which are maintained are those associated with the highest amount of minimal pairs in the lexicon. This is a case which is fully compatible with the Functional Load Hypothesis.

3.4.5 French

For French, we have 10 corpora in CHILDES containing child-directed speech, and we can rely on Lexique (New et al., 2004) for a phonemic transcription and for filtering out non-words. After the filter and the exclusion of proper nouns, I obtained 7137 lemmas. I selected the most 5000 frequent ones, which correspond to ≈1.25M tokens of adult speech.

2.4.5.1 Vowels   There are several cases of reported mergers in progress. Wedel et al. (2013) correctly report the mergers between /ɛ/-/e/ in southern Metropolitan French and /ɛ̃/-/œ̃/ in Parisian French, following Fagyal et al. (2006), but somehow fail to report the other mergers involving mid-vowels, /ɔ/-/ø/ and /œ/-/ø/, which are parallel to the merger between /ɛ/-/e/. Instead, they report a three-way merger among /o/-/ø/-/œ/, for which I

Phonemes	MinPairs	Entropy loss
ɛ-e	35	.0058
œ-ø	**0**	**.0004**
o-ɔ	**7**	.0029
u-o	83	.0053
ɔ-a	44	.0059
a-ɛ	101	.0154
a-œ	16	.0020
ø-ɛ	20	.0021
œ-ɛ	12	.0019
ø-e	28	.0033
e-i	68	.0095
y-e	27	.0055
y-i	34	.0065
	μ=43.3 σ=30.5	μ=.0057 σ=.0041

Phonemes	MinPairs	Entropy loss
œ̃-ɛ̃	**3**	.0013
œ̃-ɔ̃	3	.0016
œ̃-ɑ̃	7	.0012

Table 3.14: Functional Load for vowel mergers in French.

could not find any reference in Fagyal et al. (2006).

The merger between /a/ and /ɑ/, attested in Parisian French and in Belgium, is missing from the list because the two phonemes are represented with a unique symbol in Lexique. Apart from the merger affecting the nasal vowels, all the other mergers must be studied according to the weak point hypothesis, because they involve vowel contrasts which are related to each other (i.e., loss of tense/lax distinction, which becomes allophonic).

Table 3.14 shows that apart from the relatively high number of minimal pairs for the merger between /ɛ/ and /e/, the merger between /œ/-/ø/ yields no minimal pairs, while the merger between /ɔ/-/o/ is mostly due to noise, like onomatopoeia, loanwords and the lack of stress distinction:

- ɔ-o (POMME-PAUME): haut-hot, hockey-ok, top-taupe, portée-porter, os-hausse, pomme-paume, notre-nôtres

These last two mergers are clearly functional. The case of the merger between /ɛ/ and /e/ is more difficult to evaluate: it is true that many of the other mergers involving these

two vowels and other phonemes in their neighborhood would yield to less minimal pairs, but it is also true that since this is a change which is parallel to the other two, there is a structural pressure against the preservation of the contrast.

From the viewpoint of entropy loss, the merger between /ø/ and /œ/ is in the low end of the distribution in both cases, while for both /ɛ/-/e/ and /o/-/ɔ/, one can find mergers that yield to similar entropy losses.

In the case of the merger between the two nasal vowels, we only have three minimal pairs:

- œ̃-ɛ̃ (HEIN-UN): hein-un, hein-uns, brun-brin

In this case, merging /œ̃/ with either /ɛ̃/ or /ɔ̃/ can be considered functional, and the reason why the first merger is favored can be explained in terms of acoustic distance (since they are both front vowels). From the viewpoint of entropy loss, this merger is comparable to the others.

3.4.6 Spanish

For the study of Spanish, we can use 20 corpora from CHILDES. The corpora represent different varieties of Spanish from Europe and from Latin America, but we can use all of them as a sample of child directed speech, since the varieties mostly differ in pronunciation, but not in orthography or vocabulary. Proper nouns were excluded.

In this case, I could not find corpora of Spanish that were phonologically transcribed. However, a phonemic transcription of Spanish would be quite close to its orthography, and therefore I wrote some rules to convert words to a phonemic transcription (mostly involving the treatment of /θ/, /ʎ/, /tʃ/ and /x/). The words were filtered through a dictionary.[13] After the filter, the number of wordforms was ≈4200, for a total of 559487 tokens.

---

[13]I used the online Spanish pronunciation dictionary created by Kyubyong Park (https://github.com/Kyubyong/pron_dictionaries). Even though the dictionary contains an IPA phonetic transcription for each word, I did not use the transcribed words, because they contain phonetic details which are not relevant for the phonological distinctions that I investigated here.

Phonemes	MinPairs	Entropy loss
ʎ-j	**1**	.0014
ʎ-l	7	.0031
ʎ-ɲ	2	.0011

Table 3.15: Functional Load for consonant contrasts in Spanish.

2.4.6.1 Consonants Two consonant mergers are reported for Spanish. The first one is the well attested merger between /ʎ/ and /j/ (*yeísmo*), a merger between the consonants represented by orthographic 'll' (*lluvia* 'rain') and 'y' (*yo* 'I'). The second merger is the merger between /θ/ and /s/, the *ceceo-seseo* merger that affects varieties in southern Spain and Latin America. Wedel et al. (2013) cite Harris (1967) for this merger, but actually Harris convincingly argues that the two sounds are *not* the result of a merger, but the result of the independent development of Old Spanish sibilants (/z/ and /s/) and affricates (/dz/ and /ts/) in the 16th century, which led to one system with two distinct /s/ and /θ/ sounds, and another system with a single phoneme:

```
1500    z     s    dz    ts        z     s    dz    ts
         \   /     |     |         |     |   /     |
1600      s        z     ts        z     s        ts
          |         \   /           \   /          |
          s           θ               s           ts
                                       \         /
                                        s/θ
```
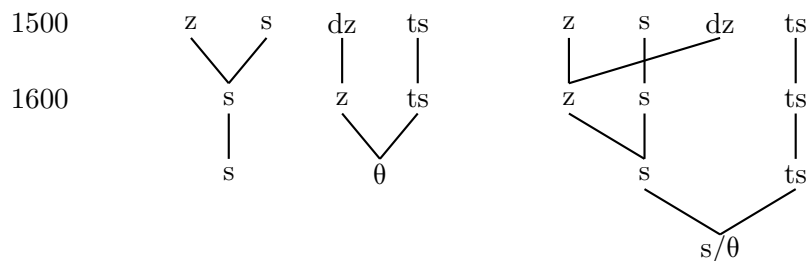
For this reason, the /θ/-/s/ merger cannot be considered a merger in progress, or at least not an endogenous change in Honeybone (2016)'s term, but it should be treated as an exogenous change. Therefore, it is excluded from further considerations.[14]

Table 3.15 shows calculations for the merger /ʎ/-/j/. The outcome of this merger is clear (ʎ > j), and therefore we can ask whether /ʎ/ could have merged with other phonemes

---

[14]Technically, the merger between historical /ts/ and the phoneme that resulted from the three-way merger among /z,s,dz/ would be a plausible merger to study. However, this cannot be done with modern Spanish orthography. After the spelling reform in 1726 (Hernando Cuadrado, 1997), the sequences /dzi/-/dze/ and /tsi/-/tse/, which were spelled *zi,ze* and *çi,çe*, have been conflated in *ce,ci* (*catorce* 'fourteen' < *catorze*). Similarly, Old Spanish /dza/-/dzo/-/dzu/ (*za,zo,zu*) and /tsa/-/tso/-/tsu/ (*ça,ço,çu*) have been conflated in *za,zo,zu* (*cabeza* 'head' < *cabeça*). For this reason, one would need to use Old Spanish orthography to estimate the functional load of this merger.

Phonemes	MinPairs	Entropy loss
ʎ-l	1	**.0018**
ʎ-j	1	.0012
ʎ-ɲ	1	.0005
ʎ-ɟ	0	.0003

Table 3.16: Functional Load for consonant contrasts in Slovak.

in its featural neighborhood, and test the least resistance hypothesis.

The merger yields only one minimal pair: *calló* 'be-quite.3SG-PAST' and *cayó* 'fall.3SG-PAST', which is the best alternative. Entropy loss would instead favor a merger with /ɲ/.

This is a merger which is compatible with functional load when we look at minimal pairs, while the evidence from entropy loss is unclear.

3.4.7 Slovak

For Slovak, child directed speech is not available in CHILDES. However, the Slovak National Corpus (Horák et al., 2004, Šimková, 2006) provides word lists calculated over lemmas and whole words. The lemma word list contains 8712 entries. By filtering the most 5000 frequent ones, we obtain a word list with 5000 lemmas for ≈850M tokens.

2.4.7.1 Consonants In Slovak, /ʎ/ is reported as completely merging with /l/ in western varieties (Horák et al., 2004). Like the Spanish case, this is a case in which we investigate the least resistance hypothesis.

Table 3.16 shows that functional load does not really explain this merger: minimal pairs are rare in general, and from the viewpoint of the entropy loss, this change is actually anti-functional.

2.4.7.2 Vowels The phoneme /æ/ is reported as merged, but the outcome varies. According to Horák et al. (2004), the phoneme is merged with /ɛ/ for most of the speakers. A merger with /a/ is also reported in Král' (1988) and Krajčovič (1988).

A problem with this case is that the two phonemes are the only targets that make sense phonetically. Table 3.17 shows that the merger with /ɛ/ seems to be the optimal one, while

Phonemes	MinPairs	Entropy loss
æ-ɛ	**0**	**.0007**
æ-a	1	.0007

Table 3.17: Functional Load for vowel contrasts in Slovak.

the merger with /a/ yields one minimal pair: *spät* 'back' and *spať* 'sleep'. In this case, we can evaluate the merger with /ɛ/ as functional by our definition, while there isn't enough evidence to evaluate the second merger, especially because it is not reported as being as frequent as the first one. For this reason, it is simply ignored.

3.4.8 Korean

Wedel et al. (2013) include the loss of laryngeal contrasts in post-vocalic obstruents in Korean in their analysis, and some additional mergers of obstruents in coda position. The two types of mergers are described in Sohn (1999). As Silverman (2010) points out, the loss of laryngeal contrast (referred to as *aplosivization*) is not a merger in progress, but a case of synchronic neutralization, and therefore a case that is not comparable with the mergers in progress examined here. For this reason, I will exclude a more detailed treatment of the case, which can be found in Silverman (2010).

3.4.9 Cantonese

For Cantonese, we can use CANCORP, in CHILDES, the same corpus used in Surendran & Niyogi (2006). The corpus contains ≈420K tokens of child-directed speech, for a total of 3769 wordforms.[15]

2.4.9.1 Consonants   Wedel et al. (2013) report two cases of consonant mergers in Hong Kong Cantonese: the merger between syllable-final /-n/ and /-ŋ/, from Zee (1985), and the well studied /n/-/l/ merger (Zee, 1999), which was also investigated in Surendran & Niyogi (2006). The two mergers are described as qualitatively different in the literature:

---

[15]The count includes homophones, that were separated using the English translations available in the corpus.

while the /n/-/l/ merger is completed for most young speakers, the merger between syllable-final /-n/ and /-ŋ/ is described as a complex sound change, with the characteristics of an allophonic split conditioned on the preceding vowel (Zee, 1999, 156), and therefore not a merger, at least in Hong Kong Cantonese. For this reason, only the first sound change will be investigated.

While counting minimal pairs for Cantonese is straightforward, measuring entropy requires changing our trigram model over phonemes, because this model ignores the presence of suprasegmental features like tones, which are important for distinguishing among syllables in tonal languages. The entropy model I adopted for this case is the same proposed by Surendran & Niyogi (2006) and described at the beginning of the chapter: a word unigram model, which calculates entropy using word frequencies, and therefore it is sensitive to the amount of homonymy introduced in the lexicon after a sound change.[16]

A problem that we need to deal with is that the merger between /n/-/l/ is typically described as occurring in onset position, but it is actually unconditioned, because the contrast does not appear outside of onset position, since /l/ cannot occur in codas. The reason why it is described in this way is that while [n] becomes [l] at the beginning of the syllable, it remains unchanged at the end of the syllable, so that after the change the two phonemes are in complementary distribution. This situation makes the change ambiguous between a conditioned and an unconditioned sound change.

Surendran & Niyogi (2006) decide to treat the merger as a word-initial merger, and therefore their entropy measure only captures the amount of information that would be lost if any consonantal pair would be neutralized in word-initial position only. This is an arbitrary choice for a couple of reasons: the merger is described in the literature as occurring at the beginning of syllables, and thus technically it could also occur word-internally if

---

[16]The languages investigated so far had stress contrasts, but these have been ignored for practical purposes, because the position of the stress was largely predictable. However, especially in the case of French, we have seen that the idealization does not always hold, because some of the minimal pairs that I reported were in fact contrasting according to the position of the stress. This implies that the entropy models described so far are not appropriate for languages that have a stress contrast and whose average word length is long, because a phoneme trigram model would ignore the stress information, and a word unigram model like the one devised for Cantonese would suffer from a sparsity problem, because minimal pairs would be very uncommon. In these cases, one would need to devise new models.

Phonemes	MinPairs	Entropy loss
n-l	**21**	.0004
n-t	20	.0004
n-t$^\text{h}$	13	.0001
n-s	11	.0003
n-ts	19	.0006
n-ts$^\text{h}$	21	.0018

Table 3.18: Functional Load for consonant contrasts in Cantonese, only word-initial.

Phonemes	MinPairs	Entropy loss
2-5	84	.0034
2-1	343	.0081
2-3	285	.0090
6-5	68	.0030
3-5	75	.0017

Table 3.19: Functional Load for tone contrasts in Cantonese.

the word has more than one syllable (for instance *naai4naai2*, 'mother-in-law'); moreover, since the contrast only appears in onsets, it might well be that the merger is actually unconditioned, and that the phonetic alternation is allophonic. Since one could argue that /n/ is still present in the language, I will temporarily treat the merger as occurring in onset position only (but cf. the discussion in 3.5 about the possibility of an unconditioned merger).

Table 3.18 shows that the number of minimal pairs associated with this merger is very high, and along with /n/-/ts$^\text{h}$/, it is the maximum number of minimal pairs we find.

If we calculate entropy for the merger as only occurring in onset position, we obtain the same result in Surendran & Niyogi (2006), with the opposition between /n/ and /l/ being quite high, but not the highest one. The merger appears to be anti-functional when we look at minimal pairs, while the evidence from entropy loss is unclear.

2.4.9.2 Tones   Some ongoing merger has been reported for tones as well. In particular, Mok & Wong (2010) report several works showing that the two raising tones ('2' and '5') are merging.

Table 3.19 shows that the contrast between the 'high-raising' tone ('2') and the 'low-

raising' tone ('5') is much lower than that between '2' and the 'high-level' ('1') and 'mid-level' ('3') tones, in terms of both minimal pairs and entropy. However, the contrasts involving ('5') are comparable. Mok & Wong (2010) report that the outcome of the merger varies among speakers: some of them merge the two tones to '2', some to '5', and others to some pattern in between. Interestingly, the paper reports no cases of mergers between '5' and other tones, even though according to our calculations they should be even more likely, given functional considerations. Entropy loss shows the same result, with the mergers involving '2' favoring the '2'-'5' merger, but the mergers between '6'-'5' and '3'-'5' both carrying a lower functional load than '2'-'5'. This is likely to be another case in which functional considerations do not play a role.

3.5   Summary and Discussion

We have examined a large number of reported ongoing mergers using electronic corpora and two functional load measures. We have identified three types of mergers: i) mergers which are functional, in the sense that they target low-contrastive phonemic pairs: ii) mergers which are neutral, in the sense that it is not clear whether they target low-contrastive pairs or whether they just happen to be among the many pairs that could have merged; iii) mergers which are anti-functional, in the sense that they target phonemic pairs which bear a lot of information compared to other mergers that could have occurred. I summarized them in Table 3.20 and Table 3.21.

In terms of minimal pairs, the mergers appear functional half of the time (14/27). In terms of entropy loss, however, for the majority of the mergers (19/27) the evidence for functional load is unclear. This is compatible with Wedel et al. (2013)'s finding that minimal pairs are better predictors of mergers than entropy loss (see Appendix D for a statistical validation of the result).

The result is interesting. The number of minimal pairs is a word type measure, and therefore it is insensitive to token frequencies. This suggests that speakers might be more sensitive to lexical contrast in its simplest implementation, rather than changes in the

Phonemes	Language	Type	Functional	Neutral	Anti-Functional
aɪ-ɔɪ	English (UK)	Unconditioned	X		
ɪə-ɛə	English (UK)	before /r/ (?)			X
ɜː-ɛə	English (UK)	before /r/ (?)		X	
t-d	North American Eng.	V́_V			X
w-ʍ	North American Eng.	Unconditioned		X	
ɑ-ɔ	North American Eng.	Unconditioned	X		
ɔɪ-ɛr	North American Eng.	Unconditioned	X		
ɪ-ɛ	North American Eng.	before [NASAL]		X	
ɪ-i	North American Eng.	before /l/			X
ʊ-u	North American Eng.	before /l/		X	
ʊ-oʊ	North American Eng.	before /l/		X	
ʌ-ɔ	North American Eng.	before /l/	X		
ʌ-oʊ	North American Eng.	before /l/	X		
ɛ-eɪ	North American Eng.	before /l/			X
ʌ-oʊ	North American Eng.	before /l/	X		
f-v	Dutch	Unconditioned	X		
s-z	Dutch	Unconditioned	X		
χ-ɣ	Dutch	Unconditioned	X		
ɛ-e	French	Unconditioned		X	
œ-ø	French	Unconditioned	X		
o-ɔ	French	Unconditioned	X		
ɛ̃-œ̃	French	Unconditioned	X		
ʎ-j	Spanish	Unconditioned	X		
ʎ-l	Slovak	Unconditioned		X	
æ-ɛ	Slovak	Unconditioned	X		
n-l	Hong Kong Cantonese	onset (?)			X
2-5	Hong Kong Cantonese	Unconditioned		X	
TOTAL			14	8	5

Table 3.20: Functional Load measured with minimal pairs.

Phonemes	Language	Type	Functional	Neutral	Anti-Functional
aɪ-ɔɪ	English (UK)	Unconditioned	X		
ɪə-ɛə	English (UK)	before /r/ (?)		X	
ɜː-ɛə	English (UK)	before /r/ (?)		X	
t-d	North American Eng.	V́_V			X
w-ʍ	North American Eng.	Unconditioned			X
ɑ-ɔ	North American Eng.	Unconditioned		X	
ɔɪ-ɛr	North American Eng.	Unconditioned		X	
ɪ-ɛ	North American Eng.	before [NASAL]		X	
ɪ-i	North American Eng.	before /l/		X	
ʊ-u	North American Eng.	before /l/		X	
ʊ-oʊ	North American Eng.	before /l/		X	
ʌ-ɔ	North American Eng.	before /l/		X	
ʌ-oʊ	North American Eng.	before /l/		X	
ɛ-eɪ	North American Eng.	before /l/		X	
ʌ-oʊ	North American Eng.	before /l/		X	
f-v	Dutch	Unconditioned	X		
s-z	Dutch	Unconditioned		X	
χ-ɣ	Dutch	Unconditioned	X		
ɛ-e	French	Unconditioned		X	
œ-ø	French	Unconditioned	X		
o-ɔ	French	Unconditioned		X	
ɛ̃-œ̃	French	Unconditioned		X	
ʎ-j	Spanish	Unconditioned		X	
ʎ-j	Slovak	Unconditioned			X
æ-ɛ	Slovak	Unconditioned	X		
n-l	Hong Kong Cantonese	onset (?)		X	
2-5	Hong Kong Cantonese	Unconditioned		X	
TOTAL			5	19	3

Table 3.21: Functional Load measured with entropy loss.

entropy of the system. This could also explain why calculating functional load on token frequency did not yield positive results in this case. In particular, finding mergers whose functional load was on average or above average was much more likely (22 vs 5) than finding mergers whose functional load was low when using entropy loss over word tokens, while the difference between the two cases was noticeably reduced when we used minimal pairs as a proxy for functional load (13 vs 14). Independently from the level of representation (whole-word units like in Surendran & Niyogi (2006), or subword units like phoneme trigrams), it might be that entropy loss is just not a good proxy for functional load, or that the formulas so far proposed in the literature are not capturing contrast in the same way as speakers do.

However, if we believe that a high number of minimal pairs is sufficient to prevent a merger, why do we have mergers which appear to be anti-functional? One way to answer this question is to state that functional load is one of many factors that influence sound change, and therefore it has weak predictive power in the single case (Kaplan, 2011); it follows that any report of anti-functional mergers should be considered a statistical accident. Clearly, though, the presence of anti-functional mergers can be used to question the very existence of a functional load factor, if functional load is stated in terms of a universal law (Sampson, 2015, 2020).

An alternative way of addressing this question is to examine the cases which appear anti-functional and to try to explain them. In this direction, from the inspection of Table 3.20, there are a couple of observations that one could make.

- most of the unconditioned mergers (11/15) are consistent with the Functional Load Hypothesis. This is not true for the conditioned mergers (3/12). These proportions are different, according to a Chi-square test (χ^2=6.24, p=0.013).

- all the mergers that violate the Functional Load Hypothesis are conditioned on a phonological environment

These observations point toward the hypothesis that functional load is a factor that is only identifiable in unconditioned mergers, namely mergers that completely remove a

contrast from the grammar. When the merger is phonetically motivated, possibly as a result of coarticulation, and limited to a phonological environment, functional load considerations are less relevant to decide whether to merge two sounds. In fact, by running a statistical analysis on conditioned mergers only, minimal pairs become not significant as a predictor (see Appendix D).[17] The only potential exception of this pattern is the merger between /ɪə/ and /ɛə/ (the NEAR-SQUARE merger), because its status as an unconditioned merger is not established.[18] One can also point to the /n/-/l/ merger in Cantonese being potentially unconditioned, but in that case the merger would not violate functional load: in fact, if we look at contrasts outside onset position, while the number of minimal pairs for /n/-/l/ stays the same (because /l/ does not appear elsewhere), the number for the other contrasts increase. For instance, the contrast with /tʰ/ would yield 73 pairs instead of 13, and therefore the merger would not violate functional load.

The question at this point is: what is the source of the correlation between lexical contrast and sound change, and why would this factor influence one kind of sound change, but not the other? Wedel et al. (2013) explain their results by arguing that functional load can be considered a bias that influences production, other than perception, in both children and adults, who would produce exaggerated tokens when using words that have minimal pairs in the lexicon (Wedel, 2012). This explanation, however, does not predict an asymmetry between conditioned and unconditioned mergers. One potential way to capture this asymmetry is to look at language acquisition.

Cui (2020) shows that a model of phonological learning based on minimal pairs can explain how children form categories based on variable acoustic input. The identification of the phonological categories must be on its way when children are very young, because experimental evidence shows that children lose sensitivity to phonetic differences very early in the course of the developmental period. While at birth they are able to distinguish among

---

[17]This is compatible with the finding that word-final lenition does not seem to be motivated in terms of functional load (Cohen Priva, 2017).

[18]The NEAR-SQUARE merger appears to be unconditioned in New Zealand English (Gordon et al., 2004, Bauer et al., 2007). This clearly violates the predictions, unless one can show that the change is not endogenous, but was triggered by contact with (imported) UK varieties in which the merger was historically limited to the rhotic environment.

different sounds, within their first year of life their perceptual system adapts to the sounds which characterize the language to which they are exposed, to the point that they lose sensitivity to fine phonetic distinctions that are not associated with the phonemic contrasts of the language they are learning (Werker & Tees, 1984, Kuhl et al., 2006).

On the other hand, when children have acquired an inventory of phonemes, they can start paying attention to the alternations among them, for instance realizing that in English the sound /t/ has a different phonetic realization depending on whether it appears at the beginning of the word or intervocalically (Klein & Altman, 2002). The developmental window for learning these allophonic rules is estimated to start around their first birthday (Jusczyk et al., 1999b, White et al., 2008), and it can last until they are 5-year old (Richter, 2018).

In this scenario, it is not implausible that unconditioned mergers occur in the first stage, while conditioned mergers occur in the second stage. Unconditioned mergers occur when a child fails to acquire a phonological boundary between two categories, which are then acquired as a unique entity. Conditioned mergers, on the other hand, do not affect the phonological inventory: the number of categories is the same before and after the merger. The only difference in the two stages is the presence of an allophonic rule, and the fact that an entire class of words (like the class of *pen* words, which contain the sequence /ɛ/ + [nasal]) is reanalyzed as containing a different phoneme.[19] While these last cases are technically described as mergers, they do not yield the loss of a phonological contrast in the grammar.

This would explain why functional load is strongly associated with unconditioned mergers, but it does not explain conditioned mergers: by the time children reach the stage of allophonic rule-learning and the phonological representation of words become adult-like, their categories are already fixed. In this second stage, the learning strategy might be different, and other properties of the lexicon, like morpho-phonological alternations or finer phonetic distinctions, can play a greater role (cf. Richter, 2018).

---

[19]This case case corresponds to a 'merger by phonological transfer', following the terminology in Dinkin (2016).

The next chapter will further consider whether this revision of the Functional Load Hypothesis, which crucially focuses on lexical contrast at the type level, can be supported by what we know about the acquisition of phonemic contrasts.

Chapter 4

Lexical contrast and child language acquisition

In the previous chapter, we have seen that historical mergers, at least those which are unconditioned, tend to target contrasts which are not well represented in the lexicon. In particular, the lack of minimal pairs turned out to be a better predictor of unconditioned mergers than functional load calculated over word tokens.

This hypothesis comes with some predictions. If mergers can result from the misperception/misproduction of a sound (Ohala, 1983), and we find that lexical contrast can be used to predict the occurrence of mergers, then we might expect to also see a correlation between lexical contrast and the misperception/misproduction of a sound in the developmental stage in which sound contrasts are learned.

Previous research pointed toward a relationship between lexical contrast and phonological development (Locke, 1988, Walley, 1993, Luce & Pisoni, 1998, Metsala & Walley, 1998, Storkel, 2002, Coady & Aslin, 2003, Swingley & Alarcon, 2018). In particular, phoneme frequency and functional load have been shown to predict the age of emergence and the production accuracy of sounds in early child speech.[1] Stokes & Surendran (2005), following some preliminary results in Pye et al. (1987) and Ingram (1988), showed that functional load is a good predictor of the age of emergence of consonants in English. Moreover, phoneme frequency turns out to be a good predictor of phonological development in Cantonese and in Dutch. Edwards & Beckman (2008a) show that within-language syllable frequencies are

---

[1]As explained in the previous chapter, functional load can be used to generalize the notion of lexical contrast to a subword level.

correlated with syllable production accuracy in child speech in four different languages.

One of the challenges in studying phonological development is that while it is easy to obtain *production* data from children, such data do not necessarily reveal whether a contrast is present at the *perception* level. Therefore, while certain factors seem to correlate with children's speech production, it is not clear whether they also influence their speech perception. Children typically perform well in discrimination tasks, even at a relatively young age (Eimas et al., 1971, Kuhl, 1979, Kuhl, 1980, Swingley, 2003). However, the literature reported cases in which children's behavior revealed the absence of a contrast in perception, for instance between [s] and [z] (Eilers & Minifie, 1975) and [f] and [θ] (Eilers & Oller, 1976) in English. In such cases, it is unclear if the difficulty in discriminating between these sounds is simply based on their acoustic properties, or if structural considerations also play a role. Another fact which suggests a relationship between structural considerations and misperception is the finding that after their first year of life, children fail at discriminating contrasts which do not distinguish among phonemes in their ambient language (Werker & Tees, 1984, Kuhl et al., 2006).

The hypothesis of a relationship between misperception of phonemic contrasts and structural considerations is interesting from the perspective of sound change. Misperception or attention errors can induce children to create a lexical representation different from their caregiver's one (Macken, 1980a, Vihman, 1982), and if the errors are not corrected within the critical period, they can result in a change in the phonological grammar.[2] Showing a relationship between errors in lexical representations and structural considerations would allow one to explain the correlation between historical mergers and functional load as a possible outcome of language acquisition: if a contrast is not acquired within the critical period, then a child might develop a phonological grammar in which that contrast is absent, an event which would constitute the actuation of a sound change.

In order to achieve this goal, in this chapter I study children's production mistakes

---

[2]'It seems reasonable to expect perceptual-encoding rules to be more difficult to change and that therefore the evidence for such a rule (i.e. incorrectly stored lexical representations) will persist longer' (Macken, 1980a, 11)

in two of the languages of the previous chapter (American English and Cantonese) and two additional languages (Greek and Japanese) using the Paidologos Corpus (Edwards & Beckman, 2008a) in CHILDES.

First, I use a regression analysis to predict the accuracy of the segments in production given structural factors associated with phoneme frequency and functional load (like in Stokes & Surendran, 2005).

Second, I perform a qualitative analysis of the errors made by the children, to see if they are uniform across different languages. In case they do not, I investigate the hypothesis that the differences in the quality of the errors are correlated with differences in the distribution of the sounds in the lexicon of the target languages.

Third, I use one of the case studies investigated (the contrast between /f/ and /θ/ in English and Greek) to show how an acquisition model can account for the correlation between lexical contrast and phonological development.

4.1 Early phonological development

4.1.1 Categorical perception

Starting from the '70s, experimental research showed that infants are remarkable at distinguishing different sounds in perception studies. Eimas et al. (1971) show, through a high-amplitude sucking paradigm, that 1-month- and 4-month-old infants learning English can discriminate between two synthetic sounds [p] and [b] along the same acoustic dimension that adults English speakers use to distinguish between the two corresponding phonemes, while they are less sensitive to variation that does not cross the boundary between the two phonemes. This finding shows that not only children are sensitive to phonetic distinctions, but they categorize them in adult-like manner. This result was replicated, and confirmed for other voicing contrasts, in Trehub & Rabinovitch (1972), and it was robust to the use of natural instead of synthetic speech. Discrimination was also found among contrasts based on place of articulation (Moffitt, 1971, Eimas, 1974, Eilers & Minifie, 1975, Miller & Morse,

1976), the contrast in the approximants [r] and [l] (Eimas, 1975), and contrasts involving manner of articulation (Eimas & Miller, 1980), using the same paradigm. Using a head-turn preference procedure, Hillenbrand et al. (1979) show that children between 6-month- and 7-month-old can distinguish between the stop [b] and the semivowel [w]. Through the same paradigm, researchers also showed that children of the same age can distinguish between [s] and [ʃ] (Kuhl, 1980) and between stop and nasal consonants (Hillenbrand, 1983). Similar studies were conducted on vowel discrimination. Trehub (1973) shows that infants between 1-month- and 4-month old can discriminate the vowels [a], [i] and [u], in the presence or in the absence of an onset. The result was confirmed in Kuhl & Miller (1975), in which it is also shown that discrimination is robust to variation in pitch contour (cf. also Kuhl & Miller, 1982). The results have been also replicated for highly confusable vowels like [i] and [ɪ] (Swoboda et al., 1976) and [a] and [ɔ] (Kuhl, 1983).

4.1.2 Failures of acquiring a contrast

Kuhl (1980) reports that negative results are limited to few cases of contrasts involving fricative consonants. Eilers & Minifie (1975) show that one of the few contrasts that infants fail to discriminate through a high-amplitude sucking paradigm is the voicing contrasts between [s] and [z]. Distinguishing this contrast, and the contrast between [f] and [θ], turned out to also be difficult for 2-year-old children tested through a head-turn procedure (Eilers et al., 1977), even though some positive results were reported in Jusczyk et al. (1979) and Levitt et al. (1988).

However, naturalistic data have been used to show that certain contrasts are not present in children's grammars until late. Macken (1980a), using longitudinal data from Smith (1973), shows that the child under investigation (in the age range 2;00-4;00) produces words containing a /təl/ sequence (like 'bottle') as [kəl], and viceversa ('circle' comes out as [sə:təl]). This cannot be the result of an articulatory error or a phonological rule, because words like 'little' and 'tickle' are produced correctly throughout the whole period. Using the same data, Grunwell (1982) shows that a similar pattern can be identified for the cluster

/tr/ and /tʃ/: initially, they are simplified to [t], but when they emerge in the speech of the child, they do not map to the target form, because words like 'chocolate' are pronounced as [tɹaklit]. Similarly, Vihman (1982) shows that her daughter uses [f] and [θ] interchangeably at the age of five, with words like 'with' being pronounced [wif] and 'wife' being pronounced [waɪθ].[3]

These findings suggest the possibility that even though infants are good at discriminating sounds in a laboratory environment, they have more difficulties in mapping speech sounds to their target categories in natural settings (MacKain, 1982, Jusczyk, 1992), and therefore in converging to an adult-like lexical representation for the words they learn.

4.1.3 The emergence of speech sounds in production

While most of the literature mentioned so far focused on perception, other works aimed instead at estimating the age of emergence of sounds in production. This type of data, especially when collected across languages, is informative to determine the extent to which phonological development is constrained by articulatory factors, which are universal by definition.

Several works followed the hypothesis in Jakobson (1941/1968) that articulatory development drives phonological acquisition (Templin, 1957, Sander, 1972, Locke, 1983, Dinnsen, 1992, Kent, 1992, MacNeilage et al., 2000). For English, a classic reference is Grunwell (1982), in which one can find a profile of phonological development based on the results of several experimental studies conducted in the previous decades (Templin, 1957, Olmsted, 1971, Sander, 1972, Prather et al., 1975, Crystal, 1976, Arlt & Goodban, 1976), in which children were typically tested on picture-naming or word repetition tasks in the age range 2;00-5;00. Another source of developmental data is the Iowa Articulation Norm Project (Smit et al., 1990), a joint effort between the Iowa Department of Education and the University of Iowa to assess the acquisition of speech sounds in preschoolers and children attending public and private schools in Iowa during the 1985-1986 school year. The study

---

[3]Vihman notes that 'wife' was pronounced correctly before the child started to produce [θ].

has been replicated in Nebraska two years later (the results have been included in Smit et al., 1990).

The data collected in Grunwell (1982) and Smit et al. (1990) have been summarized in Figure 4.1, by Graham Williamson.[4] According to the author of the graph, 'The left-hand edge of each horizontal bar represents the age at which 50% of children produce the particular consonant correctly and use it in their speech. The right-hand edge of each horizontal bar represents the age at which 90% of children have mastered the use of the particular consonant in their speech'. Even though the two references are not directly comparable, because Grunwell reports the results from the experimental literature, while the Iowa and Nebraska projects relied on transcriptions from trained clinicians, they can provide us with a general picture of phonological development in English.[5]

By the age of 2, children already master the two bilabial plosive consonants /p/ and /b/, the two nasal consonants /m/ and /n/, the labiovelar /w/ and the glottal /h/.[6] Grunwell (1982) classifies this set as Stage 1, the only difference being that also the alveolar plosive /t/ and /d/ are included, and that /h/ is excluded.

By the age of 3, children expand their plosive set to include velar and alveolar consonants (/k,g,d,t/), the velar nasal /ŋ/, the labio-dental fricative /f/ and the glide /j/.[7] The next stage is characterized by the liquid consonants /r/ and /l/ and the fricative /s/, which appear when children are past their third birthday.[8] The final stage is characterized by the

---

[4]https://www.sltinfo.com/wp-content/uploads/2014/01/speech-sound-development-chart.pdf

[5]Note that relying on speech transcriptions to estimate phonological development, rather than eliciting data in a laboratory and then running acoustic analyses, adds a layer of human intervention that might obscure interesting patterns in the data. Transcribers can have different sensitivity in whether they perceive a sound within or outside certain phonemic boundaries.

[6]This set maps to Set 1 of the articulatory hierarchy in Kent (1992), which represents an attempt to map the phonological development in English with the physiological development of the speech apparatus. The only difference is that Kent attributes the voicing distinction in plosive to a later stage. The fact that voiceless segments are easier to produce than voiced segments has been widely confirmed by experimental studies in different languages (Macken & Barton, 1980a, Macken & Barton, 1980b, Allen, 1985, Gandour et al., 1986, Pan, 1994, Davis, 1995).

[7]This set maps to Grunwell's Stage 2 and to Set 2 in Kent (1992)'s hierarchy, the only difference being that for Grunwell /f/ and /j/ appear at a later stage, which is captured in the graph by their shifted position with respect to the second set of segments.

[8]Here Kent's hierarchy disagrees with the experimental findings, because it considers the mastering of the tongue configuration required for the liquids /r,l/ easier than the fine force regulation required for alveolar frication /s/, while Grunwell (1982) reports that /s/ appears before /r/ and /l/ in child speech.
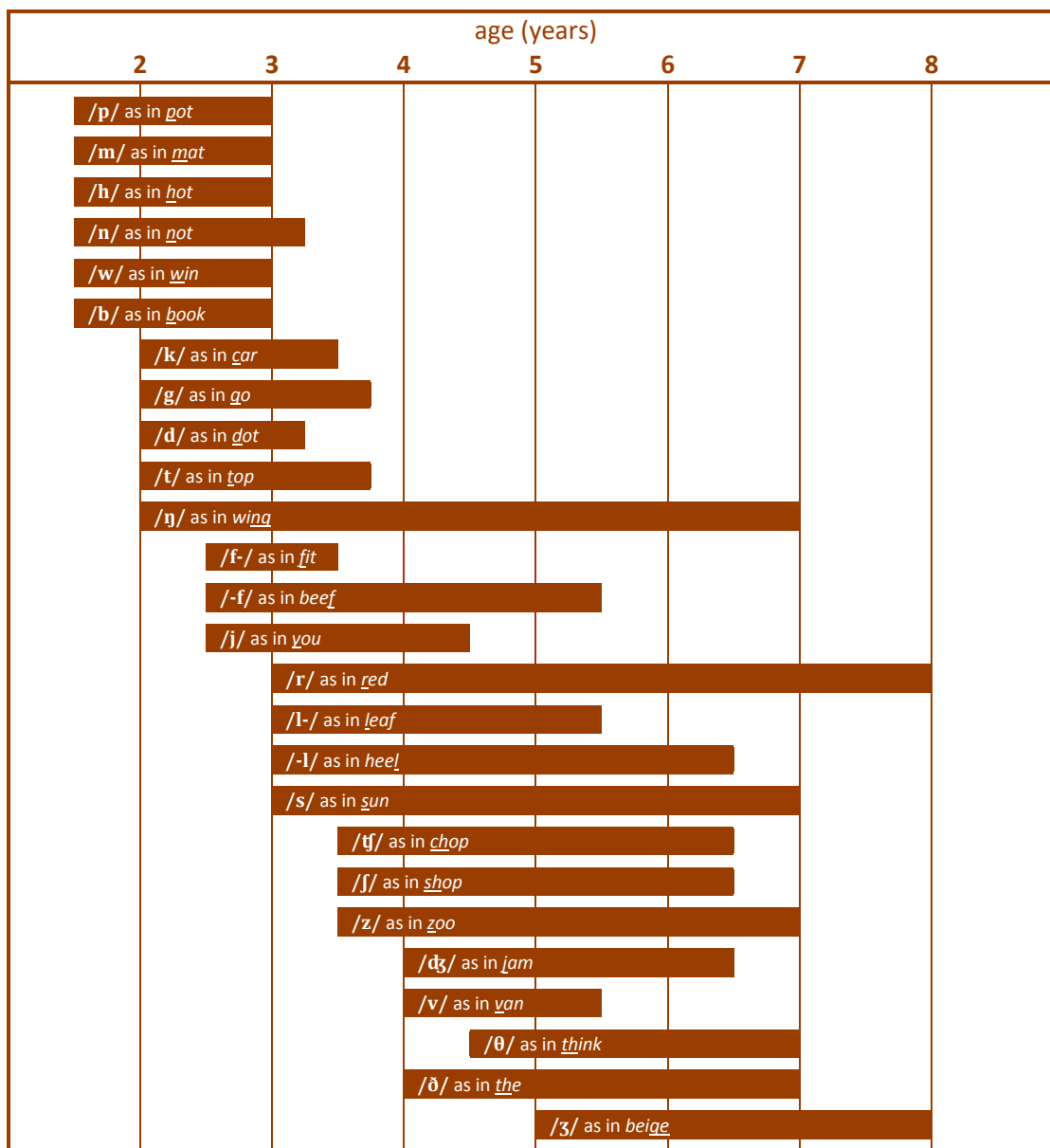
Figure 4.1: Phonological development, by Graham Williamson.

other fricative consonants and the affricate /tʃ/, which are all produced by the age of 5, but are fully mastered only at 8.

These data give us an overview of phonological development in English, and they provide us with a baseline that can be compared with findings in other languages.

4.1.4   Cross-linguistic differences in phonological development

Many works have argued against a universal account for child phonology based on articulatory development (Macken, 1980b, Menn, 1983, Pye et al., 1987, Stokes & Surendran, 2005, Edwards & Beckman, 2008a). The main argument for this position is that the analysis of languages other than English revealed some salient difference in phonological development.

Macken (1979) tracks the phonological development of a Mexican Spanish child in the age range 1;7-2;5. The consonants produced at the beginning of the period are roughly mapping to Grunwell's Stage 1. However, she notices that voiced consonants are missing in a first stage. Interestingly, the child appears to master fricative consonants earlier than what we would expect based on English: /f/ and /s/ both appear at 1;1, and /tʃ/ is contrastive by 2;1. This finding is surprising, given that fricative and affricate consonants are associated with the last developmental stage in the experimental literature on English.

Pye et al. (1987) notice a similar pattern in five children learning Quiché Mayan, who master the lateral /l/ and the affricate /tʃ/ earlier than children learning English. Chevrie-Muller & Lebreton (1973) and Vihman (1993) also report an early use of /l/ in children learning French, and Amayreh & Dyson (2000) show that the same is true for children learning Arabic. Hua & Dodd (2000) and Hua (2002) report early use of both fricative and affricate consonants for children learning Mandarin Chinese.

Ingram (1988) notices that /v/ is acquired before the age of 2;00 in Swedish, Estonian and Bulgarian, while in English it usually does not appear until 4;00. Li et al. (2009) notice that while in English /s/ is acquired earlier than /ʃ/, this is less clear in Japanese. Other differences have been noted in Turkish (Topbas, 1997) and Cantonese (So & Dodd, 1995, Stokes & To, 2002).

These differences require an explanation which must be independent from universal considerations about articulatory development.

4.2   Explaining cross-linguistic differences with the properties of the ambient language

One possible explanation for cross-linguistic differences in phonological development is that the properties of the ambient language can influence the acquisition of phonological contrasts. Pye et al. (1987) and Ingram (1988) propose that the frequency of phonemes in the target words can predict their appearance in phonological development. de Boysson-Bardies et al. (1989) show that the vowels that are first produced by children in the babbling stage match the most frequent vowels in the ambient language. The finding can be generalized to consonants (de Boysson-Bardies & Vihman, 1991). Amayreh & Dyson (2000) also point toward the high frequency of occurrence of /l/ and /j/ to justify their early acquisition in Arabic. Crucially, many of these works stress the need of developing functional load measures in order to account for the contrastiveness of phonemes in the lexicon.

The first study which aimed directly at investigating the relationship between cross-linguistic differences in developmental patterns and phoneme frequency and functional load is Stokes & Surendran (2005). The question is also addressed in Edwards & Beckman (2008a). Both studies will be reviewed in the next subsections.

4.2.1   Stokes and Surendran (2005)

The aim of Stokes and Surendran's study is to determine whether articulatory complexity, phoneme frequency, and functional load are predictors of consonant development in Cantonese, English and Dutch. In order to operationalize the notion of articulatory complexity, Stokes and Surendran use the hierarchy proposed in Kent (1992), which is based on four sets of physiologic characteristics necessary for motor control at different levels of articulatory complexity (cf. also Dinnsen, 1992). As for phoneme frequency, the authors

decide to focus on word-initial segments, a decision based on the fact that children are less sensitive to contrasts which are not word-initial. Functional load is estimated using the entropy formula in Surendran & Niyogi (2006), which was described in the previous chapter. Since production accuracy, articulatory complexity and frequencies are measures that can be calculated for each phoneme, while functional load is a measure that involves contrasts between phonemes, one needs a way to assign a single functional load value to each phoneme. Stokes and Surendran do so by calculating the average functional load between a phoneme and all the phonemes which share the same place of articulation in the language, and using that as an estimate of the functional load of the phoneme. This is motivated by the fact that when children make perception or production mistakes, those are usually associated with aspiration, manner, or voicing, but not with place (Eilers & Oller, 1976, Kuhl, 1980, Paschall, 1983), a fact which suggests that place of articulation cues are the least ambiguous, and therefore can be used by children to identify phonemic contrasts. Both frequency and functional load are calculated using word token frequencies from written corpora.

As for the dependent variables, there are two ways in which the presence or not of a consonant in an inventory is estimated. The first is 'Age of Emergence', which in this work is equivalent to the period in which a child uses a consonant twice in spontaneous speech (Amayreh & Dyson, 2000, Hua & Dodd, 2000, Stokes & To, 2002). The second is the 'Accuracy of Production' (Beers, 1995, Paschall, 1983).

Table 4.1 summarizes the statistics associated with three different linear regression models, applied to the four phonological development datasets that the authors employed. Each model contains one of the independent variable of interest (articulatory complexity, functional load, and frequency) as the only predictor.

For English, the authors ran a linear regression analysis using age-of-emergence data from Robb & Bleile (1994), production accuracy data from Stokes et al. (2005), and frequency and functional load measures from CELEX. The analysis showed that both articulatory complexity and functional load were significant predictors of age of emergence, and

111

	Articulatory Complexity	Functional Load	Frequency
Eng - Age of Emergence	0.68**	0.74**	0.52*
Eng - Accuracy of Production	0.63**	0.44	0.12
Can - Age of Emergence	0.53*	0.46	0.79**
Dut - Accuracy of Production	0.51*	0.43	0.65**

Table 4.1: Summary of the findings in Stokes & Surendran (2005). R^2 values are reported. *p<0.05, **p<0.01.

consonant frequency was moderately significant, when the three variables were used to predict age of emergence in three separate models. When all predictors were put together, only functional load turned out to be significant. The authors then concluded that functional load is the best predictor of the age of emergence of a consonant in English.

The same analysis was run to predict production accuracy in English. In this case, only articulatory complexity turned out to be a significant predictor.

For Cantonese, only age-of-emergence data were available, from Stokes & To (2002). Frequency and functional load measures were derived from a corpus of spoken Cantonese. Both frequency and articulatory complexity accounted for age of emergency, but the effect of functional load was not significant. Controlling for frequency also made articulatory complexity not significant. The authors then concluded that phoneme frequency was the best predictor of age of emergence of phonemes in Cantonese.

For Dutch, only production accuracy data were available, in Beers (1995). Frequency and functional load measures were derived from CELEX. The results were similar to those obtained for Cantonese: frequency was a significant predictor, along with articulatory complexity, while functional load was not significant. Articulatory complexity became not significant when frequency was controlled for. Like in the previous case, frequency was the best predictor, even though here the dependent variable was production accuracy rather than age of emergence.

To summarize, Stokes and Surendran showed that properties of the target languages are indeed correlated with age of emergence (for English and Cantonese) and production accuracy (for Dutch) of consonants in child speech. The results, however, are not uniform across languages. In particular, while functional load is strongly correlated with age of

emergence in English, it is not in Cantonese, for which frequency is a better predictor. Moreover, functional load fails to predict the accuracy of production for both English and Dutch, and frequency is a significant predictor in the latter case, but not in the former.

4.2.2   Edwards and Beckman (2008)

The study by Stokes and Surendran has two limitations. First, the production data they used come from different projects, and therefore is not uniform, but it varies according to the focus of the papers (age of emergence versus production accuracy) and the methods of data collection. Second, the way articulatory complexity was controlled for crucially depends on Kent (1992)'s hierarchy.

These two factors were controlled for in another study, by Edwards & Beckman (2008a). Their study addresses the interaction between universal and language-specific effects on the development of phonological inventories by collecting data through a uniform experimental paradigm in four different languages (English, Greek, Cantonese, Japanese) and then using the cross-linguistic frequency of a phoneme as a proxy for its articulatory complexity, under the assumption that sounds which are easy to produce should be more common across languages than sounds that require more articulatory effort.

The methodology they employ is a repetition paradigm: 2-year-old and 3-year-old children are exposed to a picture; then, a native speaker names the picture, and asks the children to repeat the word they just heard. The pictures are chosen in order to elicit word-initial coronal and velar obstruents in all languages, and to represent as many different nuclei as allowed by the language vocabulary and the phonotactics of the languages.[9] Two phoneticians, which were native speakers of the languages, were then asked to transcribe the elicitations and to judge them as correct or incorrect. Limiting the analysis to coronal and velar obstruents allowed the researchers to exclude sounds which are universally acquired very early (like the labial /p/) and sounds which instead appear very late (like

---

[9]Greek and Japanese have a vowel inventory that roughly maps to five nuclei, /a,e,i,o,u/. English and Cantonese have more vowels in their inventory, but they were mapped to five nuclei to make a comparison with Greek and Japanese meaningful. For instance, English /ʌ,ɑ,ɔ/ were mapped to a unique category /a/, since they are typically described as low vowels.

/r/) from further investigation, and to focus on those phonemes for which different times of acquisition have been reported, in particular the interdental fricative /θ/ and the affricates /ts/ and /tʃ/. The analysis of the production accuracy of the children led to three main results.

First, the authors tested whether phoneme frequencies, derived from adult speech corpora, were correlated among languages. Instead of focusing on word-initial segments, the authors used CV sequences, therefore controlling for consonant-vowel coarticulation effects. Out of the six pairwise correlations, only two (English-Greek and Greek-Japanese) were marginally significant (respectively, R^2=0.15, p=0.045, and R^2=0.17, p=0.01). The rest of the pairwise correlations were not significant. This result led the authors to conclude that phonemic frequencies in the lexicon do not follow from some universal pressure, but they are mostly language-specific.

Second, the frequencies were used to predict production accuracy through a linear regression analysis. Two predictors were used in the analysis: the within-language frequency of the CV sequence in the target language, and the mean frequency of the same CV sequence in the other languages. This second predictor is used to estimate the frequency we would expect if there were a universal bias that favors or disfavors particular phoneme sequences. In this setup, one can see if language-specific or universal factors emerge, independently from one another. The results yielded a significant coefficient for the language-specific CV frequency in English, but significant coefficients for the across-language average CV frequency for Cantonese and Greek, while neither coefficient was significant for Japanese (see Table 4.2 for a summary). The authors concluded that both language-specific factors, in the case of English, and universal factors, in the case of Cantonese and Greek, must account for children phonological development.

To summarize, Edwards and Beckman showed that both language-specific and universal properties can explain production accuracy in phonological development, but they cannot be generalized to all languages, confirming the results in Stokes & Surendran (2005).

114

Language	R^2	Frequency, within-lang (t)	Frequency, across-lang (t)
English	0.58**	5.79**	−1.22
Cantonese	0.41**	1.62	−3.66**
Greek	0.26**	0.69	−3.44**
Japanese	0.13*	0.84	−1.73

Table 4.2: Summary of the findings in Edwards & Beckman (2008a). R^2 of the model and the t-values associated with the coefficients are reported. *p<0.05, **p<0.01.

4.3   The role of functional load: word types and word tokens

The two studies just summarized showed that there is a correlation between the production accuracy of consonants and language-specific factors, beyond the expectations from their articulatory properties. However, it is not clear which language-specific properties best correlate with phonological development.

Stokes & Surendran (2005) focus on the consonant contrasts in word-initial position, and they quantify them in terms of phoneme frequency and functional load. Their results show that while in English the effect of functional load was significant, in Cantonese and Dutch phoneme frequency was instead correlated with the children's behavior. Edwards & Beckman (2008a) detected a frequency effect for explaining production accuracy in English, but not for other languages, and they did not test any measure of functional load. Their last experiment shows that language-specific factors were necessary to explain the differences in production of fricative and affricate consonants across languages, but they did not speculate on what these factors might be.

In this section, I expand on Edwards & Beckman (2008a) by replicating their analysis, and focusing on different structural factors that can explain cross-linguistic differences in production accuracy.

4.3.1   Calculating frequency and functional load: types and tokens

Both Stokes & Surendran (2005) and Edwards & Beckman (2008a) based their measurements on token frequencies (e.g., the frequency of a phoneme in a corpus), but they did not include word type properties (e.g., the frequency of a phoneme in the vocabulary).

Previous works like Pye et al. (1987), Ingram (1988), de Boysson-Bardies & Vihman (1991) and Cychosz (2017), and the previous chapter on sound change, suggest that word types, rather than word tokens, might be what matters for explaining language acquisition and change. In particular, Ingram (1999) motivates the choice of using word types rather than word tokens when studyng phonological development by noticing that the phoneme /ð/ is very frequent at the token level in English, because of its presence in function words, but rare at the type level (because it is uncommon outside of those words), and it is this last property that might explain its late acquisition.

The hypothesis that children are more sensitive to word types rather than word tokens when it comes to rule learning is not novel (Endress & Hauser, 2011, Richtsmeier, 2011, Cui, 2020, Kodner, 2020). In particular, Yang (2016) shows that generalization over word types can be the cognitive mechanism that explains how children acquire syntactic and morphological rules, in spite of the presence of exceptions (Yang, 2016, 67). For this reason, in replicating the previous studies, I will integrate type counts for both frequency and functional load measures.

4.3.2 Data

Edwards & Beckman (2008a) derive frequency measures from corpora, but they recommend the use of child-directed speech over adult speech when it is possible. Since in the previous chapter we already focused on two of the four languages in their study (English and Cantonese), we can use the CHILDES corpora employed in that chapter to derive frequency and functional load measures from child-directed speech. For Japanese, I extracted word counts using the corpora in CHILDES, which all contained a romanized version of adult speech. For Greek, the corpora in CHILDES are not sufficient to match the size of the other wordlists, and so I decided to use the 5000 most frequent words in SUBTLEX[10], and then use the corpus GreekLex2 (Kyparissiadis et al., 2017) to filter out non-dictionary forms and transform the words into a phonological trascription, in order to obtain a wordlist of com-

---

[10]The list has been taken from https://github.com/hermitdave/FrequentWords, and was compiled in 2018.

	English	Greek	Cantonese	Japanese
Word Tokens	5252297	128259725	422573	1278073
Word Types	5000	5000	5000	5000
CV sequences (Tokens)	3847119	87927423	362457	975188
CV sequences (Types)	3340	2705	3110	3788

Table 4.3: Word types, tokens and CV sequences for the three child-directed corpora assembled from CHILDES (plus SUBTLEX, for Greek).

parable type and size with respect to the other languages.[11] See Table 4.3 for a summary of the data extracted.

Following Edwards & Beckman (2008a), CV sequences are extracted from the child-directed speech of the corpora, for which a phonological transcription is first obtained, and used to extract frequency and functional load type and token measures.[12] Frequency measures for each consonant are derived by summing the sequences containing the consonant, and dividing them by the total number of sequences, like in Edwards & Beckman (2008a). As for functional load measures, the methodology follows Stokes & Surendran (2005), which based their entropy calculations on the contrasts among word-initial consonants, and used a word unigram model (like in Surendran & Niyogi (2006), see previous chapter). However, in the previous chapter we mentioned that this model works well for Cantonese, but does not generalize to languages in which the average word length is longer, like Japanese and Greek: in these cases, there are less minimal pairs in general, and therefore entropy loss would always be quite low. For this reason, I decided to estimate entropy loss by keeping a unigram model, but using word-initial CV sequences (like in Edwards & Beckman, 2008a), rather than full words. The main difference with Stokes and Surendran's model is that while a word unigram model captures contrasts only if they lead to complete homophony (namely, they are only sensitive to how many minimal pairs would become homonyms after

---

[11]This strategy is not optimal, because it excludes many words that do not frequently occur with their dictionary form, but it is needed to limit the redundancy of inflected forms, and it is consistent with the strategy used to obtain the English wordlist, where inflected forms were filtered out through CELEX.

[12]English was phonemicized in ARPAbet, using the CMU pronunciation dictionary (Weide, 1998). Greek, in GreekLex2, was already phonemicized. Cantonese Jyutping and Japanese Hepburn, the standards used in CHILDES, required some postprocessing to better capture phonological distinctions. Cantonese, 'ng' (ŋ), 'kw' (kʷʰ), 'gw' (kʷ), 'aa' (/aː/), 'oe' (œː), 'eo' (ɵ), 'yu' (yː) and Japanese 'ch' (tʃ), 'sh' (ʃ), and 'ts' (ts) were re-coded as unique symbols. The vowel sequences 'aa', 'ee', 'ii', 'oo', 'uu' were also associated with independent symbols.

the loss of a contrast), a unigram model based on word-initial CV sequences is sensitive to any contrast that neutralizes the difference between two word-initial CV sequences. This follows from the intuition that children should be sensitive to contrasts among word-initial sequences, and in particular word-initial syllables, regardless of whether there are other contrasts toward the middle or the end of the word (namely, near-minimal pairs can also reveal the presence of a contrast), which might be especially true for languages in which words can be long, and therefore minimal pairs are less frequent.[13]

Production accuracy was calculated using the data from the Paidologos corpus. I extracted the production accuracy of the children in the first four groups (2a, 2b, 3a and 3b, which cover the children in the age range 2;00-4;00). For each word elicited, a phonetician assigned a '1' or a '0' to each target consonant depending on the accuracy of the production, making a binary decision between correct production ('1') or mistake ('0'). Accuracy can be calculated as the ratio of '1' over the total number of elicitations of the words containing the consonant. A second phonetician marked each mistake with a specific symbol when the outcome of the mistake was identifiable ('$', for substitution), while other symbols (like '+' and '#') were instead used when the target was not a phoneme in the language, or when the mistake was clear, but the quality of the error was not. Cases in which the production was not clear were coded with an independent symbol ('m'), and therefore can be excluded from the analysis. Furthermore, I excluded a few cases where the sound that appeared after the substitution symbol was the same target phoneme, because in this case I could not tell if there was a transcription error, of if the annotations by the two phoneticians were in disagreement.

The number of words associated with each consonant are in Table 4.4. The production accuracy, which represents the dependent variable, is displayed in Table 4.5. This table contains all the target consonants in the Paidologos corpus. I excluded those with no phonemic status, like [k^j] and [k^w] in English.

Japanese represents an interesting case study, because the dataset includes five obstru-

---

[13]This point will be discussed more in depth toward the end of the chapter.

	English	Greek	Cantonese	Japanese
t	579	596	679	505
k	590	374	568	473
d	582			377
g	238			475
t^h			440	
k^h			574	
k^w			108	
k^{wh}			116	
s	569	617	431	461
ʃ	565			453
θ	117	118		
x		246		
tʃ	536			459
ç		238		
dʒ				231
ts		355	458	116
dz				459
ts^h			441	

Table 4.4: Number of elicitations for each target consonant in Paidologos.

	English	Greek	Cantonese	Japanese
t	0.795	0.788	0.622	0.672
k	0.837	0.704	0.651	0.813
d	0.784			0.516
g	0.739			0.513
t^h			0.554	
k^h			0.447	
k^w			0.352	
k^{wh}			0.190	
s	0.491	0.473	0.551	0.319
ʃ	0.509			(0.387)
θ	0.222	0.342		
x		0.578		
tʃ	0.498			(0.552)
ç		0.449		
dʒ				(0.309)
ts		0.288	0.424	(0.287)
dz				(0.212)
ts^h			0.356	

Table 4.5: Production accuracy for 2- and 3-year-old children in the Paidologos corpus.

ents (/ɕ, tɕ, dʑ, ts, dz/) whose phonemic status is unclear, because they only contrast in loanwords, while they are allophones in native words. The palato-alveolar /ɕ, tɕ, dʑ/ have been aligned with English postalveolar /ʃ, tʃ, dʒ/, because even though they are described as phonetically different, they involve the same articulators.[14]

Table 4.5 shows that the languages pattern according to some of the phonological universal previously mentioned: voiceless stops are more accurate than voiced stops, and stops are in general more accurate than fricative and affricate consonants. However, some differences are also apparent. I will postpone a detailed comparison of these numbers to the next section, because in order to provide an account of the cross-linguistic differences, it is informative to also analyze the direction of the mistakes. In this section, I examine the correlation between these values and the frequency and functional load predictors.

4.3.3 Independent variables

For each language, four independent variables were calculated. The first two were the relative frequencies of the consonants in word-initial CV sequences, at the token and type level. Then, two functional load measures applied to the word-initial CV sequences were calculated, at the token and type level.

In this case, the functional load measure for each phoneme was the average functional load of the phoneme when contrasted with all the other phonemes investigated in this study for the relative language. Even though the phonemes are distributed across two different places of articulation (coronal and velar consonants), I decided against analyzing them separately (like in Stokes & Surendran, 2005), for two reasons: first, velar consonants were

---

[14]The choice of aligning English /t,d/ with their counterpart in other languages is arbitrary, because phonetically they should be aligned with the aspirate consonants in Cantonese, given that they are aspirated word-initially as a result of an allophonic rule (cf. /ten/ > [tʰen]). However, since the annotation of the children's production which are judged as correct uniformly report [tʰ] for English, and [t] is never reported as a mistake, I think that the annotators would not have judged a [t] production as a mistake in this case, considering [t,tʰ] as instances of a unique phoneme. For this reason, a comparison with Cantonese (in which instead the aspiration contrast is phonemic, and aspiration errors are reported) would be misleading: the production in English would be more accurate, but for the wrong reason, because lack in aspiration would not be marked as a mistake. A comparison with Greek and Japanese /t/ appears to be more appropriate. As a consequence, we still expect that when counting errors for Chinese /t/ we will have aspiration errors that are not present in other languages, but this should be balanced by the fact that other languages will display voicing errors instead.

scarcely represented, and therefore the only functional load measure would have been that associated with their voicing/aspirated counterpart, which is not informative about their contrastive status with respect to the other phonemes; second, we have several examples of velar and affricate consonants merging historically as a result of palatalization (Bateman, 2007), and therefore it is worth checking whether their contrastive status was correlated with palatalization errors in the children's speech.

A fifth predictor was added to control for articulatory complexity. Starting from Kent (1992)'s hierarchy, one would obtain three categories. The first one has all the stops with the exception of [t,t^h], while the third one has all the fricatives and the affricates. Since the choice of considering [t,t^h] as more complex than other stop segments is not backed up by the experimental literature, where it is usually reported that [t] appears at the same time of other coronal stops, I found it plausible to reduce the hierarchy to a dummy variable that just distinguishes fricative and affricate consonants from stops. The facts that fricatives and affricates require more articulatory effort than stop consonants and appear later in child language has been corroborated by both experimental work (Macken, 1979, Grunwell, 1982) and by Edwards and Beckman's results.

4.3.4   Results - Linear Regression

Since the four independent variables are largely correlated (see Table 4.6), a single model in which all these variables are used simultaneously would suffer from collinearity. For this reason, following the statistical analyses in the previous works, I fitted four separate linear models with one of the four lexical predictors plus the dummy variable controlling for articulatory difficulty, using the *lm* function in R (R Core Team, 2017), and checked how much of the variance was explained by each of the models. See Table 4.7 for the R^2 of the four models, for the four languages.

If we examine the results in detail, we note that type measures are usually better than token measures at explaining the variance. In particular, functional load over type is the best predictor in English, Greek and Cantonese, while for Japanese type frequency is better than

	Freq (type)	FL (token)	FL (type)
Freq (token)	0.61	0.81	0.64
Freq (type)	-	0.74	0.83
FL (token)	-	-	0.88

Table 4.6: Correlation matrix for the four independent variable studied.

R^2	English	Greek	Cantonese	Japanese
Frequency (token)	0.74*	0.60	0.52*	0.57*
Frequency (type)	0.82**	0.59	0.84**	0.69**
Functional Load (token)	0.75*	0.59	0.74**	0.59*
Functional Load (type)	0.88**	0.63	0.92**	0.66**

Table 4.7: Adjusted R^2 values from a linear regression analysis on the production accuracy of the children in Paidologos (2;00-4;00) versus phoneme frequency or functional load measures, plus articulatory complexity. The p-values associated with the F-statistic are: *$p<0.05$, **$p<0.01$.

functional load over types, which is still significant. The measure is statistically significant English, Cantonese and Japanese. In Greek, it is associated with p=0.06 despite the R^2 is similar to Japanese, probably as a result of a smaller sample size: there are 7 consonants tested, rather than 10.

These findings are particularly interesting because with few exceptions (Pye et al., 1987, Cychosz, 2017), the literature did not consider functional load calculated at the type level as a predictor of child behavior, while here it turns out to be a good predictor in all languages.

4.3.5 Results - Mixed Effects Logistic Regression

The choice of a linear model is not fully justified given the fact that accuracy values are constrained between '0' and '1'. The appropriate analysis would be a mixed-effect logistic model, in which we try to predict whether a single production is accurate ('1') or not ('0') for each single token, given the predictors, and we consider several factors like the children, the words, and their age as factors that might have introduced additional variation in the data. Since a linear model is more interpretable, and more directly comparable to the previous experiments, it was the natural choice for a first analysis. However, I decided to re-run the same analysis using a mixed effect logistic model instead.

Fixed effects	Estimate	Std. Error	z-value	AIC
Frequency (token)	0.074	0.021	3.486*	34725
Frequency (type)	0.170	0.029	5.861*	34704
Functional Load (token)	0.134	0.026	5.129*	34711
Functional Load (type)	0.183	0.030	6.158*	34700

Table 4.8: Logistic mixed-effect regression analysis on the production accuracy of the children in Paidologos (2;00-4;00). Articulatory Complexity, Language, Age, Word and Child are random intercepts. *p<0.001

I fitted four mixed effects models using the *glmer* function from the 'lme4' package (Bates, 2007). Each model contains one of the four predictors of interest as the independent variable, in the fixed effects structure. The random effects structure, which contains factors that might have influenced production accuracy, has five random intercepts: i) articulatory complexity, ii) language, iii) age, iv) word, v) child.

The first random effect is the one used in the preceding section, and it distinguishes stop consonants ('0') from fricatives and affricates ('1'). By including language in the random effects structure, we can run a unique model for all languages, rather the repeating the analysis for each language. In this way, we can estimate a unique coefficient for the predictor of interest, and still allow production accuracy to vary across languages because of language-specific factors not captured by the independent variables. Age is operationalized as a categorical variable which covers the first four age periods for which data has been collected in Paidologos: 2a (2;0-2;6), 2b (2;6-2;12), 3a (3;0-3;6) and 3b (3;6-3;12). Word is also included in the random effect structure, in order to make sure that the predictors are robust to variation due to lexical effects. Finally, we use the child identifier number to control for variation in production accuracy among children.

The function *glmer* does not return a R^2 for the model, because the variance is divided between the fixed and the random effects. However, we can compare the statistics associated with the predictors in the four models, in particular using z-values as an estimate of the effect size and the Akaike's Information Criterion (AIC, Akaike, 1973) to evaluate how well the model fits the data. For z-values, the higher the z-values, the stronger the effect of a variable in explaining the data, while for AIC the lower values are associated with the best

models.

The four estimates yield all positive coefficients, as expected, since higher frequency or functional load should be associated with higher production accuracy. In this case, the predictor of interest is significant in all four models, but type measures are those associated with the best models, and functional load (with z=6.158 and AIC=34700) is slightly better than frequency. The analysis confirms the results of the previous section.

4.3.6   Discussion

The fact that word type measures appear the best predictors of phonological development in all four languages, and in particular the fact that functional load calculated over word types was the best predictor in the overall, is consistent with the results of the previous chapter, in which minimal pairs turned out to a better predictor of historical mergers than functional load calculated over word tokens. This result has both methodological and theoretical implications, and raises further questions.

From the methodological viewpoint, we note that functional load calculated over word tokens is usually robust to the length of the wordlist used to derive the measure, as long as the most frequent words are included: as a consequence of Zipf's law, low-frequency words do not contribute significantly to the metric. On the contrary, if we calculate functional load on word types, the length of the wordlist matters, because rare words are weighted as much as frequent words. The strategy used in this work, for the last two chapters, was that of limiting the analysis to a list of the most 5000 frequent words in child-directed speech (when available). The number was chosen arbitrarily, and it approximates the receptive vocabulary of a preschooler (Jakobson, 1931, Ingram, 1989, Hoff, 2009, McLeod & Crowe, 2018). While 5000 words might appear too little, they actually represent an overestimation of the vocabulary of a child in their third of fourth year of life. If estimating a realistic receptive vocabulary is required to derive measures which can be successfully used to study phonological development, then we need to devise strategies to do it with the existing resources (cf. Kodner, 2019).

From a theoretical viewpoint, these results suggest something that has already been proposed in the literature on morphological development (Aronoff, 1976, MacWhinney, 1978, Bauer, 2001), in particular in Yang (2016) and Kodner (2020): types matter more than tokens when it comes to learning.

However, there are at least two potential questions which are left unanswered in this section. First, if functional load at the type level is highly correlated with frequency at the type and the token level, one might ask in which of the cases employed in the statistical analysis the measures make different predictions, if any. Second, it is still unclear whether the predictors studied here only influence production, or whether they can also affect perception, and then lexical representations. These two questions will be addressed in the following section.

4.4 Perception and production in phonological development

In the last section, we have seen that language-specific factors like functional load do play a role in predicting phonological development across languages. If this is true, then does it mean that children become aware of different contrasts in some languages earlier than in others depending on the distributional properties of the contrasts, or that they are just not accurate in reproducing contrasts for which they do not have enough evidence, but of which they are fully aware? This question is a challenging one, because while production accuracy can be measured, it is more difficult to quantify perception capabilities, or to establish if children's lexical representations are target-like.[15] However, taking a look at children's production errors in detail can allow one to make some hypotheses.

In this section, the errors made by the children of the Paidologos corpus will be analyzed and compared across languages, with the aim of distinguishing those that can be explained with reference to production, and those that must be explained in terms of perception, or

---

[15]This distinction is necessary because while works like Macken (1980a) cite misperception as an explanation for the lack of adult-like lexical representations, recent works like Vihman et al. (2004) emphasize instead the role of memory, and propose that children only memorize contrasts in salient positions in early stages. This means that they might be able to perceive all the contrasts, but they only store in memory the salient ones. This will be addressed in more detail later.

non-target-like representations. The mistakes reported in this section are all those marked with the substitution symbol '$' in Paidologos. Only those mistakes that do not clearly map to a phoneme, for instance deletion or substitution with a consonant cluster, are excluded.[16] Errors which did not appear at least five times across the corpora are also excluded.

Before proceeding, it is worth noting that relying on transcriptions rather than acoustic measures to make generalizations about perception or production of segments comes at the cost of losing some information, for instance the presence of covert contrasts in child production (Li et al., 2009), or potentially obtaining results biased by transcriber effects. These methodological issues are discussed at length in Edwards & Beckman (2008b). For this reason, I will also make reference to the acoustic studies that I found in the literature when trying to explain developmental patterns.[17]

4.4.1 Plosives

In this first section, we will compare alveolar and velar stops.

4.4.1.1 /t,k/

Production accuracy for /t/ varies cross-linguistically. In particular, accuracy seems to be higher in English and Greek than in Cantonese and Japanese (Figure 4.2). The difference between English and Greek is not significant, according to a Chi-square test (χ^2=0.08, p=0.77), and the same is true for Cantonese and Japanese (χ^2=3.1, p=0.08). However, English and Greek are different from Cantonese and Japanese (p<0.01 for the four pairwise comparisons). The mistakes made by the children when producing words with a word-initial /t/ are in Table 4.9.

The most common mistake across languages is /k/. In particular, Cantonese shows

---

[16]Some of the substitution are marked as leaning toward the target: for instance when the symbol '$d' appears, it means that the child produced a 'd' rather than a 't', while when the symbol '$d:t' appears, the annotator indicates a sound which is a 'd' slightly leaning toward a 't'. I still consider this as a substitution, because the annotator had the option of using the symbol 't:$d' when the outcome was closer to the intended target than to an alternative phoneme. In the first case, the production is marked as an error ('0'), but in the second case is marked as correct ('1').

[17]For the whole section, we refer to the sounds using slashes ('//') because the transcription system is only sensitive to phonemic distinctions. Any phonetic distinction is ignored.
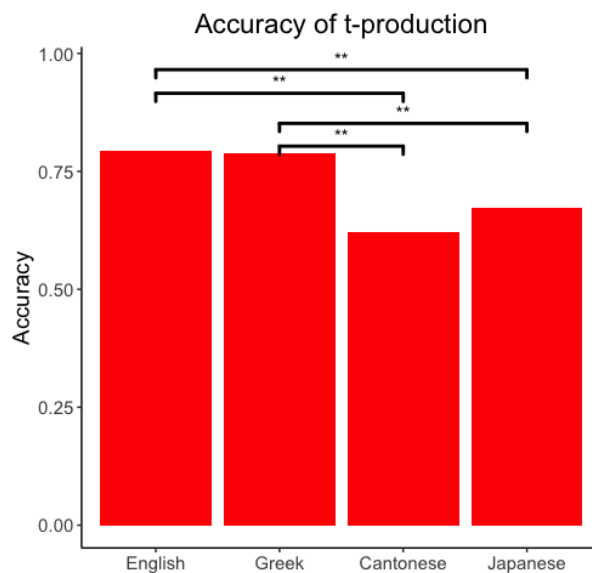
Figure 4.2: Accuracy of /t/ in production. Statistical significance is estimated using a Chi-square test. *p<0.05, **p<0.01.

/t/	English (579)	Greek (596)	Cantonese (679)	Japanese (505)
k	11	3	76	12
tʃ	30			38
kʲ		15		39
tʰ			48	
ts		7	27	3
d	8	7		10
p	4	11		7
s	1	8	5	
h			8	3
kʰ			10	
n		1	2	3
θ	2	4		
g	1			4

Table 4.9: Production errors made by 2- and 3-year-old children in the Paidologos corpus for /t/.

many mistakes toward /k/, while the number is smaller for English, Greek and Japanese. The main case in which we expect a place of articulation error for a /t/ production is via assimilation, for instance producing the word 'duck' as [gʌk] in English (Ingram, 1999). Otherwise, place of articulation errors, specially those involving /t/, are uncommon (cf. Paschall, 1983, Stokes & Surendran, 2005).

The following mistake, /tʃ/, could be explained as a production error, because it involves the same place of articulation, and it shows similar numbers for the two languages in which the affricate consonant is part of the inventory (English and Japanese).[18,19] This reasoning extends to /ts/. Interestingly, the number of /ts/-substitutions is higher in Cantonese than it is in Greek and Japanese, but we know that the phoneme is more uncommon in these two languages.[20] Similarly, /tʰ/ can be considered a production error, because aspiration errors are common in children (Eilers & Oller, 1976). The same reasoning extends to /tsʰ/ and to the voicing error /d/. Other mistakes appear more sporadic.

Interestingly, the low production accuracy of Japanese seems to depend on the presence of a palatalized [k], /kʲ/, as a common error. This error also appears in Greek, but not in English. Beckman et al. (2003) note that in Japanese velar consonants are more frequent than coronal consonants, and therefore we might expect more errors in the direction of velar consonants. In this specific case, however, the error can be explained with reference to the fact that in Japanese all consonants palatalize in front of /i,j/ (Itô & Mester, 1995), so that /ti/ surfaces as [tɕi]. In fact, almost all errors occur in contexts in which /t/ is followed by /i,j/, a pattern that is only found in loanwords (*tiikappu* 'teacup' and *tiishatsu* 't-shirt'), and that Japanese learners force into the phonotactically correct [kʲɪ], when they do not produce it as [tɕɪ]. If we exclude this case, children learning Japanese do not exhibit any pattern which appears unusual compared to those we see for children learning English and

---

[18]While the sound is also present in Cantonese, it is usually considered non-contrastive, and it is absent in some dialects. I think it is for this reason that it was not included in the Paidologos corpus.

[19]In Japanese, the affricate does not have full phonemic status, but it is an allophone of /t/. It appears as a production error for /t/ relatively more often than in English (308/505=7.5% versus 30/579=5.2%), which can be explained with reference to its allophonic status.

[20]Its type frequency in word-initial CV sequences is 0.002 in Greek and 0.034 in Japanese, versus 0.083 in Cantonese. Also, in Japanese it is limited to the sequence /tu/>[tsu], and it is in complementary distribution with /t/ in native words.

Figure 4.3: Accuracy of /k/ in production. Statistical significance is estimated using a Chi-square test. *p<0.05, **p<0.01.

Greek.

The production of /k/ also varies cross-linguistically. In this case English and Japanese show similar production accuracy (χ^2=1.01, p=0.32), and so do Greek and Cantonese (χ^2=2.85, p=0.09), while the four pairwise comparisons between the two groups show significant differences (p<0.01) See Figure 4.3.

This set of comparisons is more difficult to interpret, but the direction of the mistakes, in Table 4.10, is informative.

The most common mistake in the production of /k/ across languages is /t/. This is not unusual: Jakobson (1941/1968) and Stemberger & Stoel-Gammon (1991) argue that coronals segments tend to appear early in acquisition across languages, and therefore we might expect some velar consonants to appear as dental in production. Moreover, several works reported this mistake when the segment is followed by front vowels (cf. MacNeilage & Davis, 1993 Nicolaidis et al., Beckman et al., 2003, Levelt, 1996). Interestingly, though, the frequency of this mistake appears to vary across languages.

The number of /t/-mistakes in Greek is quite high, especially given that they account

/k/	English (590)	Greek (374)	Cantonese (568)	Japanese (473)
t	20	30	56	16
kʲ	5			26
g	16	2		10
kʰ			27	
h	4		16	2
ts			20	
p	3	7		4
tʃ	3			7
tsʰ			8	
ç				7
x		8		
w			7	
tʰ			6	
kʷ			5	

Table 4.10: Production errors made by 2- and 3-year-old children in the Paidologos corpus for /k/.

for about 8% of the productions (30/374) versus 3% in English (20/590) and Japanese (16/473). However, Greek seems to pattern closer to Japanese and to English when it comes to substitutions, if we look at how few phonemes are involved in the mistakes. A manual investigation of the errors shows that most of the errors involve noisy data, and in particular the presence of /x/ in the phoneme inventory led to the exclusion of many productions (18 tokens) for which it was difficult for the annotator to determine the quality of the sound, because it appeared as a sound included between /k/ and /x/. These facts suggest that the presence of three velar obstruents, /k,g,x/, might result in a lower production accuracy for /k/ in Greek, but otherwise the developmental trajectory looks similar to the other languages.

The situation is different in Cantonese. First, there are many more /t/-errors than those we see for the other languages, and they account for about 10% of the productions (56/568). Second, many misproductions involve place of articulation errors, in particular /ts/ and /tsʰ/, while the other languages mostly show mistakes associated with manner, voicing and palatalization.

Table 4.11 summarizes the frequencies of /t,k/ across languages and reports their func-

	English	Greek	Cantonese	Japanese
Freq(t) tokens	0.051	0.198	0.073	0.046
Freq(k) tokens	0.044	0.041	0.112	0.156
Freq(t) types	0.063	0.057	0.067	0.082
Freq(k) types	0.088	0.133	0.089	0.180
FL(t,k) tokens	0.0113	0.015	0.0264	0.0262
FL(t,k) types	0.0171	0.0164	0.0224	0.0333

Table 4.11: Comparison between /k/ and /t/ in the four languages.

tional load measures. The numbers do not seem to point toward a distributional difference among Cantonese and Greek and the other languages. In particular, both frequency and functional load are higher in Cantonese than in English, and therefore they cannot explain this case.

4.4.1.2   /d,g/

A developmental difference is also suggested by contrasting the errors for /d/ and /g/ in English and Japanese, which exhibit a significant difference (χ^2=74.3 and χ^2=33.2, respectively, and p<0.01 in both cases), see Figure 4.4.[21] Their errors are in Table 4.12.

The most common error for /d/ is a voicing error (/t/), and here Japanese shows more errors than English: 86 versus 24, corresponding roughly to 22% and 4% of the productions. Other errors in English are the more complex phonemes /ð,dʒ/, which share the same place of articulation of the target phoneme, and /g/, which is the only place of articulation error that appears. In Japanese, substitutions involve other phonemes, but /t/ is the most frequent by far. These facts might suggest that the voicing distinction is acquired later in Japanese (cf. also Kong et al., 2012).

As for /g/, the pattern is similar to that we have seen for /d/: most of the errors in English and Japanese are voicing errors, and Japanese displays many more voicing errors than English (148 versus 22, corresponding roughly to 31% and 9%). These patterns confirm that the voicing distinctions is mastered later.

In Table 4.13, I summarized the frequencies of /d/ and /g/ in English and Japanese, and
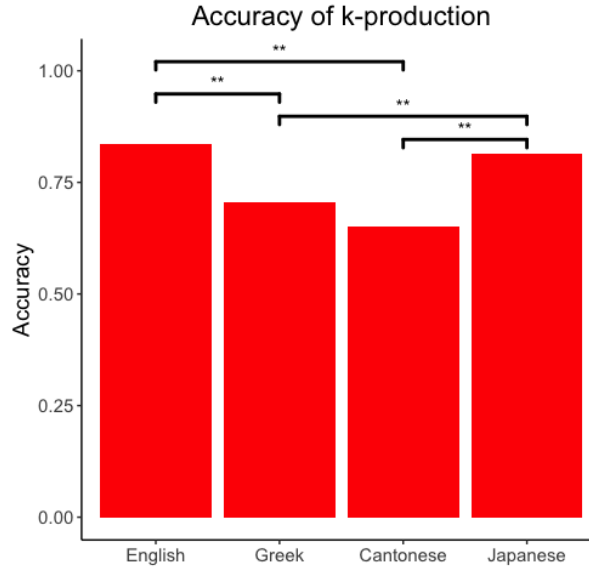
---

[21]Data for Greek /d/ were not available in Paidologos.
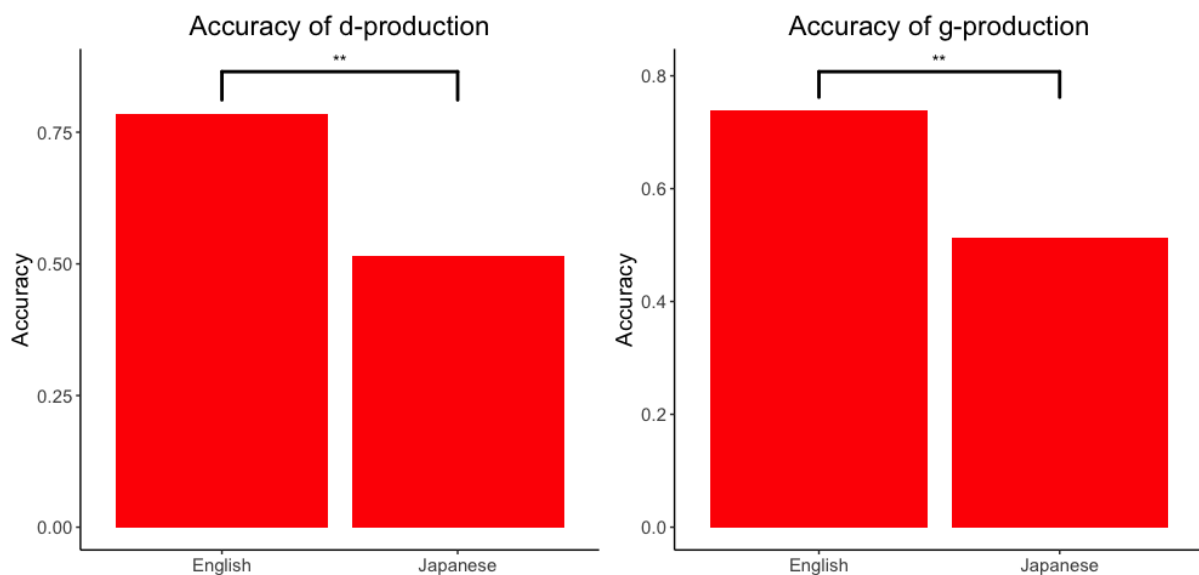
Figure 4.4: Accuracy of /d/ and /g/ in production. Statistical significance is estimated using a Chi-square test, *p<0.05, **p<0.01.

/d/	English (582)	Japanese (377)
t	24	86
g	17	7
n	3	14
dʒ	17	
k	2	7
h	2	4
ð	6	
p	1	4
r		5

/g/	English (238)	Japanese (475)
k	22	148
d	17	9
kʲ		9
h		7
dʒ	3	4
t	1	5

Table 4.12: Production errors made by 2- and 3-year-old children in the Paidologos corpus for /d/ and /g/.

	English	Japanese
Freq(d) tokens	0.060	0.102
Freq(g) tokens	0.047	0.039
Freq(d) types	0.053	0.048
Freq(g) types	0.024	0.034
FL(b,p) tokens	0.0093	0.0032
FL(t,d) tokens	0.0098	0.0191
FL(k,g) tokens	0.0058	0.0233
FL(b,p) types	0.0222	0.0092
FL(t,d) types	0.0136	0.0185
FL(k,g) types	0.0099	0.0214

Table 4.13: Voicing contrasts in English and Japanese.

the functional load associated with voicing contrasts among stops, but no clear asymmetry is detectable. In Japanese, /d/ is more frequent if we look at token frequency, and less frequent if we look at type frequency, and the opposite is true for /g/. If we look at the voicing contrast from a functional load viewpoint, labial stops are more contrastive in English, while dental and velar stops are more contrastive in Japanese. In this case as well, structural considerations do not seem to be correlated with the pattern detected.

4.4.1.3 Discussion

To summarize, in the case of stop consonants, we see at least two patterns which are different cross-linguistically.

First, Cantonese production accuracy is the only one that clearly stands out as the lowest. This pattern does not have a clear explanation: previous research on phonological development in Cantonese did not found any delay in acquisition when comparing Cantonese and English children phonological development. Conversely, it has been argued that phonological development is faster in Cantonese, because most of the target consonants appear in the children's speech earlier than in English (Tse, 1982, So & Dodd, 1995). The only way to account for these findings is that even though both /k/ and /t/ appear early in Cantonese, substitution errors are more frequent than in the other languages. So & Dodd (1995) show indeed that most of the assimilation errors in Cantonese involve backing of /t/ in the presence of a velar consonant in the coda, and they occur in the age range

2;0-3;0. They also report fronting of /k/ to /t/ until 3;6. However, given the numbers presented here, it appears that these substitutions are not attested with the same frequency for children learning English or other languages in the Paidologos corpus. In particular, /k/-substitutions for /t/ are also present in words which do not have a velar coda: only 35% of the mistakes /t/>/k/ are due to assimilation in the presence of a velar consonant in the word (27/76), a number which is comparable with the other languages (in English it is 4/11, 36%). Therefore, we cannot argue that assimilation is more frequent in Cantonese than it is in other languages.

This asymmetry cannot be explained with reference to structural properties, because the contrast between /k/ and /t/ should be clear from the distribution of the two phonemes in the lexicon, as we have seen in Table 4.11. As So & Dodd (1995) argue, functional load considerations should make contrasts less clear in English than in Cantonese, because English has more vowels in its inventory, and therefore consonant contrasts are less informative to distinguish among syllables. If the number of substitutions is higher in Cantonese than it is in English, the motivation must be independent from structural considerations.

A final point worth mentioning is that Cantonese has a sound change in progress which is removing the distinction between syllable-final /-k/ and /-t/ (Zee, 1999, To et al., 2015), which might have influenced the production of the children in the Paidologos corpus. In any case, it is clear that functional load cannot explain this asymmetry.

The second pattern we need to address is the delay in the acquisition of the voicing contrast in Japanese. Kong et al. (2012) address this pattern with reference to the differences in the acoustic properties that characterize Japanese voiced stops compared to Greek and English ones, and they argue that the differences in the children's production have a phonetic explanation. I also note that /b,g/ are weakened to fricatives in Japanese, when they appear intervocalically (Maekawa, 2018). Since this does not happen in English, children learning English have an additional cue to voicing contrasts.[22]

To conclude, functional load does not seem to explain either of the patterns identified

---

[22]The same phenomenon might explain the delay in the acquisition of the voicing contrast in Spanish, documented in Macken (1979). In Spanish, like in Japanese, voiced consonants are lenited intervocalically.

in these first two cases.

4.4.2 Fricatives

This subsection examines the acquisition of fricatives across languages.

The acquisition of sibilants has been the object of previous research (Cheon & Anderson, 2008, Li, 2008). In particular, Li et al. (2009) observe that children learning English acquire /s/ earlier than children learning Japanese, and that while it is common for them to use the alveolar sibilant instead of the postalveolar sibilant /ʃ/ in production (cf. Weismer & Elbert, 1982, Baum & McNutt, 1990), the reverse pattern is true for Japanese, in which the alveolar /s/ is substituted with the palato-alveolar [ɕ] in production, more often than the opposite (cf. Nishimura, 1980).

Something worth noting, before proceeding with the analysis, is that the sound described as [ɕ] in Japanese does not have phonemic status in Japanese native words, but it is only contrastive in loanwords. In Japanese native words, the sound results from the sequences /sj, si/ > [ɕj, ɕi] (Vance, 1987). Additionally, this sound in Japanese is often compared with English /ʃ/, even though it is represented with a different symbol. The reason why the sound is described as an alveolo-palatal sibilant rather than a postalveolar sibilant is because it is realized by speakers with their tongue closer to the palate. Despite these two differences (phonemic status and articulation), the literature has been treating [ʃ] and [ɕ] as basically the same sound, since it involves the same articulator and it appears in both languages in early child language production, in addition of being contrastive in both languages if we include Japanese loanwords in the analysis. In this section, I will adopt the same convention, and thus I will be treating /ʃ/ as if it were a phoneme in both English and Japanese.

Another phoneme which has been extensively studied in the experimental literature is /θ/. Its contrast with /f/ is among the few contrasts that are not easily learned by children (Eilers et al., 1977, Jusczyk et al., 1979, Levitt et al., 1988). Edwards & Beckman (2008a) compare the production accuracy of /θ/ in English and Greek, and show that there is a

Figure 4.5: Accuracy of /s/ in production. Statistical significance is estimated using a Chi-square test. *p<0.05, **p<0.01.

significant difference that cannot be predicted by the property of the words, and that points toward an early acquisition in Greek.

These phonemes and their error patterns will be examined in detail in the next subsections.

4.4.2.1   /s/

For this segment, cross-linguistic variation in production accuracy is more limited in Paidologos (Figure 4.5). The Chi-square tests show that the only noticeable difference is that among Japanese and the other three languages (p<0.01), with production accuracy in Japanese being significantly lower. The fact that /s/ is acquired late in Japanese has been found in previous studies (Yasuda, 1970, Li et al., 2009). Another pattern that emerges is the fact that children learning Cantonese seem to perform slightly better than children learning Greek (χ^2=6.07, p=0.01). All the other comparisons are not significant.

The mistakes produced by the children are in Table 4.14. The most popular mistake is /θ/, in English and Greek. Then, we have /t/, as in the previous case, but this time

136

/s/	English (569)	Greek (617)	Cantonese (431)	Japanese (461)
θ	43	97		
t	13	50	32	41
ʃ	46			59
ts		15	20	25
ç		25		28
tʃ	10			33
k	1		21	11
h	1		16	16
tʰ			26	
tsʰ			26	
xʲ		21		4
d	12	2		8
f	8	2	4	3
kʲ	2	1		4
kʰ			6	

Table 4.14: Production errors made by 2- and 3-year-old children in the Paidologos corpus for /s/.

clear tendencies do not emerge: children learning English are those that exhibit the least amount of /t/-errors, while Greek, Cantonese and Japanese show similar ratios. In this case, it looks like the mistakes are mostly distributed across the other fricative and affricate consonants available in the languages: /ʃ, tʃ/ for English and Japanese, /ts/ for Greek, Cantonese and Japanese, /ç/ for Greek and Japanese, and so on. Note that the number of mistakes toward /ʃ/ is higher in Japanese than it is in English, as noted by Li et al. (2009): we have 46/569=8% in English and 59/461=13% in Japanese. Interestingly, many of the phonemes that are typically described as allophones in Japanese, /ʃ, ts, ç, tʃ/, appear as production errors for a target /s/, a fact which suggests that children learning Japanese are not treating them differently from other phonemes, even though they are distributionally restricted.

A possible explanation for the differences in production accuracy of this segment is that Japanese has a wide inventory of alveolar and alveolo-palatal fricatives and affricates (8, if we include all the possible allophones). An inspection of the frequency of /s/ across languages (in Table 4.15) shows that Japanese displays a smaller value than the other languages in type frequency for /s/, which suggests that /s/ has indeed a bigger competition in the

	English	Greek	Cantonese	Japanese
Freq(s) tokens	0.047	0.053	0.060	0.065
Freq(s) types	0.080	0.102	0.108	0.077
FL(s,t) tokens	0.0106	0.0186	0.0196	0.0146
FL(s,t) types	0.0179	0.0197	0.0238	0.0209

Table 4.15: Comparison between /s/ and /t/ in the four languages.

vocabulary than it has in other languages. In terms of functional load, however, the contrast with the stop consonant /t/, which is the most frequent error in Japanese that does not involve a fricative consonant, is comparable to that of the other languages, and the numbers are higher than in English.

There are multiple possibilities to explain this case. Li et al. (2009) noted that the presence of many fricatives in the Japanese inventory might offer more possibilities for the annotator to judge a production of /s/ as a misproduction, while in English the choice is more limited. For instance, a production like [ts] can be judged as correct in English, while in Japanese the same production would lead the annotator to mark it as a mistake. However, their acoustic analysis on adult speech reveals a difference between English and Japanese regarding the /s-ʃ/ contrast: it appears that the contrast between the two sounds in English can be identified using different acoustic dimensions, while in Japanese the contrast is more ambiguous, a fact that would explain the delay in its acquisition. In any case, it seems that functional load is not relevant in this case.

4.4.2.2   /ʃ/

For this segment, the situation is similar to what we have seen for /s/. A Chi-square test reveals that the difference between the two languages having the phoneme, English and Japanese, is significant (χ^2=66.93, p<0.01). In English, the accuracy is higher (0.51 versus 0.39) See Figure 4.6.[23]

An analysis of the mistakes (in Table 4.16) shows what we would expect given the analysis of /s/, namely the errors are mostly distributed around the other fricative consonants

---

[23]Cantonese has a similar consonant, but is produced as a retroflex. The consonant is missing in certain dialects (Duanmu, 2007), and is not tested in Paidologos.
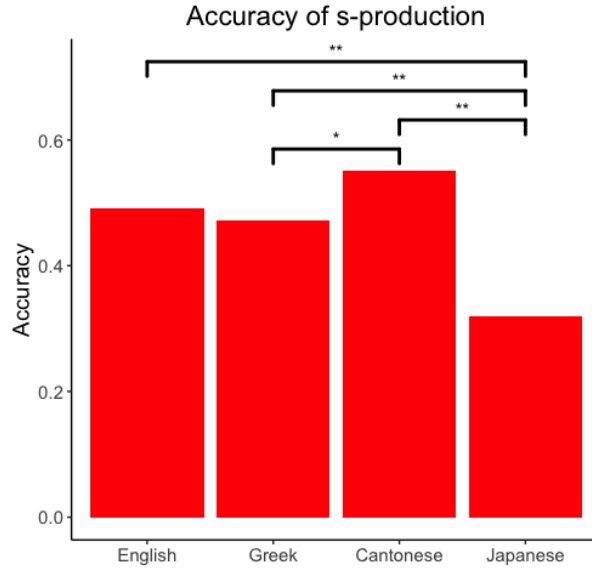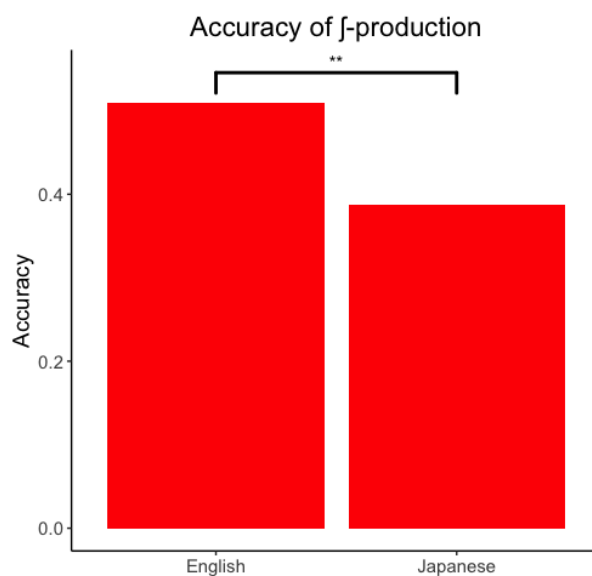
Figure 4.6: Accuracy of /ʃ/ in production. Statistical significance is estimated using a Chi-square test. *p<0.05, **p<0.01.

/ʃ/	English (565)	Japanese (461)
s	97	39
tʃ	37	54
ç		43
kʲ		19
t	4	12
dʒ	8	4
θ	12	
ts		9

Table 4.16: Production errors made by 2- and 3- children in the Paidologos corpus for /ʃ/.

	English	Japanese
Freq(ʃ) tokens	0.017	0.033
Freq(ʃ) types	0.026	0.044
FL(t,ʃ) tokens	0.0037	0.0035
FL(t,ʃ) types	0.0094	0.0077
FL(s,ʃ) tokens	0.0056	0.0020
FL(s,ʃ) types	0.0097	0.0061

Table 4.17: Frequency of /ʃ/ and its functional load in English and Japanese.

of the languages. An interesting difference, though, is that Japanese also shows several errors involving the stops /k,t/, while English does not, since the errors are mostly toward /s/. Like in Li et al. (2009), the probability of producing /ʃ/ as /s/ is higher in English (97/569=17%) than it is in Japanese (39/461=8%).

Table 4.17 shows the frequency and the contrastiveness of the segment in the two languages. Interestingly, /ʃ/ is more frequent in Japanese than it is in English. However, the contrasts with /t/ and /s/ are weaker, as a result of the fact that the phoneme is restricted to syllables with a high front vowel or with a /j/ in native words, and therefore it is only contrastive in loanwords.

As previously mentioned, Li et al. (2009) explain this difference with reference to acoustic cues, and show that the cues for the sibilant are more ambiguous in Japanese than they are in English. Another possibility is that the contrast in Japanese is acquired later than in English because its functional load with respect to other phonemes in the lexicon is lower than it is in English, and therefore children learning Japanese need more evidence to distinguish this sound from the other sounds of their phonemic inventory. This is a case in which functional load can explain the difference in children's production accuracy for the two segments in the two languages.

4.4.2.3 /θ/

We already know from the experimental studies on English that /θ/ is one of the few phonemes for which there is evidence of perceptual difficulties. However, there is less evidence on whether children learning Greek also fail to discriminate the two sounds at an
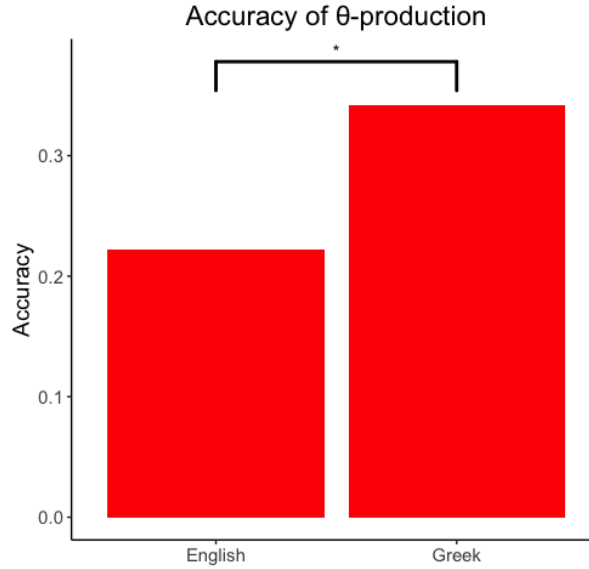
Figure 4.7: Accuracy of /θ/ in production. Statistical significance is estimated using a Chi-square test. *p<0.05, **p<0.01.

/θ/	English (117)	Greek (118)
s	14	27
f	29	
t	6	16
ç		7
d	7	

Table 4.18: Production errors made by 2- and 3-year-old children in the Paidologos corpus for /θ/.

early age. A Chi-square test shows that children learning Greek are slightly better than children learning English at producing the sound (χ^2=3.95, p=0.047, see Figure 4.7), as noted by Edwards & Beckman (2008a).

An error analysis reveals an interesting asymmetry (Table 4.18). While in both languages a fair amount of /s/ is produced, English learners also display /f/, while Greek learners prefer /t/. This pattern is interesting because both phonemes are present in the two languages.

As we see in Table 4.19, the interdental fricative in Greek is both more frequent and more contrastive with /f/ than it is in English. The fact that /f/ and /θ/ are similar

	English	Greek
Freq(θ) tokens	0.012	0.068
Freq(f) tokens	0.019	0.010
Freq(θ) types	0.007	0.021
Freq(f) types	0.056	0.048
FL(f,θ) tokens	0.0024	0.0035
FL(f,θ) types	0.0032	0.0077

Table 4.19: Frequencies of /f/ and /θ/ and their functional load in English and Greek.

acoustically has been studied since Miller & Nicely (1955) and Eilers & Oller (1976), but interestingly we see that the confusion does not arise for children learning Greek. Most of the errors that they make are limited to other alveolar consonants, while in English most of the errors are place-of-articulation errors. This suggests that while for English we have a perception error (cf. Blevins, 2004, 134-135, McGuire & Babel, 2012), for Greek we just have a production error, and the phoneme is acquired at an earlier stage.

Interestingly, acoustic studies do not show that /f/ and /θ/ can be more easily discriminated in Greek than in English. An analysis based on cepstral coefficients run by Athanasopoulou & Vogel (2011) shows that the contrast between labiodental and interdental fricatives is very ambiguous, exactly like it has been shown in similar acoustic analyses on English (Jongman et al., 2000).

In order to further explore cross-linguistic differences in the acquisition of /θ/, I decided to analyze naturalistic data from Castillian Spanish using the PAIDUS corpus (Lleó, 2002). Phonetic transcriptions are available for three of the children that are studied in the corpus (Jose, Maria and Miguel), and they cover the children's production in the first few years of life, up to 3;00, therefore matching the age of the children in the Paidologos Corpus. Production accuracy cannot be compared between the two corpora, because in the case of the PAIDUS corpus the data are conversational, and not from elicitations. However, the error patterns are comparable.

Table 4.20 shows that among the errors that children learning Spanish make there is /p/, which is unexpected, given what we see in English and Greek. The error is limited to the word *zapatos* 'shoes', which is the most common word that contains the interdental

/θ/	Spanish (99)
s	12
p	12
t	11
ʃ	6
b	6
f	3

Table 4.20: Production errors made by the children in the PAIDUS corpus for /θ/.

	Spanish
Freq(θ) tokens	0.005
Freq(f) tokens	0.009
Freq(θ) types	0.027
Freq(f) types	0.036
FL(f,θ) tokens	0.0017
FL(f,θ) types	0.0081

Table 4.21: Frequencies of /f/ and /θ/ and their functional load in Spanish.

fricative in child speech, and therefore can be explained as an assimilation error (cf. Macken & Barton, 1980b). Apart from /p/, the errors follow the Greek pattern in showing mostly /s/ and /t/, but not /f/.

A look at phoneme frequencies and functional load (Table 4.21), using the Spanish data from the previous chapter, shows that while /θ/ has lower token frequency and functional load than in English and Greek, its type frequency and functional load are higher than in English, and closer to those in Greek.

In other words, these results from the PAIDUS data show that in English the frequency and the functional load associated with /θ/ are lower than those we find in Spanish, if one looks at the type level, in addition of being lower than those we find in Greek. This could explain why in English we see errors that do not appear in other languages.[24,25]

---

[24]In this case, we cannot rule out that acoustic differences also play a role, and that to some extent /θ/ is closer in acoustic space to /f/ in English than it is in Greek and in Spanish, even though so far this has not been shown, to the best of my knowledge.

[25]It is worth noting that an alternative possibility is that /f/ is not frequent enough in Spanish and in Greek, and this is why children do not use it in production. This seems to be correct for Spanish, while in Greek its type frequency is comparable to the English one, but still lower, which leaves this possibility open.

4.4.3 Affricates

The last two segments examined are the affricate consonants /ts,tʃ/, which have been covered in detail in Edwards & Beckman (2008a), and are among the sounds whose acquisition is expected to be the latest across languages because of their articulatory complexity (Dinnsen, 1992, Kent, 1992), even though Pye et al. (1987), So & Dodd (1995), Hua & Dodd (2000) and Hua (2002) showed that this might not be true for all languages.

Edwards and Beckman showed that the production accuracy of these sounds is influenced by both articulatory and language-specific factors. In particular, while both phonemes are typically acquired later than /t/, for /ts/ the delay is narrower in Cantonese than it is in Greek, while for /tʃ/ it is narrower in Japanese than it is in English.

Like in the case of /ʃ/, the two phonemes do not have clear phonemic status in Japanese, because they are described as allophones of /t/ in native Japanese words (Vance, 1987). And like in that case, while [ts] is comparable acoustically to their counterpart in Greek and Cantonese, the second affricate is typically described as an alveolo-palatal, [tɕ]. However, the sounds are typically matched to /ts,tʃ/ in cross-linguistic comparisons, and I will do the same in the following subsections.

4.4.3.1 /ts/

The second of the sounds studied in detail in Edwards & Beckman (2008a) is the affricate /ts/. The comparison in that paper is limited to Greek and Cantonese, because in Japanese the phoneme is distributionally restricted to an environment (before /u/) in which /t/ is not allowed in native words, and therefore the two are typically described as allophones. In this work I decided to include the language, since we are just interested in the production accuracy and the error patterns of the sound.

The production accuracies in the three languages do not differ significantly: the only pair which exhibits some marginal difference is Greek and Cantonese (χ^2=4.12, p=0.042), while the pairs involving Japanese are both not significant (see Figure 4.8). Japanese and Greek exhibit similar values, but the fact that the sample size in Japanese is small (cf.

Figure 4.8: Accuracy of /ts/ in production. Statistical significance is estimated using a Chi-square test. *p<0.05, **p<0.01.

Table 4.4) makes its comparison with Cantonese not significant.[26]

The error analysis (in Table 4.22) shows that even though Japanese learners have a production accuracy comparable to that of Greek and Cantonese learners, the error pattern is different. Almost all the errors in Japanese have the other affricate consonant, /tʃ/, as the outcome. On the other hand, both Greek and Cantonese have their errors involving mostly /t/, but also the other alveolar consonants. Interestingly, in both languages /k/ is also present as a mistake, which suggests a poor representation of the sound, because it is a place of articulation error. This difference between Japanese and the other languages can be explained with the fact that Greek and Cantonese do not have affricate consonants involving a different place of articulation, and therefore this is the only affricate that has to be learned. On the other hand, Japanese learners have independent evidence for the presence of affricate consonants in their language.

A comparison of the properties of the sound in the three languages (in Table 4.23) shows

---

[26]I tried to replicate the analysis by including the production accuracy of older children, but the results did not change: most of the variation is confined to the age range 2;00-4;00, and after this period children behave quite uniformly.

/ts/	Greek (355)	Cantonese (458)	Japanese (116)
t	85	117	1
s	38	30	2
k^j	44		7
k		31	
θ	26		
tʃ			41
t^h		16	
ts^h		18	
k^h		5	

Table 4.22: Production errors made by 2- and 3-year-old children in the Paidologos corpus for /ts/.

	Greek	Cantonese	Japanese
Freq(ts) tokens	0.001	0.083	0.011
Freq(ts) types	0.010	0.084	0.034
FL(t,ts) tokens	0.001	0.0213	0.0001
FL(t,ts) types	0.0057	0.0209	0.0001
FL(s,ts) tokens	0.0009	0.0202	0.0048
FL(s,ts) types	0.0066	0.0279	0.0098
FL(k,ts) tokens	0.0008	0.0257	0.0042
FL(k,ts) types	0.0031	0.0226	0.0102

Table 4.23: Frequency of /ts/ and its functional load in Greek, Cantonese and Japanese.

that in Cantonese it is associated with higher frequency and higher functional load than in Greek and Japanese, when we look at the contrasts with /t,s,k/. However, the contrast with /t/ in Japanese is basically zero, because the two are in allophonic distribution. This makes it difficult to argue for an earlier acquisition of the sound in Japanese, where /t/-errors and /k/-errors are not present, on the basis of lexical contrast, but the analysis is compatible with an earlier acquisition in Cantonese, which exhibits the higher production accuracy.

This case is compatible with at least three alternative hypotheses. First, one could argue that the asymmetry between /tʃ/-errors and /t/-errors can be explained in terms of production: children simply use the closest phoneme available in the inventory to produce the sound in the period in which they cannot produce it for articulatory reasons. Second, one could argue based on the production errors involving velar consonants in Greek and Cantonese that the phoneme must be acquired later in these languages, because children
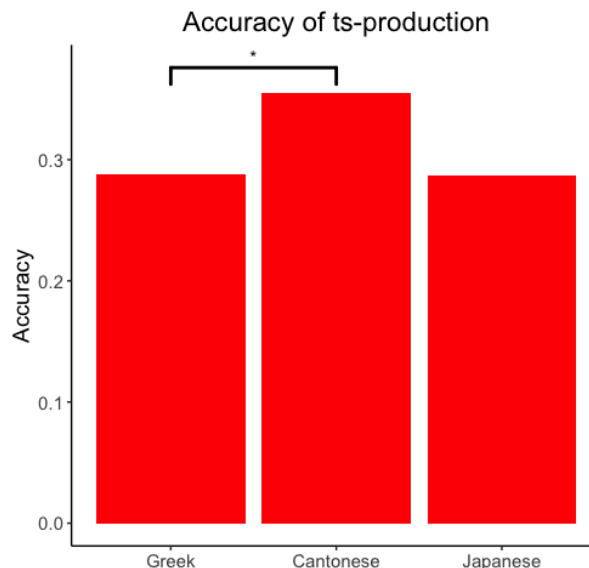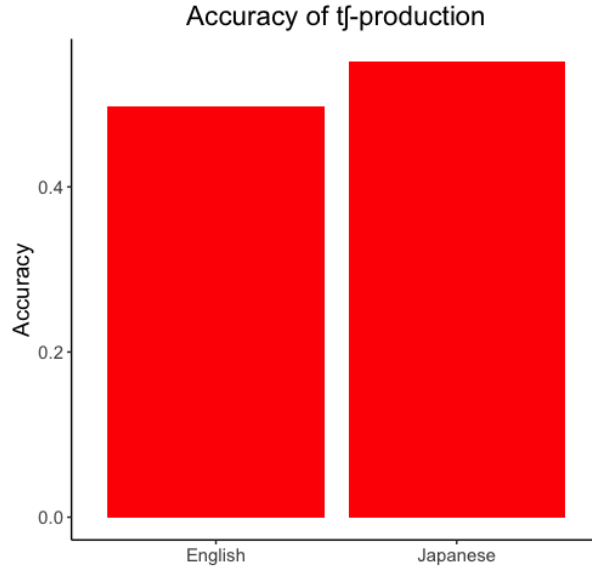
Figure 4.9: Accuracy of /tʃ/ in production. Statistical significance is estimated using a Chi-square test. *p<0.05, **p<0.01.

learning Japanese rarely produce /k/ instead of /ts/, despite the velar stop is the most frequent phoneme in their inventory. Third, one could argue that it is difficult to clearly assess whether the error patterns point toward a perception or a production error, but given the raw production accuracy data and the distributional facts, the phoneme must be acquired early in Cantonese, and later in Greek and Japanese. This last hypothesis would also be plausible given the fact that the phoneme is extremely rare in Greek and in Japanese, and that children learning Cantonese use /ts/ in words with a target /t/, /k/ or /s/ (cf. 4.4.1.1 and 4.4.2.1), a pattern which is not shown by children learning Japanese and Greek.

4.4.3.2 /tʃ/

The third of the sounds covered in detail in Edwards & Beckman (2008a) is the affricate /tʃ/. In Japanese, the distribution of this phoneme is restricted, because in native words it is the allophone of /t/ in front of /i,j/. The production accuracy between English and Japanese does not differ significantly (χ^2=2.79, p=0.09, see Figure 4.9).

/tʃ/	English (536)	Japanese (459)
t	120	26
kʲ	2	53
ʃ	20	12
s	13	3
ts		15
k	10	5
d	6	5
dʒ	6	4
ç		9

Table 4.24: Production errors made by 2- and 3-year- children in the Paidologos corpus for /tʃ/.

The error analysis (in Table 4.24) reveals an asymmetry in the direction of the mistakes. In English, the mistakes overwhelmingly show /t/, and to some extent /ʃ/, which are the two expected production errors, given the articulators involved in the production of the affricate. In Japanese, apart from /t/ and /ʃ/ we also have /ts/, which is the other affricate in the language, but in particular /kʲ/, which is instead a place of articulation error, and is the most common mistake.[27] This last sound is a palatalized velar, which is found where /k/ is followed by /i,j/ in Japanese.

This last error suggests that for children learning Japanese, there might be no clear distinction between /tʃ/ and /kʲ/ at an early stage, while children learning English can distinguish between the first sound in the word *cherry* and the first sound in the word *cube*, because they rarely produce [kʲ] for a target /tʃ/. Since Edwards & Beckman (2008a) annotate the instances in which /k/ is followed by /j/ in English separately, we can isolate the specific words which have a palatalized /k/ in the two languages. Table 4.25 shows indeed that children learning Japanese also displays an asymmetry in the opposite direction: a target /kʲ/ often triggers /tʃ/ in child speech (61/479=13%), but this does not happen as frequently in English (3/120=3%).

An analysis of the frequencies and the functional load of /kʲ/ and /tʃ/ (in Table 4.26) shows that we cannot explain this asymmetry with reference to functional load, because

---

[27]I use the notation /kʲ/ for consistency with /tʃ/ and with the fact that the sound is considered an independent phoneme in Edwards & Beckman (2008a), but actually this sound cannot be considered a phoneme if one looks at Japanese native words.

[kʲ]	English (120)	Japanese (479)
tʃ	3	61
k	18	34
t	3	12
ç		10
g	1	4
hʲ	4	
kʷ	3	
ʃ	1	2
p		2

Table 4.25: Production errors made by 2- and 3-year- children in the Paidologos corpus for /kʲ/.

	English	Japanese
Freq(tʃ) tokens	0.0048	0.0164
Freq(kʲ) tokens	0.0004	0.0197
Freq(tʃ) types	0.0236	0.0223
Freq(kʲ) types	0.0024	0.0360
FL(tʃ,kʲ) tokens	0.0001	0.0046
FL(tʃ,kʲ) types	0.0005	0.0073

Table 4.26: Frequencies of /tʃ/ and /kʲ/ and their functional load in English and Japanese.

/kʲ/ is more frequent in Japanese than it is in English, and the two sounds occur in the same environment in Japanese native words (before /i,j/), and therefore they are clearly contrastive.

Since the two sounds occur in the same environment in Japanese, the only ways in which children can realize that these two sounds are in allophonic relation with /t,k/ is by paying attentions to their morphological alternations (Vance, 1987, 77). Alternatively, they might temporarily adopt an analysis in which a unique palatal sound is associated with the /i,j/ context, which would not be incompatible with the input. This might explain their interchangeability.

This hypothesis might be less plausible for children learning English. First, [kʲ] is rare in the vocabulary (and is typically described as a consonant cluster in phonological analyses rather than a phoneme), and second, it is not really in allophonic or morphological alternation with any other phoneme. The functional load between this sound and /tʃ/ is low not because the two sounds are not contrastive: to some extent, they are more contrastive

	English	Japanese
FL(tʃ,t) tokens	0.0024	0.0063
FL(tʃ,t) types	0.0086	0.0072

Table 4.27: Functional load of /tʃ/ and /t/.

in English that they are in Japanese (because they can occur in different environments). It is the rarity of [kʲ] which makes their functional load low, and its rarity makes it not an option that children can consider until they develop a large vocabulary.

One last possible interpretation is that the reason why English learners produce many /t/-mistakes is because the contrast with /tʃ/ is not prominent in the vocabulary. This does not seem to be true (Table 4.27), at least for functional load at the type level, for which the contrast is more evident in English than it is in Japanese, in which by definition it can occur only in non-native words.

To summarize, in this case the production accuracy for English and Japanese is similar, but for different reasons. In English, there is no other phoneme competing with /tʃ/, and therefore most of the errors involve /t/, which is a candidate for a production error. In Japanese, the phonemic inventory contains more sounds which are acoustically similar to /tʃ/ (at least /kʲ/, /ts/ and /ç/), and therefore the error patterns are different, but they do not cause the production accuracy of the phoneme to be significantly lower than English: on the contrary, the production accuracy is higher (but not significantly higher). This is interesting, given the fact that while the sound has full phonemic status in English, in Japanese it has allophonic status in native words, and therefore one might have expected the opposite pattern: in fact, allophonic rules like flapping appear late in English (Klein & Altman, 2002, Richter, 2018).[28]

In this case, we do not see a clear developmental difference between the children learning the two languages: one possibility is that the phoneme is acquired later in Japanese because of the presence of similar sounds in the inventory, and especially because of the impossibility to clearly distinguish /kʲ/ and /tʃ/ distributionally. Another possibility is that in Japanese

---

[28]Yoneyama et al. (2003) suggest that since each consonant has a palatalized allophone in Japanese, there might be a structural pressure that favors an early acquisition of the palatalization rule.

it is actually acquired earlier, given the many /t/-errors that children make in English, but this cannot be explained on distributional grounds. A third possibility, suggested by the similar production accuracies, is that even though the developmental trajectories are different, children learning the two languages identify the phoneme roughly at the same time.

In any case, all the differences in the error patterns do not seem to be related to functional load.

4.4.4   Summary

To summarize, in this section I examined the relationship among production accuracy, distributional properties of the segments, and errors in the speech of 2-year-old and 3-year-old children. We concluded that:

- There is some salient difference in the production of /k,t/ in Cantonese compared to the other three languages of the sample, which however does not correlate with the structural properties of the contrast.

- The voicing distinction is acquired earlier in English than in Japanese, even though this does not seem to depend on structural properties.

- The production accuracy of /s/ is quite uniform cross-linguistically, and it only depends on the alveolar consonant inventory. The larger the inventory, the lower the accuracy.

- Functional load might explain why /ʃ/ is produced with a better accuracy in English than in Japanese, despite the fact that it is more frequent in Japanese.

- Both frequency and functional load (at the type level) might explain why /θ/ is acquired earlier in Greek (and in Spanish) than in English.

- /ts/ has higher accuracy in Cantonese than it has in Greek and Japanese, and this can be explained by both frequency and functional load considerations.

151

- /tʃ/ shows some asymmetry in the error pattern between English and Japanese, but its production accuracy is similar in the two languages, and it is not related to functional load.

These findings show that, at least in some cases, an asymmetry in the behavior of children learning different languages can be explained as a representational error motivated by structural considerations, mostly with reference to functional load. In the next section, I will elaborate on what these findings can imply for an acquisition model.

4.5   An acquisition model

In the preceding sections, we have seen that functional load calculated at the type level is the best predictor of phonological development. This finding has some implications for the models of phonological acquisition that have been proposed in the literature.

On the one hand, there are models proposing that children learn phonemes by paying close attention to the acoustic properties of the sounds they hear, and identifying a finite number of target clusters via statistical learning. In Hillenbrand et al. (1995) and Kornai (1998), it is shown that by extracting formants and vowel duration through a spectral analysis from English speech, one can reach a high accuracy in automatically identifying the underlying English vowel given a sound token, which suggests that vowel categories are discriminable using acoustic features. Maye & Gerken (2000) and Maye et al. (2002) show that the distributions of two sounds [d] and [t] in the acoustic space can be manipulated in order to prompt a categorical judgment by children in a laboratory experiment: if the sounds to which they are exposed are drawn from a unimodal distribution in the training phase, children are more likely to interpret 'extreme' instances of the sounds as variations of the same sound in the test phase, while if the distribution in the training phase is bimodal, then children are more likely to form categorical judgments when exposed to the same extreme instances. These experimental results led to the hypothesis that infants form phonetic categories by paying attention to the acoustic properties of the sounds they hear, storing them in memory, and organizing them along some 'exemplar' categories through

unsupervised clustering (Pierrehumbert, 2003). Essentially, this procedure is similar to how the popular machine learning algorithm K-means works, with the difference that the number of clusters is not a given parameter, but it must be estimated. At a later stage, as children learn words, they will refine the categories, and group them according to their phonological properties and their phonotactic distributions. This step is necessary to explain adaptation to native phonological systems (Werker & Tees, 1984, Kuhl et al., 2006) and the fact that a language phonotactic grammar can influence the perception of word boundaries (Hay et al., 2004).

On the other hand, another class of models argue that infants do not form categories by performing unsupervised clustering over acoustic features, but by focusing on the differences which are relevant to distinguish among words in their native language (Charles-Luce & Luce, 1990, Walley, 1993, Swingley & Alarcon, 2018, Cui, 2020). There are several motivations for including the notion of lexical contrast in an acquisition model: first, statistical learning accounts assume that children are endowed with at least adult-like perceptual and memory capabilities, an assumption which has not been supported by experimental studies (Vihman et al., 2004, Fikkert & Levelt, 2008); second, sounds that are acoustically similar across languages are not always recognized as distinct categories, especially if they do not contrast in the lexicon (Dresher & Zhang, 2005, Swingley, 2019); third, children with a large vocabulary perform better in tasks that aim at testing phonological development, like nonword repetition, beyond the expectations from sound or syllable frequencies (Edwards et al., 2004). The most recent attempt to formalize an acquisition model based on lexical contrast is Cui (2020), in which it is proposed that children learn phonemes by partitioning the acoustic space into separate meaningful categories each time they notice that two words, with two distinct meanings, can be distinguished according to some acoustic features. In the long run, the acoustic space will be divided along the dimensions that are useful to distinguish words in the vocabulary, and differences which do not serve this purpose are simple ignored.

These two families of models make different predictions in terms of what matters for

phonological development. Models based on exemplar theory, and in particular those that argue for statistical learning as the strategy through which categories are formed, predict than in early stages the frequency of a sound in the input (i.e., its token frequency) is the main predictor of its acquisition. A sound which is frequent will provide many exemplars of its category, while if a sound does not occur very frequently, it would be hard for the child to obtain a robust estimate of its acoustic properties and its exemplar distribution. Models which instead focus on lexical contrast predict that the frequency of a sound in the vocabulary (i.e., its type frequency), or the presence of word minimal pairs for that sound (Cui, 2020), are the best predictors of phonological development.

The results presented in this chapter seem to argue in favor of the second class of models: type measures are better predictors of production accuracy in children than token measures. In particular functional load, which is a measure that captures the contrastive properties of a phoneme, turns out to be a slightly better predictor than raw type frequency, a finding which suggest that paying attention to lexical contrast might be the strategy that children use to identify phonemes.

In this section, we elaborate on the possible ways in which lexical contrast can be integrated in a model of acquisition.

4.5.1 The acquisition of the /f/-/θ/ contrast in English and Greek

One of the cases of the preceding section for which a clear asymmetry in phonological development was found was the acquisition of the /f/-/θ/ contrast in English and in Greek. According to the data collected by Edwards & Beckman (2008a), children learning English produce /f/ when asked to repeat a word starting with /θ/ very frequently, while children learning Greek never do it. Analyzing the pattern of errors, this appears to be a case in which children learning English fail to acquire the contrast in an early stage, while children learning Greek only commit errors that involve other sounds sharing the same place of articulation of the interdental fricative (/s,t/), and that can be interpreted as production rather than perception errors (Blevins, 2004, 134-135, McGuire & Babel, 2012). Evidence

from spectral analyses do not show that the sounds /f/-/θ/ are easier to discriminate in Greek than in English (Jongman et al., 2000, Athanasopoulou & Vogel, 2011). For this reason, here we consider the hypothesis that the two sounds are more contrastive in the Greek lexicon than they are in English.

4.5.2 Modeling the acquisition of the /f/-/θ/ contrast

The acquisition model proposed by Cui requires a first stage in which word learning occurs. Word learning is modeled as a probabilistic process: words are sampled from the lexicon and assigned a familiarity index (r), which is a function of their frequency. At this point, a random acquisition threshold (t) is generated from a uniform distribution over the [0,1] interval, and if $r>t$, the word is acquired. This step ensures that frequent words have a higher likelihood to be acquired than rare words.

The second step of the model regulates the acquisition of the phonological grammar. Every time a new word is acquired, it is initially represented in the lexicon as a vector that contains the acoustic information associated with that word. Since the learner is exposed to multiple instances of the same word, the acoustic dimensions are not associated with a single instance, but represent a weighted average of all the instances that the learner has been exposed to (similarly to what happens in exemplar models). When the child acquires her first two words, she will notice one (or more) contrasts between the two words, and store the dimensions along with the contrast is manifested in her phonological module, along with an estimation of where the acoustic boundary for each contrastive dimension falls. As more and more words are acquired, the child is faced with three different choices:

1. If the word can be distinguished from the other words in the vocabulary using the contrasts and the boundaries already in the phonological grammar, the number of contrastive dimensions in the grammar is not modified, but the acoustic boundaries are adjusted in order to better differentiate the words in the vocabulary along the existing dimensions.

2. If the word cannot be distinguished from other words in the vocabulary using the

existing contrasts, a new contrast is created along the dimensions that distinguish the new word from the words already in the vocabulary.

3. If the word is too close acoustically to an existing word, it is stored as a homophone.

I will not expand on how the child decides among these options, and the reader is invited to read Cui (2020) for further details. However, the important element is that by learning new words, the child will converge to a phonological grammar that contains all the contrasts that are relevant to distinguish words in her language.

4.5.3   Implementation

In order to implement the first step, I decided to adopt a slightly different strategy than the one employed by Cui. Since the sizes of the English and Greek corpora that were used to obtain the word frequency lists differ by several magnitudes, a unique formula to derive a familiarization rate through word frequency would skew the numbers of words learned toward the bigger corpus. For this reason, I simplified this step to the two following stages:

1. Sample a number of n words according to their frequency.

2. Memorize all the words that appear at least m times.

This would guarantee that the number of words learned by the English and the Greek artificial child is comparable.

The second step of the learning algorithm is less straightforward to model in the absence of acoustic measurements. However, for the purposes of this study, we can assume that when the learner acquires words which are *only* contrastive in terms of the places of articulation labiodental vs interdental, this contrast should be *sufficient* to encode the contrast between the two places of articulation in the grammar. This is the equivalent of saying that when the learner acquires minimal pairs involving the /f/-/θ/ distinction, this constitutes a sufficient trigger for learning that the contrast is phonological. Crucially, this would be true only under the assumption that the child has the perceptual/cognitive abilities to hear the contrast and store it in memory.

We will now examine the minimal pairs involving these two phonemes in English and Greek.

4.5.3.1 Minimal pairs

The first thing to notice is that both English and Greek have minimal pairs involving the /f/-/θ/ contrast in the list of 5000 words collected for this work:

- English: *three-free, thin-fin, death-deaf, offer-other, first-thirst*

- Greek: φα-θα ('F' (the musical note)-(verb particle))[29], φάρος-θάρρος ('lighthouse'-'courage'), φυτεία-θητεία ('plantation'-'service'), φέρετρο-θέρετρο ('coffin'-'resort')

This means that, in theory, both contrasts are learnable. Crucially, there are also some minimal pairs for the parallel contrast /v/-/ð/:

- English: *than-van, leather-lever*

- Greek: ήδη-ήβη ('already'-'puberty'/'adolescence'/(proper noun))[30]

Since these pairs can be used to learn that there is a contrast between labiodental and interdental fricatives, they will also be taken in considerations.

In order to simulate the child experience, I decided to sample words from the English and Greek wordlists and keep track of the number of minimal pairs involving the contrast under study. I ran the model several times for different values of the parameter n, varying from 200K tokens to 2 million. This parameter represents the number of word tokens that the child hears. As for m, I decided to use $m=5$, so that a child will only learn a word if it appears at least 5 times in the input. This parameter did not appear to have any concrete impact on the results, other than changing the amount of words learned as a function of the word tokens heard. The graph in Figure 4.10 summarizes the amount of minimal pairs that the child is exposed to as a function of the numbers of words learned.

---

[29]The first word is not native, but it is a borrowing from Italian.

[30]The second word has different meanings: it is used to refer to 'puberty'/'adolescence', but it is also found as a proper noun. In Ancient Greek, the word referred to 'youth'. Its presence in the high-frequency word list is probably accidental.
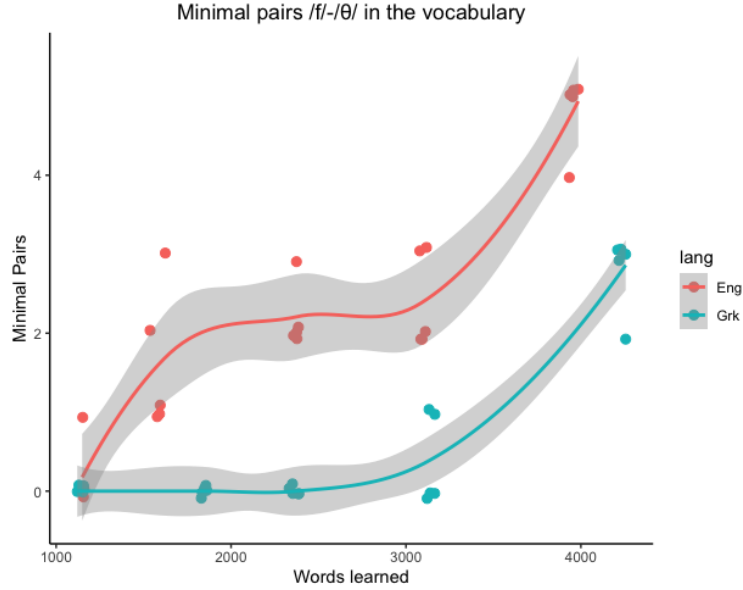
Figure 4.10: Minimal pairs acquired as a function of the number of words in the lexicon.

Interestingly, we note that for neither language minimal pairs are present when the vocabulary has approximately 1000 words. For English (red dots), some minimal pairs are present when the vocabulary reaches ≈1500 words. Many of the words involved in the minimal pairs for the contrast are quite common: in particular, the minimal pair *three-free* is present at each iteration that reaches a vocabulary of 1500 words. When the child reaches a vocabulary of 4000 words, she is guaranteed to have at least four out of the seven minimal pairs mentioned above available in the corpus that is used for the study.

On the contrary, Greek (blue dots) exhibits no minimal pairs in most of the simulations, and some are visible only when the vocabulary reaches about 3000 words. This follows from the fact that the Greek words that are involved in the minimal pairs for the contrast are uncommon, and therefore minimal pairs only appear when the child has acquired a large vocabulary.

These results are problematic for a few reasons. First, they are not compatible with our findings in the previous section, and they make the opposite prediction: that the contrast should be learned in English earlier than in Greek, and not vice versa. Second, for both English and Greek, minimal pairs appear when children acquire a vocabulary of some thou-

sands words, but we know from the experimental literature that it is unlikely for children to possess such a large vocabulary until a late stage: the children studied in the previous section are in the age range [2;00-4;00], and in this period they are expected to have learned, on average, less than a thousand words (Fenson et al., 1994, Hoff, 2009, Carlson et al., 2014). If minimal pairs were the only trigger for a phonological contrast, one would have to conclude that preschoolers of both languages would not be able to learn the contrast. This is an argument against a lexical-type-statistics approach to phonological development (cf. Maye & Gerken, 2000, Pierrehumbert, 2003).

4.5.3.2   Syllable contrast

An alternative way to address phonological development is focusing on subword units. An approach that relies on minimal pairs, like the one just described, has to make the idealization that children can identify the contrasts in all the positions in which they occur. Indeed, this idealization seems to be justified by the fact that the classical studies on categorical perception mentioned at the beginning of the chapter (Eimas et al., 1971, Trehub & Rabinovitch, 1972, Eilers & Minifie, 1975, Miller & Morse, 1976, Kuhl, 1979 and subsequent work) reported that infants can discriminate a wide range of contrasts, and in different positions (Jusczyk, 1977).

However, from some of such studies it also emerged that identifying a contrast is more difficult in certain positions of the word than in others. Trehub (1976) shows that while children can distinguish the contrast between [p] and [b] when the sounds are presented in monosyllabic (*ba/pa*) and disyllabic (*aba/apa*) words, discrimination is blocked when the contrast is embedded in the final syllable of a trisyllabic stimulus (*ataba/atapa*). Similarly, Jusczyk et al. (1999a) show that nine-months-olds' are sensitive to shared phonetic properties among syllable onsets, but not among syllable codas. The fact that children might not be able to perceive or pay attention to all the sounds that they are exposed to has been extensively studied after Macken (1980a), who provided evidence for the fact that lexical representation in children are non-target like. For instance, at the beginning of the chapter,

we mentioned that Grunwell (1982) and Vihman (1982) provided evidence for the fact that certain contrasts (like /tr/-/tʃ/ and /f/-/θ/) are not part of children lexical representation when they begin to produce words.

Another type of evidence comes from word recognition studies (Hallé & de Boysson-Bardies, 1994). Vihman et al. (2004) show that when the onset of a stressed syllable of a familiar word is changed in the stimulus, children exhibit looking times comparable to those they exhibit when exposed to rare words, which means that they do not recognize the word (or they recognize it as a different word, see Swingley & Aslin, 2000). However, when the onset of an unstressed syllable in the same familiar word is changed, children show a preference for the altered word over the rare word, which means that they recognize the familiar word. The findings hold for English and for French (cf. Hallé & de Boysson-Bardies, 1996). Similarly, Altvater-Mackensen & Fikkert (2010) show that children habituated to a monosyllabic word in Dutch cannot recognize the word if its onset is changed, but do recognize it if the coda is changed. This shows that children go through a stage in which only the contrasts in onset position (and in stressed syllables) are acquired, while others are left underspecified (cf. also Swingley, 2005, 2009).

These considerations have an implication for the model of acquisition previously described. If children go through a stage in which their lexical representation are holistic, and they store in memory only those contrasts which appear in salient positions of the word, like stressed syllables, then it might be more informative to look at contrasts among subword units, rather than minimal pairs, when studying the acquisition of a contrast.

For this reason, I replicated the experiment above by tracking the number of words whose first syllable is stressed, and that start with either of the phonemes investigated (or their voiced counterpart). I will assume that if a child learns two words whose stressed syllables share the same nucleus, but a different onset, children might learn that the two phonemes in the onset must belong to two different categories. Following the considerations in Grunwell (1982), I only keep track of contrasts that involve simple onsets, and ignore consonant clusters. Basically, this strategy makes the idealization that children initially
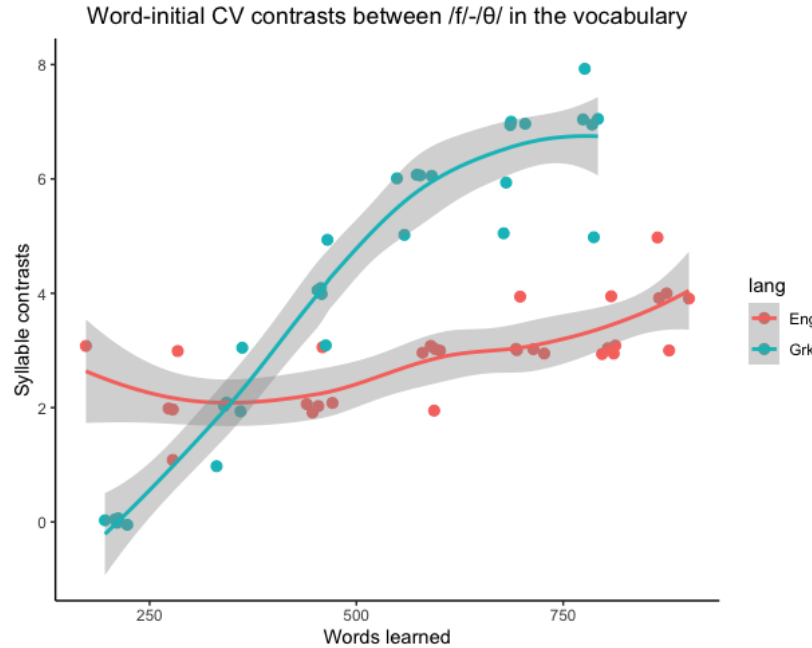
Figure 4.11: Syllable contrasts acquired as a function of the number of words in the lexicon.

have a CV lexical representation for each word they learn.[31] By replicating the previous experiment making this idealization, the results are those summarized in Figure 4.11.

The dots representing English (red) show that children notice some syllable contrasts in word-initial syllable even with a small vocabulary of a few hundred words. In particular, the contrast /fɪ/-/θɪ/ is immediately visible because of common words like *thing, think, fish, finger, fix*. Similarly, children notice a contrast between /vɛ/-/ðɛ/ because of the common words *them, there, very*. However, as the vocabulary increases, we do not see many other syllable contrasts emerging. The third contrast that emerges is the one between /fɔ/ and /θɔ/, when the words *fall* and *thought* are learned, and the fourth contrast, /fʌ/-/θʌ/, requires *fun* and *thumb*. In spite of a large vowel inventory, the evidence is mostly limited to these four cases. This is an example run at about 900 words learned:

- /fɪ/-/θɪ/ (*figure, fix, finish, finger, fish, fit, thing, think*)

---

[31]Stoel-Gammon & Cooper (1984) and Walley (1993) show that even though early lexical representations are extremely simple and limited to a single consonant and a single vowel, they do not necessarily reduce to the same CV pattern: children exhibit a lot of variation in their preference for certain position of the syllable, or certain features.

- /fʌ/-/θʌ/ (*fun, funny, thumb*)

- /fɔ/-/θɔ/ (*fall, thought*)

- /vɛ/-/ðɛ/ (*very, them, there, then*)

Note how while before /ɪ/ there is a good amount of evidence for the contrast, in the other environments the evidence is limited to a few words.

The dots representing Greek (blue) reveal a different pattern. While no contrast is visible when children have a vocabulary of a few hundred words, by the time they approach 400 words they know as many contrasts as their English peers, and when they reach 700 words they are basically exposed to almost all the contrasts before the five vowels of the Greek vowel inventory. Most of them appear to be associated with many word pairs:

- /fa/-/θa/ (φάντασμα 'ghost', φάση 'phase', θάνατος 'death')

- /fe/-/θe/ (φέρω 'carry', φεύγω 'leave', θέλω 'want', θέμα 'theme', θέση 'position')

- /fi/-/θi/ (φίλος 'friend', θείος 'uncle', θεία 'aunt', θύμα 'victim')

- /va/-/ða/ (βάρκα 'boat', βάσει 'based on', βάζω 'put', δάσκαλος 'teacher', δάσος 'forest')

- /ve/-/ðe/ (βέβαια 'surely', δέκα 'ten', δέντρο 'tree', δέρμα 'skin')

- /vi/-/ði/ (βήμα 'step', δείγμα 'sample', δίκαιο 'right', δίκη 'trial', δίνω 'give', δύναμη 'power', δύο 'two', δίπλα 'next-to', δείπνο 'dinner')

- /vo/-/ðo/ (βόλτα 'stroll', βόμβα 'bomb', βόρεια 'north', δώρο 'gift', δόση 'dose', δόξα 'glory')

The only syllable contrast which is missing is /fo/-/θo/, which requires access to the only word with a /θo/ initial syllable in the wordlist, θόρυβος 'noise'. This word is not very frequent, while many words display /fo/ at the beginning. Contrasts before /u/ do not appear in the wordlist.

162

These findings can account for the asymmetry noted in the previous section. First, it is clear that children learning Greek have a lot of evidence for a word-initial syllable contrast once they learn 600-700 words, and this observation can explain the lack of confusability between /f/ and /θ/ in production. On the other hand, for English the evidence of the contrast is mostly limited to one specific environment.

The findings are not necessarily incompatible with an exemplar model of acquisition. In the previous section, we showed that /θ/ is more frequent in Greek than in English, at the type and the token level. If we allow exemplars to be stored in a multi-level network (Wedel, 2012), children might be storing exemplars at the level of both words and phonetic categories, and therefore it would not be a surprise that if they can have access to more θ-words, then their representation of the category is more robust.

An observation that weakens an explanation in terms of exemplars, though, is the fact that the specific environment in which the contrast is evident from the beginning in English (/θɪ/), is also the environment that Edwards & Beckman (2008a) used in their elicitation task for both English and Greek. The words used to elicit the sound in English were *thinking, thin* and *thimble*, with the first two being very frequent words: *think* is the 39th most frequent word in the input, while the rank of *thin* is 1684.[32] The word *thing*, which is in their phonological neighborhood, is also very frequent (rank: 146). On the other hand, the rank of the most frequent Greek word used in the task (θείος, 'uncle') is only 568, but children exhibit a higher production accuracy for the sound in this word (17/37=46%) than in the more frequent English words *thinking* (10/29=26%) and *thin* (5/38=13%). This makes the behavior of children learning English even more puzzling: if they are able to store several exemplars of the words, why are they not able to reproduce them as faithfully as their Greek peers?

This is a case in which there is an asymmetry between word frequency and the type frequency of a phoneme that the words contain: while the words elicited are quite frequent, they contain a phoneme which is not very common in the lexicon. The only way to explain

---

[32]For practical purposes, the word *think* was elicited with the suffix *-ing*, but I do not see any reason why this would have severely altered the elicitation of the first segment.

this asymmetry would be to argue that the exemplars of the words are somehow 'down-weighted' compared to the exemplars of the phonemes in the multi-level exemplar network (Wedel, 2012), or that word exemplars are stored in memory only after the phonemic categories are acquired. This second position might be more compatible with the position by Pierrehumbert (2003), who argues that phonological categories are acquired independently from word learning, and probably in a stage in which children only know a limited amount of words. However, this position would not be compatible with the fact that type frequency and functional load are predictors of phonological development, like suggested in the previous section. While Pierrehumbert does not deny the role of type frequency in phonological acquisition, she argues that type statistics is only relevant at a later developmental stage, when categories are already formed, and therefore should not play a role in the stage of category formation.[33]

The same observation also represents a problem for a lexical contrast approach, because as we have seen, the contrast between /fɪ/ and /θɪ/ is well represented in the English lexicon, and therefore should be acquired early, according to the model by Cui (2020). One potential explanation for this pattern is the fact that the high front vowel often triggers a change in place of articulation, both diachronically and synchronically, to the point that palatalization is one of the most frequently attested cases of sound change or morphological alternation (cf. Bateman, 2007, Kochetov, 2011). Acoustic studies have also found that fricative-vowel coarticulation is very common with high vowels, but typically does not occur with low vowels (Soli, 1981). We can hypothesize that since children learning Greek are exposed to the contrast in different vowel environments, they can clearly capture the distinction between the labiodental and the interdental fricatives and make it phonological, while children learning English have evidence for the contrast only in an environment in which coarticulation is expected, and they need a large vocabulary before they are able to notice the contrast in other environments. This predicts a delay in the acquisition of the contrast, that would explain the results in Edwards & Beckman (2008a).

---

[33]'This level of analysis exceeds the capabilities of infants, since refined type statistics require a large lexicon, and syntactic and semantic development unfolds over many years.' (Pierrehumbert, 2003, 139-140).

In this case, it is clear that the frequency of the syllable /θɪ/ in the input is not a good predictor of phonological development, because if it were, then we would predict children learning English to acquire the phoneme earlier, on the basis of frequent words like *thing* and *think*.[34] Rather, it is the lack of evidence for a contrast with /f/ in the early lexicon which can account for the delay.

4.6 Conclusion

In this chapter, we have seen that lexical contrast can influence phonological development. In particular, following Stokes & Surendran (2005) and Edwards & Beckman (2008a), we focused on the production accuracy of word-initial consonants, by running regression models using two different measures of phonemic frequency and two measures of functional load.

The first finding was that measures associated with word types are better predictors of phonological development than measures based on word tokens. In particular, functional load calculated over word types turned out to be the best predictor of production accuracy in children. This finding is compatible with the idea that children are not only sensitive to the frequency of the particular linguistic input they receive, but they use considerations about the diversity of the input to form categories.

In particular, in the cases of the fricative /s/ and the affricate consonants /ts,tʃ/, the errors seem to be influenced by the phonological inventory: children show similar production accuracy, which is compatible with a similar path of phonological development, and they replace sounds which are difficult from the articulatory viewpoint with some of the sounds they have already acquired. However, in the case of the contrasts involving the fricative consonants /ʃ/ and /θ/, there is some evidence for different learning paths in children of different languages, and they seem to be related to functional load considerations. The factor might also explain why children learning Cantonese have a better production accuracy of /ts/ with respect to their Greek and Japanese peers.

---

[34] A similar argument can be made for the voiced interdental fricative, /ð/, which is very frequent in the input because of its use in fuction words, but it's not very contrastive, and it's acquired late (Ingram, 1999).

Finally, we have seen that by modeling the acquisition of phonological categories, while a model based on minimal pairs cannot account for the asymmetry in the production of /θ/ in children learning English and those learning Greek, a model based on contrasts in word initial stressed syllables makes the right prediction. This is compatible with the proposals that children go through a stage in which words are represented holistically (Hallé & de Boysson-Bardies, 1996, Vihman et al., 2004, Altvater-Mackensen & Fikkert, 2010), and that a phonological grammar emerges in response to the contrasts that children notice in their early representations (Walley, 1993). The developmental pattern presented in not clearly explained by exemplar models, which instead assume rich representations and a direct correlation between word/phoneme/syllable frequency and their acquisition. In this case, one would have to explain why the low production accuracy of words like *think* and *thin* is not explained by their frequency of use, but by the rarity of their word initial sound in the vocabulary (i.e., its type statistics).

This final finding provides a potential explanation for the results in the last chapter: the reason why the number of minimal pairs can influence historical mergers is because lexical contrast is one of the factors that drives the acquisition of phonology in child-language acquisition. The failure of acquiring a contrast, if not corrected in the course of the developmental period, can result in the actuation of a sound change.

Chapter 5

Final remarks

In this chapter, I summarize the main findings of this work, and I discuss their theoretical implications, along with the questions that they leave open.

5.1 Lexical change over time

In Chapter 2, we have seen that regular sound change creates a bias toward small phonological inventories, and a lexicon in which the average number of distinctions among words tends to be low. I argued that these two properties are the consequence of the irreversibility of mergers. A system in which splits can be reversed, but mergers cannot, inevitably creates an attraction toward states of low phonological dispersion. In the extreme case, regular sound change can create homonymy, and once two words become homonyms, there is no sound change that can divide them. The fact that languages tend to display words which are very similar among each other, then, is not sufficient to argue in favor of some functional pressure acting on the lexicon of languages as to make them more compact. On the other hand, it is not necessary to postulate functional mechanisms acting on languages in order to avoid states of massive homonymy and high confusability: once we allow lexical replacement, and speakers have the option to replace words with loanwords from other languages or dialects, or to coin new words by recombining the words in their vocabulary, this problem is resolved.

While this work shows that sound change does not need to be modeled with reference to language communicative functions in order to account for lexical change over time, it does

not rule out the possibility that functional pressures exist, in this domain and in others.

For instance, while I argued that functional pressures on loanwords are not needed to model lexical change over time, the proposal in Martin (2007) and Graff (2012) that the process of lexical borrowing is not entirely neutral, but it is influenced by considerations about the phonological shape of the loanwords and their competition with existing words in the vocabulary, has not been discussed in this work. A way to evaluate this claim would be to follow Martin (2007) in collecting data about loanwords, and to see if their distribution is better explained by a neutral model or by a selection model. This question has not been further explored in the literature, to the best of my knowledge, and this dissertation does not fill the gap. More research in this direction is needed to further address this question.

Similarly, this work has not explored the role of word coining in shaping the lexicon over time. A hypothesis which is compatible with the findings of this work is that while regular sound change is essentially a neutral force, when speakers coin new words, they tend to fill gaps in the vocabulary. Martin (2007) presents some data in favor of this hypothesis, by noticing that the amount of neologisms in French starting with 'b-' is very high, even though the phoneme was rare in Latin, and therefore he argues that 'b-' words were selected for. However, a systematic study on English neologisms by Caplan et al. (2020) shows the opposite results: namely, that stochastic processes of word formation can account for their form better than models optimized for communication (*contra* Piantadosi et al., 2012). This is another domain that will benefit for more research.

5.2   Functional Load

In Chapter 3, we have seen that the Functional Load Hypothesis can be related to mergers which have a direct impact on the phonemic inventory of a language, but it has little predictive power when it comes to sound changes which do not cause any change in the phonemic inventory of the speaker, but just affect the lexical representations in the vocabulary or the phonological rules in the grammar. This result provides support for theories which link the acquisition of the phonological inventory to the notion of lexical contrast. Crucially,

though, this is only true if we use measures like minimal pairs as a proxy of functional load, while this is not true if we use an entropy model: in this case, functional load has no predictive power (as already shown by Surendran & Niyogi, 2006, Wedel et al., 2013, and Cohen Priva, 2017). This suggests that children are more sensitive to local contrasts that emerge within subsets of the vocabulary, rather to global measures of informativity.

The findings of this chapter are not conclusive, because they are based on the study of ≈30 instances of sound change, and only 7 language varieties, all Indo-European (with the exception of Cantonese). It is likely that different patterns will emerge by looking at more language families. In particular, it could be that outside of Indo-European languages, unconditioned mergers which neutralize well-supported contrasts are more common, or that conditioned mergers that lead to homonymy are less common than the cases we found in the data examined here. Either of these findings would force another reassessment of the Functional Load Hypothesis. More empirical work in this direction is needed, because revising the Functional Load Hypothesis might help us understand the cognitive mechanisms that underlie the factor, and ultimately explain how functional load relates to phonological acquisition.

In particular, Wedel et al. (2013) explain their results by arguing that functional load could be a bias that influences production and perception in children and adults. In sum, when adults produce a word which is potentially ambiguous, they 'disambiguate' the word by enhancing its contrastive cues, with the effect of making it more understandable for the listener (Wedel, 2012, 326-327). This is why the presence of minimal pairs blocks mergers over time.

This conclusion is not compatible with the scenario depicted in this work, where we see that in the case of unconditioned mergers, lexical contrast seems to be indeed weak, but this does not seem to be true in other cases of sound change. This apparently odd conclusion is not easily explained with reference to communicative biases: if adults and children were biased in their perception and production against homonymy, we should see an effect independently from the diachronic change that caused homonymy in the first place.

However, this asymmetry is compatible with a scenario in which lexical contrast is relevant in the stage of category formation, but it is less relevant once the children have converged on a phonemic inventory. The changes that happen after this stage, which could involve the refinement of lexical representations and phonological rules in the grammar, and could lead to conditioned mergers, might be blind to lexical contrast.

Another difference between this account and Wedel et al. (2013)'s account is that this account has more predictive power, and therefore is also more easily falsifiable. Under Wedel et al. (2013)'s account, functional load has no predictive power in the single case: it is just one of many factors, like social factors and phonological factors, that can contribute to determine the occurrence of a sound change. This means that cases of sound changes that result in homonymy cannot be used to falsify the hypothesis, unless one adds them to a large corpus and runs a regression analysis. On the contrary, the account of functional load presented here predicts that in cases of endogenous sound changes, unconditioned mergers cannot be anti-functional. If we see cases of anti-functional mergers, and these cases cannot be explained with reference to language contact (cf. Honeybone, 2016), then these cases can be used to falsify the hypothesis.

5.3 Phonological acquisition

In Chapter 4, we have seen that lexical contrast can be used to explain errors in phonological development. The results of the chapter show that children are more sensitive to contrasts among words, in particular to syllable contrasts, rather than to the frequency of a specific sound or a specific word in speech. This is not predicted by models which assume rich representations in the children's memory.

The findings in Chapter 4 support the hypothesis advanced in the previous chapter, namely that functional load is a factor that can be explained with reference to phonological acquisition. The level at which this factor operates, however, is not the level of utterances or conversations between children and adults, but must be the level of children lexical representation, because children appear to be more sensitive to word type statistics than

to word token statistics. This poses some problems for exemplar models of phonological acquisitions (Pierrehumbert, 2001, 2003, Wedel, 2012).

In particular, the comparison between the learning of the contrast between /θ/ and /f/ in English and Greek cannot be easily explained with an exemplar model. Children learning English are exposed to frequent words exhibiting the sequence /θɪ/. They also have access to many minimal pair contrasts, a fact that should have the effect of enhancing the contrast with /f/ (Wedel, 2012), and give them an advantage over children learning Greek. Yet, they are unable to faithfully reproduce the sequence in speech. This fact has been traditionally explained with reference to articulatory constraints on production. However, children learning Greek at the same age perform better on the task, in spite of having less evidence for word contrasts that rely on this acoustic cue for disambiguation. The only way in which this phenomenon can be explained is by assuming a weak representation of the words in the lexicon of the children, and the development of a phonological inventory along the lines of syllabic contrasts, starting from an early age. This suggests that a great part of phonological development occurs even before children begin to formulate sentences or acquire a large vocabulary: at these stages, phonological learning is already on its way.

5.4 Conclusion

This work started as an attempt to model regular sound change over time, and to study its relation to language communicative functions. In particular, this work used computational simulations to study the interactions of different types of sound change, and examined large written corpora of contemporary languages, to quantify the amount of information provided by sound contrasts in the lexicon.

The model of lexical change over time proposed here provided results which are compatible with the classical Neogrammarian position, with Labov (1994)'s program, and also with more 'functionalist' models of lexical change (Martin, 2007, Graff, 2012, Dautriche et al., 2017). Sound change is essentially a neutral process, that typically has the effect of introducing more ambiguity from the communication viewpoint. However, speakers have

several repair strategies that they can adopt in case disambiguation becomes necessary. Most of these strategies involve the use of elements internal to the system to disambiguate: the use of optional elements (like adjectives and optional articles), or the increased use of a synonym to replace an old word that becomes opaque, are both strategies that can serve this purpose. Another option is borrowing words from a different language variety. In any case, these are phenomena that are independent from sound change, and that occur after sound change has already happened: they are not factors that condition its occurrence.

A similar conclusion was reached from the revision of the Functional Load Hypothesis. This factor has traditionally been presented as a constraint on sound change, motivated by the fact that speakers would not endanger the language communicative function by merging two sounds that are easily confusable, if they are contrastive in the lexicon. This treatment of the hypothesis has survived in recent exemplar-theoretic interpretations of functional load as resulting from cue-enhancement in adult speech (Wedel, 2012), which has the effect of disambiguating words that have minimal pairs in the lexicon. However, the empirical findings of this work led to a reassessment of the hypothesis.

First, we have seen that children are sensitive to lexical contrast early on. Under this premise, the decision of merging or not merging two sounds must be grounded in their ability to identify a contrast distributionally in their early lexicon. This correctly predicts that if a contrast is salient in the lexicon, then it is also likely to be acquired. It follows that the presence of minimal pairs must be correlated with the absence of mergers.

However, this does not prevent homonymy to appear in the lexicon by other means. For instance, we have seen that neutralization of vowel contrasts before a liquid is not prevented among the vowel contrasts that lead to homonymy in English varieties. Word final lenition in North American English has also been shown to be independent from homonymy considerations (Cohen Priva, 2017). Similarly, simplification of consonant clusters and consonant codas did not prevent homonymy to arise in Middle Chinese (Sampson, 2013, 2015). I argued that these are not statistical accidents, but cases that appeared because they did not involve the loss of a salient contrast in the grammar. Based on the evidence

from experimental studies on infants, we have reasons to believe that these changes might have occurred because children failed to reach the target lexical representations of a class of words, or decided to restructure their phonological grammar after being exposed to an input that was different from that of their caregivers.

All these considerations lead me to conclude that functionalism, to the extent in which it can be defined and studied, is still being 'overestimated' in the modern literature on sound change.

Appendix A

Mergers and Splits in Uralic and Altaic languages

A.1 Uralic

The following paragraphs summarize the historical phonological development of Uralic (Sammallahti, 1988). For the Proto-Finno-Ugric branch, the report stops at the level of Proto-Ugric and Proto-Finno-Permic, because further developments are controversial.

- **Proto-Uralic to Proto-Samoyed, consonant inventory**: for Proto-Uralic, 17 consonants can be reconstructed, but internal reconstruction from Samoyed languages points to only 13 segments. Two phonemes which were probably representing fricative consonants, /d/ and /d'/, merged in all positions with /r/ and /j/. The same happened with the phoneme /s'/, which merged with the unmarked /s/. The status of Proto-Uralic /x/ is less clear, and therefore I ignored it. Overall, it is clear that at least three phonemes were lost as a consequence of mergers.

- **Proto-Uralic to Proto-Samoyed, vowel inventory**: Proto-Uralic had a eight-vowel system, with four high vowels. Two new vowels, /ĕ/ and /ɵ/, resulted from splits targeting high vowels, which added two phonemes to the inventory. The presence of a tentative phoneme /ö/ is also reported, but its sources are not clear.

- **Proto-Uralic to Proto-Finno-Ugric, consonant inventory**: the inventory is stable. Internal reconstruction points to three new phonemes, /š/, /c'/ and /l'/, whose presence is postulated only for words that have no cognates with other branches, and

174

therefore their origin is not clear.

- **Proto-Uralic to Proto-Finno-Ugric, vowel inventory**: contractions are reported from the interaction of Proto-Uralic /x/ with five different vowels, so that the vowel inventory of Proto-Finno-Ugric is reconstructed with 13 total vowels. In this case, contractions resulted in five new phonemes.

- **Proto-Samoyed to Proto-South-Samoyed, consonant inventory**: no changes.

- **Proto-Samoyed to Proto-South-Samoyed, vowel inventory**: a merger affected the two back vowels /ä/ and /å/, therefore yielding only one back vowel for Proto-South-Samoyed. A second merger is reported for /ö/, but since the status of the vowel in Proto-Samoyed is not clear, this merger is ignored.

- **Proto-Samoyed to Proto-North-Samoyed, consonant inventory**: /c/ and /t/ merge in all environments.

- **Proto-Samoyed to Proto-North-Samoyed, vowel inventory**: no changes.

- **Proto-Finno-Ugric to Proto-Ugric, consonant inventory**: /s/ and /š/ merge into /θ/.

- **Proto-Finno-Ugric to Proto-Ugric, vowel inventory**: four of the long vowels (/uu/, /oo/, /ɨɨ/, /ee/) are merged with their short correspondents. Interestingly, the merger between /uu/ and /u/ has the consequence that the allophonic alternation between /u/ and /ŭ/ is phonemicized. Here, we have four mergers and one split in total.

- **Proto-Finno-Ugric to Proto-Finno-Permic, consonant inventory**: no changes.

- **Proto-Finno-Ugric to Proto-Finno-Permic, vowel inventory**: /ɨɨ/ merge with /oo/.

A.2 Altaic

The following paragraphs summarize the historical phonological development of Altaic (Robbeets, 2003).

- **Proto-Japanese to Old Japanese, consonant inventory**: The main difference between the consonant inventory of Proto-Japanese and Old Japanese is that internal comparison would not allow to reconstruct voiced obstruents for Proto-Japanese. The fact that in Old Japanese we find the voiced obstruents /b/, /d/, /z/ and /g/, but they do not appear in word-initial position, can be explained through a development of consonant clusters with nasals (*np, *nt, *ns, *ng) into voiced obstruents. In this case, contractions resulted in four new phonemes.

- **Proto-Japanese to Old-Japanese, vowel inventory**: While Old Japanese has eight vowels, scholars agree that only four vowels (/a/, /o/, /i/, /u/) should be reconstructed for Proto-Japanese. Old Japanese developed one high central vowel and three new mid vowels as a consequence of vowel hiatus. Like in the previous case, contractions resulted in four new phonemes.

- **Proto-Korean to Middle Korean, consonant inventory**: The Middle Korean alphabet encodes many phonemes that cannot be reconstructed for Proto-Korean. While Robbeets acknowledges that the three phonemes associated with voiced fricatives might be allophones of the obstruents, there is solid evidence that the aspirated obstruents /p^h/, /t^h/, /c^h/ and /k^h/ all developed from consonant clusters *pk, *tk, *ck and *kk. This would imply that also for Middle Korean, we have four new phonemes resulting from contractions.

- **Proto-Korean to Middle Korean, vowel inventory**: The question of the phonetic nature of Proto-Korean seven-vowel system is debated, but there is an agreement that the seven-vowel system of Middle Korean has a direct reflex in Proto-Korean, either as a direct descent or as a result of a vowel shift.

- **Proto-Tungusic to Manchu, consonant inventory**: The phonemic inventories of Tungusic languages are quite similar, and therefore reconstruction is not controversial. The reference here is Benzing (1955). If we follow the development into Manchu, we would find a /p/ and /f/ contrast that is not reconstructed for Proto-Tungusic: since a change /p/ > /f/ is reported in the history of Manchu, modern /p/ can be explained through dialectal contact with other Tungusic languages in which the sound change did not occur. Another feature of modern Manchu is the contrast between /s/ and /ʃ/, which following Robbeet's report resulted from a secondary split. In this case, we have two new phonemes resulting from one instance of borrowing and one of split.

- **Proto-Tungusic to Manchu, vowel inventory**: An eight-vowel system is reconstructed for Proto-Tungusic, but modern Manchu is described as a six-vowel system, as a result of the loss of a length contrast for /o/ and /u/. Here, we have two mergers.

- **Proto-Mongolic to Mongolian, consonant inventory**: The reconstruction follows Poppe (1954), and the changes in the phonemic inventory until contemporary Mongolian can be tracked in detail through Svantesson et al. (2005). Proto-Mongolic lacked /w/ and /ʃ/, which entered the language as allophonic splits from *p and *s, and in the second case this might have been facilitated by the presence of loanwords, where the allophonic alternation was not respected. Many of the consonants in Mongolian are associated to the palatalized versions of the obstruents (eight), the nasals (two), the liquids (two) and the approximant /w/, which contrary to Proto-Mongolic exhibit contrast with their non palatalized version before pharyngeal vowels, and therefore add a total of 13 phonemes to the language via contractions.

- **Proto-Mongolic to Mongolian, vowel inventory**: The only point of debate is whether the modern Mongolian seven-vowel system can be traced back to Proto-Mongolic, or if the proto-language had a vowel /ï/ that was lost. Following Robbeets, I assign this phoneme to Proto-Mongolic, and I postulate a merger with /i/. In Mongolian, vowel length is contrastive as a result of contractions affecting short vowels

in the proto-language. Therefore, we have one merger and seven contractions in this case.

- **Proto-Turkic to Turkish, consonant inventory**: The reconstruction of Proto-Turkic is less clear than the others. In particular, there is no agreement about whether Turkish /z/ and /ʃ/ result from secondary splits of reconstructed liquid consonants, or from contractions. The only patterns which are uncontroversial are the emergence of /h/ in result of a secondary split of /p/ in word initial position, while /f/ and /ʒ/ entered the language through loanwords.

- **Proto-Turkic to Turkish, vowel inventory**: The eight-vowel Turkish system can be traced back to Proto-Turkic, but while vowel length was contrastive in Proto-Turkic (and therefore the language had originally sixteen distinct contrastive vowels), in modern Turkish the length contrast has been lost. For this reason, we must attribute eight mergers to Proto-Turkic.

Appendix B

Implementing the notion of phonological feature

In order to make sound changes more realistic, one needs to implement the level of representation of phonological features in the system. For the purposes of sound change, two are the relevant generalizations that should be captured.

First, mergers and splits tend to target phonemes that are close in a phonological space defined by feature vectors. For instance, while for vowels like /i/ and /e/ mergers are attested, and the same is true for vowels like /e/ and /a/, mergers between /i/ and /a/ in a language in which /e/ is also present are not attested. This is because while /i/ and /e/ are usually defined as segments that contrast only on a [±high] feature, and /e/ and /a/ are usually defined as segments that contrast only on a [±low], segments like /i/ and /a/ instead contrast on two features ([±high, ±low]). This makes a merger between them more unlikely.

Second, the conditioning environment for a merger and a split is also determined by their features. For instance in the case of palatalization, languages in which a phoneme /k/, which is [+back], splits into a [-back] phoneme like /c/ or /tʃ/, the conditioning environment is represented by the [-back] phonemes /i/ and /e/. The conditioning environment agrees with the outcome of the split. This generalization is less clear-cut, because it does not account for cases of dissimilation, in which the outcome of the conditioned merger has an opposite feature than its conditioning environment, but it can still account for the majority of the sound changes typically attested.

Since our conditioning environments are limited to vowel-consonant interactions, there

179

aren't many features, apart from place of articulation, that are shared by both consonants and vowels. Other features, therefore, would not be useful in a lexicon of CVC syllables, because they would only become relevant for longer sequences. For instance, in real languages consonants can become voiced intervocalically, because vowels are voiced segments, but this change in voicing cannot be encoded in a model in which words have only one nucleus. For these reasons, our model will be sensitive to distinctions along a single feature vector.

By assigning an index to each symbol, we can create two unidimensional feature vectors.

b	c	ch	d	f	g	h	j	k	l	m	n	P	r	s	sh	t	th	tw	v	w	wh	x	y	z	'	'
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	

a	ai	e	ea	ee	ei	i	ie	io	o	oa	oi	oo	ou	u
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14

These vectors serve two purposes. First, when a sound change is applied to the lexicon, we can force the requirement that the two segments targeted for the change (X and Y) are adjacent when ordered according to their index: in this case, in an inventory like ['b', 'd', 't'], mergers like 'b' (0) and 't' (16) are avoided, as long as other segments in the range 1-12 are available as potential targets. However, 'd' can merge with either of them, as long as other closer segments are not present in the inventory. Second, these unidimensional feature vectors are meant to force the notion of 'directionality': when a sound change has the result of changing a segment towards a higher index, it must do so because the conditioning environment has high indexes, and viceversa.

Here is an example. A sound change that targets 'k' and forces a split into 'c' in onsets will work as follows.

First, we have to split the vowel space in a conditioning environment and a conservative environment. This can be done by arbitrarily choosing a symbol among those available as an environment boundary, and assigning all the segments which are in the direction of the change to the conditioning environment, while the rest is part of the conservative environment. Let us assume that the symbol selected is 'o' (9). The directionality of the

change ('k' to 'c' is a change from '8' to '1', and therefore from high to low indexes) forces all the indexes lower than 'o' to be the conditioning environment.

a	ai	e	ea	ee	ei	i	ie	io	o	oa	oi	oo	ou	u
0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	9	10	11	12	13	14

This would result in changes like the following:

- ken, ket, kot > **c**en, **c**et, kot

From the formal viewpoint this is the process through which Latin /k/ is palatalized in Romance languages. A more precise representation of the features and the implementation of mergers as feature assimilation would lead these functions to mimic realistic sound changes, but for the purposes of a null model, this idealization captures the mechanical aspect of feature assimilation that motivates the majority of the attested sound changes.

The question that we can ask at this point is: does implementing the mechanic of features alter the prediction of the null model? Figure B.1 shows that once we constrain sound change as to target symbols which are close to each other, the different runs of the simulation are more similar among each other, and the final outcome is still the decrease of both the size of phonemic inventories and phonological dispersion in the lexicon, which in this case happens much faster.

This means that implementing the notion of phonological feature does not prevent lexical saturation over time. On the contrary, it appears to facilitate it.
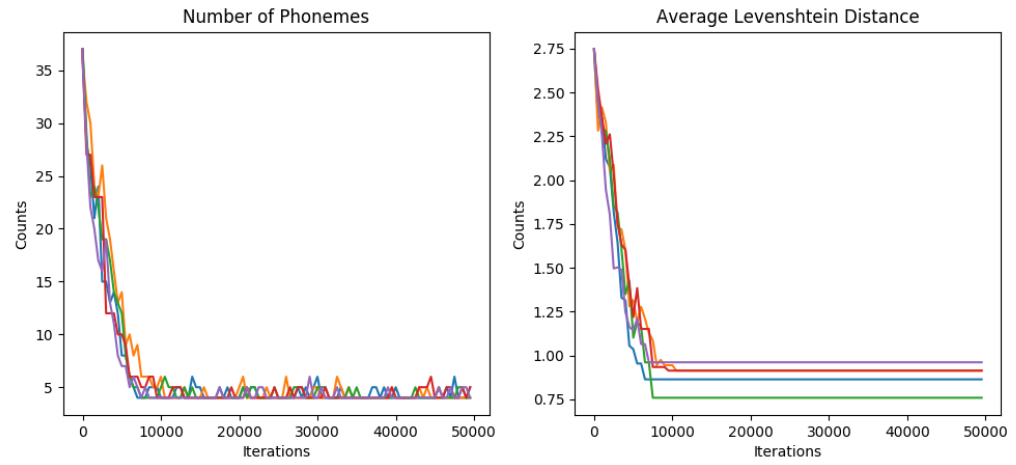
Figure B.1: The output of five sound change simulations, where 50000 sound changes are applied to the 150-word English list. Number of possible segments: 41. Phonological features are implemented.

Appendix C

Yule's distribution

A potential application of a sound change model is the study of phonemic frequencies within a language. Tambovtsev & Martindale (2007) show that in 95 languages, within-language phonemic frequencies are distributed according to a power-law distribution (see also Macklin-Cordes & Round, 2020). Specifically, the distribution proposed is the Yule-Simon distribution, a superset of Zipf's distribution that is found for word frequencies (Simon, 1955). This is an empirical fact that requires an explanation (cf. Sayeed & Ceolin, 2019, Nichols & Kauhanen, 2019).

Even though power-law distributions are usually associated with stochastic processes (Baayen, 1991, Banavar et al., 2004), one could also justify them as a byproduct of some functional pressure for optimizing communication (cf. the discussion in Piantadosi, 2014). For instance, Mandelbrot (1953), while reasoning on power law distributions, states that 'a quite statistical structure, entirely independent of meaning, appears, underlying meaningful written languages. This fact is to be considered as a very strong argument in favour of a thesis that language is a message intentionally if not consciously produced in order to be decoded word-by-word in the easiest possible fashion'. Or again, when describing specifically Yule's distribution for phonemic frequencies, Tambovtsev & Martindale (2007) suggest that 'it is interesting that the same equation describes the frequency distributions of DNA codons and phonemes. This may be purely a coincidence. However, it might imply a similarity between linguistic and genetic information transmission'.

In this case, a null model of sound change can be used to address the question, and investigate the possibility that also Yule's distributions result from regular sound change.
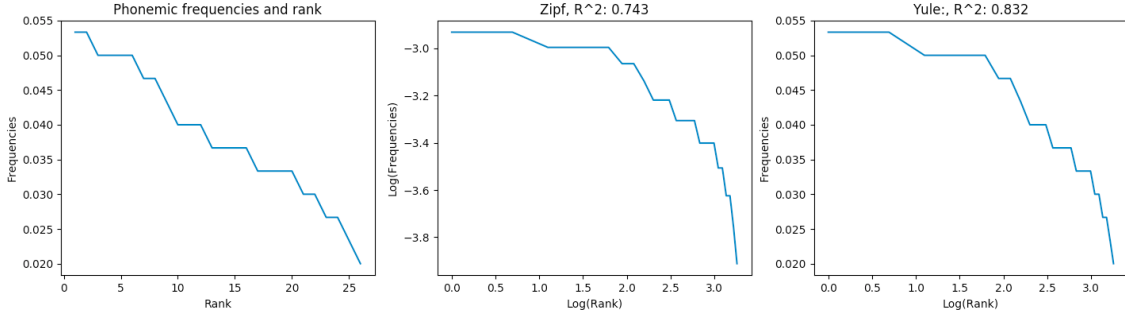
Figure C.1: Correlation between frequency and rank of consonants frequencies derived from an artificial list of 150 CVC words generated by sampling over the possible segments using a uniform distribution.

In order to address this hypothesis, I created an artificial list of 150 words that follow the same CVC pattern of the English wordlist, but the difference is that words are created by random sampling over consonants and vowels, so that their distribution is uniform (like in section 2.3.4).

The first graph in Figure C.1 shows that if we try to plot the consonant rank versus consonant frequency, the distribution does not resemble a power law distribution.

In the second graph, we have the log-transformed frequencies and ranks, which should be linearly correlated in case of a Zipfian distribution. In this case, there doesn't seem to be a linear fit graphically, even though the correlation is not null (R^2: 0.743).

In the third plot, the frequencies are plotted versus log-ranks of consonants. The situation is similar to the preceding one: there is no linear fit, but the correlation is not null (R^2: 0.832). According to Tambovtsev & Martindale (2007), a linear correlation between frequencies and log-ranks is a property of Yule's distribution. Since Yule's distribution has an extra parameter compared to the Zipfian, the better fit is not surprising: one would need to take into account the difference in degrees of freedom to actually evaluate Yule's distribution against the Zipfian (cf. Nichols & Kauhanen, 2019 and Macklin-Cordes & Round, 2020).

Next, I applied the artificial model of sound change on the lexicon, and I visualized the outcome of five parallel runs after 50000 sound changes (Figure C.2).

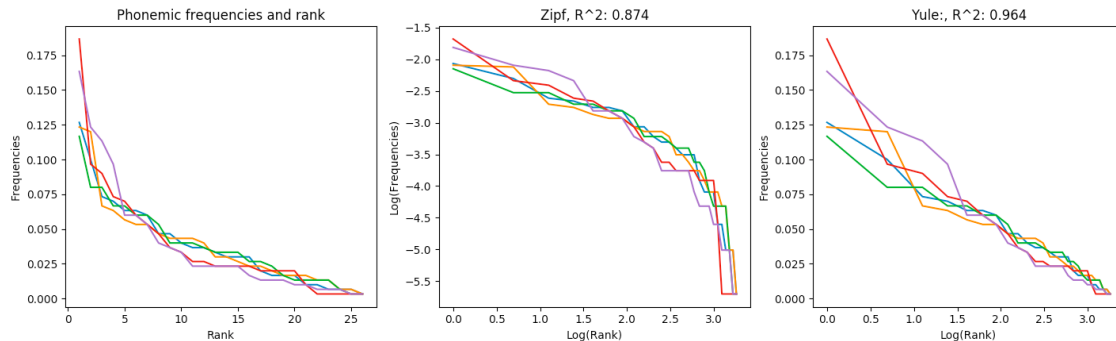In the first plot, we obtain what looks like a power law distribution.

Figure C.2: Correlation between frequency and rank of consonants frequencies derived from an artificial list of 150 CVC words generated by sampling over the possible segments using a uniform distribution, on which 50000 sound changes are applied. Number of possible segments: 41. Contractions and lexical replacement are included. Lexical replacement occurs with p=1/7.

In the second plot, we notice a linear correlation that seem to hold among the highest ranked consonants (R^2: 0.874), matching the results presented in Macklin-Cordes & Round (2020).

In the third plot, we show that the result improved if we correlate frequencies and log(rank) (R^2: 0.964), even though in this case the correlation appears stronger for the lowest ranked consonants.

This result is compatible with previous literature results that pointed to other distributions rather than the Zipfian as a better fit for phoneme frequencies (Sigurd, 1968, Good, 1969, Gusein-Zade, 1987, Borodovsky & Gusein-Zade, 1989).

For our purposes, though, the simulations show that the final lexicon is surprisingly similar, in terms of the internal distribution of consonants, to those found in real languages: it shows a power law distribution if we plot frequencies versus rank, and a good fit for Yule's distribution.[1] This is another case where we do not need to appeal to communication to explain a statistical property of languages: the simulation shows, in fact, that the presence of a power low distribution and of a linear correlation between frequencies and log(ranks) is a property that emerges following regular sound change, as proposed in Ceolin & Sayeed

---

[1] A caveat: here we are calculating phonemic *type* frequencies, while in the literature cited one can find claims of a fit for both *type* and *token* frequencies, depending on the source material (whether it was vocabulary lists or written texts).

(2019), even if one uses a lexicon where the distribution of the segments is uniform at the initial state. We note that the results did not change by using the English 150-word lexicon instead of the artificial lexicon as the starting point.

This result suggests that regular sound change might be the explanation for power laws that emerge from the distribution of phonemes in the lexicon of real languages.

Appendix D

Functional Load: Statistical analysis

In order to statistically validate the results of Chapter 3, I decided to replicate the regression analysis in Wedel et al. (2013).

D.1   Wedel et al.'s analysis

Wedel et al. are interested in evaluating the interplay between two different factors, phonemic frequency and number of minimal pairs, in predicting the realization or the absence of a merger between two phonemes. For this reason, they use a simple Mixed Effects Logistic Regression model.

The dependent variable of the model is a binary variable that has '1' for each attested merger, and '0' for each merger which is not attested in the literature.

The number of minimal pairs associated to each phonemic pair is the first independent variable they choose. Choosing a measure for phonemic frequency is more complicated: one can decide to take the sum of the frequencies of the two phonemes of the pair, or other measures such as the product, or the maximum. Moreover, one can decide to count frequencies based on tokens in a corpus or types in a dictionary. The way they choose the measure to use is empirical: after different experiments, they found that the frequency of the most common of the two phonemes in the pair is the measure that leads to the better fit for the model. They found no significant differences in using token or type frequencies. For my replication, I decided to use type frequencies. The factor is described as 'Phoneme

	Estimate	Std. Error	z value	Pr(>z)
(Intercept)	-3.33	0.43	-7.82	*<0.001
MinPairs	-0.34	0.92	-3.65	*<0.001
Phoneme Probability	0.40	0.23	1.77	0.076
NoMP	-0.51	0.50	-1.02	0.306
Phon. prob. by NoMP	1.53	0.72	2.11	*0.035

Table D.1: Wedel et al's Mixed Effect model.

Probability' (which is equivalent to phoneme frequency).

Since they note that many pairs show no minimal pairs at all, they add to the model a dummy variable to encode for the absence of minimal pairs. The dummy variable has '1' when no minimal pairs are present in the analysis. This is equivalent of keeping the pairs for which no minimal pairs are present separated in the analysis.

The setup of a Mixed Effect model allows them to add each merger comparison in the random effect structure, as a random intercept. Since it is natural to have large numbers by comparing vowel contrasts in a whole corpus and low numbers by comparing only the vowel contrasts that appear before /l/ or /r/, having a random effect for merger group takes into account the variability among classes, making sure that the differences across classes are not biasing the analysis.

The model shows that the number of minimal pairs is the best predictor of a merger (Table D.1). Phoneme probability is marginally significant, and only in the absence of minimal pairs: this means that in all the cases in which minimal pairs are not present, a phonemic contrast which involves a frequent phoneme is more likely to merge than a contrast that involves less common phonemes.

D.2   A replication

I replicated the previous analysis on the data in Table 4.4, using the function *glmer* in R (R Core Team, 2017). The number of minimal pairs was log-transformed, because it increases exponentially with respect to the amount of words considered.[1] Frequency was

---

[1]I had to increase each count by one to avoid calculating log over zeros. The measure was not log-transformed in Wedel et al.'s analysis.

	Estimate	Std. Error	z value	Pr(>z)
(Intercept)	-0.41	0.55	-0.76	0.449
MinPairs	-0.51	0.24	-2.11	*0.035
Phoneme Probability	0.63	0.51	1.24	0.216
NoMP	0.29	0.78	0.36	0.716
Phon. prob. by NoMP	-0.93	0.59	-1.57	0.117

Table D.2: Replication of Wedel et al. (2013), with Minimal Pairs.

standardized. The results of the new model are in Table D.2. From what we see in the table, the results are largely confirmed. The minimal pair coefficient is significant, and predicts a decrease in the likelihood of a merger as minimal pairs increase. Conversely, phoneme probability increases the likelihood of a merger, but it is not significant.

The other coefficients instead point toward different directions. While in Wedel et al. (2013)'s experiment, in the absence of minimal pairs, frequency was increasing the likelihood of a merger, our model shows the opposite: if there are no minimal pairs at all, it is the least frequent phonemes that tend to disappear. The coefficient, however, does not appear to be significant.

The fact that the results of Wedel et al. (2013) are robust to corpora and phoneme selection is a strong indication of the presence of a functional load effect.

D.3   Entropy

I also tried to replicate the analysis using the entropy loss coefficient instead of the minimal pairs one. The entropy coefficient was standardized. Merger group is still the only variable in the random effect structure. The results are in Table D.3.

The coefficient for entropy has the the expected sign (a higher entropy loss is negatively correlated with the presence of mergers), but the effect is not significant (z=-1.27), even though it is not much different than the effect we detected for minimal pairs. This confirms Wedel et al. (2013)'s finding that the number of minimal pairs is better correlated with the presence or the absence of mergers than entropy loss.

	Estimate	Std. Error	z value	Pr(>z)
(Intercept)	-0.88	0.62	-1.40	0.161
Entropy	-0.41	0.33	-1.23	0.220
Phoneme Probability	0.94	0.57	1.65	0.100
NoMP	-0.46	0.67	-0.69	0.490
Phon. prob. by NoMP	-1.14	0.63	-1.81	0.071

Table D.3: Replication of Wedel et al. (2013), with Entropy loss.

	Estimate	Std. Error	z value	Pr(>z)
(Intercept)	-0.07	0.90	-0.076	0.94
MinPairs	-0.99	0.34	-2.88	*0.004
Phoneme Probability	-0.07	0.41	-0.18	0.856
NoMP	1.10	1.15	0.96	0.716

Table D.4: Analysis limited to unconditioned mergers.

D.4   Conditioned vs Unconditioned mergers

As a final experiment, I ran two other models that keep conditioned mergers separated from unconditioned mergers, under the hypothesis, suggested at the end of Chapter 3, that they are two different phenomena, and that minimal pairs are only correlated with the latter.

Since by splitting the data into two different datasets we end up with less data points, I decided to remove the interaction between Phoneme Probability and the dummy variable encoding the presence of minimal pairs, in order to make the models converge.

As expected, the coefficient associated with minimal pairs is significant if we run the model on unconditioned mergers only (Table D.4), in spite of the loss of statistical power due to the smaller sample size.[2] On the other hand, the coefficient becomes not significant if we run the model on conditioned mergers, and it is associated with the wrong sign (Table D.5). This negative result could be due to the small sample size (only 5 merger groups, and 12 sound changes, while we have 11 and 16 for the unconditioned set, respectively). However, the fact the sign is positive suggests that the number of minimal pairs is not correlated with the absence of mergers.

---

[2]Another possibility is that it is the absence of the interaction term the reason for the increase of the effect size, but as I said in the previous footnote, removing the term does not substantially improve the model.

	Estimate	Std. Error	z value	Pr(>z)
(Intercept)	-0.91	0.70	-1.29	0.197
MinPairs	0.51	0.49	1.042	0.298
Phoneme Probability	0.24	0.36	0.68	0.497
NoMP	-1.41	1.33	-1.056	0.291

Table D.5: Analysis limited to conditioned mergers.

The models validate the hypothesis that only unconditioned mergers are blocked by the presence of minimal pairs.

Bibliography

Akaike, Htrotugu. 1973. Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika* 60(2). 255–265.

Allen, George D. 1985. How the young French child avoids the pre-voicing problem for word-initial voiced stops. *Journal of Child Language* 12(1). 37–46.

Altvater-Mackensen, Nicole & Paula Fikkert. 2010. The acquisition of the stop-fricative contrast in perception and production. *Lingua* 120(8). 1898–1909.

Amayreh, Mousa M. & Alice T. Dyson. 2000. Phonetic inventories of young Arabic-speaking children. *Clinical Linguistics & Phonetics* 14(3). 193–215.

Arlt, Phyllis B. & Marjorie T. Goodban. 1976. A comparative study of articulation acquisition as based on a study of 240 normals, aged three to six. *Language, Speech, and Hearing Services in Schools* 7(3). 173–180.

Aronoff, Mark. 1976. *Word formation in generative grammar*. Cambridge, MA: MIT Press.

Athanasopoulou, Angeliki & Irene Vogel. 2011. The classification of Greek fricatives with cepstral coefficients. *The Journal of the Acoustical Society of America* 129(4). 2422.

Atkinson, Quentin D. 2011. Phonemic diversity supports a serial founder effect model of language expansion from Africa. *Science* 332(6027). 346–349.

Baayen, R. Harald. 1991. A stochastic process for word frequency distributions. In *Proceedings of the 29th annual meeting on Association for Computational Linguistics*, 271–278. Association for Computational Linguistics.

Baayen, R. Harald, Richard Piepenbrock & Leon Gulikers. 1996. CELEX2. *Linguistic Data Consortium, Philadelphia* .

Babinski, Sarah & Claire Bowern. 2018. Mergers in Bardi: Contextual probability and predictors of sound change. *Linguistics Vanguard* 4(s2).

Baerman, Matthew. 2011. Defectiveness and homophony avoidance. *Journal of Linguistics* 47(1). 1–29.

Banavar, Jayanth R., Paolo De Los Rios, Alessandro Flammini, Neal S. Holter & Amos Maritan. 2004. Scale-free behavior and universality in random fragmentation and aggregation. *Physical Review E* 69(3). 036123.

Baranowski, Maciej Andrzej. 2006. *Phonological variation and change in the dialect of Charleston, South Carolina*: University of Pennsylvania dissertation.

Bard, Ellen G. 1991. Competition, lateral inhibition, and frequency: Comments on the chapters of Frauenfelder and Peeters, Marslen-Wilson, and others. In Gerry T. M. Altmann (ed.), *Cognitive models of speech processing*, 185–210. Cambridge, MA: MIT Press.

Bateman, Nicoleta. 2007. *A crosslinguistic investigation of palatalization*: University of California San Diego dissertation.

Bates, Douglas. 2007. The lme4 package. *R package version* 2(1). 74.

Bauer, Laurie. 2001. *Morphological productivity*. Cambridge University Press.

Bauer, Laurie, Paul Warren, Dianne Bardsley, Marianna Kennedy & George Major. 2007. New Zealand English. *Journal of the International Phonetic Association* 37(1). 97–102.

Baum, Shari R. & James C. McNutt. 1990. An acoustic analysis of frontal misarticulation of /s/ in children. *Journal of Phonetics* 18(1). 51–63.

Baxter, Gareth J., Richard A. Blythe, William Croft & Alan J. McKane. 2009. Modeling language change: an evaluation of Trudgill's theory of the emergence of New Zealand English. *Language Variation and Change* 21(2). 257–296.

Beckman, Mary E., Kiyoko Yoneyama & Jan Edwards. 2003. Language-specific and language-universal aspects of lingual obstruent productions in Japanese-acquiring children. *Journal of the Phonetic Society of Japan* 7(2). 18–28.

Beers, Wilhelmina. 1995. *The phonology of normally developing and language-impaired children*: Universiteit van Amsterdam dissertation.

Beguš, Gašper. 2018. *Unnatural phonology: A synchrony-diachrony interface approach*: Harvard University dissertation.

Behr, Wolfgang. 2015. Discussion 6: G. Sampson, 'A Chinese phonological enigma: four comments'. *Journal of Chinese Linguistics* 43(2). 719–732.

Bentz, Christian. 2018. *Adaptive languages*. Berlin: De Gruyter Mouton.

Benzing, Johannes. 1955. *Die tungusischen Sprachen: Versuch einer vergleichenden Grammatik*. Mainz: Akademie der Wissenschaften und der Literatur.

Berg, Thomas & Ulrich Schade. 1992. The role of inhibition in a spreading-activation model of language production. I. The psycholinguistic perspective. *Journal of Psycholinguistic Research* 21(6). 405–434.

Blevins, Juliette. 2004. *Evolutionary phonology: The emergence of sound patterns*. Cambridge University Press.

Blevins, Juliette. 2006. New Perspectives on English Sound Patterns: "Natural" and "Unnatural" in Evolutionary Phonology. *Journal of English Linguistics* 34(1). 6–25.

Blevins, Juliette & Andrew Wedel. 2009. Inhibited sound change: An evolutionary approach to lexical competition. *Diachronica* 26(2). 143–183.

Bodmer, Walter F. & Luigi L. Cavalli-Sforza. 1968. A migration matrix model for the study of random genetic drift. *Genetics* 59(4). 565.

Boland, Julie E. & Allison Blodgett. 2001. Understanding the constraints on syntactic generation: Lexical bias and discourse congruency effects on eye movements. *Journal of Memory and Language* 45(3). 391–411.

Borodovsky, M. Yu & S. M. Gusein-Zade. 1989. A general rule for ranged series of codon frequencies in different genomes. *Journal of Biomolecular Structure and Dynamics* 6(5). 1001–1012.

Bouchard-Côté, Alexandre, David Hall, Thomas L. Griffiths & Dan Klein. 2013. Automated reconstruction of ancient languages using probabilistic models of sound change. *Proceedings of the National Academy of Sciences* 110(11). 4224–4229.

de Boysson-Bardies, Bénédicte, Pierre Hallé, Laurent Sagart & Catherine Durand. 1989. A crosslinguistic investigation of vowel formants in babbling. *Journal of Child Language* 16(1). 1–17.

de Boysson-Bardies, Bénédicte & Marilyn May Vihman. 1991. Adaptation to language: Evidence from babbling and first words in four languages. *Language* 67(2). 297–319.

Bybee, Joan L. 2002. Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change* 14(3). 261–290.

Bybee, Joan L. 2017. Grammatical and lexical factors in sound change: A usage-based approach. *Language Variation and Change* 29(3). 273.

Calabrese, Andrea & Leo Wetzels. 2009. *Loan phonology*. Amsterdam: John Benjamins.

Campbell, Lyle. 1996. On sound change and challenges to regularity. In Malcolm Ross & Mark Durie (eds.), *The comparative method reviewed: Regularity and irregularity in language change*, 72–89. Oxford University Press.

Campbell, Lyle. 2013. *Historical linguistics*. Edinburgh University Press.

Caplan, Spencer, Jordan Kodner & Charles Yang. 2020. Miller's monkey updated: Communicative efficiency and the statistics of words in natural language. *Cognition* 205. 104466.

Carlson, Matthew T., Morgan Sonderegger & Max Bane. 2014. How children explore the phonological network in child-directed speech: A survival analysis of children's first word productions. *Journal of Memory and Language* 75. 159–180.

Carter, David M. 1987. An information-theoretic analysis of phonetic dictionary access. *Computer Speech & Language* 2(1). 1–11.

Cavalli-Sforza, Luigi L. & Anthony W.F. Edwards. 1967. Phylogenetic analysis. Models and estimation procedures. *American Journal of Human Genetics* 19(3.1). 233–257.

Ceolin, Andrea, Cristina Guardiano, Monica Alexandrina Irimia & Giuseppe Longobardi. 2020. Formal syntax and deep history. *Frontiers in Psychology* 11. 2384.

Ceolin, Andrea & Ollie Sayeed. 2019. Modeling markedness with a split-and-merger model of sound change. In Nina Tahmasebi, Lars Borin, Adam Jatowt & Yang Xu (eds.), *Proceedings of the 1st International Workshop on Computational Approaches to Historical Language Change*, ACL.

Charles-Luce, Jan & Paul A. Luce. 1990. Similarity neighbourhoods of words in young children's lexicons. *Journal of Child Language* 17(1). 205–215.

Cheon, Sang Yee & Victoria B. Anderson. 2008. Acoustic and perceptual similarities between English and Korean sibilants: Implications for second language acquisition. *Korean Linguistics* 14(1). 41–64.

Chevrie-Muller, C. & M.T. Lebreton. 1973. Étude de la réalization des consonnes au cours d'une épreuve de mots, sur des groupes d'enfants de 3 ans et 5 ans 1/2. *Revue de layngologie* 94. 109–152.

Chomsky, Noam & Robert Berwick. 2016. *Why only us.* Cambridge, MA: MIT Press.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.

Coady, Jeffry A. & Richard N. Aslin. 2003. Phonological neighbourhoods in the developing lexicon. *Journal of Child Language* 30(2). 441–469.

Cohen Priva, Uriel. 2012. *Sign and signal: Deriving linguistic generalizations from information utility*: Stanford University dissertation.

Cohen Priva, Uriel. 2017. Informativity and the actuation of lenition. *Language* 93(3). 569–597.

Creanza, Nicole, Merritt Ruhlen, Trevor J. Pemberton, Noah A. Rosenberg, Marcus W. Feldman & Sohini Ramachandran. 2015. A comparison of worldwide phonemic and genetic variation in human populations. *Proceedings of the National Academy of Sciences* 112(5). 1265–1272.

Crystal, David. 1976. *Child language, learning and linguistics: an overview for the teaching and therapeutic professions*. London: E. Arnold.

Cui, Aletheia. 2020. *The emergence of phonological categories*: University of Pennsylvania dissertation.

Cychosz, Margaret. 2017. Functional load and frequency predict consonant emergence across five languages. *UC Berkeley PhonLab Annual Report* 13(1).

Dautriche, Isabelle, Kyle Mahowald, Edward Gibson, Anne Christophe & Steven T. Piantadosi. 2017. Words cluster phonetically beyond phonotactic regularities. *Cognition* 163. 128–145.

Davies, Mark. 2009. The 385+ million word Corpus of Contemporary American English (1990–2008+): Design, architecture, and linguistic insights. *International Journal of Corpus Linguistics* 14(2). 159–190.

Davis, Katharine. 1995. Phonetic and phonological contrasts in the acquisition of voicing: Voice onset time production in Hindi and English. *Journal of Child Language* 22(2). 275–305.

Dell, Gary S. 1986. A spreading-activation theory of retrieval in sentence production. *Psychological Review* 93(3). 283.

Dinkin, Aaron J. 2016. Phonological transfer as a forerunner of merger in Upstate New York. *Journal of English Linguistics* 44(2). 162–188.

Dinnsen, Daniel A. 1992. Variation in developing and fully developed phonetic inventories. In Charles A. Ferguson, Lisa Menn & Carol Stoel-Gammon (eds.), *Phonological development: Models, research, implications*, 191–210. Timonium, MD: York Press.

Doherty, Martin J. 2004. Children's difficulty in learning homonyms. *Journal of Child Language* 31(1). 203–214.

Donohue, Mark & Johanna Nichols. 2011. Does phoneme inventory size correlate with population size? *Linguistic Typology* 15(2). 161–170.

Dresher, B. Elan. 2009. *The contrastive hierarchy in phonology*. Cambridge University Press.

Dresher, B. Elan. 2011. The phoneme. In Marc Van Oostendorp, Colin J. Ewen, Elizabeth Hume & Keren Rice (eds.), *The Blackwell Companion to Phonology*, 1–26. Hoboken, NJ: John Wiley & Sons.

Dresher, B. Elan & Xi Zhang. 2005. Contrast and phonological activity in Manchu vowel systems. *The Canadian Journal of Linguistics/La revue canadienne de linguistique* 50(1). 45–82.

Duanmu, San. 2007. *The phonology of standard Chinese*. Oxford University Press.

Dyen, Isidore, Joseph B. Kruskal & Paul Black. 1992. An Indoeuropean classification: A

lexicostatistical experiment. *Transactions of the American Philosophical society* 82(5). iii–132.

Edwards, Jan & Mary E. Beckman. 2008a. Some cross-linguistic evidence for modulation of implicational universals by language-specific frequency effects in phonological development. *Language learning and development* 4(2). 122–156.

Edwards, Jan & Mary E. Beckman. 2008b. Methodological questions in studying consonant acquisition. *Clinical Linguistics & Phonetics* 22(12). 937–956.

Edwards, Jan, Mary E. Beckman & Benjamin Munson. 2004. The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language, and Hearing Research* 47(2). 421–436.

Eilers, Rebecca E. & Fred D. Minifie. 1975. Fricative discrimination in early infancy. *Journal of Speech and Hearing Research* 18(1). 158–167.

Eilers, Rebecca E. & D. Kimbrough Oller. 1976. The role of speech discrimination in developmental sound substitutions. *Journal of Child Language* 3(3). 319–329.

Eilers, Rebecca E., Wesley R. Wilson & John M. Moore. 1977. Developmental changes in speech discrimination in infants. *Journal of Speech and Hearing Research* 20(4). 766–780.

Eimas, Peter D. 1974. Auditory and linguistic processing of cues for place of articulation by infants. *Perception & Psychophysics* 16(3). 513–521.

Eimas, Peter D. 1975. Auditory and phonetic coding of the cues for speech: Discrimination of the [r-l] distinction by young infants. *Perception & Psychophysics* 18(5). 341–347.

Eimas, Peter D. & Joanne L. Miller. 1980. Discrimination of information for manner of articulation. *Infant Behavior & Development* 3. 367–375.

Eimas, Peter D., Einar R. Siqueland, Peter W. Jusczyk & James Vigorito. 1971. Speech perception in infants. *Science* 171(3968). 303–306.

Endress, Ansgar D. & Marc D. Hauser. 2011. The influence of type and token frequency on the acquisition of affixation patterns: Implications for language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 37(1). 77–95.

Eychenne, Julien & Tae-Yeoub Jang. 2015. On the merger of Korean mid front vowels: Phonetic and phonological evidence. *Journal of the Korean Society of Speech Sciences* 7(2). 119–129.

Fagyal, Zsuzsanna, Douglas Kibbee & Frederic Jenkins. 2006. *French: A linguistic introduction*. Cambridge University Press.

Fedzechkina, Maryia, T. Florian Jaeger & Elissa L. Newport. 2012. Language learners restructure their input to facilitate efficient communication. *Proceedings of the National Academy of Sciences* 109(44). 17897–17902.

Felsenstein, Joseph. 1976. The theoretical population genetics of variable selection and migration. *Annual Review of Genetics* 10(1). 253–280.

Fenson, Larry, Philip S. Dale, Steven J. Reznick, Elizabeth Bates, Donna J. Thal, Stephen J. Pethick, Michael Tomasello, Carolyn B. Mervis & Joan Stiles. 1994. Variability in early communicative development. *Monographs of the Society for Research in Child Development* 59(5). i–185.

Fikkert, Paula & Claartje Levelt. 2008. How does Place fall into Place? In Peter Avery, Elan Dresher & Keren Rice (eds.), *Contrast in phonology: Theory, perception, acquisition*, 231–268.

Finegan, Edward. 2014. *Language: Its structure and use*. Boston: Cengage Learning.

Fisher, Ronald A. 1930. *The genetical theory of natural selection: a complete variorum edition*. Oxford University Press.

Fitch, W. Tecumseh, Marc D. Hauser & Noam Chomsky. 2005. The evolution of the language faculty: clarifications and implications. *Cognition* 97(2). 179–210.

Flemming, Edward. 2004. Contrast and perceptual distinctiveness. In Bruce Hayes, Robert Kirchner & Donca Steriade (eds.), *Phonetically based phonology*, 232–276. Cambridge University Press.

Gahl, Susanne. 2008. Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language* 84(3). 474–496.

Gandour, Jack, Soranee H. Petty, Rochana Dardarananda, Sumalee Dechongkit & Sunee Mukngoen. 1986. The acquisition of the voicing contrast in Thai: A study of voice onset time in word-initial stop consonants. *Journal of Child Language* 13(3). 561–572.

Gilliéron, Jules. 1918. *Généalogie des mots qui désignent l'abeille d'après l'atlas linguistique de la france.* Paris: Champion.

Good, I. J. 1969. Statistics of language. In A. R. Meetham & Richard A. Hudson (eds.), *Encyclopaedia of linguistics, information and control*, 567–581. Oxford: Pergamon Press.

Gordon, Elizabeth, Lyle Campbell, Jennifer Hay, Margaret Maclagan, Andrea Sudbury & Peter Trudgill. 2004. *New Zealand English: its origins and evolution.* Cambridge University Press.

Graff, Peter Nepomuk Herwig Maria. 2012. *Communicative efficiency in the lexicon*: MIT dissertation.

Greenberg, Joseph. 1959. A method for measuring functional yield as applied to tone in African languages. In Richard S. Harrell (ed.), *Report of the Tenth Annual Round Table Meeting on Linguistics and Language Studies*, vol. 12, 7–16.

Greenberg, Joseph. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In Joseph Greenberg (ed.), *Universals of language*, 73–113. Cambridge, MA: MIT Press.

Grimm, Jacob. 1848. *Geschichte der deutschen sprache*, vol. 1. Leipzig: Hirzel.

Grunwell, Pamela. 1982. *Clinical phonology.* London: Croom Helm.

Gurevich, Naomi. 2004. *Lenition and contrast: The functional consequences of certain phonetically conditioned sound changes.* New York: Garland.

Gusein-Zade, S. M. 1987. On the frequency of meeting of key words and on other ranked series. *Science-Technical Information, Series 2: Information Processes and Systems* 2. 28–32.

Gylfadóttir, Duna. 2018. *The effective borrowing of a phonemic contrast*: University of Pennsylvania dissertation.

Halle, Morris. 1970. Is Kabardian a vowel-less language? *Foundations of Language* 6(1). 95–103.

Hallé, Pierre A. & Bénédicte de Boysson-Bardies. 1994. Emergence of an early receptive lexicon: Infants' recognition of words. *Infant Behavior and Development* 17(2). 119–129.

Hallé, Pierre A. & Bénédicte de Boysson-Bardies. 1996. The format of representation of recognized words in infants' early receptive lexicon. *Infant Behavior and Development* 19(4). 463–481.

Harris, James W. 1967. *Spanish phonology*: MIT dissertation.

Hauser, Marc D., Noam Chomsky & W. Tecumseh Fitch. 2002. The faculty of language: What is it, who has it, and how did it evolve? *Science* 298(5598). 1569–1579.

Hay, Jennifer & Laurie Bauer. 2007. Phoneme inventory size and population size. *Language* 83(2). 388–400.

Hay, Jennifer, Janet B. Pierrehumbert & Mary E. Beckman. 2004. Speech perception, well-formedness and lexical frequency. In John Local, Richard Ogden & Rosalind Temple (eds.), *Papers in Laboratory Phonology VI*, Cambridge University Press.

Hayes, Bruce. 2011. *Introductory phonology.* Hoboken, NJ: John Wiley & Sons.

Hernando Cuadrado, Luis Alberto. 1997. El diccionario de autoridades (1726-1739) y su evolución. *Verba* 24. 387–401.

Herzog, Marvin. 1965. *The Yiddish language in northern Poland*: Indiana University dissertation.

Hillenbrand, James. 1983. Perceptual organization of speech sounds by infants. *Journal of Speech, Language, and Hearing Research* 26(2). 268–282.

Hillenbrand, James, Laura A. Getty, Michael J. Clark & Kimberlee Wheeler. 1995. Acoustic characteristics of American English vowels. *The Journal of the Acoustical society of America* 97(5). 3099–3111.

Hillenbrand, James, Fred D. Minifie & Thomas J. Edwards. 1979. Tempo of spectrum change as a cue in speech-sound discrimination by infants. *Journal of Speech, Language, and Hearing Research* 22(1). 147–165.

Hockett, Charles F. 1955. *A manual of phonology.* Baltimore: Waverly Press.

Hockett, Charles F. 1967. The quantification of functional load. *Word* 23(1-3). 300–320.

Hoenigswald, Henry M. 1960. *Language change and linguistic reconstruction.* University of Chicago Press.

Hoff, Erika. 2009. Language development at an early age: Learning mechanisms and outcomes from birth to five years. *Encyclopedia on early childhood development* 1–5.

Holman, Eric W., Søren Wichmann, Cecil H. Brown, Viveka Velupillai, André Müller & Dik Bakker. 2008. Explorations in automated language classification. *Folia Linguistica* 42(3-4). 331–354.

Honeybone, Patrick. 2016. Are there impossible changes? θ > f but f > θ. *Papers in Historical Phonology* 1. 316–358.

Horák, Alexander, Lucia Gianitsová, Mária Šimková, Martin Šmotlák & Radovan Garabík. 2004. Slovak National Corpus. In Petr Sojka, Ivan Kopeček & Karel Pala (eds.), *Text, Speech and Dialogue*, 89–93. New York: Springer.

Hua, Zhu. 2002. *Phonological development in specific contexts: Studies of Chinese-speaking children*. Bristol: Multilingual Matters.

Hua, Zhu & Barbara Dodd. 2000. The phonological acquisition of Putonghua (modern standard Chinese). *Journal of Child Language* 27(1). 3–42.

Hualde, José Ignacio & Pilar Prieto. 2014. Lenition of intervocalic alveolar fricatives in Catalan and Spanish. *Phonetica* 71(2). 109–127.

Huttenlocher, Daniel P. 1985. Exploiting sequential phonetic constraints in recognizing spoken words. *MIT AI Lab Memo 867* .

Ingram, David. 1988. The acquisition of word-initial [v]. *Language and Speech* 31(1). 77–85.

Ingram, David. 1989. *First language acquisition: Method, description and explanation.* Cambridge University Press.

Ingram, David. 1999. Phonological acquisition. In Martyn Barrett (ed.), *The development of language*, 73–97. East Sussex, UK: Psychology Press.

Itô, Junko & Armin Mester. 1995. Japanese phonology. In John Goldsmith, Jason Riggle & Alan C. L. Yu (eds.), *The Handbook of Phonological Theory*, 816–838. Oxford: Blackwell.

Jackendoff, Ray & Steven Pinker. 2005. The nature of the language faculty and its implications for evolution of language (Reply to Fitch, Hauser, and Chomsky). *Cognition* 97(2). 211–225.

Jakobson, Roman. 1931. Prinzipien der historischen phonologie. *Travaux du Cercle Linguistique de Prague* 4. 247–267.

Jakobson, Roman. 1941/1968. *Child language, aphasia and phonological universals.* Berlin: De Gruyter.

Jongman, Allard, Ratree Wayland & Serena Wong. 2000. Acoustic characteristics of English fricatives. *The Journal of the Acoustical Society of America* 108(3). 1252–1263.

Jusczyk, Peter W. 1977. Perception of syllable-final stop consonants by 2-month-old infants. *Perception & Psychophysics* 21(5). 450–454.

Jusczyk, Peter W. 1992. Developing phonological categories from the speech signal. In Charles A. Ferguson, Lisa Menn & Carol Stoel-Gammon (eds.), *Phonological development: Models, research, implications*, 17–64. Timonium, MD: York Press.

Jusczyk, Peter W., Mara B. Goodman & Angela Baumann. 1999a. Nine-month-olds' attention to sound similarities in syllables. *Journal of Memory and Language* 40(1). 62–82.

Jusczyk, Peter W., Elizabeth A. Hohne & Angela Bauman. 1999b. Infants' sensitivity to allophonic cues for word segmentation. *Perception & Psychophysics* 61(8). 1465–1476.

Jusczyk, Peter W., Janice Murray & J. Bayly. 1979. Perception of place of articulation in fricatives and stops by infants. Presented at the Biennial meeting of the Society for Research in Child Development, San Francisco .

Kaplan, Abby. 2010. *Phonology shaped by phonetics: The case of intervocalic lenition*: University of California, Santa Cruz dissertation.

Kaplan, Abby. 2011. How much homophony is normal? *Journal of Linguistics* 47(3). 631–671.

Kaplan, Abby. 2015. Discussion 4: A highly improbable data point. *Journal of Chinese Linguistics* 43(2). 710–713.

Kauhanen, Henri. 2017. Neutral change. *Journal of Linguistics* 53(2). 327–358.

Kent, Ray D. 1992. The biology of phonological development. In Charles A. Ferguson, Lisa Menn & Carol Stoel-Gammon (eds.), *Phonological development: Models, research, implications*, 65–90. Timonium, MD: York Press.

Kessler, Brett. 2007. Word similarity metrics and multilateral comparison. In *Proceedings of Ninth Meeting of the ACL Special Interest Group in Computational Morphology and Phonology*, 6–14. Association for Computational Linguistics.

King, Robert D. 1967a. Functional load and sound change. *Language* 43(4). 831–852.

King, Robert D. 1967b. A measure for functional load. *Studia linguistica* 21(1). 1–14.

Kiparsky, Paul. 2016. Labov, sound change, and phonological theory. *Journal of Sociolinguistics* 20(4). 464–488.

Kissine, Mikhail, Hans Van de Velde & Roeland Van Hout. 2003. An acoustic study of standard dutch /v/ , /f/, /z/ and /s/. *Linguistics in the Netherlands* 20(1). 93–104.

Klein, Harriet B. & Elaine K. Altman. 2002. The acquisition of medial /t,d/ allophones in bisyllabic contexts. *Clinical Linguistics & Phonetics* 16(3). 215–232.

Kochetov, Alexei. 2011. Palatalization. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume Hume & Keren Rice (eds.), *The Blackwell Companion to Phonology*, 1–25. Hoboken, NJ: John Wiley & Sons.

Kodner, Jordan. 2019. Estimating child linguistic experience from historical corpora. *Glossa: a journal of general linguistics* 4(1). 122.

Kodner, Jordan. 2020. *Language acquisition in the past*: University of Pennsylvania dissertation.

Kodner, Jordan & Christopher Cerezo Falco. 2018. A framework for representing language acquisition in a population setting. In *Proceedings of the 56th annual meeting of the association for computational linguistics (volume 1: Long papers)*, vol. 1, 1149–1159.

Kong, Eun Jong, Mary E. Beckman & Jan Edwards. 2012. Voice onset time is necessary but not always sufficient to describe acquisition of voiced stops: The cases of Greek and Japanese. *Journal of Phonetics* 40(6). 725–744.

Kornai, Andras. 1998. Analytic models in phonology. In Jacques Durand & Bernard Laks (eds.), *The organization of phonology: Constraints, levels and representations*, 395–418. Oxford University Press.

Krajčovič, Rudolf. 1988. *Vývin slovenského jazyka a dialektológia*. Slovenské pedagogické nakl.

Král', Ábel. 1988. *Pravidlá slovenskej výslovnosti*. Slovenské pedagogické nakladateľstvo.

Kroch, Anthony S. 1989. Reflexes of grammar in patterns of language change. *Language Variation and Change* 1(3). 199–244.

Kučera, Henry. 1963. Entropy, Redundancy and Functional load in Russian and Czech. In *American Contributions to the Fifth International Conference of Slavists*, 191–219. The Hague: Mouton.

Kuhl, Patricia K. 1979. Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. *The Journal of the Acoustical Society of America* 66(6). 1668–1679.

Kuhl, Patricia K. 1980. Perceptual constancy for speech-sound categories in early infancy. In Grace H. Yeni-Komshian, James F. Kavanagh & Charles A. Ferguson (eds.), *Child phonology, Vol. 2: Perception*, 41–66. New York: Academic Press.

Kuhl, Patricia K. 1983. Perception of auditory equivalence classes for speech in early infancy. *Infant Behavior and Development* 6(2-3). 263–285.

Kuhl, Patricia K. & James D. Miller. 1975. Speech perception in early infancy: Discrimination of speech-sound categories. *The Journal of the Acoustical Society of America* 58(S1). S56–S56.

Kuhl, Patricia K. & James D. Miller. 1982. Discrimination of auditory target dimensions in the presence or absence of variation in a second dimension by infants. *Perception & Psychophysics* 31(3). 279–292.

Kuhl, Patricia K., Erica Stevens, Akiko Hayashi, Toshisada Deguchi, Shigeru Kiritani & Paul Iverson. 2006. Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental science* 9(2). F13–F21.

Kuipers, Aert Hendrik. 1960. *Phoneme and morpheme in Kabardian.* The Hague: Mouton.

Kyparissiadis, Antonios, Walter J. B. van Heuven, Nicola J. Pitchford & Timothy Ledgeway. 2017. GreekLex 2: A comprehensive lexical database with part-of-speech, syllabic, phonological, and stress information. *PloS one* 12(2).

Labov, William. 1966. *The social stratification of English in New York City.* Cambridge University Press.

Labov, William. 1987. The overestimation of functionalism. In René Dirven & Vilém Fried (eds.), *Functionalism in linguistics*, 311–332. Amsterdam: John Benjamins.

Labov, William. 1994. *Principles of linguistic change: Internal factors.* Oxford: Blackwell.

Labov, William. 2010. *Principles of linguistic change: Cognitive and cultural factors.* Hoboken, NJ: Wiley & Blackwell.

Labov, William, Sharon Ash & Charles Boberg. 2008. *The Atlas of North American English: Phonetics, phonology and sound change.* Berlin: De Gruyter.

Labov, William, Mark Karen & Corey Miller. 1991. Near-mergers and the suspension of phonemic contrast. *Language Variation and Change* 3(1). 33–74.

Lakoff, Robin. 1972. Another look at drift. In Robert P. Stockwell & Ronald K. S. Macaulay (eds.), *Linguistic change and generative theory*, 172–198. Bloomington: Indiana University Press.

Lass, Roger. 1981. *On explaining language change.* Cambridge University Press.

Lass, Roger. 1992. Phonology and morphology. In Richard M. Hogg, Norman F. Blake, Roger Lass & R. W. Burchfield (eds.), *The Cambridge history of the English language*, 1066–1476. Cambridge University Press.

Lass, Roger. 1997. *Historical linguistics and language change.* Cambridge University Press.

Levelt, Clara. 1996. Consonant-vowel interactions in child language. In Barbara Bernhardt, John Gilbert & David Ingram (eds.), *Proceedings of the UBC International Conference on Phonological Acquisition*, 209–239.

Levitt, Andrea, Peter W. Jusczyk, Janice Murray & Guy Carden. 1988. Context effects in two-month-old infants' perception of labiodental/interdental fricative contrasts. *Journal of Experimental Psychology: Human Perception and Performance* 14(3). 361–368.

Li, Fangfang. 2008. *The phonetic development of voiceless sibilant fricatives in English, Japanese and Mandarin Chinese*: The Ohio State University dissertation.

Li, Fangfang, Jan Edwards & Mary E. Beckman. 2009. Contrast and covert contrast: The phonetic development of voiceless sibilant fricatives in English and Japanese toddlers. *Journal of Phonetics* 37(1). 111–124.

Lieberman, Erez, Jean-Baptiste Michel, Joe Jackson, Tina Tang & Martin A. Nowak. 2007. Quantifying the evolutionary dynamics of language. *Nature* 449(7163). 713–716.

Lightfoot, David. 2006. *How new languages emerge*. Cambridge University Press.

Liljencrants, Johan & Björn Lindblom. 1972. Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language* 48(4). 839–862.

Lleó, Conxita. 2002. The role of markedness in the acquisition of complex prosodic structures by German-Spanish bilinguals. *International Journal of Bilingualism* 6(3). 291–313.

Locke, John L. 1983. *Phonological acquisition and change*. New York: Academic Press.

Locke, John L. 1988. The sound shape of early lexical representations. In Michael D. Smith & John L. Locke (eds.), *The emergent lexicon: The child's development of a linguistic vocabulary*, 3–22. San Diego: Academic Press.

Longobardi, Giuseppe, Cristina Guardiano, Giuseppina Silvestri, Alessio Boattini & Andrea Ceolin. 2013. Toward a syntactic phylogeny of modern Indo-European languages. *Journal of Historical Linguistics* 3(1). 122–152.

Luce, Paul A. & David B. Pisoni. 1998. Recognizing spoken words: The neighborhood activation model. *Ear and Hearing* 19(1). 1–36.

MacKain, Kristine S. 1982. Assessing the role of experience on infants' speech discrimination. *Journal of Child Language* 9(3). 527–542.

Macken, Marlys A. 1979. Developmental reorganization of phonology: A hierarchy of basic units of acquisition. *Lingua* 49(1). 11–49.

Macken, Marlys A. 1980a. The child's lexical representation: the 'puzzle-puddle-pickle'evidence. *Journal of Linguistics* 16(1). 1–17.

Macken, Marlys A. 1980b. Aspects of the acquisition of stop systems: A cross-linguistic perspective. In Grace H. Yeni-Komshian, James F. Kavanagh & Charles A. Ferguson (eds.), *Child phonology, Vol. 1: Production*, 143–168. New York: Academic Press.

Macken, Marlys A. & David Barton. 1980a. The acquisition of the voicing contrast in English: A study of voice onset time in word-initial stop consonants. *Journal of Child Language* 7(1). 41–74.

Macken, Marlys A. & David Barton. 1980b. The acquisition of the voicing contrast in Spanish: A phonetic and phonological study of word-initial stop consonants. *Journal of Child Language* 7(3). 433–458.

Macklin-Cordes, Jayden L. & Erich R. Round. 2020. Re-evaluating phoneme frequencies. arXiv preprint:2006.05206 .

MacNeilage, Peter F. & Barbara L. Davis. 1993. Motor explanations of babbling and early speech patterns. In Bénédicte de Boysson-Bardies, Scania de Schonen, Peter W. Jusczyk, Peter McNeilage & John Morton (eds.), *Developmental neurocognition: Speech and face processing in the first year of life*, 341–352. New York: Springer.

MacNeilage, Peter F., Barbara L. Davis, Ashlynn Kinney & Christine L. Matyear. 2000.

The motor core of speech: A comparison of serial organization patterns in infants and languages. *Child Development* 71(1). 153–163.

MacWhinney, Brian. 1978. The acquisition of morphophonology. *Monographs of the Society for Research in Child Development* 43(1/2). 1–123.

MacWhinney, Brian. 2014. *The CHILDES project: Tools for analyzing talk, Volume II: The database*. Mahwah, NJ: Erlbaum.

Maekawa, Kikuo. 2018. Weakening of stop articulation in Japanese voiced plosives. *Journal of the Phonetic Society of Japan* 22(1). 21–34.

Mahowald, Kyle, Isabelle Dautriche, Edward Gibson & Steven T. Piantadosi. 2018. Word forms are structured for efficient use. *Cognitive Science* 42(8). 3116–3134.

Mandelbrot, Benoit. 1953. An informational theory of the statistical structure of language. *Communication theory* 84. 486–502.

Martin, Andrew T. 2007. *The evolving lexicon*: University of California, Los Angeles dissertation.

Martinet, André. 1955. *Economie des changements phonétiques*. Berne: Francke.

Mathesius, Vilém. 1929. La structure phonologique du lexique du tchèque moderne. *Travaux du Cercle linguistique de Prague* 1. 67–84.

Maye, Jessica & LouAnn Gerken. 2000. Learning phonemes without minimal pairs. In S. C. Howell, S. A. Fish & T. Keith-Lucas (eds.), *Proceedings of BUCLD 24*, vol. 2, 522–533. Somerville, MA: Cascadilla Press.

Maye, Jessica, Janet F. Werker & LouAnn Gerken. 2002. Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* 82(3). B101–B111.

Mazzocco, Michèle M.M. 1997. Children's interpretations of homonyms: A developmental study. *Journal of Child Language* 24(2). 441–467.

McGuire, Grant & Molly Babel. 2012. A cross-modal account for synchronic and diachronic patterns of /f/ and /θ/ in English. *Laboratory Phonology* 3(2). 251–272.

McLeod, Sharynne & Kathryn Crowe. 2018. Children's consonant acquisition in 27 languages: A cross-linguistic review. *American Journal of Speech-Language Pathology* 27(4). 1546–1571.

Menn, Lise. 1983. Development of articulatory, phonetic, and phonological capabilities. In Brian Butterworth (ed.), *Language production*, 3–50. New York: Academic Press.

Metsala, Jamie L. & Amanda C. Walley. 1998. Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability. In Jamie L. Metsala & Linnea C. Ehri (eds.), *Word recognition in beginning literacy*, 89–120. Hillsdale, NJ: Erlbaum.

Meyer-Benfey, Heinrich. 1901. Über den Ursprung der germanischen Lautverschiebung. *Zeitschrift für deutsches Altertum und deutsche Litteratur* 45. 101–128.

Miller, Cynthia L. & Philip A. Morse. 1976. The "heart" of categorical speech discrimination in young infants. *Journal of Speech and Hearing Research* 19(3). 578–589.

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

Moffitt, Alan R. 1971. Consonant cue perception by 20- to 24-week-old infants. *Child Development* 42(3). 717–731.

Möhle, Martin & Serik Sagitov. 2001. A classification of coalescent processes for haploid exchangeable population models. *The Annals of Probability* 29(4). 1547–1562.

Mok, Peggy Pik-Ki & Peggy Wai-Yi Wong. 2010. Perception of the merging tones in Hong Kong Cantonese: Preliminary data on monosyllables. In *Speech Prosody 2010-Fifth International Conference*, Chicago: International Speech Communication Association.

Mondon, Jean-François. 2009. *The nature of homophony and its effects on diachrony and synchrony*: University of Pennsylvania dissertation.

Moran, Steven, Daniel McCloy & Richard Wright. 2012. Revisiting the population vs phoneme-inventory correlation. *Language* 88(4). 877–893.

Morley, Rebecca L. 2015. Can phonological universals be emergent?: Modeling the space of sound change, lexical distribution, and hypothesis selection. *Language* 91(2). e40–e70.

Moulton, William G. 1962. Dialect geography and the concept of phonological space. *Word* 18(1-3). 23–32.

Munson, Benjamin & Nancy P. Solomon. 2004. The effect of phonological neighborhood density on vowel articulation. *Journal of Speech, Language, and Hearing Research* 47(5). 1048–1058.

New, Boris, Christophe Pallier, Marc Brysbaert & Ludovic Ferrand. 2004. Lexique 2: A new French lexical database. *Behavior Research Methods, Instruments, & Computers* 36(3). 516–524.

Newberry, Mitchell G., Christopher A. Ahern, Robin Clark & Joshua B. Plotkin. 2017. Detecting evolutionary forces in language change. *Nature* 551(7679). 223.

Newmeyer, Frederick J. 2003. Formal and functional motivation for language change. In Raymond Hickey (ed.), *Motives for language change*, 16–36. Cambridge University Press.

Nichols, Stephen & Henri Kauhanen. 2019. Vowel-pair rank–frequency distributions are polylogarithmic. Presented at LAGB, Queen Mary University of London .

Nicolaidis, Katerina, Jan Edwards, Mary Beckman & Georgios Tserdanelis. ???? Acquisition of lingual obstruents in Greek. In Georgia Katsimali, Alexis Kalokairinos, Elena Anagnostopoulou & Ioanna Kappa (eds.), *Proceedings of the 6th International Conference of Greek Linguistics*, Rethymno: Linguistics Lab.

Nishimura, Bensaku. 1980. Emergence of functional articulation disorders. In David Ingram, Fred C. C. Peng & Philip Dale (eds.), *Proceedings of the first international congress for the study of child language*, 439–447. Lanham, MD: University Press of America.

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

Nowak, Martin A., David C. Krakauer & Andreas Dress. 1999. An error limit for the evolution of language. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 266(1433). 2131–2136.

Ohala, John J. 1983. The origin of sound patterns in vocal tract constraints. In Peter F. MacNeilage (ed.), *The production of speech*, 189–216. New York: Springer.

Olmsted, D. L. 1971. *Out of the Mouth of Babes: Earliest Stages in Language Learning*. The Hague: Mouton.

Osthoff, Hermann & Karl Brugmann. 1880. *Morphologische Untersuchungen auf dem Gebiete der indogermanischen Sprachen*, vol. 3. Leipzig: S. Hirzel.

Oswalt, Robert L. 1971. Towards the construction of a standard lexicostatistic list. *Anthropological Linguistics* 13(9). 421–434.

Pagel, Mark, Quentin D. Atkinson, Andreea S. Calude & Andrew Meade. 2013. Ultraconserved words point to deep language ancestry across eurasia. *Proceedings of the National Academy of Sciences* 110(21). 8471–8476.

Pan, Ho-Hsien. 1994. *The voicing contrasts of Taiwanese (Amoy) initial stops: Data from adults and children*: The Ohio State University dissertation.

Paschall, Leslie. 1983. Development at 18 months. In John V Irwin & Seok P Wong (eds.), *Phonological development in children 18 to 72 months*, 27–54. Carbondale: Southern Illinois University Press.

Paul, Hermann. 1891. *Grundriss der Germanischen Philologie*. Strasbourg: Trübner.

Peters, Ann M. & Eran Zaidel. 1980. The acquisition of homonymy. *Cognition* 8(2). 187–207.

Phillips, Betty. 2006. *Word frequency and lexical diffusion*. New York: Springer.

Piantadosi, Steven T. 2014. Zipf's word frequency law in natural language: A critical review and future directions. *Psychonomic bulletin & review* 21(5). 1112–1130.

Piantadosi, Steven T., Harry Tily & Edward Gibson. 2011. Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences* 108(9). 3526–3529.

Piantadosi, Steven T., Harry Tily & Edward Gibson. 2012. The communicative function of ambiguity in language. *Cognition* 122(3). 280–291.

Pierrehumbert, Janet B. 2001. Exemplar dynamics: Word frequency. In Joan L. Bybee & P. Hopper (eds.), *Frequency and the emergence of linguistic structure*, 137–157. Amsterdam: John Benjamins.

Pierrehumbert, Janet B. 2003. Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech* 46(2-3). 115–154.

Pinker, Steven & Paul Bloom. 1990. Natural language and natural selection. *Behavioral and Brain Sciences* 13(4). 707–727.

Pinker, Steven & Ray Jackendoff. 2005. The faculty of language: what's special about it? *Cognition* 95(2). 201–236.

Plotkin, Joshua B. & Martin A. Nowak. 2000. Language evolution and information theory. *Journal of Theoretical Biology* 205(1). 147–159.

Poppe, Nicholas. 1954. *Grammar of Written Mongolian*. Leipzig: Otto Harrassowitz.

Prather, Elizabeth M., Dona Lee Hedrick & Carolyn A. Kern. 1975. Articulation development in children aged two to four years. *Journal of Speech and Hearing Disorders* 40(2). 179–191.

Pye, Clifton, David Ingram & Helen List. 1987. A comparison of initial consonant acquisition in English and Quiché. In Keith E. Nelson & Ann Van Kleeck (eds.), *Children's Language*, 175–190. Hillsdale, NJ: Erlbaum.

R Core Team. 2017. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing Vienna, Austria. `https://www.R-project.org/`.

Reali, Florencia & Thomas L. Griffiths. 2009. Words as alleles: connecting language evolution with Bayesian learners to models of genetic drift. *Proceedings of the Royal Society B: Biological Sciences* 277(1680). 429–436.

Richter, Caitlin. 2018. Learning allophones: What input is necessary? In Anne B. Bertonili & Maxwell J. Kaplan (eds.), *Proceedings of BUCLD 42*, 659–672. Somerville, MA: Cascadilla Press.

Richtsmeier, Peter T. 2011. Word-types, not word-tokens, facilitate extraction of phonotactic sequences by adults. *Laboratory Phonology* 2(1). 157–183.

Ringe, Don. 2011. A pilot study for an investigation into Atkinson's hypothesis. *Linguistic Typology* 15(2). 223–231.

Ringe, Don & Joseph F. Eska. 2013. *Historical linguistics: Toward a twenty-first century reintegration*. Cambridge University Press.

Rischel, Jørgen. 1962. On functional load in phonemics. *Statistical Methods in Linguistics* 1. 13–23.

Robb, Michael P. & Ken M. Bleile. 1994. Consonant inventories of young children from 8 to 25 months. *Clinical Linguistics & Phonetics* 8(4). 295–320.

Robbeets, Martine Irma. 2003. *Is Japanese related to the Altaic Languages?*: Leiden University dissertation.

Roberts, Ian. 2007. *Diachronic syntax*. Oxford University Press.

Sammallahti, Pekka. 1988. Historical phonology of the Uralic languages with special reference to Samoyed, Ugric and Permic. In Denis Sinor (ed.), *The Uralic Languages: Description, History and Foreign Influences*, 478–554. Leiden: Brill.

Sampson, Geoffrey. 2013. A counterexample to homophony avoidance. *Diachronica* 30(4). 579–591.

Sampson, Geoffrey. 2015. A Chinese phonological enigma. *Journal of Chinese Linguistics* 43(2). 679–691.

Sampson, Geoffrey. 2020. An unaddressed phonological contradiction. *Journal of Chinese Linguistics* 6(2). 221–237.

Sander, Eric K. 1972. When are speech sounds learned? *Journal of Speech and Hearing Disorders* 37(1). 55–63.

Sapir, Edward. 1921. *Language: An introduction to the study of speech*. NY: Harcourt Brace.

Saunders, Ian W., Simon Tavaré & G. A. Watterson. 1984. On the genealogy of nested subsamples from a haploid population. *Advances in Applied probability* 16. 471–491.

Sayeed, Ollie & Andrea Ceolin. 2019. 'Markedness' is an epiphenomenon of phonetically grounded sound change. In Katherine Hout, Anna Mai, Adam McCollum, Rose Sharon & Matt Zaslansky (eds.), *Proceedings of the Annual Meetings on Phonology*, Washington DC: Linguistic Society of America.

Scheer, Tobias & Philippe Ségéral. 2019. What the Neogrammarians knew, and what they ignored. Presented at The Fourth Edinburgh Symposium on Historical Phonology .

Shannon, Claude E. 1951. Prediction and entropy of printed English. *Bell system technical journal* 30(1). 50–64.

Sigurd, Bengt. 1968. Rank-frequency distributions for phonemes. *Phonetica* 18(1). 1–15.

Silverman, Daniel. 2010. Neutralization and anti-homophony in Korean. *Journal of Linguistics* 46(2). 453–482.

Šimková, Mária. 2006. Slovak National Corpus - History and current situation. In Mária Šimková (ed.), *Insight into the Slovak and Czech corpus linguistics*, 151–159. Bratislava: Veda Publishing.

Simon, Herbert A. 1955. On a class of skew distribution functions. *Biometrika* 42(3/4). 425–440.

Sindi, Suzanne S. & Rick Dale. 2016. Culturomics as a data playground for tests of selection: Mathematical approaches to detecting selection in word use. *Journal of theoretical biology* 405. 140–149.

Smit, Ann Bosma, Linda Hand, J. Joseph Freilinger, John E. Bernthal & Ann Bird. 1990. The Iowa articulation norms project and its Nebraska replication. *Journal of Speech and Hearing Disorders* 55(4). 779–798.

Smith, Neilson V. 1973. *The acquisition of phonology: A case study.* Cambridge University Press.

Smits, Roel, Natasha Warner, James M. McQueen & Anne Cutler. 2003. Unfolding of phonetic information over time: A database of Dutch diphone perception. *The Journal of the Acoustical Society of America* 113(1). 563–574.

So, Lydia K. H. & Barbara J. Dodd. 1995. The acquisition of phonology by Cantonese-speaking children. *Journal of Child Language* 22(3). 473–495.

Sohn, Ho-Min. 1999. *The Korean language.* Cambridge University Press.

Soli, Sigfrid D. 1981. Second formants in fricatives: Acoustic consequences of fricative-vowel coarticulation. *The Journal of the Acoustical Society of America* 70(4). 976–984.

Sóskuthy, Marton. 2013. *Phonetic biases and systemic effects in the actuation of sound change*: University of Edinburgh dissertation.

Stemberger, Joseph Paul & Carol Stoel-Gammon. 1991. The underspecification of coronals: Evidence from language acquisition and performance errors. In Carole Paradis & Jean-François Prunet (eds.), *The special status of coronals: Internal and external evidence*, 181–199. Elsevier.

Stoel-Gammon, Carol & Judith A. Cooper. 1984. Patterns of early lexical and phonological development. *Journal of Child Language* 11(2). 247–271.

Stokes, Stephanie F., Thomas Klee, Cecyle K. Carson & David K. Carson. 2005. A phonemic implicational feature hierarchy of phonological contrasts for English-speaking children. *Journal of Speech, Language, and Hearing Research* 48(4). 817–833.

Stokes, Stephanie F. & Dinoj Surendran. 2005. Articulatory complexity, ambient frequency, and functional load as predictors of consonant development in children. *Journal of Speech, Language, and Hearing Research* 48(4). 577–591.

Stokes, Stephanie F. & Carol Kit Sum To. 2002. Feature development in Cantonese. *Clinical Linguistics & Phonetics* 16(6). 443–459.

Storkel, Holly L. 2002. Restructuring of similarity neighbourhoods in the developing mental lexicon. *Journal of Child Language* 29(2). 251–274.

Straka, George. 1979. *Les sons et les mots. Choix d'études de phonétique et de linguistique.* Paris: Klincksieck.

Surendran, Dinoj. 2003. The Functional Load of Phonological Contrasts. University of Chicago Master Thesis .

Surendran, Dinoj & Partha Niyogi. 2006. Quantifying the functional load of phonemic oppositions, distinctive features, and suprasegmentals. In Ole Nedergaard Thomsen (ed.), *Competing models of linguistic change: Evolution and beyond*, 43–58. Amsterdam: John Benjamins.

Svantesson, Jan-Olof, Anna Tsendina, Anastasia Karlsson & Vivan Franzén. 2005. *The phonology of Mongolian*. Oxford University Press.

Swadesh, Morris. 1955. Towards greater accuracy in lexicostatistic dating. *International Journal of American linguistics* 21(2). 121–137.

Sweet, Henry. 1990. *History of language*. London: J. M. Dent and Co.

Swingley, Daniel. 2003. Phonetic detail in the developing lexicon. *Language and Speech* 46(2-3). 265–294.

Swingley, Daniel. 2005. 11-month-olds' knowledge of how familiar words sound. *Developmental science* 8(5). 432–443.

Swingley, Daniel. 2009. Onsets and codas in 1.5-year-olds' word recognition. *Journal of Memory and Language* 60(2). 252–269.

Swingley, Daniel. 2016. Two-year-olds interpret novel phonological neighbors as familiar words. *Developmental Psychology* 52(7). 1011–1023.

Swingley, Daniel. 2019. Learning phonology from surface distributions, considering dutch and english vowel duration. *Language Learning and Development* 15(3). 199–216.

Swingley, Daniel & Claudia Alarcon. 2018. Lexical learning may contribute to phonetic learning in infants: A corpus analysis of maternal spanish. *Cognitive Science* 42(5). 1618–1641.

Swingley, Daniel & Richard N. Aslin. 2000. Spoken word recognition and lexical representation in very young children. *Cognition* 76(2). 147–166.

Swoboda, Philip J., Philip A. Morse & Lewis A. Leavitt. 1976. Continuous vowel discrimination in normal and at risk infants. *Child Development* 47(2). 459–465.

Tadmor, Uri, Martin Haspelmath & Bradley Taylor. 2010. Borrowability and the notion of basic vocabulary. *Diachronica* 27(2). 226–246.

Tambovtsev, Yuri & Colin Martindale. 2007. Phoneme frequencies follow a yule distribution. *SKASE Journal of Theoretical Linguistics* 4(2). 1–11.

Templin, Mildred C. 1957. *Certain language skills in children; their development and interrelationships.* MN: University of Minnesota Press.

To, Carol K. S., Sharynne McLeod & Pamela S. P. Cheung. 2015. Phonetic variations and sound changes in hong kong cantonese: Diachronic review, synchronic study and implications for speech sound assessment. *Clinical Linguistics & Phonetics* 29(5). 333–353.

Topbas, Seyhun. 1997. Phonological acquisition of Turkish children: implications for phonological disorders. *European Journal of Disorders of Communication* 32(4). 377–396.

Trehub, Sandra E. 1973. Infants' sensitivity to vowel and tonal contrasts. *Developmental Psychology* 9(1). 91–96.

Trehub, Sandra E. 1976. *Auditory linguistic sensitivity in infants*: McGill University dissertation.

Trehub, Sandra E. & M. Sam Rabinovitch. 1972. Auditory-linguistic sensitivity in early infancy. *Developmental Psychology* 6(1). 74.

Trnka, Bohumil. 1931. Bemerkungen zur homonymie. *Travaux du Cercle linguistique de Prague* 4(1931). 152–156.

Trubetzkoy, Nikolai S. 1939. *Principles of phonology.* 1969 ed. Berkeley, CA: University of California Press.

Trudgill, Peter. 2002. *Sociolinguistic Variation and Change.* Edinburgh University Press.

Tse, Sou-Mee. 1982. The acquisition of phonology: English versus Cantonese. *Acta Psychologica Taiwanica* 24. 1–7.

Upward, Christopher & George Davidson. 2011. *The history of English spelling.* Hoboken, NJ: John Wiley & Sons.

Vance, Timothy J. 1987. *An introduction to Japanese phonology.* Albany, NY: SUNY Press.

Vennemann, Theo. 1973. Explanation in Syntax. In John Kimball (ed.), *Syntax and Semantics,* 1–50. New York: Seminar Press.

Vihman, Marilyn May. 1982. A note on children's lexical representations. *Journal of Child Language* 9(1). 249–253.

Vihman, Marilyn May. 1993. Variable paths to early word production. *Journal of Phonetics* 21(1-2). 61–82.

Vihman, Marilyn May, Satsuki Nakai, Rory A. DePaolis & Pierre Hallé. 2004. The role of accentual pattern in early lexical representation. *Journal of Memory and Language* 50(3). 336–353.

Wallace, Douglas C., Michael D. Brown & Marie T. Lott. 1999. Mitochondrial DNA variation in human evolution and disease. *Gene* 238(1). 211–230.

Walley, Amanda C. 1993. The role of vocabulary development in children's spoken word recognition and segmentation ability. *Developmental review* 13(3). 286–350.

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

Wang, William S-Y. 1967. The measurement of functional load. *Phonetica* 16(1). 36–54.

Wang, William S-Y. 1969. Competing changes as a cause of residue. *Language* 45(1). 9–25.

Wedel, Andrew. 2012. Lexical contrast maintenance and the organization of sublexical contrast systems. *Language and Cognition* 4(4). 319–355.

Wedel, Andrew, Abby Kaplan & Scott Jackson. 2013. High functional load inhibits phonological contrast loss: A corpus study. *Cognition* 128(2). 179–186.

Weide, Robert. 1998. *The CMU pronunciation dictionary, release 0.6*. PA: Carnegie Mellon University.

Weinreich, Uriel, William Labov & Marvin Herzog. 1968. *Empirical foundations for a theory of language change*. University of Texas Press.

Weismer, Gary & Mary Elbert. 1982. Temporal Characteristics of "Functionally" Misarticulated /ʃ/ in 4-to 6-Years-Old Children. *Journal of Speech, Language, and Hearing Research* 25(2). 275–287.

Wells, John C. 1982. *Accent in English: An Introduction*. Cambridge University Press.

Werker, Janet F. & Richard C. Tees. 1984. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant behavior and development* 7(1). 49–63.

White, Katherine S., Sharon Peperkamp, Cecilia Kirk & James L. Morgan. 2008. Rapid acquisition of phonological alternations by infants. *Cognition* 107(1). 238–265.

Wiese, Richard. 2000. *The phonology of German*. Oxford University Press.

Wright, Richard, Bruce Hayes, Robert Kirchner & Donca Steriade. 2004. *Phonetically based phonology*. Cambridge University Press.

Wright, Sewall. 1931. Evolution in Mendelian populations. *Genetics* 16(2). 97.

Yang, Charles. 2000. Internal and external forces in language change. *Language Variation and Change* 12(3). 231–250.

Yang, Charles. 2009. Population structure and language change. Ms. University of Pennsylvania .

Yang, Charles. 2016. *The price of linguistic productivity: How children learn to break the rules of language.* Cambridge, MA: MIT Press.

Yasuda, Akiko. 1970. Articulatory skills in three-year-old children. *Studia Phonologica* 5. 52–71.

Yoneyama, Kiyoko, Mary E. Beckman & Jan Edwards. 2003. Phoneme frequencies and acquisition of lingual stops in Japanese. Ms. Ohio State University .

Zee, Eric. 1985. Sound change in syllable final nasal cononants in Chinese. *Journal of Chinese Linguistics* 13(2). 291–330.

Zee, Eric. 1999. Change and variation in the syllable-initial and syllable-final consonants in Hong Kong Cantonese. *Journal of Chinese Linguistics* 27(1). 120–167.

Zipf, George Kingsley. 1939. *Human behavior and the principle of least effort: An introduction to human ecology.* Cambridge, MA: Addison-Wesley.