universität wien

DISSERTATION / DOCTORAL THESIS

Titel der Dissertation /Title of the Doctoral Thesis

„Modeling the evolutionary ecology of phonotactics: cognitive, linguistic, and non-linguistic determinants“

verfasst von / submitted by

Mag. phil. Mag. rer. nat. Andreas Baumann

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Doktor der Philosophie (Dr. phil.)

Wien, 2018 / Vienna 2018

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 792 425
Dissertationsgebiet lt. Studienblatt / field of study as it appears on the student record sheet:	Cognitive Science/Kognitionswissenschaften
Betreut von / Supervisor:	Univ.-Prof. Mag. Dr. Nikolaus Ritt

Danke

Anfang 2014 habe ich im Grunde keine Ahnung gehabt, worum es in der Phonotaktik eigentlich geht (also in dem Teilbereich der Sprachwissenschaft, der sich um das Aneinanderreihen von Lauten dreht). In der Zwischenzeit ist recht viel passiert. Ich habe mitgeholfen, einen Projektantrag zu dem Thema zu schreiben; ich war ein Weilchen in Südafrika und habe mich dort eher mit der Ausbreitung von Krankheiten als mit Lautfolgen beschäftigt; der Projektantrag wurde (Gerlinde und Horst sei Dank) angenommen; ich bin an die Anglistik zurückgekommen und habe schließlich gemeinsam mit Chrisi und Niki begonnen, zu Konsonantenfolgen zu arbeiten. Auf irgendeine Art und Weise hat das alles dann doch seinen Weg in diese Dissertation gefunden (sogar der Klippschliefer, den wir in Bettysbaai gesehen haben; den findet man an prominenter Stelle auf Seite 2).

Auf dem Weg dorthin, also zum Abschluss dieses Dissertationsprojektes, war ich nicht alleine. An allererster Stelle bedanke ich mich bei Niki, der nicht müde wurde, sich seit 2011 für mich einzusetzen. Ohne seine inspirierenden und ansteckenden Ideen würde ich jetzt wahrscheinlich schon längst mit Sterbetafeln und Versicherungsprämien herumhantieren oder mich mit ähnlich (oder weniger) spannenden Dingen beschäftigen. Auch alle anderen in Nikis Truppe haben mich redlich unterstützt, nicht nur durch das Korrekturlesen von Teilen dieser Arbeit, in dessen Genuss eigentlich jeder der hier genannten irgendwann gekommen ist: Chrisi war eine immense Hilfe bei der Datenarbeit; Tezi hat mit mir gemeinsam zu demografischen Fragestellungen gearbeitet; Klaus hat mir in unseren vielen Unterhaltungen immer wieder aus gedanklichen Sackgassen geholfen; und Kamil verdanke ich vor allem, dass er mir während meiner vielen Aufenthalte in Poznań das psycholinguistische Experimentieren beigebracht hat (und auch wie man einen Kaffee auf Polnisch bestellt). Ein großes Dankeschön geht an Eva, Lotte und Lotti, die immer ein offenes Ohr für alle meine Fragen und Anliegen jeglicher Art und, was viel wichtiger ist, motivierende Antworten parat gehabt haben. Meiner Familie danke ich dafür, mir in den vergangenen Jahren den Rücken freigehalten zu haben, sodass ich mich immer auf meine Arbeit konzentrieren konnte.

Theresa danke ich vor allem für ihr Verständnis, wenn ich mal wieder länger am Institut geblieben bin, am Wochenende arbeiten oder wegen Tagungen verreisen musste. Nicht nur einmal hat sich unsere Urlaubsplanung nach meinen Konferenzreisen gerichtet anstatt umgekehrt. Das Schöne ist, dass ich auch dabei nie alleine war.

Contents

Modeling the evolutionary ecology of phonotactics: a companion paper

1. Introduction

When we speak we concatenate sounds. However, we do not do so arbitrarily. For instance, if I want to name the animal depicted in Figure 1a, a rock hyrax, I produce a specific string of sounds. Each of these sounds is associated with some articulatory gesture which usually involves exhaling and simultaneously moving my tongue and lips while sometimes letting my vocal cords vibrate. For example, the very final sequence of sounds in the word *hyrax* is produced by blocking the airstream with my tongue somewhere at the back of my mouth, subsequently releasing this closure and finally moving the tip of the tongue relatively close to my teeth so that the airstream is pushed through this gap (Figure 1b). By proceeding like that, I produce acoustic sound waves. They can be physically measured as illustrated by the spectrogram shown in Figure 1c. The box roughly captures the temporal sequence which belongs to the final two sounds in the word *hyrax*. More importantly, the sequence can be perceived by some nearby listener. It shall be denoted by the transcription /ks/. This dissertation is about how we use sequences like /ks/, and about how and why they change over time.

The production of /ks/ is quite complicated as it involves very fine grained motor patterns comprising different parts of my body; particularly a multitude of muscular movements I am not even aware of. However, I am very good at producing the /ks/ sequence, not only when I talk about hyraxes, but also when using words like *hoax*, *box*, *jokes*, *likes*, and many others. That is, I have knowledge of how this sequence is produced and how it should sound like. As a matter of fact, I am not alone with this knowledge. Many other people produce /ks/, for instance when referring to hyraxes, and they have no major problem with understanding someone using /ks/. Actually, it is the people I communicate with who made me learn and use the /ks/ sequence when talking about hyraxes, hoaxes, jokes, etc. (Figure 1d), at least this holds true for members of the English speech community.

Speakers of English, of course, also know many other sound sequences. In the language sciences, this knowledge is referred to as 'phonotactic' knowledge. That is, phonotactics is the part of grammatical knowledge of a shared linguistic system (a language) that covers which sequences of sounds are used in that language. As such, phonotactic knowledge is part of (human) cognition. Implicitly, it also covers which sequences are not possible in certain positions. For instance, there is no English word which begins with /ks/. More generally about one third of all combinatorically possible pairs of sounds are never used in English speech at all.[1]

Phonotactics, as any grammatical knowledge, is shared among speakers of a language. To the extent that it is not biologically inherited, it needs to be culturally transmitted through populations of speakers and across generations. The production and perception of sound sequences is learned, and this learning may take place either when a language is acquired as a child or later, for instance, when one learns a second language.

[1] This was demonstrated by Christiansen et al. (2009) by analyzing sound pairs in a large English text corpus (about 5.5 million words in 1.4 million utterances, which amounts to over 40 million sound pairs). Given the large size of this corpus, it is hard to imagine that over one third (954) of all possible sound pairs (3,025) does not show up in the data just by pure chance.

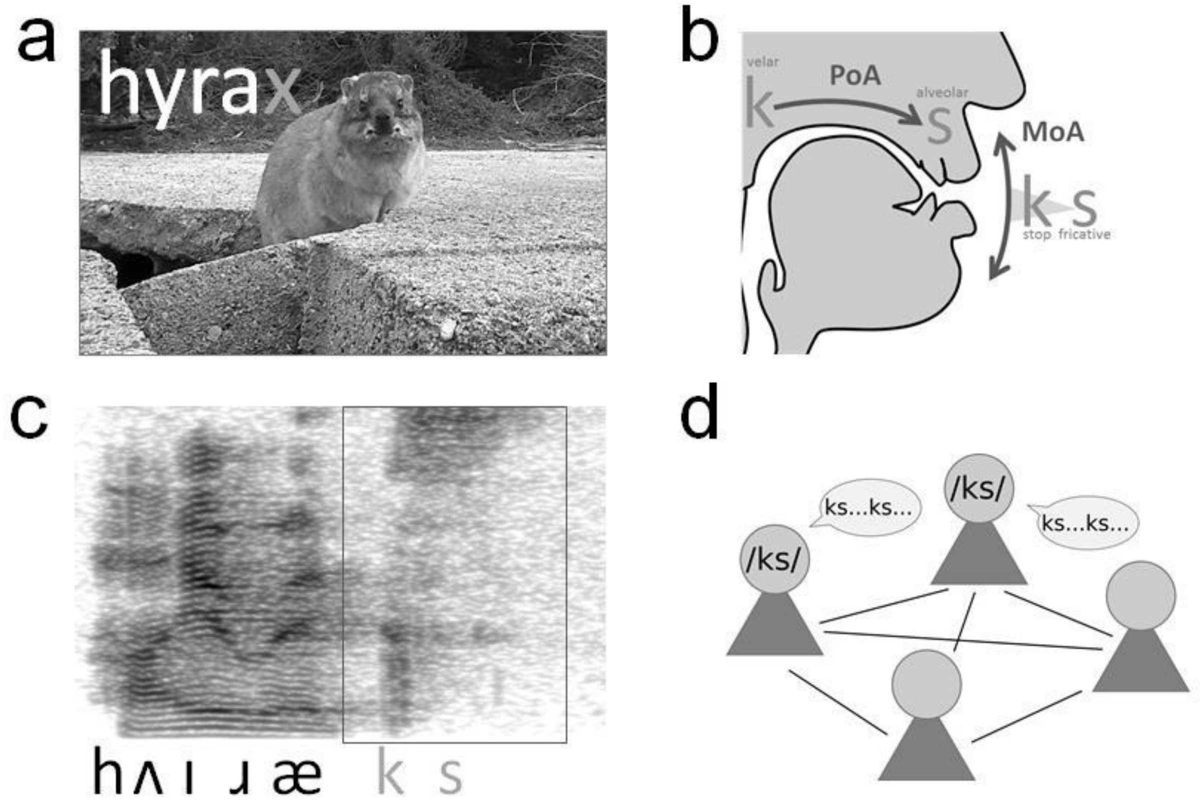


Figure 1. (a) A rock hyrax, *Procavia capensis*. The final letter in the written form of its name, <x>, represents the sound sequence /ks/. (b) Schematic representation of the articulatory movements involved in the production of /ks/. The sketch shows dynamics both in where the constriction of the airstream occurs (place of articulation, PoA) and how tight it is (manner of articulation, MoA). (c) Sonogram of the word *hyrax*, together with its phonological transcription. Time (s) is measured on the horizontal axis, and frequency (Hz) on the vertical axis (black indicates high energy in a given point). The box roughly depicts the sequence corresponding to /ks/. (d) Phonotactic knowledge of /ks/ is transmitted through a population of speakers in multiple production and perception events.

Since phonotactic knowledge is transmitted, it is subject to evolutionary mechanisms. Thus, sound sequences are sometimes not transmitted faithfully. Sounds in a sequence may for instance be articulated in a different way. For example, I could vibrate with my vocal folds when saying the final sequence in *hyrax*, which would be unusual. At the same time, it could be perceived as a different sequence from the one I intended it to be by some listeners. Also, sound sequences differ in how easily they are articulated and perceived, which in turn affects their ease of transmission. Sound sequences that are transmitted more easily have higher chances of being shared among many speakers and used more frequently than, for instance, articulatorily cumbersome sequences. Overall, this influences which sound sequences belong to a language's phonotactic inventory at a given point in time and which do not. Indeed, phonotactic knowledge is not static and so it might happen that over the years, certain sound sequences go extinct (like word final /mn/ which used to be present in English *hymn* a couple of centuries ago, or /mb/ in *womb*; both are still represented in writing). That is, speakers may lose certain parts of phonotactic knowledge (or indeed never acquire them) to the effect that the corresponding sequences of sounds are not used any more.

How phonotactic knowledge changes over time in a population of speakers and which factors are relevant to its evolution is the central question of this thesis. This is a very broad question, and one might wonder why it is feasible to adopt a diachronic – or evolutionary – approach at all rather than just studying phonotactic knowledge synchronically at one point in time. The question is what a diachronic perspective on how phonotactic knowledge changes over time can add to the study of human cognition?

The answer is the following: as I outlined above phonotactic knowledge needs to be transmitted between speakers through speech events in order to be shared by a speech community. This transmission is constrained, among other things, by cognitive biases. As a matter of fact, these biases can be very weak and it can be difficult to detect them by just looking at a single generation of speakers. However, since language use consists of a vast number of linguistic interactions and transmission events, weak biases can become visible on a larger time scale. That is, vanishingly small effects caused by weak cognitive biases can accumulate over many generations of speakers to yield strong observable tendencies (which may be eventually described as 'universal' properties of language; Kirby, 2012; Kirby et al., 2007; Thompson et al., 2016).

Of course, *a priori* we can only speculate about which biases determine successful transmission. The crucial point is that different sound sequences are adapted differentially to these biases. By comparing the histories of a large number of sound sequences in a language, and more precisely by investigating which sound sequences do well on a larger time scale and which ones eventually go extinct, we can learn about the determinants that govern their evolution.

This is exactly what I intend to do in this thesis. By adopting tools from the study of ecology and evolution I study a broad range of determinants on different levels of organization. Cognition is not only physiologically constrained within individuals but also by interactions among several individuals and ultimately entire populations of agents. Due to the diversity of these levels, various methods need to be employed. I rely on experimental methods as well as quantitatively empirical analyses of diachronic linguistic long-term data as well as data from language acquisition research. Most prominently, however, I make use of mathematical models. That is, I do not only look at empirically attested histories of sound sequences but I also simulate their evolution. In doing so, I build specific cognitive biases into my models (such as 'frequency of exposure facilitates learning'). If the simulated history of the phonotactic inventory matches the empirical one this ultimately corroborates the relevance of that bias. The models which will be used have originally been developed for studying population dynamics in biological evolution, ecology and epidemiology. Nevertheless, since I embrace an evolutionary and population based approach to cognition, these tools can be fruitfully applied to my questions.

In this companion paper, I provide a brief introduction to and summary of my dissertation project. I begin with embedding phonotactics into the theory of cognition and cognitive phonology. I discuss why phonotactics is a cognitive phenomenon (2.1) and conceptual models of how sounds (2.2) and sound strings (2.3) may be cognitively represented. Based on this discussion, it is concluded that speakers indeed make use of self-contained phonotactic representations, i.e. constituents of phonotactic knowledge. After that, I describe how cognitive phonotactics can be investigated in the light of evolutionary theory by applying well established concepts of evolutionary linguistics (3.1). Most importantly, I will conceptualize phonotactic representations as culturally transmitted replicators. In 3.2, methodological aspects will be linked to the evolutionary approach. That is, I give a short introduction to the mathematical toolkit used in this thesis. Subsequently, I focus on the selection of determinants which I cover in my project on the different levels of organization mentioned above. Together with this, I present a number of more specific research questions which I will address in my research (3.3). Finally, Section 4 briefly sketches the subprojects of this thesis together with their specific results and in which way they relate to the questions specified in 3.3. Section 5 finally discusses the main results and provides an outlook.

As we go along, I will make a few excursions to related topics in cognitive research. These excursions are organized as separate mini-chapters (gray boxes). Here, without going too much into detail, I provide sketchy analyses which link different strands of research in cognitive science with topics covered in this thesis.

2. Cognitive phonotactics

When studying the transmission of phonotactic knowledge one needs a working hypothesis on how that knowledge is cognitively represented. After all, at least partially it is the cognitive representation of a phonotactic item which determines the way in which it is transmitted as a sound sequence. There are various views on how language in general and phonotactics in particular is cognitively represented. These views range from conceptualizing phonotactic knowledge as (potentially innate) abstract rules to episodic memories of articulatory and acoustic patterns. In what follows, I will discuss and contrast these views to finally map out the working hypothesis on phonotactic representations this thesis builds on.

2.1 Sounds and cognition

Language is part of cognition and a central principle of cognitive approaches to language is that language and any other part of cognition are subject to the very same cognitive mechanisms (Lakoff, 1993; Langacker, 2008). This principle is referred to as the 'cognitive commitment' (Lakoff, 1990). This commitment also includes the notion of embodiment, which entails that "language is embodied and situated in the sense that it is embedded in the experiences and environments of its users" (Mompeán, 2006, vii), or as Thelen et al. (2001, p. 1) put it more generally

> "[t]o say that cognition is embodied means that it arises from bodily interactions with the world. From this point of view, cognition depends on the kinds of experiences that come from having a body with particular perceptual and motor capabilities that are inseparably linked and that together form the matrix within which reasoning, memory, emotion, language, and all other aspects of mental life are meshed."

Consequently, the properties of cognitive representations must depend on these bodily interactions. If a cognitive representation "arises from bodily interactions with the world", it will obviously be influenced by at least to factors: (i) the body and (ii) the world. The production and perception of language obviously covers aspects of both (i) and (ii). From the perspective of the speaker (i) includes all parts of the body which are necessary for producing and perceiving sound (most evidently the articulatory and perceptual organs) while (ii) subsumes the air, at least, and a potential listener. What empirical support do we have that the cognitive representation of a sound is influenced by or indeed "arises from" interactions among (i) and (ii)? In the following couple of paragraphs, I collect pieces of evidence for the previous hypothesis.

Let us begin with evidence from research on sound symbolism. A well-known example is the so-called bouba-kiki effect (Kristiansen, 2006; Ramachandran and Hubbard, 2001). If people are presented drawings of a spikey and a roundish shape and asked to allocate the labels 'bouba' and 'kiki' to them, they tend to match 'bouba' with the roundish shape and 'kiki' with the spikey shape. As an explanation for this, Ramachandran and Hubbard (2001, p. 19) have argued that the spikey shape resembles "the sharp phonemic inflections of the sound kiki, as well as the sharp inflection of the tongue on the palate". If they are right then the association is established both on perceptual and articulatory grounds.

Along with this, it has been shown that vowel quality is associated with meaning in that low vowels like /a/ correspond to large and spacious objects while high vowels like /i/ correspond to tiny and slim objects (Tsur, 2006). High vowels are characterized by a high acoustic frequency while low vowels have a low frequency. Crucially, acoustic frequency is an immediate physical reflex of the amount of space between the tongue and the palate which the airstream is transmitted through: large cavity for /a/ and narrow gap for /i/. Thus, it can be argued that this sound-size correspondence in vowels is an example of linguistic embodiment established through articulation (and perhaps to some extent also through perception). Considering consonants, Winter et al. (2017) found that words containing /r/ are associated with 'rough' meanings and words containing /l/ more likely than chance have 'smooth' meanings. They argue that this correlation can be explained by "a cross-modal correspondence between the roughness of surfaces and the intermittent airflow in the production of trilled /r/" (2017: 9).

So far, we have only discussed single sounds. There is, however, evidence for sound symbolism in sequences of sounds. For instance, Topolinski et al. (2015) found that the sequence of places of articulation in a word affects its meaning. Nonce words in which consonantal places of articulation proceed from the front to the back (such as "bodika") are rated (semantically) more positive than items in which places of articulation proceed from the back to the front ("kodiba"). They argue that this correspondence is established because front-to-back articulation resembles ingestion (of edible food) while back-to-front articulation resembles expectoration (of non-edible substances). Similarly, Laham et al. (2012) demonstrated that people with difficult to pronounce names are rated as less likeable than people whose names are easy to pronounce. This hints at a correspondence between physical labor during articulation and attitude. Interestingly, the ratings for ease of pronunciation of the names used in Laham et al.'s experiments correlate strongly with the number of consonant clusters, i.e. sequences of consonants, and hence phonotactic complexity (see Excursion 1).

Another example of symbolic features of sounds that is linked to articulatory labor and complexity is provided by Naturalness Theory (Donegan and Nathan, 2014; Dressler, 1989, 2003). For instance, plural forms tend to be phonologically longer than their corresponding singulars (often because plural is marked by an additional affix, e.g. *-s* in English *rabbits*). The idea is that phonological substance and concomitant articulatory (bodily) effort corresponds to (semantic) quantity.

The embodiment hypothesis, clearly goes beyond bodily constraints and actions alone, as it also encompasses interactions with the world. In the production-perception loop of (a string of) sounds this includes the transmitting medium, typically air, in the first place, and a set of perceiving individuals. Both aspects affect the setup of a phonological system. For example, Everett et al. (2016) have demonstrated that air humidity exerts an impact on phoneme quality. They argue that tonal contrasts are difficult to produce and perceive in dry environments, because of the dry air inhibiting the accuracy of the vocal cords which in turn affects acoustic quality. As a consequence, languages whose speakers are located in dry areas are more likely to lack tonal contrast.

Obviously, phonotactic knowledge depends on interactions with other individuals. Exposure to a string of sounds which one has not yet learned and used before increases the likelihood of acquiring that sound string. This is self-evident in first-language acquisition but equally clearly visible in adult speakers who acquire a novel sound structure. For instance, this can happen by importing some non-native loan word like German *Schnitzel* ‘schnitzel’ to English together with the word-final sequence /ʃn/ (Figure 2).

Figure 2. Although there are only about 5,000 native speakers of German left in Texas, the schnitzel together with its initial consonant cluster /ʃn/ is entrenched in the Texan population and cuisine.

There is also additional evidence that phonotactic knowledge is shaped by interactions between the speaker and other listeners. The duration of diphones, i.e. sequences of two sounds, has been shown to be optimized in such a way that articulatory effort is minimized and ease of perception maximized at the same time (Kuperman et al., 2008). The relationship between sound-sequence duration and utterance frequency adopts the shape of an inverse U across languages: it is sound sequences in the middle of the duration spectrum which are produced most frequently. This is interesting, as it contrasts with the inverse relationship between duration and frequency found in single sounds. This suggests that properties of phonotactic items are to a larger extent constrained by perception during interactions than single phonemes are.

On a less fine-grained level, phonotactic knowledge is probably also constrained by the size and architecture of the social network speakers are embedded in (see Section 3). The ease at which phonotactic knowledge is transmitted depends on the number of acquaintances and linguistic contacts, which in turn depends on the network architecture of the speaker population. A fair share of this thesis is dedicated to illuminating the relationship between population structure and phonotactic knowledge. I will come back to this issue later.

Excursion 1. Phonotactics and emotion: too many clusters are unappealing

The phonological shape of a word is connected to the attitude towards it. This has non-negligible consequences. Stocks with complicated labels are less successful (Alter and Oppenheimer, 2006), difficult to pronounce drugs are considered risky (Song and Schwarz, 2009), and people having names that can be easily pronounced are more likely to adopt top positions than people with articulatory cumbersome names. The latter was demonstrated by Laham et al. (2012) in a study in which they also experimentally tested people's affective attitudes towards names with differential degree of pronunciation difficulty. Overall, participants (all of them speakers of English) turned out to have better attitudes towards names which were rated as easy to pronounce, while names rated as difficult scored significantly worse. Crucially, this even holds if controlling for linguistic origin of the name (which may correlate with social factors).

The result is taken to support the notion that names are pieces of embodied cognition as the articulatory effort that must be allocated to the production of a name is inversely mapped to emotional comfort. Having less labor is more comfortable, and being more comfortable arguably is an emotionally preferred state. This then feeds the attitude towards the name and ultimately the individual behind it. But which phonological properties of a name are responsible for its ease or difficulty of pronunciation in the first place?

To dig a bit deeper, I took a closer look at the list of names which were used as stimuli in Laham et al.'s (2012) experiments. To measure phonological complexity, I looked at three properties, which are often associated with it. First, phonological length, i.e. the number of phonemes in a word, as a measure of overall phonological substance (already controlled for in Laham et al. 2012). Second, the fraction of consonants in the word. Third, the number of syllable-internal consonant clusters. The latter two measures correspond to phonotactic complexity.

An analysis of the respective effects of the three properties on the subjective ease-of-pronunciation ratings in Laham et al. (2012) reveals that, as expected, length reduces ease of pronunciation (as already pointed out by the authors). The behavior of the phonotactic measures is more revealing. The fraction of consonants did not show any robust effect on ease of pronunciation. If anything, the analysis shows that names featuring roughly two consonants per vowel are preferred (speaking against dominance of strict consonant-vowel alternation; rather closed syllables seem to be more ideal). Most interestingly, names with multiple clusters are rated significantly worse than words with at most one cluster.

There are two messages to be taken home from this investigation. First, syllable-internal phonotactic complexity affects articulatory effort, and increased effort elicits negative emotions (dislike, antipathy, skepticism). Second, speakers of English are fine with a certain amount of phonotactic complexity. Consonant clusters, which are considered as phonotactically complex items, adopt a central role in this thesis.

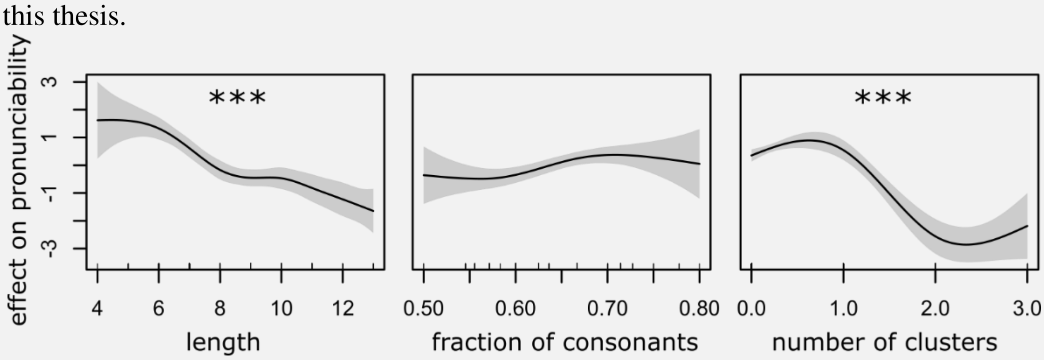


Figure E1. Effects of length, fraction of consonants and number of clusters on pronunciability ratings of names in Laham et al. (2012). Generalized additive model with three smooth terms. Length and number of clusters represent significantly nontrivial (roughly negative) effects on ease of pronunciation.

2.2 Cognitive representations in phonology

Before we turn to phonotactics, let us briefly look at different conceptualizations of how single sounds are mentally represented. Cognitive representations can be generally defined as "the way in which information is encoded when it is inside an agent, and processing amounts to converting information from one form of encoding to another" (Bryson, 2009, p. 78). The overarching question is this: if sounds have mental representations, are these representations abstract or detailed (Ernestus, 2014; Moreton, 2002; Pierrehumbert, 2016)?

Traditional accounts of cognitive phonology, which have their origins in Natural Phonology (Stampe, 1979), assume the former (Nathan, 2006; Taylor, 2002; Taylor, 1995).[2] Here, phonemes are abstract categories of sounds. They consist of a set of variants, or allophones. These allophones form an equivalence class and the underlying equivalence relationship is defined as something like 'x equals y if changing x to y in a word does not change the word's meaning'. This is often visualized as a radial network in which variants, so-called extensions, are grouped around a prototypical proponent (Nathan, 2007). An example is shown in Figure 3a: the voiceless alveolar stop can have many realizations in English, such as [t] (the prototype), aspirated [tʰ], palatalized [tʲ], or even a glottal stop [ʔ]. The radial members of the network are derived from the central prototype, and this derivation depends on the phonological environment (cf. complementarily distributed allophones; e.g. aspiration of English voiceless stops only in initial position) or socio-pragmatic context (such as foregrounding in formal speech). The prototypical proponent usually is the most frequent member of the network. The radial network model was extended to also include a schema which is the intersection of all members of the network in that it encompasses the features which are common to all extensions and the prototype (Figure 3b).

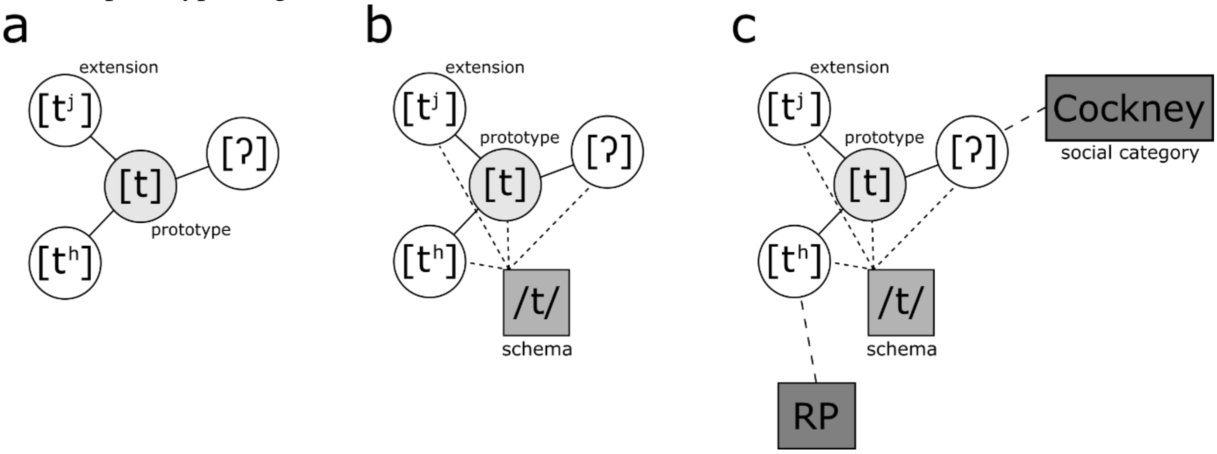


Figure 3. Models of cognitive phonological representations: (a) In the radial network model, extensions are grouped around a prototypical sound (Nathan, 2007; Taylor, 2002). (b) A phonological schema is associated with the nodes in the network. It captures any features which are shared among all nodes (Mompeán-González, 2004). (c) Social information is associated with some extensions (Kristiansen, 2006).

A couple of remarks are in order. First, this model of phonological representation does not explicitly encompass entrenchment or any other form of frequency dependence. The choice among variants in the network is simply a function of the linguistic and socio-pragmatic context. Information about frequency as such is not stored. Second, the model implicitly includes a meaning component, because the choice of representatives is a function of socio-pragmatic information. Kristiansen (2006) made this relationship explicit by linking subsets of extensions to social information. For example, some extensions in Figure 3c are linked to social

[2] These accounts actually go back to work by Trubetzkoy (1969).

(in this case dialectal) categories like ‘Cockney’ or ‘RP’. Third, networks of different phonemes may overlap, i.e. they may share extensions (Mompeán-González, 2004; Nathan, 2007). For example, /d/ in the English past-tense suffix can be devoiced to surface as [t] in voiceless contexts (e.g. in *kiss-ed* /kɪst/). This somewhat contradicts the defining feature of phonemes as minimal meaning distinguishing units. If phonemes are defined as equivalence classes of items that do not change lexical meaning, then it is impossible for an extension to belong to two non-identical equivalence classes, i.e., phonemes. This is usually repaired by arguing that this overlap is possible because extensions are always derived by additional external information, i.e., the context, thus relaxing the classical assumption of disjoint categories (Nathan, 2007).

An account which promotes detailed knowledge as opposed to abstraction is that of exemplar theory (Pierrehumbert, 2001, 2016; Wedel, 2006). In exemplar models of phonology, the speaker is assumed to store episodic memories of every single encountered instance of a sound, called exemplars. Each of these episodic memories contains articulatory, acoustic, social, pragmatic and other context dependent details. For instance, speakers are supposed to memorize acoustic duration of encountered instances of [n] (Figure 4a). All exemplars belong to an exemplar cloud which is associated with a label. Exemplar clouds are updated immediately whenever new utterances are perceived. The criterion for adding a newly encountered exemplar to an existing cloud is similarity. If the novel exemplar fits well to the presently stored exemplars it is added to the cloud. If not, it may end up in a different cloud.

Since speakers store every single occurrence of a sound they implicitly have knowledge about frequency. For instance, they implicitly know the frequency distribution of durations of the sound [n], i.e. they know which duration is prototypical in the sense that it is used most often and which durations are rarely employed (Figure 4b). This is a crucial difference to the abstract models of phonological representation outlined before. In exemplar models, the perceived frequencies translate into entrenchment. For production, this means that highly entrenched durations (in our example) are in turn produced most often while less entrenched ranges of duration are employed less frequently.

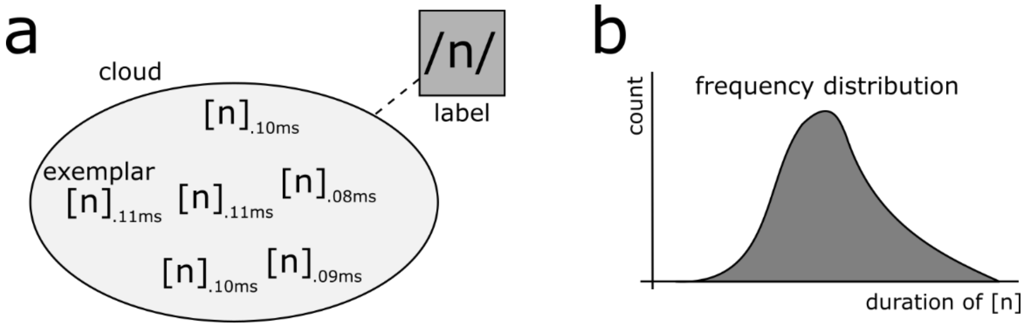


Figure 4. Exemplar model of cognitive phonological representations: (a) Episodic memories of encountered sounds (exemplars) are stored in a cloud which is tagged with an abstract label. Here, the depicted stored information is acoustic duration, but the nature of memorized details can be more complex. (b) The frequency distribution of durations in the exemplar cloud can be interpreted as relative entrenchment of durations. The same distribution is employed during production.

As can be seen, also the exemplar model does not work without processes of abstraction or schematization. Exemplars, although representing episodic memories, are collected into labeled categories, i.e. analogues of phonemes. Not all proponents of exemplar theory adopt this notion, however. For instance, Välimaa-Blum (2009b) argues that there are clouds of episodic memories of lexical or morphological items, but not for phonemes. Rather, phonemes are generalizations across lexical exemplar clouds. Apart from these abstract generalizations, phonemes do not have self-contained representations, she says, her main argument being that any cognitive representation requires a meaning component and sounds are, as she argues, intrinsically meaningless units. The latter follows from the (widely assumed) hypothesis of duality of patterning, which in a nutshell asserts that in human language, meaningless units (phonemes) are always combined to yield meaningful units (morphemes; words).

The question is, if the latter hypothesis and particularly the assumption of sounds lacking any kind of meaning is universally true or rather a statistical tendency (cf. de Boer et al. 2012, Blevins, 2012 or Ladd, 2014: chapter 5 for a thorough discussion). For one, we have seen in the previous section that there is evidence for phonemic iconicity and sound symbolism. Further, the choice of actual realizations greatly depends on the socio-pragmatic context and listeners can easily infer social information from phonetic detail (Eckert and Labov, 2017). Thus, sounds can be argued to carry socio-pragmatic information. Finally, sounds can play a functional role in speech processing. For instance, certain sounds only occur in word-initial or word-final position (such as /h/ in Greek) and acoustic duration of sounds depend, among other things, on the morphological complexity of words (Kemps et al., 2005; Plag et al., 2015). Thus, these phonetic details might in turn help the listener in decomposing the speech stream into lexical units. Once word boundaries are identified, phonetic details (duration and vowel formants; van Bergem, 1993) of word-internal sounds and phonological setup of the word boundaries (Christiansen et al., 2009) provide reliable information about the syntactic class of a word. In that sense, sounds can have grammatical meaning (similar to complementizers like *that* or *because* which indicate the presence and/or syntactic nature of subordinate clauses). If taken for granted, then these components of meaning and information must be part of the cognitive representation of sounds.

2.3 Phonotactic representations

The concepts described in the previous section can be extended to the phonotactic level to provide a cognitive model of phonotactic representations. There are two central questions: (a) Do speakers have separate mental representations of sound sequences, to begin with? And (b), if speakers do have mental representations of phonotactic items, are these representations abstract generalizations or collections of detailed episodic memories, or both?

The answer to the first question is not a priori clear. If one assumes speakers to have access to phonological as well as lexical representations, it may be argued that having knowledge of the entire linear organization of all sounds in a given word renders any separate storage of sublexical sequences, i.e. any phonotactic knowledge, redundant (Välimaa-Blum, 2009b). There are several arguments against this. Speakers are sensitive to whether they are familiar with a particular sound sequence in a given word. Familiar sound sequences are produced more easily and perceived more accurately than (relatively) unknown sound sequences, even if these sound sequences surface in nonce words (Berent et al., 2007). Unfamiliar sound sequences are recognized and rejected easily (Shatzman and Kager, 2007). They also have a higher chance of undergoing repair strategies, such as reduction of one of the segments involved or insertion of an additional sound (e.g. schwa) to transform one unfamiliar diphone into two familiar ones. These strategies also apply to novel words (such as imported loans) which do not fit well to previously encountered phonotactic patterns. Thus, native speakers of Japanese repair non-native diphones /tr/ in, say, three-syllabic *Austria* by inserting an epenthetic vowel to yield the four-syllabic item /ɑʊsturiɑ/[3] (Dupoux et al., 1999); native speakers of Italian often extend word-final consonants by a schwa to yield word final /Cə/ diphones, e.g. English *business* pronounced as /bisinisse/ (Repetti, 2012: 178; Grice et al., 2015); and some clusters in English loans imported into Finnish undergo segmental deletion, such as word initial /bl/ reduced to /l/ in *blurb* (Välimaa-Blum, 2009a, p. 7). This strengthens the notion that speakers have phonotactic knowledge derived from previous linguistic experience.

The question now is whether this knowledge is abstract or detailed. Abstract phonotactic knowledge consists of schematic representations of sound sequences, analogous to phonemes in phonology, phonotactic rules, or constraints (Blevins et al., 2003; McCarthy, 2004; Vennemann, 1988). For instance, the relative under-representation of sequences of segments with similar or identical place of articulation has been modeled by means of the Obligatory Contour Principle (OCP); it is basically a set of constraints which restrict, for instance, sequences of two labials like /b/ and /w/ (Frisch et al., 2004; Shatzman and Kager, 2007). Such constraints are abstract in the sense that they apply to classes of sound sequences, namely those which show a particular phonological property (such as 'all segments are labial'). Shatzman and Kager (2007) showed that speakers of Dutch indeed process sequences which do not conform to the OCP (in that they have multiple labial segments) slower than OCP-well-formed sequences. Crucially, these sequences were embedded into nonce words and other factors such as similarity to phonologically related lexical items were carefully controlled for.

Similarly, Moreton (2002) tested the sensitivity of English speakers towards sequences of two coronals (/dl/) and two labials (/bw/), which both violate OCP.[4] None of these diphones occur word initially in English. The important difference between these two sequences is that /d/ and /l/ show a smaller difference in sonority than /b/ and /w/ do. Since small differences in sonority are dispreferred, speakers of English are expected to favor word initial /bw/ over /dl/ (this is referred to as the Sonority Sequencing Principle, SSP; we will come back to it in Section 4, but see Excursion 2 for some related insights from neurological research, and a perception experiment on OCP/SSP conducted by Kamil Kaźmierski and myself). This preference of /bw/ over /dl/ is exactly what Moreton (2002) shows. He concludes that this supports the existence of abstract feature based representations of sound sequences.

In contrast, proponents of the unit- or exemplar-based approach argue that phonotactic knowledge is mainly statistical and dependent on utterance frequency: Saffran et al. (1996) detected that 8-month old infants are sensitive to phonotactic transitional probabilities. Vitevich et al. (1997) found that nonce words composed of high-frequency syllable structures are processed faster and yield higher well-formedness ratings than low-frequency syllable structures. Moreover, Hay et al. (2004) show that well-formedness ratings of nasal-obstruent clusters can be predicted from and correlate positively with utterance frequency, even if these clusters stretch across syllable boundaries.

[3] Which, given the non-phonemic difference between /r/ and /l/ in Japanese, makes *Austria* and *Australia* hardly distinguishable from each other in conversations in which English is used as lingua franca, to give some personal anecdotal evidence.
[4] In this experiment, participants listened to sound strings and were asked to notify which string they perceived.

Excursion 2. Phonotactics and neuroscience: good diphones show dispersed processing

Diphones are composed of segments. These segments are not processed at the very same place in the brain. In a neurological study, Mesgarani et al. (2014) tried to find where exactly phonological segments are processed. They focused on the superior temporal gyrus, a brain area covering Wernicke's area which is involved in perceptual processing of speech (Figure E2a). By measuring neural activity with a large set of electrodes in this brain area on a relatively large set of subjects that underwent brain surgery during the measurement, processing areas of phonological segments could be located at high resolution.

Their key finding was this: sounds which share similar manner of articulation (MoA) are processed closely to each other, sounds sharing similar place of articulation (PoA) are distributed all over the analyzed area. Thus, sounds with similar MoAs form patches, and each of these patches includes sounds with different PoA. Put differently, MoA has a better discriminating function when it comes to the location of perceptual processing than PoA does.

The finding converges with the established notion in phonology that sonority, which is tightly linked with MoA fulfils a key discriminating role in perception. This is particularly relevant to phonotactics, where large intersegmental sonority differences facilitate perception of diphones (Berent et al., 2007; Ulbrich et al., 2015). Thus, it seems that diphones are processed more easily if their segments differ in MoA. Can we quantify the relative impact of intersegmental difference in MoA and difference in PoA in the perception of consonant diphones?

Together with Kamil Kaźmierski, I conducted an experiment in which speakers of Polish were exposed to consonant diphones that do not actually exist in Polish (Baumann and Kaźmierski, 2017). Crucially, the diphones differed as to their intersegmental differences in MoA and PoA. In a discrimination task (ABX), we estimated how fast the respective diphones are processed. We found that large differences in MoA uniformly increased processing speed, while both small and large differences in PoA led to lower response times, i.e. quick processing (Figure E2b). This agrees with the findings referred to above: MoA functions as a good discriminator in sound perception.

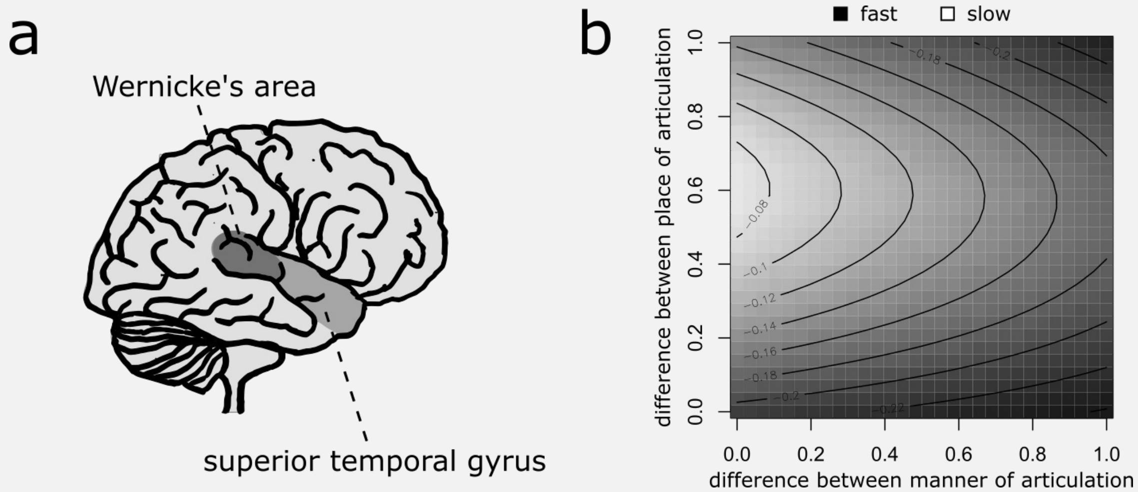


Figure E2. (a) Location of the superior temporal gyrus (gray) examined by Mesgarani et al. (2014). (b) Effects of difference in MoA and PoA on response time in Baumann & Kaźmierski (2017), dark regions denoting fast responses (exponential GAM with inverse link and significantly non-trivial smooth tensor-product term of difference in MoA and difference in PoA).

These small-scale effects may accumulate over many speech events and generations so that diphones with MoA-wise too similar segments are selected against. In this thesis, I have a look at the impact of differential articulatory differences on the acquisition, frequency, and diachronic success of diphones (Baumann and Wissing, submitted).

Under the latter approach, speakers store episodic memories of phonotactic items similar to memories of phonological items described in the previous section. As a consequence, speakers have probabilistic phonotactic knowledge, which in turn can be exploited for the segmentation of the speech stream into words. Since speakers know how often a particular diphone occurs across a boundary or within a word, respectively, diphones which only rarely surface word internally function as markers of word boundaries (Daland and Pierrehumbert, 2011; Jusczyk et al., 1999; McQueen, 1998). For example, /lw/ hardly ever occurs word internally. This knowledge can be exploited to decompose the string /kaɹlwɪnz/ into two words *Carl* /kaɹl/ and *wins* /wɪnz/. This function of signaling boundaries can be extended to the morphemic (sublexical) level as well, so that phonotactics helps the speaker in decomposing words into morphemes (Dressler and Dziubalska-Kołaczyk, 2006; Hay and Baayen, 2005). Here, the string /wɪnz/ is decomposed into the morphemes *win* /wɪn/ and *-s* /z/ based on the information that /nz/ rarely occurs morpheme-internally in English (word-internal /nz/ spans a morpheme boundary in more than 95% of all tokens; estimate based on Ritt et al. 2017). The boundary-signaling function on the morphological level figures prominently in this thesis, as we will see in Section 4.

We see that as with single sounds discussed in the previous section, there are arguments for abstract as well as for detailed phonotactic knowledge (Ernestus, 2014; Pierrehumbert, 2016). Consequently, the model of phonotactic representations adopted in this thesis will assume both (Figure 5a). This is legitimate since there is no reason why speakers might not have both kinds of knowledge. It has been shown that lexical and syntactic knowledge is highly redundant (Beekhuizen et al., 2013; Bod, 2009; Tremblay et al., 2012). In this light, it is difficult to argue that below the lexical level constraints of representational economy apply.

In other words, speakers are assumed to have detailed knowledge consisting of clouds of stored exemplars which encapsulate information about articulatory and acoustic details (such as the overall acoustic duration of a diphone or movement of the articulators). At the same time speakers know details about the external context (such as the social context a diphone was used in), or details about the linguistic context of encountered items (such as whether or not a diphone instance spans a word or morpheme boundary). The statistical information stored in this way is illustrated in Figure 5b. Since speakers implicitly have probabilistic knowledge of properties of phonotactic exemplars (like the distribution of acoustic duration of the /ks/ diphone in the lower left corner of Figure 5b), they can infer which value of the feature is most prototypical for that property (e.g. the range of duration which occurs most often in speech; cf. Pierrehumbert, 2001).

On the other hand, speakers are assumed to have abstract or schematic phonotactic knowledge. That is, they have access to the categorical information of the segments a diphone is composed of (dark gray box in Figure 4a) and any information linked to these segments (like corresponding phonological features such as rough articulatory categories which have been made available through previous linguistic interactions; Grodzinsky and Nelken, 2014; Moreton, 2002; Taylor, 1995).

Why is it relevant to commit oneself to a model of phonotactic representations like the one sketched in Figure 5? In this thesis, I focus on how phonotactic knowledge spreads through populations of speakers and how phonotactic knowledge and use changes over time. Clearly, then, it is necessary to agree on a working hypothesis of the mental setup of these bits of phonotactic knowledge which are subject to change. A framework for modeling the evolution of phonotactic knowledge will be discussed in the next section.

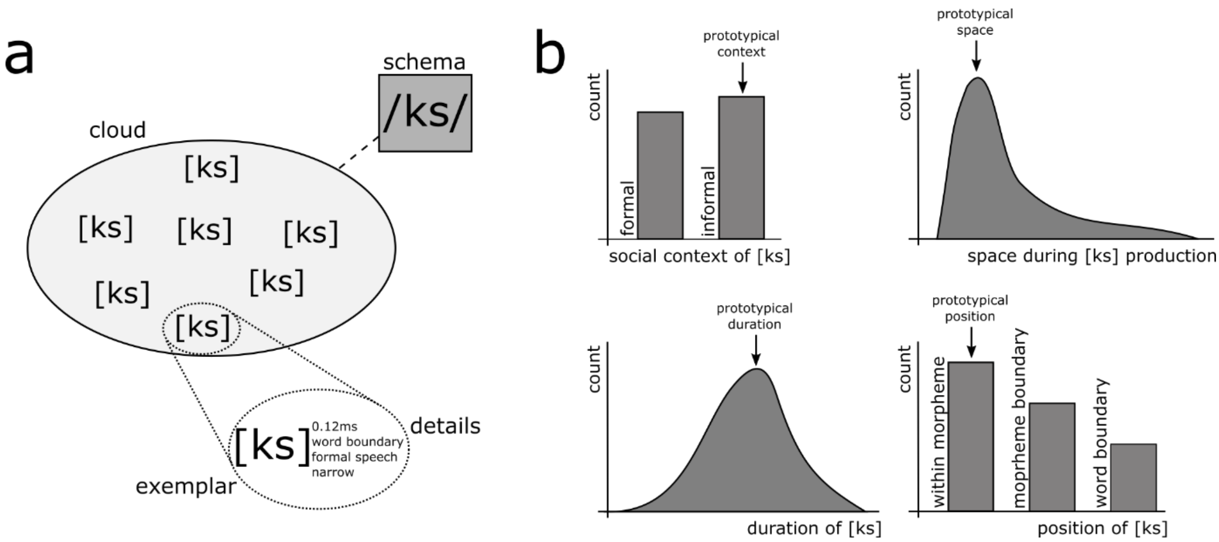


Figure 5. Model of phonotactic representations: (a) Speakers have episodic memories (exemplars) of diphone utterances (here: utterances of the [ks] diphone). Each exemplar captures detailed information. The set of exemplars is associated with an abstract phonotactic schema which encapsulates similarities shared among all exemplars. (b) Speakers have access to distributional knowledge in various dimensions, for example social context, articulatory information like the space left between the tongue and the palate, acoustic information like duration, and morphological information like the position of the diphone. The prototypical value for each dimension is implicitly given by some measure of central tendency of the respective distribution.

3. Evolutionary linguistics and phonotactics

3.1 Diphones as replicators

In this thesis, I adopt the framework of evolutionary linguistics to model phonotactic change (Brighton et al., 2005; Croft, 2000; McMahon and McMahon, 2013; Ritt, 2004). Evolutionary linguistics builds on 'generalized Darwinism' which generalizes the originally biological concepts of replication, mutation and selection to make them applicable to domains other than biology, such as economics, culture and cognition (Aldrich et al., 2008; Dawkins, 1976; Hull, 1988). Applied to linguistics, the framework represents an opposition to the primarily essentialist approach to language which predominates linguistic research (Croft, 2002). While essentialist approaches (like structuralism and generativism) analyze synchronically static linguistic systems as entities (whose knowledge is at the same time shared by all speakers in a linguistic community), evolutionary linguistics adopts a population based approach which foregrounds diachrony and dynamic processes of interacting linguistic items.[5] Languages are used the way they are not because of any universal rules (that are either innate or imposed by some higher authority) but because the constituents they are built of have been successfully transmitted through many generations in a vast number of linguistic interactions in the past. It is a logical consequence of this that fully understanding linguistic knowledge presupposes a diachronic – or evolutionary – approach. In the past decades, the evolutionary study of language has become increasingly relevant. Many conferences and workshops have been devoted to the

[5] Evolutionary linguistics is of course not the only research strand to emphasize diachrony and interactions among speakers. Many cognitive and/or functional linguists and historical socio-linguists share this view, e.g. Bybee (1994); Heine and Kuteva (2007); Trudgill (2001).

evolutionary approach (such as the Evolang and Protolang conference series), and a few years ago, the research field launched its own journal (Journal of Language Evolution).[6]

A fundamental assumption in the framework of evolutionary linguistics is that linguistic items, like words for instance, are culturally transmitted replicators. In general, replicators are units which make sufficiently similar copies of themselves. In biology, these units may be genes (which replicate by generating copies of DNA) or prions (which replicate by reshaping particular proteins). In linguistics, replicators are linguistic entities. There are diverging ideas as to whether linguistic replicators (or 'linguemes', Croft, 2000) are mainly external units (i.e. utterances of linguistic entities; Croft, 2000, 2013) orinternal (i.e. mental bits of linguistic competence; Jäger and Rosenbach, 2008; Ritt, 2004; Rosenbach, 2008). Some scholars argue that the replicator notion includes both (McCrohon, 2012). Moreover, both relatively constrained linguistic items (like words or phonemes) as well as entire languages (or their grammars) have been modeled as linguistic replicators (Jäger, 2008b; Nowak and Komarova, 2001; Pagel, 2009).

The process of replication is not always exact, i.e. it is subject to variation. In genetic replication, this is known as mutation (parts of the copied DNA differ from the original one). In linguistic replication, variation can occur at multiple stages. During production, items like words can be pronounced in a novel way either completely accidentally or partially accidentally due to some factors imposed by the linguistic context (like a previously uttered word from which certain articulatory features are erroneously transferred to the subsequent word). Similarly, linguistic items can be perceived inaccurately, for instance due to interfering acoustic noise.

The key component of generalized Darwinism is that replicators are subject to environmental constraints, that is, they have to consume resources or energy to persist and to minimize degradation. Most fundamentally, there is only a finite amount of space available to populations of replicators. In linguistics, it is clear, for instance, that the number of speakers ('interactors') is finite, as is the number of utterances each speaker can produce, and each speaker only has limited memory. Thus, this constrained environment can accommodate only a limited number of replicators, and each replicator type can produce only a limited number of copies of itself. For example, if words are conceptualized as replicators then it is clear that the growth of a word type, say *stalagmite*, is limited. It does not make sense, semantically speaking, to utilize *stalagmite* in every single utterance; saying *stalagmite* takes time; speakers can forget the word *stalagmite*; and there is clearly only a limited number of speakers who can adopt that word at the moment. As a result of this, (populations of) different replicator types compete for these limited resources, and this competition may lead to the extinction of one competitor.

The latter observation is important, since inaccurate copying gives rise to new replicator types which compete with their resident progenitors. If a new mutant replicator copes better with the environmental constraints than its resident version, i.e. if the mutant replicator is better adapted to the environment than the resident, it may replace the latter. This is precisely the principle of natural selection driven by differential replication in constrained environments generalized from biological evolution.

[6] It should not be unmentioned that the field of evolutionary linguistics also covers biological aspects of language evolution. Here, the focus is on biological traits that enable humans (and other animals) to communicate. In this thesis, however, I restrict myself to the cultural evolution in human language. See for instance Kershenbaum et al. (2016) for a review article on animal communication with a particular focus on phonotactic and syntactic structures.

The notion of conceptualizing cognitive representations as replicators has been criticized. Points of criticism often belong to one of the following three categories: the linguistic version of the replicator concept is flawed because (a) unlike biological replicators like genes or prions, linguistic replicators are very likely not discrete clear cut units, because (b) replication in the cultural and cognitive domain is said to be considerably less accurate than in biology, and because (c) linguistic replicators cannot be physically identified (Deumert, 2003; Sperber, 1985).

As to (a) it has been demonstrated by Henrich and Boyd (2002), however, that modeling evolutionary systems does not necessarily require discrete replicating units, as long as selective forces are strong enough. In that case, models of continuous replicators can be approximated by discrete replicator models (Henrich and Boyd, 2002). In other words, working with replicators allows for a certain degree of fuzziness.

Criticism (b) does not pose a problem either. For one, there is no a priori reason not to apply the principles of Darwinian selection to a certain domain (e.g. cultural evolution) only because copying fidelity is low (Aldrich et al., 2008). Second, replication is the process of creating sufficiently similar copies, and it clearly depends on the specific problem and its level of abstraction what ‘sufficiently similar’ means. In particular, acknowledging that human speakers have the capability of inferring abstract and schematic categories we could for instance say that two speakers share the same word if they use it in a largely similar way (note incidentally, that also in the biological domain, successful replication of, say, a prion is exact only on a certain level of abstraction: proteins are flexible and there will always be fine grained physical differences between a prion and its copy; what matters, however, is structural similarity, i.e. its molecular architecture).

The third group of counter arguments, (c), questions the validity of the linguistic replicator concept since we do not know enough about how exactly they are physically (e.g. neurologically) encoded (see discussion in Deumert 2003, section 3). It is certainly correct that our current knowledge of how exactly cultural replicators are encoded and that research on this matter is difficult since neurological processes are relatively widely distributed (but see Excursion 2). However, this does not prevent us from building theories on that concept (as a matter of fact, in biology genes were discovered in the mid-20th century, long after Darwin’s death).

For the purpose of this thesis, I will treat mental representations of sound sequences (as described in the previous section) as linguistic replicators. Diphone representations are assumed to be transmitted from one speaker to another either during the process of first-language acquisition or in any other linguistic interaction through production and perception. During this replication process, changes may occur, to the effect that novel mutant versions of a particular diphone type enter the arena (for instance, because a diphone is uttered much shorter or with an articulatorily different segment than usual, or because it spans a morpheme boundary for the first time). Diphone versions compete, and the one type which copes better with its environment (e.g. the articulatory or perceptual organs, or the linguistic context like frequently co-occuring items) will spread more successfully through the speaker population, surface more frequently and remain within speaker memories more easily.

At this point, it should have become clear that an evolutionary approach to language in which linguistic representations are conceptualized as replicating pieces of knowledge which subject to (a) pressures imposed by bodily constraints, (b) interactions among individuals and (c) the physical environment fits remarkably well with the cognitive commitment and the embodiment hypothesis outlined in Section 2.1. Also there, the interaction of the body with the world and the dynamics of populations of interacting individuals are considered central (Cowley, 2014; Mompeán, 2006; Steffensen and Fill, 2014; Välimaa-Blum, 2009b; but see Bryson, 2008 for some contrasting discussion).

If one wants to study dynamics as those outlined in the previous paragraphs more thoroughly, some computational tools are required. Darwinian evolution is characterized by competitive interaction among replicator populations. As soon as multiple populations of replicators interact, things get too complicated to make reliable eyeballing predictions. Fortunately, the field of mathematical ecology and evolution provides many useful tools which can be transferred to the study of linguistic evolution in general and phonotactic evolution in particular. These tools will be briefly sketched in the following section.

3.2 Population dynamics of phonotactics

One of the advantages of the evolutionary approach to language is that it allows for easy transfer of methodology. Thus, phylogenetic methods have been fruitfully applied to investigate diachronic relationships among (groups of) languages (e.g. Bowern, 2010; Gray and Atkinson, 2003), and agent-based simulation techniques are widely used to analyze complex interactions among several individuals (see Steels, 2011 for a review). In this thesis, I focus on analyzing linguistic evolution with tools from dynamical-systems theory. More precisely, this means that conceptual models of interactions among populations are translated into sets of mathematical equations (Hofbauer and Sigmund, 1998; Nowak, 2006; Otto and Day, 2007; Solé, 2011). The dynamics of these equations (or better: of the populations they are supposed to model) are then inspected analytically to learn about whether any particular events are possible, likely, unlikely or, in fact, impossible under the previously formulated assumptions. For instance, models can predict that one replicator will out-compete another one, or that multiple replicator types may stably coexist.

The application of dynamical systems theory to linguistic evolution is not new. There is a large body of research on modeling linguistic dynamics with tools from game theory and replicator equations (Baumann and Ritt, 2017; Jäger, 2008a, 2008b; Mitchener and Nowak, 2004; Nowak, 2006; Nowak and Komarova, 2001). Moreover, ecological as well as epidemiological differential-equation models have been applied extensively to model linguistic and cultural change (Abrams and Strogatz, 2003; Cavalli-Sforza and Feldman, 1981; Kandler et al., 2010; Niyogi, 2006; Nowak, 2000; Solé et al., 2010; Sonderegger and Niyogi, 2013; Wang and Minett, 2005; Yang, 2000).

When formulating a mathematical model of a linguistic dynamics, one clearly needs to know precisely what exactly it is that is supposed to change. If we study the dynamics of, say, the diphone /ks/, what exactly is dynamic about that diphone? Let us distinguish between three levels: (a) speaker/interactor level, (b) token/utterance level, and (c) feature/trait level (Figure 5a).

On the speaker level (a), we model how many speakers use a particular linguistic item, e.g. /ks/. Thus, the unit of the population is the speaker rather than a linguistic token. What changes is the abundance of a certain representation in the population of individuals. The population is split into multiple subpopulations, for instance one which consists of individuals who know and use /ks/ and another subpopulation of learners which do not yet know and use that diphone. Mathematically, this is modeled by defining the rate of change in the number of users of an item as a function of the current abundance of the item and, potentially, some additional environmental factors like interacting items (Figure 5b, upper box). Mathematical (differential-equation) models which operate on the speaker level are often formally equivalent with compartment models of epidemiological spread where populations are divided into disjoint subpopulations of infected individuals ('users' of a disease) and susceptible individuals (potential 'learners' of the disease), respectively (Anderson and May, 1991; Cavalli-Sforza and Feldman, 1981; Nowak, 2000). I will exploit this conceptual similarity in order to transfer insights from mathematical epidemiology to the field of language evolution (Baumann, forthc; Baumann and Ritt, submitted; see Section 4).

On the token level (b) it is instances (or exemplars) of linguistic items which are the units of the population. That is, we analyze the output produced by individuals carrying a certain representation without explicitly modeling the individuals themselves. Under this view, what changes is the usage frequencies of linguistic items. Thus, models are given by equations in which the rate of change in frequency of an item is a function of its current frequency (and potential environmental factors like interacting items; Figure 5b, middle box). Relative token frequencies of competing linguistic items can then be translated into relative strengths of entrenchment. From an empirical point of view, the token level is more straight forward than the speaker level since utterance frequencies can be measured more easily than numbers of speakers using a certain linguistic item; utterance frequencies are usually approximated by measuring normalized frequencies in large text corpora (consisting of transcribed spoken or written material). The token level suggests itself for modeling interactions among linguistic items (such as effects of priming or inhibition among items within utterances).

Both, level (a) and (b) refer to ecological dynamics. The feature level (c) is substantially different. Here, the question we ask is whether a linguistic item such as the diphone /ks/ evolves over time, i.e. whether it changes one of its properties or 'traits' such as acoustic duration. In that scenario, we model the rate of change in a quantitative continuous property as a function of the fitness of a replicator with a given trait value in a given environment (possibly defined by competing variants of the replicator with different trait values). The computation of this fitness function is not always straight forward, but we can use the mathematical framework of adaptive dynamics, or evolutionary invasion analysis, to derive it from population dynamical models defined on the speaker or token level (Dercole and Rinaldi, 2008; Dieckmann and Law, 1996; Doebeli, 2011; Geritz et al., 1998); see Figure 5b, lower box. The approach is eco-evolutionary in the sense that it accounts for evolution based on previously defined ecological models.

The linguistic applications of evolutionary invasion analysis are, as far as I know, relatively rare (but see Doebeli, 2011 for an application to socially driven linguistic diversification and Page and Nowak, 2002 for discussion of the link between adaptive dynamics, game theory and ecology). Nevertheless, I will make extensive use of this methodology to model evolutionary dynamics of continuous phonotactic traits (namely the evolution of the ratio of boundary spanning instances of a diphone and acoustic duration; Baumann and Kaźmierski, 2016; Baumann et al., 2016a; Baumann and Sommerer, submitted; see Section 4).

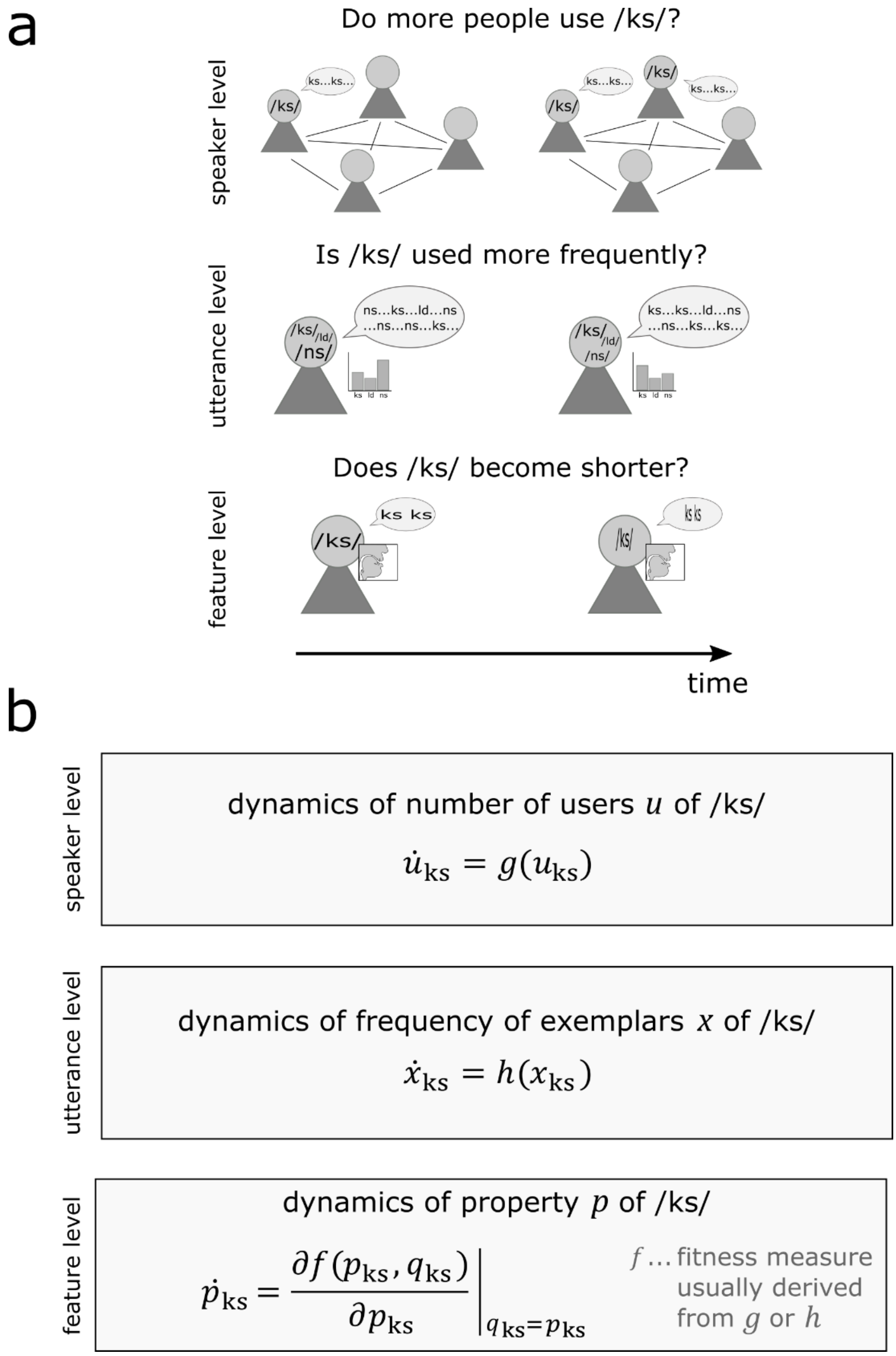


Figure 6. (a) Different levels on which evolutionary dynamics can be modeled. On the speaker/interactor level, populations consist of interactors carrying cognitive representations (replicators). The focus is on abundances of replicators. On the token/utterance level, frequencies of utterances corresponding to linguistic replicators are subject of investigation. The feature level addresses dynamics in continuous traits of linguistic replicators. (b) Ordinary differential equation models which can be used to investigate ecological and evolutionary dynamics on different levels. On the speaker and token level, rates of change are defined by functions depending on current abundances/frequencies (and potential environmental factors). On the feature level, dynamics can be analyzed by defining the rate of change in a trait value as a function of fitness. In the adaptive dynamics framework, fitness is derived from underlying ecological dynamics (e.g. on the speaker or token level).

3.3 Selection pressures in phonotactic evolution

Through the past couple of sections, I have discussed the relationships between phonotactics and cognition and different approaches to defining mental representations of phonotactic items. I have then outlined in which way phonotactic change can be conceptually and computationally modeled by concepts and tools from evolutionary theory. The goal of this thesis is to learn about the various factors – or selection pressures – that drive phonotactic evolution. These pressures may be weak, but they eventually constrain any possible evolutionary pathways (Kirby, 2012; Kirby et al., 2007). That is, they entail which phonotactic systems are more likely to be observed than others.

In what follows, I will give a short description of the various factors that I look at in my studies. These factors can be coarsely grouped into three categories: (a) general cognitive determinants, (b) linguistic determinants, and (c) external determinants, although the boundaries between these categories are fuzzy and most of the factors are intertwined. In fact, in the light of embodied, extended and situated cognition (Clark, 2013), any of the factors described below can be classified as relevant to cognition. Nevertheless, the classification makes it easier to identify the level of organization at which a certain selection pressure applies.

The first group encompasses cognitive factors and mechanisms which are not specific to language, such as learning, frequency and entrenchment, analogy and priming. Before phonotactic items are used they obviously have to be learned, either during the process of first language acquisition or at a later age. The question is, what promotes the learnability of phonotactic items, i.e. what factors lead to early acquisition and which diphones are acquired late? Studies like MacNeilage and Davies (2000) have shown that historically old phonotactic items which are evolutionarily successful in that they show high diachronic stability are those which occur very early during first-language acquisition. This begs the questions whether there is a systematic relationship between phonotactic acquisition and evolution?

Learning is often related with utterance frequency. In general, linguistic items which are frequently used are also acquired relatively early (Brown, 1973; Diessel, 2007; Kuperman et al., 2012). Frequently employed representations are thought to be more entrenched. Repeated activation leads to routinized processes like articulation of a word in a certain context (Bybee, 2010; Croft, 2000; Diessel and Hilpert, 2016; Schmid, 2016). Representations (and their corresponding routines) which are more entrenched are thought to be diachronically more stable, i.e. they show a certain resistance against processes of linguistic change. Pagel et al. (2007) have demonstrated that frequency of usage predicts the extent to which words are subject to phonological change. In their account, utterance frequency has a conserving function. The role of frequency is not straight forward, however. Indeed, it has been argued that frequency can promote change through erosion and deletion effects (Bybee, 2007). Here, frequently employed words undergo reduction effects because their high probability of occurrence makes them predictable, and predictable items do not need to be that explicit. Furthermore, uttering words which are long, (i.e. which carry much formal substance), requires effort, and if these words are produced frequently, the pressure of reducing this effort increases, which may lead to reduction. These facts directly lead to the following questions: What effect does frequency have on the phonotactic level? Is frequency promoting or inhibiting phonotactic change and what is the relationship between acquisition and change in phonotactic evolution?

Finally, similarity and concomitant analogy and priming effects are relevant to phonotactic evolution (Ferreira and Bock, 2006; Loebell and Bock, 2003; Pickering and Ferreira, 2008). Specifically, similarities between diphones that span boundaries (like /ks/ in *she likes*) and those that occur within morphemes (/ks/ in *box*) are relevant to phonotactic research. Do speakers transfer phonotactic patterns which frequently occur across word or morpheme boundaries to the morpheme-internal level (Hogg and McCully, 1987)? Or do word-internal phonotactic patterns support the production of structurally similar boundary spanning patterns? With regards to the latter, acoustic duration may play a crucial role. Diphone types are typically shorter if they span boundaries than if they surface morpheme internally (Kemps et al., 2005; Plag et al., 2015). Additionally, short linguistic items have been argued to benefit from structurally similar long items via asymmetric priming effects (Jäger and Rosenbach, 2008). What would that imply for the coevolution of boundary spanning and morpheme internal phonotactic items?

The second group of factors, linguistic determinants, is related to what was discussed in the previous paragraph. Diphones are embedded into linguistic systems which may consist of sounds, morphemes (meaningful building blocks of words), words, constructions of words etc. Much research has been done on the relationship between phonotactic and word segmentation, the main result being that speakers make use of their phonotactic knowledge to segment the continuous speech stream they perceive into discrete words (Daland and Pierrehumbert, 2011; Jusczyk et al., 1999; Vitevitch et al., 1997). More recently, the same idea was applied to the morphemic level: diphones assist the speaker in decomposing words into morphemes, e.g. stems and attached affixes (Calderone et al., 2014; Celata et al., 2015; Dressler and Dziubalska-Kołaczyk, 2006; Hay and Baayen, 2005). Thus, diphones have the function of signaling boundaries. The latter strand of research, referred to as morphonotactics, has led to the hypothesis that diphone types which are ambiguous in signaling boundaries (like /ks/ in *box* and *likes*) are dispreferred, i.e. they show inhibited processing. The evolutionary consequence of the pressure of confidently signaling boundaries then would be that diphone types evolve in such a way that they either occur exclusively across boundaries or exclusively within morphemes (Dressler et al., 2010; Ritt and Kaźmierski, 2015). The question now is whether diphones that signal boundaries in an ambiguous way are really processed slower than confidently signaling diphone types and whether the evolutionary long-term effects are as predicted.

Phonotactic items are composed of single sounds, i.e. phonemes with their own representations (see Section 2.2), and a diphone's ease of production and perception obviously depends on its components. One major factor which has been subject to phonotactic research is that of articulatory contrast between the segments of a phonotactic item. As discussed in Sections 2.3 and Excursion 2, phonotactic items are processed more easily if they exhibit a certain amount of contrast between their building blocks. Three dimensions are particularly relevant: place of articulation (Frisch et al., 2004; Shatzman and Kager, 2007), manner of articulation together with sonority (Berent et al., 2007; Clements, 1990; Ulbrich et al., 2015), and phonation or voicing (Coetzee, 2014). The question I am primarily interested in is which of these dimensions is most relevant to phonotactic evolution and whether effects of articulatory differences are similar in phonotactic acquisition and change.

The first two groups discussed above collect phonotactic selection pressures which are restricted to individual speakers (within-speaker mechanisms like entrenchment and priming) or interactions among few individuals (e.g. articulation and perception). The third group of selection pressures addressed in this thesis consists of those which can be located on a higher level of organization, namely the population level. Linguistic evolution is thought to be affected by demographic characteristics. So, population size has been shown to be positively correlated with the richness and diversity of the lexical and phonological system of the language spoken by that population (see Nettle, 2012 for a review). Likewise, grammatical complexity has been suggested to be lower in large populations (Bentz and Winter, 2013; Lupyan and Dale, 2010). In this regard, research on the phonotactic level, which is located between the lexicon and phonology, is considerably less explored (but see Maddieson, 2013; Rama, 2013). The question is whether the diversity and size of the inventory of phonotactic representations of a language is determined by demographic factors such as population size, population density or populated area? What role does the network structure of the speaker population play and in what way are phonotactic replicators influenced by fluctuations on the demographic level?

4. Description of subprojects and results

The studies collected in this thesis try to tackle the questions outlined in the previous section. In the following I describe each of the studies briefly. An overview is provided in Table 1. It also provides information about their current status in the publication process (a more detailed list of contributions can be found in the appendix).

Table 1. Subprojects of this thesis together with their central topic and status in the dissemination process

Study	Topic	Status
Baumann and Ritt, submitted	learning, frequency	Revision submitted (11/2017), Cognition
Baumann and Sommerer, submitted	priming, frequency	Revision submitted (01/2018), Language Dynamics and Change
Baumann and Kaźmierski, 2016	analogy, boundary signaling	Published (09/2016), Yearbook of the Poznan Linguistics Meeting
Baumann et al., 2016b	analogy, boundary signaling	Published (11/2016), Papers in Historical Phonology
Baumann and Kaźmierski, submitted	boundary signaling	Revision submitted (01/2018), Language Sciences
Baumann and Wissing, submitted	articulatory contrast, learning	Submitted (10/2017), Stellenbosch Papers in Linguistics
Baumann, forthc.	population size, stochasticity	Accepted (12/2017), Proceedings Evolang 12
Baumann and Matzinger, submitted	demographics, network structure	Submitted (11/2017), Journal of Language Evolution

Baumann and Ritt (submitted) investigates the relationship between acquisition, frequency and change in phonotactics. This is done by analyzing a population-dynamical model of linguistic spread (on the speaker level, cf. Section 3.2). Standard results of mathematical epidemiology are exploited to provide a link between language acquisition and change. Through this link (the so-called basic reproductive ratio; Dietz, 1993; Nowak, 2000), data from diachronic research can be directly compared with data from language-acquisition research. It is shown theoretically and empirically that diphones with low age-of-acquisition ratings (i.e. early acquired diphones) show higher rates of diachronic growth than diphones with high age-of-acquisition ratings. Moreover, by controlling for utterance frequency we show that it is only items which belong to the phonotactic periphery which suffer from frequency effects. Phonotactic core items (i.e. those which are acquired early) are not negatively affected by utterance frequency. We argue that the mechanism behind this is entrenchment. Core phonotactic items are highly entrenched and therefore not affected by frequency-driven deletion effects, very much in contrast to weakly entrenched periphery items. We suggest that this is because late acquired items simply have less time to get entrenched or because of higher cognitive plasticity at early ages (Bybee, 2010; Monaghan, 2014).

A study which addresses priming effects among diphones is Baumann and Sommerer (submitted). Here, we formulate a population dynamical model (on the token level; cf. 3.2) in which two variants of a linguistic replicator, a longer one and a shorter one, compete. This competitive interaction is subject to asymmetric priming so that short items suffer less from the presence of long items than the reverse (Jäger and Rosenbach, 2008). Additionally, the model features a term which accounts for the trade-off imposed by articulatory and perceptual effort attested in diphones (Kuperman et al., 2008). Subsequently the model is analyzed with tools from evolutionary invasion analysis (see Section 3.2) to investigate the long-term evolution of duration. A key finding is that under certain conditions long and short variants of a linguistic replicator can stably coexist. We argue that this provides a mechanism for explaining the empirically attested but semiotically unexpected coexistence of boundary spanning and morpheme internal variants of diphones (because the latter are typically longer than the former ones). Also the respective roles of frequency ease of production and perception, and their relationship to acoustic duration in phonotactics (Kuperman et al., 2008) figures centrally in this study.

The interplay of analogy effects and morphological boundary signaling in phonotactic evolution is studied in Baumann and Kaźmierski (2016) and Baumann et al. (2016b), albeit from two different angles. The latter study empirically investigates whether patterns in the evolution of the English consonant cluster inventory provide evidence for mutually supporting effects among boundary spanning and morpheme internal clusters (via structural similarity). This is done by inspecting Middle and Early Modern English diachronic corpus data (Ritt et al., 2017). In contrast, the former study (Baumann and Kaźmierski, 2016) assumes mutually supporting effects and boundary signaling functions of diphones *a priori*. These factors are then built into a structured-population model formulated on the token level (cf. 3.2; see also Baumann et al., 2016a). The model is then analyzed on the feature level to simulate the distributional evolution of morpheme internal vs. boundary spanning versions of diphones. A key finding of this study is that languages which are very sensitive to ambiguous phonotactic boundary signaling (presumably languages which make much use of morphology such as Polish) yield phonotactic inventories consisting of either exclusively boundary spanning or exclusively morpheme internal diphone types. In contrast, languages which are less sensitive to ambiguous boundary signaling stably accommodate ambiguous diphone types (ultimately as a reflex of structural similarity).

Ambiguity in morphological boundary signaling is also subject of Baumann and Kaźmierski (submitted), but in this case approached from an experimental point of view. Here, we analyze the processing of cluster types which have different degrees of ambiguity. That is, some of these clusters almost always occur morpheme internally while others also surface across morpheme boundaries and yet others are predominantly used across boundaries. In a discrimination task (AX) we show that if participants are primed for analyzing a nonce word as morphologically complex then cluster types which are perfectly ambiguous show the slowest responses. We also show that ambiguity in signaling boundaries has the strongest effect on phonotactic processing if it is operationalized by means of token frequencies (fractions of utterance frequencies of diphone types) rather than type frequencies (fractions of number of words a diphone type occurs in). We suggest that this illustrates that speakers do have probabilistic morphonotactic knowledge (which fits well with an exemplar approach to phonotactics, cf. Section 2.3) and that speakers make use of self-contained phonotactic representations rather than always inferring phonotactic patterns from word types.

Related to this, we show in Baumann and Kaźmierski (2016) that the long-term effects of ambiguity in signaling boundaries systematically depend on how sensitive speakers are with respect to ambiguity. We suggest that languages which make less use of morphological operations (e.g. English) are characterized by weak ambiguity pressures while languages which are highly inflectional/derivational (e.g. Polish) are characterized by strong ambiguity pressures. Diachronically, the former establish phonotactic inventories in which ambiguous cluster types can stably exist while the latter establish largely non-ambiguous inventories. This weakens one of the hypotheses central to morphonotactic research (Dressler and Dziubalska-Kołaczyk, 2006; Korecky-Kröll et al., 2014): the applicability of the phonotactic function of indicating morphological complexity depends on the overall structure of a language rather than being uniform across languages.

In Baumann and Wissing (submitted), we investigate the articulatory (and perceptual) dimension of phonotactics. The study consists of two parts. The first one addresses the effects of intersegmental articulatory differences (in terms of place of articulation, manner of articulation and phonation) of consonant clusters on their acquisition in Dutch. In the second part, we inspect the effects of these articulatory differences on the diachronic success of consonant clusters in the formation of the Afrikaans consonant-cluster inventory (which evolved from Dutch). In both studies it is shown that manner of articulation seems to represent the strongest selection pressure, while phonation and place of articulation show less clear effects.

The role of demographic factors is finally addressed in the two remaining studies Baumann (forthc.) and Baumann and Matzinger (submitted). In Baumann (forthc.), I again adopt a modeling approach. I formulate a population-dynamical model of linguistic spread on the speaker level which is restricted to finite (and relatively small) populations. I then derive a measure of the stability of linguistic items and show that stability increases with population size. Subsequently, I extend the model to capture stochastic effects during the process of linguistic transmission (i.e. during learning). I show that variability during the learning process (e.g. fluctuating exposure to an item or fluctuating number of informants per learner) negatively affects the stability of linguistic items. As a consequence, linguistic items (such as diphones) are assumed to be more likely to get lost in small populations or if the learning process is disturbed by some external ecological factors such as migration (Atkinson, 2011; Bromham et al., 2015; Trudgill, 2004).

The second study (Baumann and Matzinger, submitted) takes an empirical angle. Here, we analyze the co-development of phonotactic diversity and richness in English on the one hand, and several related linguistic and demographic factors on the other hand. Crucially, we also include properties of the speaker network which can be derived from population size into our analysis (Barabási, 2016). We show that it is factors which are linked to speaker density (population density; network clustering) and population spread (populated area; network diameter) which are most relevant to phonotactic diversity. The results are interesting as they provide a more direct link between linguistic evolution and population size (Bybee, 2011; Nettle, 2012).

The specific key findings of my dissertation project are summarized in Table 2 below.

Table 2. Specific key findings of my research

Level	Insight	Studies
Cognitive	Frequency has diminishing effects only in less entrenched items in phonotactic periphery.	Baumann and Ritt, submitted
Cognitive	Weak asymmetric priming can lead to diversification of phonotactic items into long and short variants.	Baumann and Sommerer, submitted
Linguistic	Speakers show gradual sensitivity with respect to the reliability at which diphones signal morphological structure.	Baumann and Kaźmierski, submitted
Linguistic	Weakly morphological languages can stably accommodate phonotactic items that unreliably signal morpheme boundaries.	Baumann and Kaźmierski, 2016; Baumann et al., 2016b
Linguistic	Manner of articulation is more relevant to phonotactic evolution than place of articulation.	Baumann and Wissing, submitted
Population	Phonotactic diversity is determined by population density and the extent to which people form clustered groups.	Baumann and Matzinger, submitted
Population	Long-term stability of linguistic items depends on fluctuations during acquisition.	Baumann, forthc.

5. Conclusion and outlook

In the first section of this companion paper, I pointed out that the goal of this dissertation project is to investigate several determinants of phonotactic evolution. A selection of specific determinants can be found in Table 2 above, and I assume these findings to be of interest for people conducting research in phonotactics. On a more general level, the following insights can be gained from my research:

I demonstrate in my work that language can be studied from an evolutionary point of view. This insight is certainly not new, as I already outlined in section three. This approach is very fruitful as it allows to explain synchronic phenomena from a diachronic point of view by just relying on a small set of general mechanisms: variation and differential reproduction. Many studies have analyzed change in phonemes or words through the evolutionary lens, i.e. linguistic items which are by and large seen as chunks. The point I would like to make is that it makes sense to study the cultural evolution of language at various different levels of grammar, i.e. that strings of linguistic items can be studied as constituents in their own right. This is legitimate as long as there is good evidence that there are sufficiently strong pressures acting on the whole string of items rather than just acting on its parts separately. We have seen that this is arguably so in the case of phonotactics (e.g. articulatory differences between segments, boundaries between segments, perception-articulation trade-off). Above the word level, strings of items have been studied in evolutionary terms (see for instance Zehentner, 2017 for an evolutionary approach to construction grammar). This project can be understood as an attempt to help filling the gap between single sounds and words on the sublexical level (see Rama, 2013 for a related approach).

We have seen that investigating linguistic diachrony makes it possible to learn about cognition in general. This is so because diachrony helps us to detect even very weak biases. These biases become visible through a vast amount of linguistic interactions across multiple generations. What this dissertation project makes clear is that cultural evolution is subject to biases on many different levels. First, the transmission of (linguistic) knowledge is influenced by other (linguistic) knowledge. I illustrated that there is a bias against unreliable boundary signaling. Crucially, this signaling function depends on other parts of the linguistic system as well (morphology). Priming effects imposed by nearby linguistic items during transmission also come to mind. Second, transmission of knowledge depends on how well it is entrenched. In particular, I have shown that phonotactic knowledge are transmitted more reliably if it is deeply entrenched. Third, transmission of knowledge depends on physiological factors. In phonotactics, there seems to be a bias for perceptual contrast. Fourth, transmission of knowledge depends on the mix and number of people one interacts with. A relatively instable or encapsulated local network of interacting individuals does not seem to promote transmission of knowledge.

All of this illustrates that cognition "arises from bodily interactions with the world", just to repeat one part of the quote from the beginning of this companion paper (Thelen et al., 2001, p. 1). Evidently, this converges with an approach towards cognition that encompasses these dimensions and which embraces cognition as embodied and situated rather than as a computational device which manipulates isolated symbols (Clark, 2013; Cowley, 2014).

The evolutionary approach to the study of cognition (and language, in particular) benefits from the rich methodological toolkit provided by the life sciences. In my project, I have demonstrate that mathematical methods from dynamical-systems theory can be fruitfully applied to the study of cognitive phenomena. They help us t (Baumann and Matzinger, submitted)o unravel complex interactions among multiple cognitive biases and their effects on the long-term evolution of cognitive systems. I want to highlight two of my contributions. First, the application of tools from the adaptive-dynamics framework can be used to study evolutionary long-term developments of cognitive traits. In this project, for instance, I study the long-term evolution of duration as well as distributional properties of phonotactic items. What makes this framework so powerful is that it allows to link the evolution of cognitive traits to properties of the environment (e.g. complex interactions with co-occurring phenomena). In doing so, it combines concepts from ecology and evolutionary game theory. At the moment, non-biological applications are relatively limited, as far as I know (except for a couple of studies on technological evolution, evolution of religion, and dialectal diversification; see Dercole and Rinaldi, 2008; Doebeli, 2011). I think that this set of mathematical tools can find many other applications in the study of cultural and cognitive evolution. In linguistics, the evolution of properties of words that are embedded into linguistic systems immediately come to mind. Baumann and Sommerer (submitted) already goes into that direction by investigating the evolution of the grammatical status of words (thereby addressing problems of grammaticalization theory; Heine and Kuteva, 2007; Hopper and Traugott, 2003).

Second, the usefulness of methods and insights from mathematical epidemiology should be mentioned. Although standard epidemiological models are relatively established in the quantitative study of cultural and cognitive evolution (Cavalli-Sforza and Feldman, 1981; Nowak, 2000) I believe that they can be exploited much further. One result that I exploit is the relationship between age of infection and dynamic behavior provided by the quantity known as the basic reproductive ratio (Dietz, 1993; Heffernan et al., 2005; Hethcote, 1989). Its advantage is that it allows to link data from different domains. In my application (Baumann and Ritt, submitted), I use this quantity to link linguistic age-of-acquisition ratings with growth data from historical linguistics. Importantly, the model is based on a mechanistic link between these domains. That is, it is more informative than straight forward correlation measures. I have demonstrated that this link can be established on the phonotactic level. It would be interesting to see if it works on the word level, for instance by investigating the acquisition and diachronic behavior of lexical innovations (cf. Monaghan, 2014; Pagel et al., 2007). Are early acquired words diachronically more successful and does this relationship hold across languages? More generally, does it hold for acquired cognitive skills other than language?

Another result I exploit is the effect of environmental fluctuations during transmission. Epidemiological theory has shown that fluctuations during transmission inhibit spread of disease through populations (Gray et al., 2011; Greenhalgh et al., 2015). I use this result to show that fluctuating exposure to linguistic input during language acquisition can lead to loss of linguistic constituents. This converges with recent findings from Newberry et al. (2017) who show that words which exhibit much fluctuation in usage are prone to extinction. A systematic analysis of this matter including empirical data from language acquisition and change in multiple languages suggests itself.

Let us come back to phonotactics. At first glance, studying sequences of sounds like /ks/ in *hyrax* may not strike one as particularly spectacular. However, seeing the multitude of factors that are responsible for the successful transmission of these items one has to acknowledge that phonotactic evolution is a surprisingly complex phenomenon. This dissertation is meant to disentangle a share of its complexity.

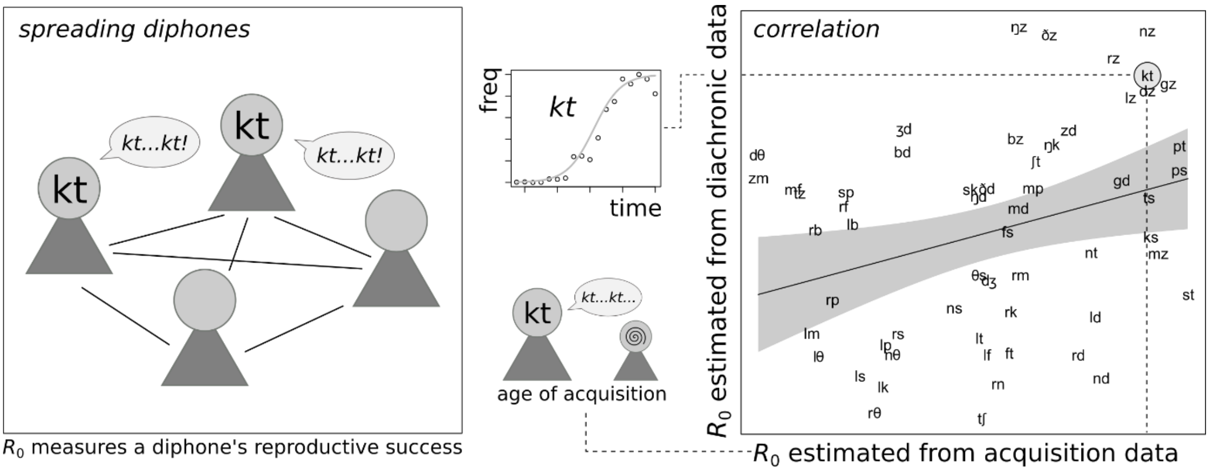
The basic reproductive ratio as a link between acquisition and change in phonotactics

Andreas Baumann & Nikolaus Ritt

Abstract

Language acquisition and change are thought to be causally connected. We demonstrate a method for quantifying the strength of this connection in terms of the 'basic reproductive ratio' of linguistic constituents. It represents a standardized measure of reproductive success, which can be derived both from diachronic and from acquisition data. By analyzing English data, we show that the results of both types of derivation correlate, so that phonotactic acquisition indeed predicts phonotactic change, and *vice versa*. After drawing that general conclusion, we discuss the role of utterance frequency and show that the latter only exhibits destabilizing effects on late acquired items, which belong to phonotactic periphery. We conclude that – at least in the evolution of English phonotactics – acquisition serves conservation, while innovation is more likely to occur in adult speech and affects items that are less entrenched but comparably frequent.

Keywords: diachronic linguistics, language acquisition, reproductive success, basic reproductive ratio, phonotactics, dynamical systems



1 Introduction

Languages are systems of mental instructions that are shared by their speakers. They are instantiated in the mind-brains of many individuals and transmitted across generations through communicative interaction and language acquisition. For a constituent of linguistic knowledge to be successfully transmitted across generations, it needs to be used and expressed by adult speakers in such a way that new generations can acquire it successfully. Thus, the history of language constituents depends on language use and language acquisition and is likely to reflect constraints on both of them. This paper focusses on the relation between history and acquisition.

That language acquisition is crucial for language history is trivially true and generally acknowledged (Briscoe, 2008; Smith & Kirby, 2008). After all, constituents that are not acquired cannot survive. However, the matter is both more complex and more interesting than that. On the one hand, there is considerable disagreement about how much language acquisition contributes to linguistic change, and on the other hand, some correlations between acquisition and diachronic stability appear to be quite specific. For instance, Monaghan (2014), demonstrates that the age at which a lexical item is acquired predicts the diachronic stability of its phonological form. The finding has inspired various attempts to account for it, but no consensus has been reached. On one interpretation, early acquisition is thought to cause diachronic stability: early acquired items become strongly entrenched, get to be used frequently, and are therefore more likely to be historically stable than items that are acquired later (MacNeilage & Davis, 2000; Monaghan, 2014). On another view, early acquisition and diachronic stability are thought to have common causes: items will both be acquired early and remain diachronically stable if they are easily produced, perceived, or memorized, for example.

This paper explores the relation between the diachronic stability of linguistic constituents and the age at which they are acquired. To determine how systematic that relation is, we introduce and test a rigorous quantitative model that relates patterns attested in historical language development to patterns attested in language acquisition. More specifically, we show how age-of-acquisition and diachronic stability can be related to each other in terms of a standardized measure of reproductive success, namely their 'basic reproductive ratio' (henceforth R_0) (Dietz, 1993; Heffernan, Smith, & Wahl, 2005). That measure (more on it below, see 2.1) has proved useful in the study of population-dynamics. We use a population dynamic model[1] that has already been applied to explain linguistic phenomena (Nowak, 2000; Nowak, Plotkin, & Jansen, 2000) and show in which way estimates of R_0 can be derived for linguistic constituents. Crucially, they can be derived both from age-of-acquisition data and from diachronic corpus evidence. By comparing the two estimates, one can then put numbers on the relation between language acquisition and language history. Thus, the model provides a method for relating data of different origins mechanistically.

Empirically, our discussion is based on English word-final CC diphones (i.e. consonant clusters containing two segments). They are short, yet clearly structured linguistic constituents (Kuperman, Ernestus, & Baayen, 2008), and have had long and diverse histories. For instance, the word final cluster /nd/ as in English *land* is likely to have existed already more than 5000 years ago in Indo-European, the ancestor of English. It still thrives today. Many others, however, such as /gz/ or /vz/ as in English *legs* or *loves*, emerged much more recently, i.e. about 800 ago in the Middle English period. There are also considerable differences among the histories of individual clusters as far as their frequencies are concerned. Some of them, such as /xt/ – graphically still reflected in words like *kni<u>ght</u>* or *lau<u>ghed</u>* – have disappeared altogether.

Since (a) there is considerable diversity among the historical developments of final consonant clusters, and since (b) the ages at which they are acquired are similarly diverse, English consonant clusters are highly suitable for our purpose. They allow us to see clearly whether the reproductive ratios that population dynamic models derive from historical evidence and acquisition data actually correlate or not. We show that they do and interpret this as proof of the concept that models which derive R_0 for linguistic constituents are capable of relating language acquisition and language history in a meaningful way.

Thus – and although we are interested in the specific phenomena we investigate – our primary concern is in fact more general. In the context of testing the usefulness of population dynamic models for linguistic purposes, we address questions such as the following: (a) Does the age at which consonant clusters are acquired correlate with their historical stability? (b) Is there a single measure that relates these two properties? (c) What can be learnt from such measurements about causal relations between language acquisition and language history?

For (a) and (b), our study suggests positive answers: models developed in the study of evolutionary dynamics do indeed provide systematic and quantifiable correlations between the

---

[1] That model we use is similar to mathematical models of cultural and linguistic change (Cavalli-Sforza and Feldman (1981); Wang and Minett (2005); Niyogi (2006)) and equivalent to basic epidemiological models (Anderson and May (1991); see also Sperber (1985)).

historical development of final clusters and the age at which are acquired. With regard to (c), we ask if the correlation between acquisition and diachronic stability differs between morpheme internal clusters (such as /mp/ in *lamp*) and morphologically produced ones (such as /gz/ in *eggs*), and whether the correlation between age-of-acquisition and historical stability is affected by utterance frequency. We show that the morphological status of clusters does not seem to matter much, but that the correlation between age-of-acquisition and historical stability is tighter among frequent than among rare clusters. Our results corroborate the view that phonological change may be more strongly driven by frequent use in adult speech (Bybee, 2007), and that early acquired core items are more resistant against frequency-driven effects like reduction, assimilation, or deletion. Thereby, our study contributes to the debate on the role which language acquisition plays in language change.

In terms of its general approach, our paper relates to a growing body of research that views culturally transmitted knowledge in evolutionary terms and models it accordingly (Cavalli-Sforza & Feldman, 1981; Dawkins, 1976; Henrich & Boyd, 2002; Newberry, Ahern, Clark, & Plotkin, 2017). It is also based on the view that the repeated learning events involved in cultural history can amplify and make visible cognitive biases that are too weak to be traceable in the behavior of individuals (Reali & Griffiths, 2009; Smith et al., 2017; Smith & Wonnacott, 2010).

We describe our modeling approach together with both ways of estimating the basic reproductive ratio in Section 2. After that, we introduce the statistical tools (3) which are used to empirically test our model against data from phonotactic acquisition and diachrony. The results of our analysis (4) are finally discussed in Sections 5 and 6, thereby particularly focusing on the effect of utterance frequency.

2 Data and methods

2.1 Standardizing reproductive success: basic reproductive ratio

Our analysis employs a modified version of the population dynamical model of linguistic spread proposed by Nowak and colleagues (Nowak, 2000; Nowak et al., 2000; Solé, 2011). For each linguistic constituent, i.e. in our case for each cluster, the model consists of two differential equations that track the growth of the number of 'users' U (speakers that know and use the cluster), and the number of 'learners' L that do not (yet) know or use it.

When users and learners meet, learners acquire the cluster at a rate $\alpha > 0$, whereby they become users (i.e. switch from class L to class U). Conversely, at a rate $\gamma = 1/G$, where $G > 0$ is linguistic generation time, users 'die' (i.e. are removed from class U) and learners are 'born' (i.e. added to class L). The respective rates of change thus read

$$\begin{aligned}\dot{L} &= -\alpha LU + \gamma U \\ \dot{U} &= \alpha LU - \gamma U\end{aligned}$$

where we set $L + U = 1$.[2]

The expected number of learners that acquire a cluster from a single user introduced into a population of learners is $R_0 = \alpha/\gamma$ (Hethcote, 1989). R_0 represents what has been labelled 'basic reproductive ratio' (Anderson & May, 1991; Nowak, 2000). It figures centrally in epidemiological research due to its straightforward properties: whenever it holds for a population (e.g. a subpopulation of infected individuals) that $R_0 > 1$, that population increases in size and spreads.

In our model, $R_0 > 1$ entails that the population of users approaches a stable equilibrium $\hat{U} = 1 - \gamma/\alpha = 1 - 1/R_0$, so that $\hat{L} = 1/R_0$. If, on the other hand, $R_0 < 1$, the fraction of users approaches 0. The linguistic item vanishes.

R_0 represents a standardized measure of reproductive success that reflects the diachronic stability of linguistic items. Its greatest asset is that it can be derived from different types of data and that all derived estimates are situated on the same scale. Thus, estimates derived from different data types can be compared directly and without further transformation. In our paper, we exploit this for comparing the R_0 derived from diachronic frequency data to the R_0 derived from language-acquisition data. We show that such a comparison yields interesting perspectives on the relation between age of acquisition and historical stability.

2.2 Estimating reproductive success from diachronic growth

The model of linguistic spread outlined in the previous section can be reformulated in terms of a logistic equation (Hethcote, 1989; Solé, Corominas-Murtra, & Fortuny, 2010) with an intrinsic (potentially negative) growth rate $\rho = \alpha - \gamma$. Thus, if the linguistic generation time $G := 1/\gamma$ and the growth rate ρ are known, then α and $\alpha/\gamma = 1 + \rho G =: R_0^{\mathrm{GR}}$ can be determined. We approximate G, i.e. the average time it takes for new language learners to enter the population, by biological generation time, so that $G \cong 30$ years (Worden, 2008). This leaves the intrinsic growth rate ρ to be determined.

In order to estimate the intrinsic growth rates ρ of final CC clusters, we use logistic growth rates r_{lg} obtained from diachronic frequency data as a proxy (see also the discussion in section 5). For that purpose, we determine a trajectory of normalized token frequencies f from 1150 to 2012 for each word-final CC cluster. The token frequencies were retrieved from various historical and contemporary language databases and corpora (see Table 1, which also indicates who carried out the phonological interpretation). The collected data were divided into periods of 50 years, yielding 18 data points for each final CC cluster.

[2] For $\gamma = 1$, the above system is exactly the model of word dynamics in Nowak (2000). In his model, α depends on the utterance frequency and learnability of a word, as well as on the number of informants a learner is exposed to (network density).

Table 1. Diachronic data covering the lineage from Early Middle English to Contemporary American English. Data were binned into periods of 50 years each (e.g. 1200 denoting 1200-1250 below). In the case of overlapping data sets (e.g. PPCMBE2 and COHA in the 19th century) weighted averages based on both corpus sizes were used to compute frequencies. Since we trace the American English lineage (COHA, COCA), phonological transcriptions for the late periods were taken from CMPD.

Sources for frequencies	Covered periods	Phonological interpretation
PPCME2 (Kroch & Taylor, 2000)	1150,1200,…,1450	ECCE (Ritt et al., 2017)
PPCEME (Kroch, Santorini, & Delfs, 2004)	1500,1550,…,1700	
PPCMBE2 (Kroch, Santorini, & Diertani, 2016)	1700,1750,…,1900	CMPD (Carnegie Mellon Speech Group, 2014)
COHA (Davies, 2010)	1800,1850,…,1950	
COCA (Davies, 2008)	2000	

We chose 1150 to 2012 as our observation period because word final CC clusters were rare before (i.e. in Old English). The vast majority of them was only first produced by schwa loss in final syllables, which started roughly at this time (Minkova, 1991). Note that although the phonological process of schwa loss affected word final sequences quite uniformly in the early Middle English period, the different cluster types it produced developed relatively independently of each other after schwa loss was completed (in the 15th century). This reflects the post-medieval influx of loans ending in CC clusters as well as phonological processes other than schwa loss – for instance final devoicing – that produced new clusters. For most of the observation period the dynamics of the individual cluster types can thus be considered as relatively independent from each other.

The derived trajectories were normalized to the unit interval with respect to their maximum values, and subsequently fit to a logistic model given by $f(t) = 1/(1 + \exp(-r_{\mathrm{lg}}(t - t_0)))$, where t_0 was set at the middle of the observation period. Non-linear least-squares regression was used to estimate r_{lg} for each cluster. The quality of this estimate depends on the actual shape of the empirical trajectory. Since the model presupposes (positively or negatively) unidirectional development, r_{lg} estimates can be unreliable for clusters who show (inverse) U-shaped developments. Therefore, we also computed Spearman's Rho ($\mathrm{P_{sp}}$) for each cluster. We excluded clusters for which $|\mathrm{P_{sp}}|$ scored below the threshold of 0.1, to rule out clearly non-monotonous developments.[3] This also eliminated

[3] We are grateful to an anonymous reviewer for addressing the issue of non-monotonous patterns. The employed threshold $|\mathrm{P_{sp}}| > 0.1$ is relatively mild, as we wanted to keep our

clusters that occurred only sporadically in a few periods. Finally, we did not consider final cluster types that are absent in Present Day English such as /mb/ in *limb* because there are no data on the age at which they are acquired. Thus, a total of 58 final CC types entered our analysis (Table A1 in the appendix). For the purpose of illustration, Figure 1 shows logistic models for nine different cluster types: for instance, /kt/ exhibits a sigmoid increase in frequency (i.e. $r_{\text{lg}} > 0$ and $R_0^{\text{GR}} > 1$), while /rn/ becomes less frequent ($r_{\text{lg}} < 0$ and $R_0^{\text{GR}} < 1$).

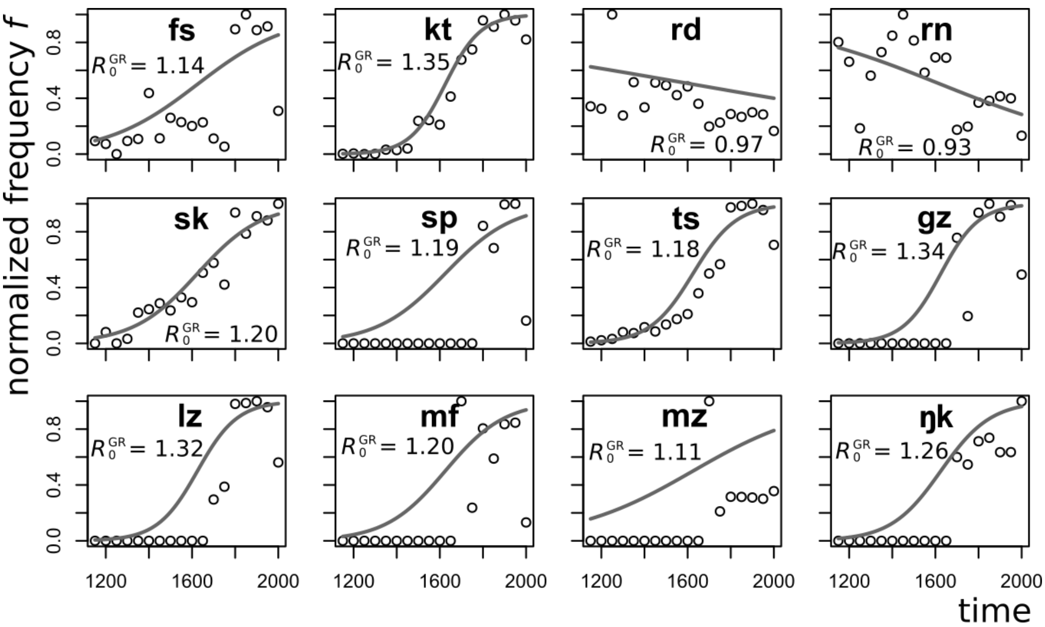


Figure 1. Logistic growth curves for a set of English word-final CC-clusters. All clusters show a non-trivial monotonous development (decreasing or increasing). The graphs were selected in order to represent a large variety of diachronic patterns. In some cases (e.g. /sk/, /ts/, /sk/) trajectories fit the logistic pattern remarkably well. In other cases (e.g. /rn/, /fs/, /sp/) they don't. Some clusters feature extremely low frequencies in early periods.

2.3 Estimating reproductive success from age of acquisition

Next, we derived R_0 estimates from language acquisition data. Here, our derivation follows Dietz (1993). The population of linguistic agents is once again split into a fraction L of 'learners' and a fraction U of 'users' for each linguistic item. AoA denotes the age of acquisition of that item and LE denotes the life expectancy of an individual. Under the assumption of a roughly rectangular age structure (Dietz 1993), at equilibrium $\text{LE}/\text{AoA} = (\hat{L} + \hat{U})/\hat{L} = R_0 =: R_0^{\text{AoA}}$. It is therefore sufficient to estimate AoA, as long as LE is known. For the

data set reasonably large. It excludes only trajectories that are strongly non- monotonous. The qualitative results of this paper still apply up to a threshold of $|\text{P}_{\text{sp}}| \sim 0.3$.

sake of simplicity, we assume a constant life-expectancy of LE $\cong 60$ years (Lancaster, 1990: 8).[4]

Our estimates for the AoAs of 58 final clusters are based on Kuperman et al.'s (2012) AoA ratings for 30,000 English words. These ratings were collected in a broad crowdsourcing study among speakers of American English and correlate highly with ratings obtained under laboratory conditions (see also Monaghan 2014). The AoA of a cluster type was operationalized as the mean of the AoA ratings of the three earliest-acquired word-forms containing it. Averaging over the first three acquired items containing a cluster yields a more robust measure of its AoA than considering only the very earliest word containing it. Since we treat CC clusters as linguistic constituents in their own right (and not just as properties of words), we consider their acquisition to require exposure to more than a single word containing them. Nevertheless, we operationalize the AoA of a cluster as a point estimate that divides the life of a speaker into a period before and a period after acquisition of that cluster (i.e. the transition date from L to U).[5]

Word-forms in which final CC clusters result from morphological operations (such as /gz/ in the plural egg+s) received the AoA rating of the base forms contained the data set (e.g. egg). There are two reasons why this is likely to yield plausible estimates. First, the lowest AoA rating in our data is 2.74, and the majority of English inflectional morphology is acquired during between 2.25 to 3.75 years (Brown, 1973). Furthermore, it has been shown that in languages which are morphologically poor (such as English as opposed to Polish) there is no significant difference between the ages at which morphologically produced and morpheme-internal clusters are acquired (Korecky-Kröll et al., 2014, p. 48). Transcriptions were once again taken from CMPD.

2.4 Utterance frequency

Frequency has often been argued to affect the diachronic stability of linguistic items (Bybee 2007). Thus, Pagel et al. (2007) show that the rate of phonological change in the lexicon can be predicted from the frequency of word use. At the same time, frequent words are acquired earlier than rare ones (Kuperman et al. 2012). This suggests that frequency increases reproductive success. On the other hand, utterance frequency has also been shown to drive

---

[4] Note that the results presented in Section 4 are qualitatively robust with respect to altering life expectancy since R_0^{AoA} scales linearly with LE. Nevertheless, incorporating time dependent LE would represent an interesting but substantially more complex extension of our method.

[5] This operationalization of AoA is most compatible with the underlying population dynamical model. We found that the exact operationalization of AoA is crucial to the comparison of the two derived R_0 estimates. AoA ratings for clusters that are derived from the AoAs of all words containing it get implausibly high because some of those words are inevitably acquired extremely late and unlikely to play any role in the acquisition of a cluster.

phonological erosion. Frequent words are also comparably expectable and therefore more tolerant of reduction (Bybee & Hopper 2001; Diessel 2007). Thus, it is unclear if frequency should increase or decrease the diachronic stability of CC clusters.

In order to investigate that issue, our study takes frequency into consideration as an additional factor. Since cluster-specific utterance frequencies fluctuate during the observation period, we first extracted per million normalized token frequencies for all cluster types in every single period of 50 years. In addition, we computed average token frequencies for each cluster type across all 18 periods, denoted as $\langle \text{frequency} \rangle$ in order to obtain a more compact summary measure (see Table A1 in the appendix).

2.5 Morphology

While syntax or pragmatics have little immediate influence on word internal phonotactics, morphology affects it strongly. Thus, many word-final CC clusters result from morphological operations (Dressler, Dziubalska-Kołaczyk, & Pestal, 2010; Hay & Baayen, 2005). As far as the acquisition of morpheme-internal phonotactics is concerned, however, we do not expect morphology to contribute much (see 2.3). In our observation period, English syntheticity (i.e. the amount of morphological operations) underwent a non-uniform development which exhibits a U-shaped curve, as demonstrated by Szmrecsanyi (2012). Thus, the interaction of morphology and the diachronic dynamics of word-final phonotactics is a priori not so clear. In order to account for morphological effects in our analysis, we classified final CC types as (a) (exclusively) morphologically produced (and 'illegal' within morphemes, e.g. /md/ in *seemed*), (b) (exclusively) morpheme internal ('legal', /lp/ in *help*), or (c) both ('mixed', /nd/ in *hand* and *planned*).

3 Calculation

To explore the relative impact and the interaction of the different factors, we employed linear models (LM) and generalized additive models (GAM, Wood, 2006a). First, z-normalized estimates of R_0^{GR} (the reproductive ratio derived from diachronic growth data) and R_0^{AoA} (the reproductive ratio derived from age-of-acquisition data) entered a LM as dependent and independent variables (Model 1a). No transformation (e.g. log) was needed for either variable. The effect of morphology ('illegal'; 'mixed'; 'legal'; the latter as default) was analyzed by adding a linear interaction term to the previous model (Model 1b).

Analyzing the interaction of frequency with the derived R_0 measures is more complicated because it involves time as an additional factor. Initially (Model 2), normalized (i.e. z-transformed) log-transformed average frequency, $\langle \text{frequency} \rangle$, was integrated as an interacting variable into a GAM, in which R_0^{AoA} figures as predictor and R_0^{GR} as dependent variable. The interaction between R_0^{AoA} and logged $\langle \text{frequency} \rangle$ was modeled by means of a tensor-product term (Wood, 2006b). The effects of logged $\langle \text{frequency} \rangle$ on R_0^{GR} and R_0^{AoA} were then evaluated in two separate GAMs (Model 3a and 3b, respectively). In both of them, logged $\langle \text{frequency} \rangle$ figures as predictor (smooth term). Finally, the interaction of time and logged frequency – both affecting R_0^{GR} and R_0^{AoA} respectively –, was modeled as a tensor product term in two additional GAMs (model 4a and 4b, respectively).[6]

4 Results

The direct comparison of the two estimates of R_0 (model 1a, Fig. 2) reveals a non-trivial linear relationship between the two variables (standardized coefficient $\beta_{\mathrm{AoA}} = 0.31 \pm 0.13SE$ at $p = 0.016$). Adding morphology (model 1b) does not reveal a statistically significant interaction and decreases the explanatory power of the model ($\beta_{\mathrm{AoA}} = 0.20 \pm 0.23SE$; $\beta_{\mathrm{AoA\times mixed}} = -0.04 \pm 0.33SE$; $\beta_{\mathrm{AoA\times illegal}} = 0.48 \pm 0.37SE$).[7] Thus, we can assume the discovered correlation to hold irrespective of morphological status.

Model 2 (Fig. 3a, right) reveals that the relationship between R_0^{GR} and R_0^{AoA}, established in model 1, is much tighter for frequent clusters (e.g. /ns/ as in *hence* vs. /st/ as in *best*) than for infrequent ones, where it is approximately constant (/rp/ as in *harp* vs. /lk/ as in *milk*; interaction term: $df = 4.33$, $F = 4.76$, $p < 0.001$). Another way of looking at Fig. 3a is this: in the phonotactic core inventory (i.e. among early acquired clusters), frequency does not affect diachronic stability, while in the phonotactic periphery (among late acquired clusters), frequency reduces it significantly (Fig 3a, left).

In model 3a (Fig. 3b), $\langle \mathrm{frequency} \rangle$ correlates negatively with R_0^{GR} (smooth term: $df = 1$, $F = 4.20$, $p = 0.045$; linear effect $\beta = -0.24$, $\mathrm{CI}_{0.95} = (-0.50, -0.01)$). Thus, clusters that have been relatively abundant in the history of English have not become more frequent.[8] In contrast, model 3b (Fig. 3b) shows that R_0^{AoA} positively correlates with average frequency (smooth term: $df = 1$, $F = 33.57$, $p < 0.001$; linear effect $\beta = 0.61$, $\mathrm{CI}_{0.95} = (0.42, 0.75)$). Frequent CC clusters are acquired significantly earlier than rare ones. Model 4a (Fig. 3c) shows that frequency and R_0^{GR} were inversely related in the beginning of the observation period but not during more recent periods. The relationship between frequency and R_0^{AoA} (model 4b, Fig. 3c) was slightly negative in the early part of the observation period but

[6] All models based on Gaussian distribution with identity link. The number of knots in smooth terms was deliberately kept low in order to detect monotone and easy to interpret (but still possibly nonlinear) relationships.

[7] Model 1a: $R^2(\mathrm{adj}) = 0.08$, $F = 6.13$, $p = 0.016$, $AIC = 163.56$; model 1b: $R^2(\mathrm{adj}) = 0.10$, $F = 3.05$, $p = 0.04$, $AIC = 164.5$; model 2: $R^2(\mathrm{adj}) = 0.11$, 16.5% explained deviance; model 3a: $R^2(\mathrm{adj}) = 0.05$, 7.00% explained deviance; model 3b: $R^2(\mathrm{adj}) = 0.36$, 37.5% explained deviance; model 4a: $R^2(\mathrm{adj}) = 0.20$, 20.7% explained deviance; model 4b: $R^2(\mathrm{adj}) = 0.33$, 34.1% explained deviance.

[8] Model 3a was additionally fit to all clusters with $R_0^{\mathrm{AoA}} > 1$ ('core' items) and $R_0^{\mathrm{AoA}} < -1$ ('periphery' items), respectively, in order to make the effect of frequency more clearly visible. Core items: smooth term at $df = 1$, $F = 0.58$, $p = 0.47$ ($n = 12$, $R^2(\mathrm{adj}) = -0.04$, 5.47% explained deviance). Periphery items: significantly decreasing smooth term at $df = 3.06$, $F = 25.3$, $p < 0.001$ ($n = 12$, $R^2(\mathrm{adj}) = 0.90$, 92.5% explained deviance).

evolved towards a strongly positive interaction later on (interaction term: $df = 4.6$, $F = 81.8$, $p < 0.001$).

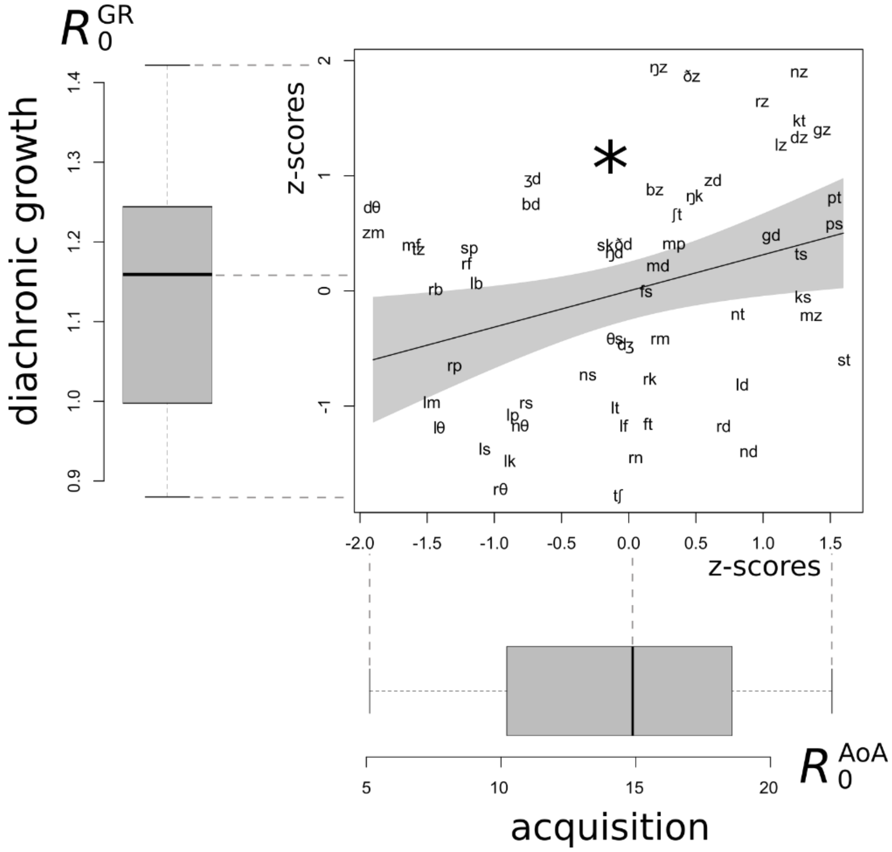


Figure 2. Linear relationship between normalized estimates of R_0^{GR} (vertical axis) and R_0^{AoA} (horizontal axis) (model 1; $p < 0.05$). Gray areas denote 95% confidence regions. Boxplots next to the vertical and horizontal axis indicate the distribution of R_0^{GR} and R_0^{AoA}, respectively. Scores derived from acquisition data are considerably higher than scores estimated from diachronic data.

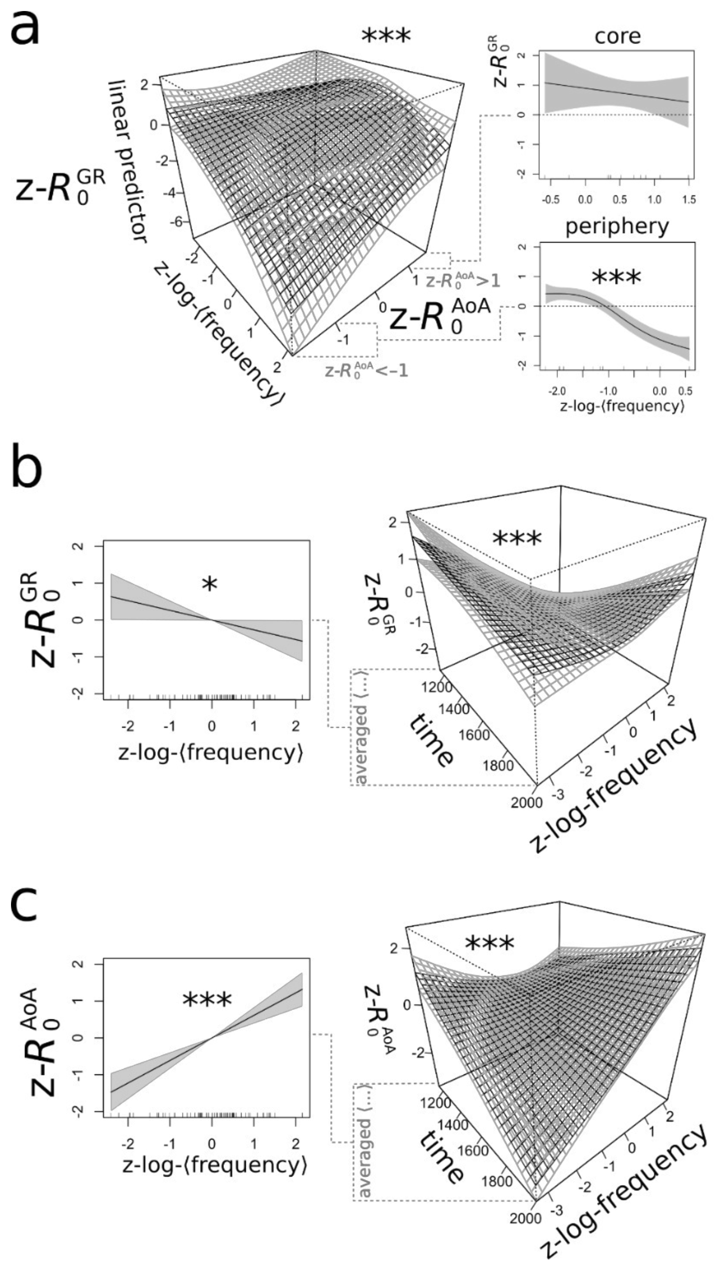


Figure 3. (a) Left: The effect of cross-temporally averaged frequency, ⟨frequency⟩, on the relationship between R_0^{GR} and R_0^{AoA} (z-scores; ⟨frequency⟩ log-transformed; model 2). The positive relationship becomes stronger as ⟨frequency⟩ increases and vanishes in low-frequency items. Right: ⟨frequency⟩ decreases R_0^{GR} significantly when looking at periphery items ($z\text{–}\, R_0^{\mathrm{AoA}} < -1$) but not in the core inventory ($z\text{–}\, R_0^{\mathrm{AoA}} > 1$) (model 3a with restricted data set). (b) Left: ⟨frequency⟩ decreases R_0^{GR} (model 3a). Right: Frequency (log- and z-transformed) computed for each period of 50 years separately and related with R_0^{GR} and time (model 4a). (c) Left: Same as in (b) with R_0^{GR} replaced by R_0^{AoA}, which correlates positively with ⟨frequency⟩ (model 3b). Right: Over the past 800 years, a strongly positive relationship between frequency and R_0^{AoA} established itself (model 4b). Recall that R_0^{AoA} is based on contemporary AoA estimates.

5 Discussion

We have shown that a simple population-dynamical model of linguistic spread derives correlating estimates of reproductive success from age-of-acquisition data on the one hand, and from diachronic corpus data on the other. At least for English final CC clusters, this means that the basic reproductive ratio[9] R_0 qualifies as a standardized measure of reproductive success which allows to the relate AoA with diachronic growth. It has a clear linguistic interpretation and permits the direct comparison of data of various origins (Heffernan et al., 2005).

The correlation between the estimates derived from acquisition data and diachronic evidence supports the widely shared view that age of acquisition and diachronic stability are causally linked. Concurring with Monaghan (2014), our study suggests that what is acquired early is diachronically more stable (and *vice versa*). Interestingly, however, the tightness of this relationship increases with the frequency of CC clusters. This means that frequent clusters are not simply acquired before rare ones, but that the historical stability of a cluster can be more confidently predicted from the age at which it is acquired when that cluster is frequent. Among rare clusters the correlation is not as tight. At the same time, these results show that late acquired items from the phonotactic periphery suffer most from frequency driven effects such as assimilation, reduction, or deletion. In that respect, they differ strongly from early acquired – and highly entrenched – core items. Thus, the notion than utterance frequency reduces historical stability still applies (e.g. via erosion in adult speech; Bybee, 2007), but we have demonstrated it to be restricted to the periphery.

The correlation between frequency and R_0 estimated from AoA is not surprising. It reflects the way in which the (linguistic version of the) basic reproductive ratio is derived. According to Nowak (2000), R_0 depends on (a) the ease with which a linguistic item is learnt and memorized, (b) utterance frequency, and (c) the density of the speaker network. Thus, our results highlight the importance learnability for the successful replication of phonotactic items (Ritt, 2004; Croft, 2000; Smith & Kirby, 2008). In that sense, age of acquisition seems to reflect linguistic and cognitive constraints on the production and the perception of clusters, and on their role in further cognitive processing. These constraints may act on articulatory and perceptual properties of clusters, such as (differences in) the manner or the place of their articulation (Berent, Steriade, Lennertz, & Vaknin, 2007; Mesgarani, Cheung, Johnson, & Chang, 2014), or on their semiotic functionality (such as boundary signaling, see McQueen, 1998, Dressler et al., 2010).

It is interesting that there is no simple positive correlation between R_0 estimated from historical data and utterance frequency. That would have been expected given the way in which Nowak (2000) defines the basic reproductive ratio. It would also have been expected from previous empirical findings, e.g. by Pagel et al. (2007) or Lieberman et al. (2007). In fact, taking frequency averaged over the entire observation period into account the opposite seems to be the case, very much in line with the view that high utterance frequency decreases an

[9] Defined as the expected number of learners that acquire an item from a single user.

item's phonological stability (Bybee, 2007, 2010; Diessel, 2007). So why do our data not reveal such a correlation? First, as discussed above, the effect of frequency on the relationship between both R_0 estimates show that frequency affects diachronic stability negatively among late acquired items, but does not do so among early acquired items. Since Pagel et al. (2007) focused exclusively on core vocabulary (200 lexical core items), which is acquired early, they would not have seen the destabilizing effects of frequency on late acquired items. Lieberman et al. (2007) analyze the loss of 177 irregular verbal forms and find that their stability is positively correlated with frequency. The divergence between their result and ours is noteworthy. We suspect that it reflects that the frequencies employed in Lieberman et al. (2007) were derived from contemporary data (CELEX) rather than historically layered sources: in the slice representing most recent periods in Figure 2b (right), a negative interaction between stability and frequency is not visible either. We think that averaged frequencies, which cover the entire observation period, provide a more robust picture.[10]

Alternatively, there might be fundamental differences between phonotactics and the lexical domain. In the sublexical domain, the destabilizing effect of frequency might be stronger than in the lexical domain, because for the recognition of lexical items listeners can rely on the syntactic, semantic and pragmatic context, and may therefore recognize them even in phonetically reduced forms (Ernestus, 2014). In this regard, cluster perception is supported at best by morphological cues and benefits much less from linguistic redundancy. Therefore, weakly entrenched phonotactic items may be more vulnerable to the destabilizing effects of frequency than weakly entrenched lexical items.

In summary, it appears that linguistic entrenchment is a function of both age of acquisition and frequency rather than just the latter (Ellis, 2012; Schmid, 2016). If we operationalize entrenchment by means of diachronic stability (because of the conserving function of routinization) then our analysis suggests that the relative age at which an items is acquired plays a key role in linguistic entrenchment. One straightforward mechanistic explanation is this: an item that happens to be acquired early has more time for being routinized than an item that is acquired late. Crucially, this holds irrespectively of how frequent an item is. Another mechanism discussed by Monaghan (2014: 533), applies to the lexical domain and involves higher plasticity of the cognitive system at early ages. Lexical items that are acquired early (for whatever reason) are more easily entrenched because the cognitive system is still more flexible. This, then, should also apply to complex processes of cognitive planning, articulation and perception relevant in the sublexical domain (Cholin, Dell, & Levelt, 2011; Levelt & Wheeldon, 1994).[11]

[10] We would like to thank an anonymous reviewer for raising this issue.

[11] According to Nowak (2000), there is a third factor that influences the spread of items, namely network density. It is reflected in the number of users to which a learner is exposed. Thus, changes in the number of communicative contacts could cause socially motivated change in phonotactics (Trudgill (2001)), because R_0 decreases as the social network gets sparse. This relates to studies about the relationship between social structure

Finally, the comparison between the reproductive ratios derived from our two data sets, sheds light on the question how much acquisition contributes to language change. To see this, note that the ratios derived from AoA data are considerably larger than the ones derived from diachronic data (Fig. 2, boxplots). While that difference may partly be an artefact of our method[12], it may also be revealing. Thus, it might plausibly be interpreted as reflecting the different contributions which first-language learners and proficient speakers make to the actuation of linguistic change (Bybee, 2010; Croft, 2000). Since age-of-acquisition data predict greater diachronic stability than is derivable from actual diachronic evidence, this potentially suggests that language use by adults may play a more important role in causing linguistic innovation than language acquisition by new generations of children (Diessel, 2012). Of course, further research is still needed to corroborate this suspicion, but the methods we have demonstrated in this paper may help to make the question addressable in quantitative terms.

6 Outlook

Although our case study has been restricted to a very specific set of phonotactic constituents and to a single language, namely English, there is no *a priori* reason why our approach should not work in other domains (e.g. modeling the spread of single phonemes or words), and for other languages. The two operationalizations of R_0, however, require (a) diachronic data that cover the complete histories of constituents (ideally from the period of their first emergence),

---

and linguistic evolution (e.g. Wichmann, Stauffer, Schulze, and Holman (2008); Nettle (2012)), but based on the data that we analyzed in this study we cannot add to this discussion at this point.

[12] To some extent, the difference may reflect the way in which R_0^{GR} has been estimated, because linguistic tokens and speakers represent two different dimensions in the first place. We suppose our token-frequency based proxy r_{lg} to represent a lower bound for the intrinsic growth rate ρ in the population-dynamical model. This is because the spread of an item in a population of tokens involves both its spread through a population of speakers (i.e. ρ), and its spread through the linguistic system and the lexicon (Kroch (1989); Croft (2000); Denison (2003); Wang and Minett (2005); Blythe and Croft (2012)). The two dimensions are hard to disentangle on the basis of the limited number of historical texts available. Only quantitative empirical and computational approaches that incorporate both dimensions can shed more light on this issue.

As to R_0^{AoA}, one possible reason why it might be overestimated is that our measure of AoA is based on lexical acquisition. Of course, the first form of a word that a child uses may not be the one containing the relevant cluster, nor will a child's first productions of what is a cluster in the target form always be accurate. Moreover, considering only AoA for estimating R_0 neglects the possibility that clusters, once acquired, may disappear again in adult speech – not only through language attrition and articulatory loss (see Seliger and Vago (1991); Ballard, Robin, Woodworth, and Zimba (2001); Torre and Barlow (2009)), but also through natural phonological backgrounding and deletion processes. If the proportion of individuals abandoning a particular cluster is underestimated, this will result in R_0^{AoA} being overestimated.

as well as (b) corresponding acquisition data. As so often, English enjoys a privileged status in this regard. A large number of historical sources have been digitized, and also research on acquisition has produced a large amount of data. Testing the methods described in this study against other languages is likely to face difficulties, although it would of course be important. At least on the lexical level, however, the prospects are not so bad. For core-vocabulary items in 25 languages a set of AoA ratings has been compiled by Łuniewska et al. (2016), and diachronic resources such as the Google Books Ngram Corpus, currently featuring eight languages, may serve as good starting points.

Acknowledgements: This work has been supported by FWF (grant no. P27592-G18).

Appendix

Table A1. Derived scores for each English type of final CC cluster used in empirical analysis: logistic growth rate r_{lg} (2.2); goodness-of-fit measure $\mathrm{P_{sp}}$ (2.2); basic reproductive ratio estimated from logistic growth R_0^{GR} (2.2); age-of-acquisition AoA (2.3); basic reproductive ratio estimated from AoA R_0^{AoA} (2.3); total per million normalized frequency across all periods $\Sigma = 18 \times \langle \mathrm{frequency} \rangle$ (2.4); average frequency across all periods $\langle \mathrm{frequency} \rangle$; morphological status (2.5).

cluster	AoA	R_0^{AoA}	r_{lg}	$\mathrm{P_{sp}}$	R_0^{GR}	Σ	$\langle \mathrm{frequency} \rangle$	morph
bd	5.51	10.88	0.0083	0.86	1.25	2875.39	159.74	illegal
bz	3.9	15.38	0.0089	0.83	1.27	3577.02	198.72	illegal
ðd	4.23	14.18	0.0066	0.76	1.2	1035.56	57.53	illegal
dθ	11.7	5.13	0.0081	0.77	1.24	182.59	10.14	mixed
dz	2.91	20.64	0.0111	0.83	1.33	16066.49	892.58	illegal
dʒ	4.17	14.38	0.0024	0.86	1.07	17120.47	951.14	legal
ðz	3.6	16.67	0.0137	0.86	1.41	624.26	34.68	illegal
fs	3.98	15.08	0.0046	0.7	1.14	4236.11	235.34	illegal
ft	3.96	15.14	-0.001	-0.16	0.97	18692.94	1038.5	mixed
gd	3.06	19.63	0.0069	0.8	1.21	2462.6	136.81	illegal
gz	2.79	21.48	0.0113	0.83	1.34	5024.83	279.16	illegal
ks	2.89	20.79	0.0044	0.86	1.13	47399.45	2633.3	mixed
kt	2.91	20.64	0.0118	0.93	1.35	33376.3	1854.24	mixed
lb	6.74	8.9	0.0049	0.75	1.15	156.01	8.67	legal
ld	3.23	18.58	0.0007	0.47	1.02	127823.96	7101.33	mixed
lf	4.21	14.25	-0.0011	-0.27	0.97	21867.05	1214.84	legal
lk	5.94	10.11	-0.0025	-0.84	0.92	10516.45	584.25	legal
lm	8.26	7.27	-0.0001	0.12	1	4858.57	269.92	legal
lp	5.87	10.22	-0.0007	-0.16	0.98	4273.8	237.43	legal
ls	6.53	9.19	-0.002	-0.56	0.94	25955.21	1441.96	mixed
lt	4.3	13.94	-0.0003	0.12	0.99	18907.59	1050.42	mixed

lθ	7.92	7.57	-0.0011	-0.64	0.97	8198.53	455.47	legal
lz	3	19.98	0.0108	0.84	1.32	40839.21	2268.85	illegal
md	3.87	15.5	0.0057	0.81	1.17	12894.59	716.37	illegal
mf	9.21	6.51	0.0066	0.86	1.2	581.9	32.33	legal
mp	3.73	16.09	0.0065	0.66	1.19	4675.2	259.73	legal
mz	2.85	21.08	0.0035	0.81	1.11	22968.2	1276.01	illegal
nd	3.19	18.81	-0.0021	-0.35	0.94	623823.11	34656.84	mixed
ŋd	4.33	13.86	0.0062	0.84	1.19	1339.24	74.4	illegal
ŋk	3.58	16.78	0.0086	0.86	1.26	10257.91	569.88	legal
ns	4.63	12.95	0.001	0.21	1.03	94903.51	5272.42	legal
nt	3.26	18.4	0.0036	0.97	1.11	133291.44	7405.08	mixed
nθ	5.7	10.52	-0.0011	-0.8	0.97	6894.34	383.02	mixed
nz	2.91	20.64	0.0138	0.83	1.41	71827.44	3990.41	illegal
ŋz	3.88	15.48	0.0141	0.84	1.42	12585.83	699.21	illegal
ps	2.74	21.92	0.0073	0.94	1.22	16989.12	943.84	mixed
pt	2.74	21.92	0.0085	0.95	1.25	15427.24	857.07	mixed
rb	8.1	7.41	0.0047	0.71	1.14	773.34	42.96	legal
rd	3.35	17.89	-0.0011	-0.59	0.97	115745.44	6430.3	mixed
rf	7.04	8.53	0.0058	0.79	1.17	402.81	22.38	legal
rk	3.95	15.2	0.0009	0.27	1.03	11891.15	660.62	legal
rm	3.85	15.58	0.0025	0.89	1.08	9209.52	511.64	legal
rn	4.08	14.69	-0.0025	-0.54	0.93	23164.88	1286.94	legal
rp	7.41	8.09	0.0013	0.29	1.04	1957.53	108.75	legal
rs	5.61	10.7	-0.0002	-0.28	1	51490.02	2860.56	legal
rθ	6.13	9.78	-0.0037	-0.91	0.89	20723.15	1151.29	mixed
rz	3.11	19.29	0.0125	0.83	1.38	23445.87	1302.55	illegal
sk	4.42	13.58	0.0065	0.96	1.2	4500.53	250.03	legal
sp	6.95	8.63	0.0063	0.76	1.19	860.12	47.78	legal
st	2.69	22.28	0.0017	0.75	1.05	164960.88	9164.49	mixed
ʃt	3.73	16.09	0.0078	0.95	1.24	14280.96	793.39	illegal
ts	2.9	20.71	0.0062	0.92	1.18	71384.23	3965.79	mixed
tʃ	4.24	14.16	-0.004	-0.6	0.88	96962.87	5386.83	legal
θs	4.32	13.9	0.0026	0.4	1.08	62.73	3.49	illegal
tz	8.85	6.78	0.0064	0.76	1.19	90.09	5	illegal
zd	3.43	17.51	0.0093	0.94	1.28	22371.96	1242.89	illegal
ʒd	5.51	10.9	0.0093	0.92	1.28	6219.11	345.51	illegal
zm	11.66	5.14	0.007	0.74	1.21	152.89	8.49	legal

References

Anderson, R., & May, R. (1991). *Infectious diseases of humans: Dynamics and control*. New York: Oxford University Press.

Ballard, K. J., Robin, D., Woodworth, G., & Zimba, L. (2001). Age-related changes in motor control during articulator visuomotor tracking. *Journal of Speech, Language, and Hearing Research*, *44*, 763–777.

Berent, I., Steriade, D., Lennertz, T., & Vaknin, V. (2007). What we know about what we have never heard: Evidence from perceptual illusions. *Cognition*, *104*(3), 591–630. https://doi.org/10.1016/j.cognition.2006.05.015

Blythe, R. A., & Croft, W. (2012). S-curves and the mechanism of propagation in language change. *Language*, *88*(2), 269–304.

Briscoe, T. (Ed.). (2008). *Linguistic Evolution through Language Acquisition*. Cambridge: Cambridge University Press.

Brown, R. (1973). *A first language: The early stages*. Harvard: Harvard University Press.

Bybee, J. (2007). *Frequency of use and the organization of language*. Oxford: Oxford University Press.

Bybee, J. (2010). *Language, Usage and Cognition*. Cambridge: Cambridge University Press.

Carnegie Mellon Speech Group. (2014). *Carnegie Mellon Pronouncing Dictionary*: Carnegie Mellon University. Retrieved from http://www.speech.cs.cmu.edu/cgi-bin/cmudict

Cavalli-Sforza, L. L., & Feldman, M. W. (1981). *Cultural transmission and evolution: aquantitative approach*. Princeton: Princeton University Press.

Cholin, J., Dell, G. S., & Levelt, W. J. M. (2011). Planning and articulation in incremental word production: syllable-frequency effects in English. *Journal of experimental psychology. Learning, memory, and cognition*, *37*(1), 109–122. https://doi.org/10.1037/a0021322

Croft, W. (2000). *Explaining language change: An evolutionary approach*. *Longman linguistics library*. Harlow, England, New York: Longman.

Davies, M. (2008). *The corpus of contemporary American English (COCA): 450 million words, 1990-2012*: Brigham Young University.

Davies, M. (2010). *The corpus of historical American English: COHA*. Brigham: BYE, Brigham Young University.

Dawkins, R. (1976). *The selfish gene*. New York, Oxford: Oxford University Press.

Denison, D. (2003). Log(ist)ic and simplistic S-curves. In R. Hickey (Ed.), *Motives for Language Change* (pp. 54–70). Cambridge: Cambridge University Press.

Diessel, H. (2007). Frequency effects in language acquisition, language use, and diachronic change. *New ideas in psychology*. (25), 108–127.

Diessel, H. (2012). New Perspectives, Theories and Methods: Diachronic change and language acquisition. In A. Bergs & L. J. Brinton (Eds.), *Historical linguistics of English* (pp. 1599–1613). Berlin [a.e.]: De Gruyter Mouton.

Dietz, K. (1993). The estimation of the basic reproduction number for infectious diseases. *Statistical methods in medical research*, *2*, 23–41.

Dressler, W. U., Dziubalska-Kołaczyk, K., & Pestal, L. (2010). Change and variation in morphonotactics. *Folia Linguistica Historica*, *31*, 51–68. https://doi.org/10.1515/flih.2010.003

Ellis, N. C. (2012). What can we count in language, and what counts in language acquisition, cognition, and use? In S. T. Gries & D. Divjak (Eds.), *Frequency Effects in Language Learning and Processing.* Berlin, Boston: De Gruyter. https://doi.org/10.1515/9783110274059.7

Ernestus, M. (2014). Acoustic reduction and the roles of abstractions and exemplars in speech processing. *Lingua*, *142*, 27–41.

Hay, J., & Baayen, H. (2005). Probabilistic phonotactics and morphological productivity. Retrieved from http://www.ling.canterbury.ac.nz/jen/documents/haybaayenh_esse.htm

Heffernan, J., Smith, R., & Wahl, L. (2005). Perspectives on the basic reproductive ratio. *Journal of The Royal Society Interface*, *2*(4), 281–293. https://doi.org/10.1098/rsif.2005.0042

Henrich, J., & Boyd, R. (2002). On Modeling Cognition and Culture: Why cultural evolution does not require replication of representations. *Journal of Cognition and Culture*, *2*(2), 87–112. https://doi.org/10.1163/156853702320281836

Hethcote, H. (1989). Three basic epidemiological models. In S. Levin, T. Hallam, & L. Gross (Eds.), *Biomathematics: Vol. 18. Applied Mathematical Ecology.*

Korecky-Kröll, K., Dressler, W. U., Freiberger, E. M., Reinisch, E., Mörth, K., & Libben, G. (2014). Morphonotactic and phonotactic processing in German-speaking adults. *Language Sciences*, *46*, 48–58.

Kroch, A. (1989). Reflexes of Grammar in Patterns of Language Change. *Language Variation and Change*, *1*, 199–244.

Kroch, A., Santorini, B., & Delfs, L. (2004). *Penn-Helsinki Parsed Corpus of Early Modern English*. Retrieved from http://www.ling.upenn.edu/hist-corpora/

Kroch, A., Santorini, B., & Diertani, A. (2016). *The Penn Parsed Corpus of Modern British English (PPCMBE2)*: Department of Linguistics, University of Pennsylvania. Retrieved from http://www.ling.upenn.edu/ppche/ppche-release-2016/PPCMBE2-RELEASE-1

Kroch, A., & Taylor, A. (2000). *Penn-Helsinki Parsed Corpus of Middle English*. Retrieved from http://www.ling.upenn.edu/hist-corpora/

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, *44*, 978–990.

Kuperman, V., Ernestus, M., & Baayen, H. (2008). Frequency distributions of uniphones, diphones, and triphones in spontaneous speech. *The Journal of the Acoustical Society of America*, *124*, 3897–3908.

Lancaster, H. O. (1990). *Expectations of Life: A Study in the Demography, Statistics, and History of World Mortality*. New York, NY: Springer New York.

Levelt, W. J., & Wheeldon, L. (1994). Do speakers have access to a mental syllabary? *Cognition*, *50*(1-3), 239–269. https://doi.org/10.1016/0010-0277(94)90030-2

Lieberman, E., Michel, J.-B., Jackson, J., Tang, T., & Nowak, M. A. (2007). Quantifying the evolutionary dynamics of language. *Nature*, *449*(7163), 713–716. https://doi.org/10.1038/nature06137

Łuniewska, M., Haman, E., Armon-Lotem, S., Etenkowski, B., Southwood, F., Anđelković, D.,. . . Ünal-Logacev, Ö. (2016). Ratings of age of acquisition of 299 words across 25 languages: Is there a cross-linguistic order of words? *Behavior Research Methods*, *48*(3), 1154–1177. https://doi.org/10.3758/s13428-015-0636-6

MacNeilage, P. F., & Davis, B. L. (2000). On the origin of internal structure of word forms. *Science*, *288*, 527–531.

McQueen, J. (1998). Segmentation of Continuous Speech Using Phonotactics. *Journal of Memory and Language*, *39*, 21–46.

Mesgarani, N., Cheung, C., Johnson, K., & Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science*, *343*, 1006–1010.

Minkova, D. (1991). *The history of final vowels in English: The sound of muting*. *Topics in English linguistics: Vol. 4*. Berlin, New York: M. de Gruyter.

Monaghan, P. (2014). Age of acquisition predicts rate of lexical evolution. *Cognition*, *133*, 530–534.

Nettle, D. (2012). Social scale and structural complexity in human languages. *Phil. Trans. R. Soc. B*, *367*, 1829–1836.

Newberry, M. G., Ahern, C. A., Clark, R., & Plotkin, J. B. (2017). Detecting evolutionary forces in language change. *Nature*, *551*(7679), 223–226. https://doi.org/10.1038/nature24455

Niyogi, P. (2006). *The computational nature of language learning and evolution*. Cambridge, MA.

Nowak, M. A. (2000). The basic reproductive ratio of a word, the maximum size of a lexicon. *Journal of theoretical biology*, *204*(2), 179–189. https://doi.org/10.1006/jtbi.2000.1085

Nowak, M. A., Plotkin, J., & Jansen, V. (2000). The evolution of syntactic communication. *Nature*, *404*(6777), 495–498.

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

Ritt, N. (2004). Selfish sounds and linguistic evolution: A Darwinian approach to language change. Cambridge University Press, Cambridge.

Ritt, N., Prömer, C., & Baumann, A. (2017). Evolution of Consonant Clusters in English (ECCE): a diachronic phonotactic database. Department of English and American Studies,University of Vienna, Vienna.

Schmid, H.-J. (2016). Linguistic entrenchment and its psychological foundations. In H.-J. Schmid (Ed.), *Entrenchment and the psychology of language learning* (pp. 435–452). Berlin: Mouton de Gruyter.

Seliger, H., & Vago, R. (Eds.). (1991). *First Language Attrition*. Cambridge: CUP.

Smith, K., & Kirby, S. (2008). Cultural evolution: implications for understanding the human language faculty and its evolution. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *363*(1509), 3591–3603. https://doi.org/10.1098/rstb.2008.0145

Smith, K., Perfors, A., Fehér, O., Samara, A., Swoboda, K., & Wonnacott, E. (2017). Language learning, language use, and the evolution of linguistic variation. *Philosophical Transactions of the Royal Society B.* Advance online publication. https://doi.org/10.1098/rstb.2016.0051

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

Solé, R. V. (2011). *Phase transitions*. Princeton: Princeton University Press.

Solé, R. V., Corominas-Murtra, B., & Fortuny, J. (2010). Diversity, competition, extinction: the ecophysics of language change. *Journal of The Royal Society Interface*, *7*(53), 1647–1664. Retrieved from http://groups.lis.illinois.edu/amag/langev/paper/sole2010diversitygsc.html

Sperber, D. (1985). Anthropology and psychology: Towards an epidemiology of representations. *Man*, 73–89.

Szmrecsanyi, B. (2012). Analyticity and syntheticity in the history of English. In T. Nevalainen & E. C. Traugott (Eds.), *The Oxford handbook of the history of English* (pp. 654–665). Oxford: Oxford University Press.

Torre, P., & Barlow, J. A. (2009). Age-related changes in acoustic characteristics of adult speech. *Journal of Communication Disorders*, *42*(5), 324–333.

Trudgill, P. (2001). *Sociolinguistic variation and change*. Edinburgh: Edinburgh University Press.

Wang, W., & Minett, J. (2005). The invasion of language: Emergence, change and death. *Trends in Ecology and Evolution*, *20*, 263–269.

Wichmann, S., Stauffer, D., Schulze, C., & Holman, E. (2008). Do language change rates depend on population size? *Advances in Complex Systems*, *11*(3), 357–369.

Wood, S. (2006a). *Generalized Additive Models: an introduction with R*. Boca Raton, Florida: Chapman & Hall/CRC.

Wood, S. (2006b). Low-Rank Scale-Invariant Tensor Product Smooths for Generalized Additive Mixed Models. *Biometrics*, *62*, 1025–1036. https://doi.org/10.1111/j.1541-0420.2006.00574.x

Worden, R. (2008). Linguistic structure and the evolution ofwords. In T. Briscoe (Ed.), *Linguistic Evolution through Language Acquisition* (pp. 75–100). Cambridge: Cambridge University Press.

Linguistic diversification as a long-term effect of asymmetric priming: an adaptive-dynamics approach

Andreas Baumann & Lotte Sommerer

Abstract: This paper tries to narrow the gap between diachronic linguistics and research on population dynamics by presenting a mathematical model which corroborates the notion that the cognitive mechanism of asymmetric priming can account for observable tendencies in language change. The asymmetric-priming hypothesis asserts that items with more substance are more likely to prime items with less substance than the reverse. Although these effects operate on a very short time scale (e.g. within an utterance) it has been argued that their long-term effect might be reductionist, unidirectional processes in language change. In this paper, we study a mathematical model of the interaction of linguistic items which differ in their formal substance, showing that in addition to reductionist effects, asymmetric priming also results in diversification and stable coexistence of two formally related variants. The model will be applied to phenomena in the sublexical as well as in the lexical domain.

Keywords: asymmetric priming, diversification, unidirectionality, population dynamics, phonotactics, grammaticalization

1 Introduction

This paper introduces a mathematical population-dynamical model on the interaction of closely related linguistic items which factors in the psychological mechanism of 'asymmetric priming' and the relationship between formal substance and utterance frequency. The model can not only successfully predict reductionist tendencies in linguistic change but also diversification, i.e. the stable coexistence of two historically related and formally similar albeit not entirely identical linguistic variants. With this paper we want to contribute to the recent interdisciplinary discussion whether and to which extent asymmetric priming – which is a cognitive mechanism that can also be found in other cognitive domains – can explain aspects of long-term linguistic change.

Hilpert and Correia Saavedra (2016: 3) define asymmetric priming as "a pattern of cognitive association in which one idea strongly evokes another, while that second idea does not evoke the first one with the same force". More explicit items (e.g. semantically and phonologically richer forms) are more likely to prime less explicit items (e.g. semantically bleached and phonologically reduced forms) than the reverse (Shields & Balota 1991); in short 'more substance primes less substance. Although these neurological/cognitive effects operate on a very short time scale, it has been suggested that they are not transient effects but – via implicit learning – can have potential long-term diachronic effects by permanently modifying cognitive representations (Loebell & Bock 2003; Kaschak 2007).

In a programmatic paper, Jäger and Rosenbach suggest that asymmetric priming might be the "missing link" to solve the puzzle of how "performance preferences may come to be encoded in grammars (i.e. on the competence level) over time" (2008: 86). They claim that "what appears as diachronic trajectories of unidirectional change is decomposable into atomic steps of asymmetric priming in language use" (2008: 85). The 'priming triggers language change' argument could be summarized in the following way: asymmetric priming favors the repeated production of certain reduced linguistic forms and supports their successful entrenchment, which diachronically promotes these reduced variants (see section 2 for details on the 'asymmetric priming hypothesis').

Although we do not believe that asymmetric priming is the only driving force in change, we are in favor of Jäger and Rosenbach's idea. We suggest that asymmetric priming can help to explain the long-term reduction of form in a more sophisticated way than the traditional, rather simplistic 'ease of effort' argument (Zipf 1949; André Martinet 1955; Hawkins 2007). Additionally, we will show that our model can also account for the phenomenon of stable diversification on the sublexical as well as on the lexical level if other factors next to asymmetric priming are also considered.

So far, not much has been written on the potential link between asymmetric priming and diachronic change (e.g. Hilpert & Correia Saavedra 2016). Our contribution to the debate is the development of a mathematical model. Our analysis unfolds in two steps. First, we formulate a population-dynamical model of the competition between linguistic items with different degrees of formal substance (Law et al. 1997; Kisdi 1999). The architecture of the model looks roughly like this: On the one hand, it features a term that accounts for the functional relationship between formal substance and frequency (e.g. Zipfian inverse duration-frequency relationship). On the other hand, in order to account for asymmetric priming, the model also features an asymmetric competition term which models the interaction of formally similar items. In a second step, we conduct an evolutionary invasion analysis of the model (Dieckmann & Law 1996; Geritz et al. 1998; Page & Nowak 2002) investigating whether new and formally reduced variants replace their formally rich counterparts. This procedure allows for a simulation of the diachronic long-term development of linguistic items with respect to their formal substance.

We will apply our model to two linguistic domains in order to demonstrate the flexibility of the model: (i) sublexical and (ii) lexical. In our first (sublexical) application, we model the interaction among pairs of sound sequences (more precisely, consonant diphones), in which one sequence is more reduced in terms of duration than its counterpart. Pairs of diphones that are phonemically identical (except for their duration) are an attested phenomenon. For instance, consonant diphones which occur across morpheme boundaries such as /nd/ in *join-ed* are typically shorter than phonemically identical morpheme internal pairs of consonants such as /nd/ in *wind*. The coexistence of morphonotactic (more reduced) and lexical (less reduced) variants of the same consonant-diphone type can be explained well with our model by integrating empirically plausible functional relationships between duration and token frequency.

In the second (lexical) application we investigate grammaticalization. For example, the form *going* evolved from a lexical verb (*I am going to town*) into an auxiliary (*I am going to stay in town*), where the auxiliary is said to be a more grammaticalized (reduced) variant of the lexical verb. Both forms coexist in a stable manner (Hopper & Traugott 2003). With regards to grammaticalization, two hypotheses have been formulated. While Jäger and Rosenbach (2008) claim that more lexical variants of a word asymmetrically prime their more grammaticalized counterparts ('lexical supports grammaticalized', and consequently 'more substance supports less substance'), Hilpert and Correia Saavedra (2016: 15-16) argue that this directionality is in fact reversed in the sense that lexical items are inhibited less by grammatical variants than the reverse. We will investigate both hypotheses. Our model builds on the empirically plausible assumption that substance and frequency in use are inversely related: words are more frequent if they are less explicit (i.e. if they are phonologically short or semantically bleached), and *vice versa*. We argue that neither Jäger and Rosenbach (2008) nor Hilpert and Correia Saavedra (2016) take this inverse relationship into account. If interaction among items unfolds in a way suggested by Jäger and Rosenbach, words are always diachronically reduced in a unidirectional manner, without any possibility of stable coexistence. If, however, the directionality of asymmetric interaction is reversed, then stable diversification of formally similar words can occur under certain conditions.

This paper is structured as follows: In section (2) we inform the reader about the cognitive mechanism of asymmetric priming and its link to linguistic change. Section (3) presents the mathematical model in all its detail. In (3.1) we introduce the general dynamical-systems model, after which we concentrate specifically on the asymmetric competition term in (3.2). This is followed by an introduction to evolutionary invasion analysis (3.3), which is applied to the model in (3.4) in order to derive formal conditions for stable diversification to occur. The model will be applied to the sublexical (mor)phonotactic domain in (4.1) and on the lexical domain (grammaticalization) in (4.2). By means of analytical analyses and simulations, we show that its predictions match with previous empirical observations. We conclude with a discussion of what the model is capable of, but also its limitations.

2 Explaining diachronic change via asymmetric priming

Several typologically universal tendencies can be observed in language change; one being grammaticalization. Grammaticalization has been defined as a development "whereby lexical terms and constructions come in certain linguistic contexts to serve grammatical functions" (Hopper & Traugott 2003: 1). Many scholars see it as an epiphenomenon; an umbrella term for a bundle of composite processes where "linguistic units lose in semantic complexity, pragmatic significance, syntactic freedom and phonetic substance" (Heine & Reh 1984: 15). One major characteristic feature of grammaticalization is the unidirectional[1] erosion of formal substance.[2]

Reductionist tendencies also affect sublexical linguistic items such as strings of sounds within words. For example, the stop /b/ is lost in final /mb/ clusters in words like *thumb* or *limb*, and word final consonant+/s/ clusters are shortened in certain morphological configurations: morphologically produced /rs/ as in *she hears* is more reduced than /rs/ in *Mars* (Plag et al. 2015). Also in this domain, speaker friendly reduction or lenition processes have been shown to be more abundant than their listener friendly strengthening or fortition counterparts (Honeybone 2008).

Another well-known fact is that diachronic change leads to diversification, i.e. the development of new variants, which either compete until one ousts the other or which coexist peacefully. In both cases, the emergence of new variants leads to (temporary or stable) synchronic variation and the existence of formally related variants. Similar to reductionist tendencies, examples of diversification can be found in more than one linguistic domain. Diversification on the lexical level is evident in pairs like $[\text{have}]_{\text{verb}}$ (as in *I have a cake*) or $[\text{have}]_{\text{auxiliary}}$ (as in *I have struggled*), where the two items clearly have different functions (and where the latter is more likely to be reduced; e.g. *I've struggled*). Similarly, we can conceptualize the coexistence of reduced and unreduced ('short' and 'long') homophonous

[1] Although exceptional cases have been listed which contradict unidirectionality claims (e.g. Brinton & Traugott (2005); Himmelmann (2004); Norde (2009)), unidirectionality "is generally accepted as a strong statistical tendency that is in need of an explanation" (Hilpert & Correia Saavedra 2016: 2; Heine & Kuteva (2002)).

[2] We can also observe unidirectional reductionist processes on the semantic level. For example, during grammaticalization, relatively rich, concrete and specific meanings develop more abstract and schematic meanings (but not the other way round).

sound sequences as cases of diversification on the phonotactic (sublexical) level. For example, above-mentioned instance of /rs/ in *she hears* (short) and /rs/ in *Mars* (long).

Diversification has been explained in functionalist terms, by employing discourse-pragmatic arguments like functional necessity; the speaker's wish for 'expressivity' (Lehmann 1985: 10) or 'extravagance' (Haspelmath 1999). Similar expressions are said to survive because they find a semantic niche with a specific function (Breban et al. 2012). On the other hand, reductionist tendencies have most often been explained via the 'ease of effort' principle; signal simplicity (Langacker 1977: 105); or a preference for 'structural simplification' or 'economy' (Roberts & Roussou 2003; van Gelderen 2004). However, many usage-based, cognitive historical linguists have also looked at cognitive motivations for change. For example, analogical or metaphorical thinking are seen as cognitive processes which steer the direction of grammaticalization (Heine et al.; Bybee et al. 1994; Fischer 2007; Smet 2013; Sommerer 2015)[3]. On top of that and rather recently, a very small group has started to discuss and research the potential influence of another cognitive mechanism, namely asymmetric priming.

Priming is a phenomenon and – at the same time – a method in psycholinguistics. As a phenomenon it is defined as "an improvement in performance in a perceptual or cognitive task, relative to an appropriate base line, produced by context or prior experience" (McNamara 2005: 3). Jäger and Rosenbach provide a more 'linguistic' definition: priming is a kind of "preactivation in the sense that the previous use of a certain linguistic element will affect (usually in the sense of facilitating) the subsequent use of the same or a sufficiently similar element (i.e. the 'target')" (2008: 89).

Psychological research on semantic and syntactic priming is extensive and mostly experimental in lexical decision tasks or naming tasks (Bock 1986; Bock & Loebell 1990; Loebell & Bock 2003; Tooley & Traxler 2010; McNamara 2005). Importantly, (forward and backward) priming is often 'asymmetrical'. For example, a concept like [eagle] strongly primes [bird] but less so the other way round. In a similar vein, [Lamp] primes [light] but not the other way round (e.g. Koriat 1981; Neely 1991; McNamara 2005; but also see Thompson-Schill et al. 1998). Note that in all the mentioned cases the prime is semantically 'richer/concrete' and more specific than the target.

Other studies have shown priming effects on the phonetic/phonological level. In their study, Shields and Balota (1991) show that a full form is more likely to prime a phonetically reduced form than the other way round, which is why it has been concluded that "prime targets are more likely to be phonologically reduced than primes" (Jäger & Rosenbach 2008: 98).[4]

This lead to the following hypothesis: more explicit items (e.g. semantically and phonologically richer forms) are more likely to prime less explicit items (e.g. semantically bleached and phonologically reduced forms) than the reverse. With regards to language change, the main point is that this cognitive asymmetry shows the same skewed directionality as frequently observed unidirectional developments in diachrony. Research has shown that priming effects do not always decay immediately right after the target is produced but

[3] Also see Haiman (1994); Diessel & Hilpert (2016); Schmid (2016) for grammaticalization as 'stimulus weakening' triggered by automatization/ routinization and strong entrenchment.

[4] This is supported by other experimental research Fowler & Housom (1987); Diessel (2007); Jurafsky et al. (2001); Ernestus (2014) which shows that there is a general relation between phonetic reduction and expectedness. Expected or more probable items are more likely to be reduced phonetically than unlikely items. Both identity and semantic relatedness of the prime leads to reduction in duration and amplitude of the target and this is strongest under identity.

sometimes persists over various trials (Bock & Griffin 2000); this represents a kind of cumulative priming effect: with repeated trials there is an increased preference of a certain structure (Chang et al. 2006). Thus, "via implicit learning the effects of structural priming may become entrenched in speaker's grammar over time" (Jäger & Rosenbach 2008: 100; Kaschak 2007).

However attractive the hypothesis about the diachronic reflex of asymmetric priming may be, its premise does not seem to hold on the lexical level when facing empirical data, as demonstrated by Hilpert and Correia Saavedra (2016) in a recent experimental study. In fact, they show that the effect of asymmetric priming among related words is reversed, so that phonologically reduced and semantically bleached words are inhibited to a larger extent by lexical and thus phonologically rich and semantically more explicit relatives than the reverse.

With regards to this contradiction, we argue that Jäger and Rosenbach's hypothesis still holds, but only on the formal level. In fact, we will show two things in this paper. First, we demonstrate that *asymmetric priming among phonotactic items* in the directionality suggested by Jäger and Rosenbach (2008), i.e. 'richer forms prime reduced forms', can explain diachronic patterns observable in phonotactic change. Second, we show that if *asymmetric priming among words* works the way which Hilpert and Correia Saavedra (2016) suggest then, under certain conditions, reduction of formal substance still takes place among formally explicit forms. On top of that, asymmetric priming (in either direction) functions as a mechanism that drives diversification without the need of additional explanations like expressiveness or the presence of a semantic niche.

3 The model

3.1 A general Lotka-Volterra model of asymmetric linguistic competition

We model the dynamics of linguistic items as a dynamical system. More specifically, we simultaneously track the token frequencies $x_1, x_2, \dots, x_N$ of $N \geq 1$ formally related linguistic items indexed from 1 to *N*, which are characterized by a formal substance s_1 to s_N, respectively. In our model, formal substance is defined as a one-dimensional continuous positive trait, i.e. $s_i \in \mathbb{R}^+$ for all $i = 1, \dots, N$. For instance, s_i could denote the duration of a linguistic item measured in seconds or the number of phonemes of a word.

As introduced above, we model the development of the abundance $x_1, x_2, \dots, x_N$ of *N* formally related linguistic types numbered from 1 to *N*, depending on their respective formal properties $s_1, s_2, \dots, s_N$ as well as on the interaction among the *N* linguistic items. $x_i \in \mathbb{R}^+$ can be thought of as token frequencies in language use. So, we model the development of continuous traits $s_1, s_2, \dots, s_N$ affecting the development of continuous frequencies $x_1, x_2, \dots, x_N$. This makes it possible to apply our model to linguistic theories which build on detailed memories of linguistic items, often referred to as 'exemplar clouds' or 'extension networks' (Pierrehumbert 2001, 2016; Mompeán-González 2004; Wedel 2006; Nathan 2006; Kristiansen 2006). See Jäger and Rosenbach (2008: 101–103) for similar considerations.

Linguistic types can be thought of as equivalence classes of variants, 'labels' or 'labeled exemplar clouds' of sufficiently similar exemplars (Pierrehumbert 2001), or cognitive 'prototypes' that are associated with various 'extensions' in a network (Mompeán-González 2004). In our case, s_i would be considered as an equivalence class of variants that share a similar amount of formal substance. In this conceptualization, the value s_i denotes the prototypical amount of formal substance in an equivalence class.

The following two factors drive the dynamics of $x_1, x_2, \dots, x_N$. First, the dynamics of item i depends on its 'intrinsic growth rate' which does not depend on any interactions among different items but solely on linguistic properties of i. Crucially, this rate is assumed to depend on the item's formal substance s_i so the intrinsic growth rate r is formulated as a function of s_i: $s_i \mapsto r(s_i)$, $\mathbb{R}^+ \to \mathbb{R}^+$. The rate is defined as the number of new tokens that are produced per token per time unit and thus functions as a measure of 'productivity' or 'reproductive success' of an item. Token production, as defined here, depends on a number of processes. In the production-perception loop, tokens, as objects on the utterance level, are (i) perceived, (ii) learned, (iii) memorized, (iv) accessed, and finally (v) articulated so that new tokens of the same (or sufficiently similar) type are produced. We take $r(s_i)$ to encompass all of these steps at once. At this point, there are no constraints on the shape of the functional dependency between growth rate and substance, since the relationship between r and s can be arguably complicated. For instance, formal substance may be positively related with perception, because long forms are perceived more easily, but negatively with articulation because it takes more effort to utter long forms.

Second, we assume that linguistic items cannot grow unrestrictedly. This is plausible because (i) time, (ii) memory, (iii) the number of possible opportunities to produce utterances, (iv) the number of possible slots within an utterance, (v) articulatory energy, and not least (vi) the number of speakers represent limited resources. Thus, the growth of a linguistic item is constrained by its environment. In some cases ($N > 1$) the environment of a linguistic item also contains other linguistic items which have a major impact on each other. This might happen, for instance, if two linguistic items compete for similar slots in speech. If one item is used very frequently, this leaves less room for other linguistic items on one or more of the levels (i) to (vi).

The interaction of an item with its environment shall be formalized as a coefficient $c \geq 0$. In the case of a single item, it accounts for the limiting factors (i-vi) above. In the case of more than one item, the term models their interaction. In that case c functions as a competition coefficient. If two items i and j co-occur within an utterance, then the overall number of i tokens produced per i token per time unit in the above described manner is decreased by c tokens per time unit. This is a simplifying assumption because it ignores any specific ordering of i and j. That is, we do not account for any structure within utterances and just assume that items i and j are randomly mixed. In other words, the probability of i occurring before j equals the probability of j occurring before i. While structural details could be implemented into models like the one we are studying, it makes their analysis considerably more complicated (up to a point at which analytical results cannot be derived any more).[5] For that reason, we stick to this simplification and leave the analysis of more complicated models open for future research.

In our model, this competition coefficient is not constant but modeled as a function of formal substance s_i and s_j of i and j, in order to account for the differential effects of asymmetric priming. We define c as a function of the difference between s_i and s_j. This is done in such a way that competition among items with little formal substance and items with more formal substance is asymmetric: short items are inhibited less by long items than the reverse because short items benefit more from the presence of long items via asymmetric priming than the reverse. A shorter item i is inhibited less by the presence of a longer item j, than j is by the presence of i. Formally, we define the coefficient c as a function $s_i - s_j \mapsto c(s_i - s_j)$, $\mathbb{R} \to \mathbb{R}^+$, so that $s_i < s_j$ implies $c(s_i - s_j) < c(s_j - s_i)$.

[5] Note that equivalent assumptions are made in game-theoretical models as well. We will comment on the relationship between the model family we use and game theoretical models below.

As we will see, the coefficient c enters our model with a negative sign which means that items are always constrained by their environment. This is done to make sure that the environmental constraints (i-vi) are realistically represented in the model. For our case this is relevant because it means that there is no formal difference between asymmetric inhibition and asymmetric priming in our model. That is we do not differentiate between these two cognitive mechanisms (cf. Hilpert & Correia Saavedra 2016): i is inhibited more by j than j is inhibited by i exactly if j is primed more by i than i is primed by j. In both cases, the coefficient c is larger for i than it is for j so that i suffers more from its interaction with the environment than j does.

The two factors described above, intrinsic growth and asymmetric competition, determine the overall rate of change of the frequency x_i of item i, i.e. the derivative of x_i with respect to time t, $\mathrm{d}x_i/\mathrm{d}t$. Thus, the set of (ordinary) differential equations defining the dynamical system reads

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = r(s_i) \cdot x_i - \sum_{j=1}^{N} c\left(s_i - s_j\right) \cdot x_j \cdot x_i \qquad (1)$$

where $i = 1, \dots, N$. It simultaneously defines the change of all N items.

For $N = 1$, i.e. in the absence of any competing variant, the system reduces to a one-dimensional logistic dynamical system

$$\frac{\mathrm{d}x_1}{\mathrm{d}t} = r(s_1) \cdot x_1 \cdot \left(1 - \frac{c(0)}{r(s_1)} x_1\right) \qquad (2)$$

where $r(s_1)$ is the intrinsic growth rate and $r(s_1)/c(0) = K$ the carrying capacity of the linguistic item. The carrying capacity can be interpreted as the amount of possible slots in speech, which is determined by factors mentioned above (limited number of speakers; limited time; limited number of slots in an utterance; etc.).

This system is well-known in the study of language dynamics. If $K = 1$ then this equation is equivalent with models that describe the spread of lexical items through speaker populations (Nowak 2000; Nowak et al. 2000; Solé et al. 2010; Solé 2011). Likewise, competition models of grammatical rules (or grammars) which are driven by triggered learning reduce to a logistic map (Niyogi 2006: 164–166). More generally, logistic models have been assumed to model the progress of linguistic change (Altmann 1983; Kroch 1989; Denison 2003; Wang & Minett 2005), thereby typically measuring token frequencies. These studies do not necessarily involve competition among variants in an explicit way, in the sense that one linguistic variant replaces another. Rather, the growth of populations of tokens is constrained by interspecific competition: tokens of a particular type thereby compete for slots in utterances and speakers. If everyone knows a linguistic type and uses it in every possible utterance, then there is simply no potential to grow any further in frequency. This is what the carrying capacity K accounts for. Since patterns of logistic – or S-shaped – spread are relatively abundant in diachronic change of linguistic items, different mechanisms have been studied that account for it (also in more realistic network structures) (Blythe & Croft 2012).

The dynamical system outlined above belongs to the Lotka-Volterra model family, which is widely used in ecological research. One key result in mathematical ecology is that any Lotka-Volterra system can be transformed into a system of replicator equations that model the dynamics of an evolutionary game (Hofbauer & Sigmund 1998; Nowak 2006). This is relevant, since evolutionary game theory has been facing growing acceptance in linguistic research (de Boer 2000; Pietarinen 2003; Nowak 2006; Jäger 2008a, 2008b).

Just like game-theoretical systems, the Lotka-Volterra system in (1) can converge to an ecological equilibrium. We are only interested in non-trivial equilibria, i.e. equilibria which are different from the zero point corresponding to the absence of all items i (details can be found in Appendix A1). In the one dimensional special case (2), this non-trivial equilibrium is given by the carrying capacity K. The two-dimensional case $N = 2$ is of particular relevance, because it can be used to model the competition among an old and a new variant of an item, with frequencies x_1 and x_2, respectively (which will be described in more detail in 3.3 and 3.4). If $N = 2$, leaving the non-trivial equilibrium aside, it can either be the case that only one of the two items stably exists in the long run, while the other one gets lost. Or, under certain conditions both items may stably coexist (again, see Appendix A1 for more details). This observation will become important when we discuss evolutionary dynamics and diversification in 3.3 and 3.4. Before that, however, we need to take a closer at the competition coefficient.

3.2 Asymmetric competition term

As described above, the competition term c is defined as a function of the difference between s_i and s_j: $\Delta = s_i - s_j \mapsto c(s_i - s_j)$, $\mathbb{R} \to \mathbb{R}^+$, which fulfils that $s_i < s_j$ implies $c(s_i - s_j) < c(s_j - s_i)$. Instead of monotone functions such as the family of sigmoid curves employed by Kisdi (1999) and Law et al. (1997) to model asymmetric competition in biology, we opt for a Gaussian function which decreases for large differences Δ (Fig. 1). This shape models the interaction among linguistic items more realistically, which we assume to become weaker if items are extremely dissimilar. The function defining the asymmetric competition term reads

$$c(\Delta) = c_{\max} \cdot e^{-\frac{(\Delta-\mu)^2}{2\tau^2}} \qquad (3)$$

where $c_{\max}$ is the maximal competitive disadvantage among interacting linguistic items, which is assumed if $\Delta = \mu$. The parameter $\mu > 0$ can be interpreted as similarity threshold, where similarity refers to how close two substances are to each other (e.g. to what extent two durations match).[6] Beyond μ competition among two items becomes less severe. This assures that items which are extremely dissimilar do not significantly affect each other through priming (Rueckl 1990; Snider 2009). Thus, μ operationalizes the scope of priming. The parameter τ the extent to which priming is asymmetric (it determines the steepness of the curve). If τ is large both items have a relatively similar impact on each other. If τ is small, in contrast, the impact of the item carrying more substance on the one with less substance is strong. That is, there is a severe asymmetric effect. Figure 1 shows the shape of the curve defined by the competition coefficient. Technical details relevant to our analysis can be found in Appendix A2. Box 1 summarizes the model parameters together with their cognitive interpretation.

[6] Note that in our account, substance is always measured by a one-dimensional real-valued parameter s. Hence, similarity in substance can be measured by means of the difference between two substance scores.

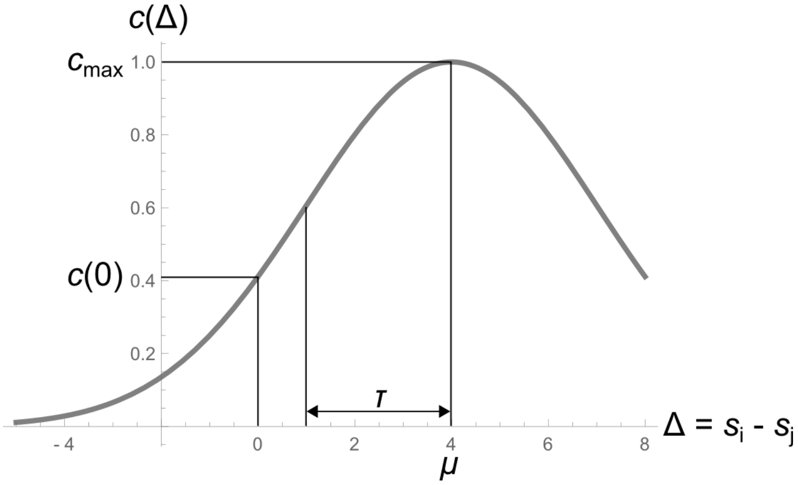


Figure 1. Gaussian function underlying the asymmetric competition term with $c_{\max} = 1, \mu = 4, \tau = 3$.

Box 1. Cognitive interpretation of model parameters

s	prototypical formal substance of a linguistic item; evolving parameter
g	prototypical degree of grammaticality related to s; evolving parameter (see 4.2)
r	intrinsic growth rate; measure of productivity independent of interactions with similar variants but depending on s
c	asymmetric competition coefficient; depends on interaction via priming among variants that differ in s; restricts growth in the one-dimensional case
$c_{\max}$	maximal competitive disadvantage imposed by a related variant
μ	similarity threshold for asymmetric priming (scope of priming); beyond a difference of μ, priming effects become weaker
τ	measure of the strength of asymmetric priming; if τ is small/large priming has strong/weak effects on processing
α	language specific articulatory effort; small α corresponds to a speaker friendly linguistic system (see 4.1)
π	language specific perceptual effort; small π corresponds to a listener friendly linguistic system (see 4.1)
κ	language specific strength of the inverse relationship between substance and productivity of words (see 4.2)

3.3 *Adaptive dynamics*

Let us go back to the case of a single linguistic type, henceforth 'item 1', specified by substance s_1. As sketched above, item 1 could for instance be a construction, a word type, a diphone, or even a single phoneme. We assume that the value s_1 merely represents the prototypical amount of substance of item 1, and that variants featuring slightly less and slightly more substance are associated with the prototype labeled as 'item 1'. We assume that variant substances within that class are distributed around the prototypical substance s_1. If a speaker picks a variant (exemplar; extension), say 'item 2', with substance s_2 slightly smaller or larger than s_1 as a new competing prototype (or label), what are the chances that item 2 replaces item 1 if we take the effect of asymmetric priming into account?

This question is tackled by the mathematical toolkit of 'adaptive dynamics' (Dieckmann & Law 1996; Geritz et al. 1998). As an extension of evolutionary game theory (Maynard Smith 1982; Nowak 2006), this framework has been developed to analyze biological phenotypic evolution, e.g. the evolution of fertility, body weight or the size of particular body parts, in ecologically complex setups like geographically, biologically or socially structured populations (Cushing 1998). A key feature of adaptive dynamics is the eco-evolutionary feedback loop. Emerging mutant populations do not occur in isolation but rather face an environment which is determined by the resident population, the mutant is a variant of. If the mutant population successfully invades and replaces the resident, it becomes the new resident population and thereby shapes an environment that future mutants have to cope with. By applying a number of mathematical techniques to a given population dynamical model, one can determine whether or not successful invasion and substitution occurs. If applied iteratively, the long-term evolution of a phenotypic trait can be predicted. In addition to evolutionarily stable configurations this can result in more complicated evolutionary dynamics such as Red-Queen dynamics, evolutionary suicide (Dercole & Rinaldi 2008), or, as of primary interest to the present study, evolutionary branching and stable coexistence (Geritz et al. 1998).

The adaptive dynamics toolkit rests on two technical assumptions about evolution: (i) mutations are sufficiently small and (ii) mutations are sufficiently rare. What these assumptions ensure is that the ecological timescale is separated from the evolutionary timescale, that is, mutations occur only if populations are close to their population-dynamical equilibrium. These assumptions arguably hold for biological evolution (Dercole & Rinaldi 2008: 65). Let us see if they apply to linguistic evolution as well. The first assumption, that linguistic variation occurs in small steps, is consistent with the wide spread notion in usage-based linguistics that linguistic change is gradual (Croft 2000; Pierrehumbert 2001; Hopper & Traugott 2003; Bybee 2010).[7] The validity of second assumption in linguistics is less obvious. As mentioned above, we assume that variation is always present in speech production. However, under our conceptualization a 'linguistic mutation' (Ritt 2004; Croft 2000) occurs only if a speaker reorganizes the cognitive setup by employing a new prototypical variant, an event which we assume to occur much rarer. In summary, we do not consider it problematic to apply the framework of adaptive dynamics to diachronic change in linguistics (see also Doebeli 2011 and Baumann & Kazmierski 2016 for other linguistic applications).

For our endeavor, assumptions (i) and (ii) have the following consequences. First, they ensure that mutations, i.e. new variants of a linguistic item, do not differ much in terms of substance from the old versions they were derived from. That is, steps of reducing or enhancing

---

[7] It applies less directly to generative approaches to language change Roberts (2007); Niyogi (2006), unless considering probabilistically weighted (or fuzzy) generative grammars (e.g. Yang (2000)).

substance are relatively small so that large jumps are not possible.[8] In other words, formal evolution is modeled as a continuous process. Second, since mutations (events of adopting new prototypes) are rare, we only have to concern ourselves with the dynamics of two populations at most in mutant-resident interactions (because under a new variant either vanishes or replaces the old variant; see Geritz et al. 2002 for more technical details). Both assumptions make mathematical computations much easier.

3.4 Conditions for stable diversification

As pointed out above, we seek to determine if a slightly different variant of item 1 (characterized by substance s_1), labeled item 2, can become more frequent and perhaps even replace the resident item 1. In order to do so, we must calculate the 'invasion fitness' of item 2, which is defined as the expected growth-rate of item 2 under the assumption that item 2 is relatively rare (since it is new) and exposed to an environment in which item 1 is already present. If invasion fitness is positive, item 2 can invade and (under certain conditions) replace item 1. If it is negative, it cannot do so. Invasion fitness can be computed directly from the underlying population-dynamical model (system (1)) for any pair of formal substances s_1 and s_2. Thus, if an item specified by formal substance s_1 is replaced by an item specified by formal substance s_2, the latter may in turn be invaded by yet another item specified by formal substance s_3. In this way, the evolutionary trajectory of formal substance s can be determined. Formal details about how this trajectory can be derived can be found in the appendix (A3).

Sometimes, evolution of formal substance can – temporarily – come to a halt, which is referred to as an 'evolutionary singularity' (because at such a point the rate of change in s becomes zero), denoted by s^*. A variety of things can happen at such a point. Formal substance could for instance reach an evolutionary optimum, a 'continuously stable strategy' (CSS). Such an evolutionary optimum cannot be invaded by nearby strategies, and evolution drives formal substance always towards that CSS.

Under certain conditions, evolution can drive formal substance towards an 'evolutionary branching point' (BP) at which a population consisting of a single item type is divided into a population consisting of two different item types. Crucially, these two types stably coexist rather than ousting each other. This scenario is interesting as it corresponds to linguistic diversification.

If we implement the asymmetric priming term as defined in (3) into the dynamical system defined in (1) it can be shown that in our model evolutionary branching occurs at an evolutionary singularity s^* if

$$r'(s^*) \cdot \underbrace{\frac{\mu}{\tau^2}}_{>0} \underbrace{>}_{\text{(i)}} r''(s^*) \underbrace{>}_{\text{(ii)}} (\mu^2 - \tau^2) \cdot \underbrace{r(s^*) \cdot \frac{\mu}{\tau^6}}_{>0}. \tag{4}$$

Details about the derivation of these inequalities can be found in the appendix. In summary, two criteria can be identified that promote stable diversification, both of which have an immediate linguistic interpretation. First, the slope of the intrinsic growth rate r as a function of formal substance must be sufficiently large at the evolutionary singularity (ideally increasing in s). That is, if reproductive success of an item increases if it is larger, then diversification as a reflex of asymmetric priming becomes more likely. Second, τ in the asymmetric-priming term should not be much smaller than μ (ideally $\tau > \mu$). If this is the case then the curve defining

[8] In fact, the adaptive-dynamics framework provides methods for dealing with scenarios where this assumption is relaxed. But it makes computations much more complicated and can lead to completely different predictions. See Appendix A3 and Geritz et al. (2002).

the effect of asymmetric priming is relatively broad. This means that asymmetric priming is relatively weak. If the effect of asymmetric priming is too strong so that the curve becomes very steep (i.e. such that inequality (ii) is reversed), then the evolutionary singularity becomes stable, resulting in an evolutionary optimum (continuously stable strategy, CSS). This is one of our key results: asymmetric priming only leads to stable diversification if it is mild. Strong priming effects, in contrast, entail optimization of formal substance.

Let us consider an example.[9] Figure 2 illustrates the evolution of s under the hypothetical assumption of a strictly increasing and mildly convex intrinsic growth rate $r(s) = s^{3/2}$. This function, for instance, models the plausible linguistic assumption that items benefit from having much formal substance, e.g. because formally explicit items are easier to perceive by the listener, and that this benefit gets less relevant the shorter an item is. No other pressures are supposed to apply in this example (which is, of course, less plausible). Thus, we investigate evolution in an extremely listener-friendly scenario in which asymmetric priming still applies. If τ is small, the asymmetric-priming curve is much steeper than if τ is large (left vs. right plot in Fig. 2a, respectively). As a consequence, formal substance s approaches an optimal strategy under strong asymmetric competition, while it undergoes evolutionary branching under sufficiently weak asymmetric competition (left vs. right plot in Fig. 2b, respectively). In the latter case, the item undergoes formal reduction until it reaches a threshold at which it is divided into two similar and stably coexisting items. The one which is more reduced maintains its formal substance, while its competing variant increases its substance again to a point at which the formal difference between the two competing populations of items is sufficiently large. Since the dynamics in this example are largely driven by the listener the result reflects a configuration in which the two items are sufficiently different so that they can be easily distinguished from another in perception.

In what follows we investigate the evolutionary behavior of formal substance in two substantially different linguistic domains: phonetic reduction of (mor)phonotactic diphones on the sublexical level and grammaticalization on the lexical level.

[9] All evolutionary invasion analyses and evolutionary trajectories in this paper were computed with Mathematica 10.3, Wolfram Research (2016), with a modified version of a script by Stefan Geritz (2010).

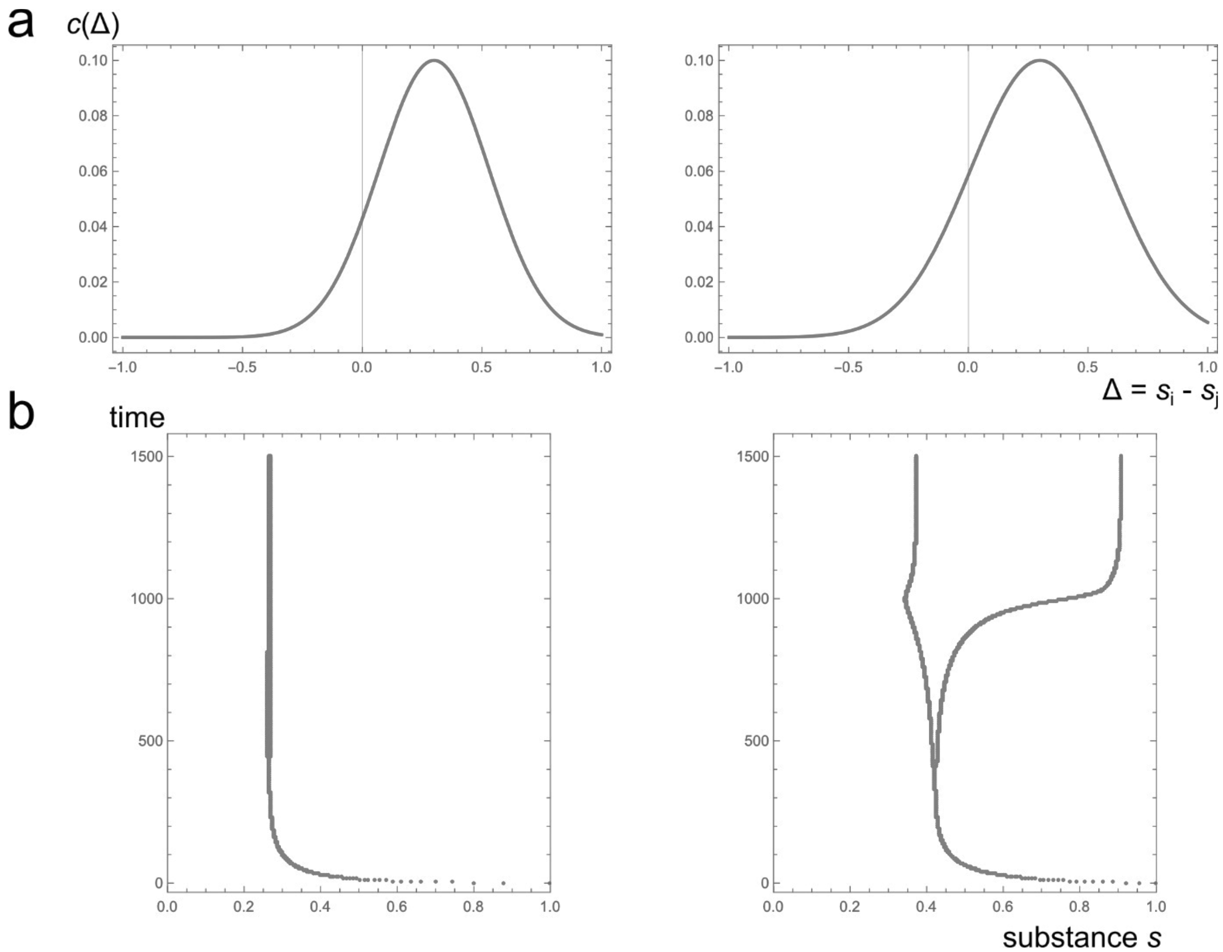


Figure 2. (a) Asymmetric competition terms with $\mu = 0.3$ and $c_{\max} = 0.1$ assuming strong (left; $\tau_{\text{strong}} = 0.23$) and weak (right; $\tau_{\text{weak}} = 0.29$) priming effects, respectively. (b) Evolutionary trajectory of formal substance s based on the canonical equation of adaptive dynamics assuming $r(s) = s^{3/2}$. If priming effects are strong, items undergo formal reduction thereby approaching an optimal degree of formal substance (left). Under weak priming effects, diversification occurs followed by stable coexistence of two items occurs that differ as to their degree of formal substance (right).

4 Applications of the model

4.1 Sublexical: asymmetric priming in phonotactics

Diphones, i.e. strings of two sounds, have been suggested to support segmentation of speech strings into words (Daland & Pierrehumbert 2011). Similarly, diphones apparently help the listener in the decomposition of words into morphemes when they span a morpheme boundary. The latter are referred to as 'morphonotactic' or 'low-probability' diphones (Hay & Baayen 2003, 2005; Dressler & Dziubalska-Kołaczyk 2006; Dressler et al. 2010). Consonant diphones are especially useful for this purpose due to their markedness. While for instance word final diphones like /md/ in *seemed* function as perfect markers of morphological complexity, other diphones such as word final /nd/ as in *banned* or /ks/ as in *clocks* are less reliable indicators of morpheme boundaries: both diphone types are also found word finally within morphemes, such as *hand* or *box*. Thus, these diphone types suffer from ambiguity in signaling complexity, evidently a dispreferred feature from a semiotic point of view (Kooij 1971; Dressler 1990).

Consequently, it has been argued that diphones should diachronically evolve in such a way that they either occur exclusively 'lexically' within morphemes, or purely 'morphonotactically' across morpheme boundaries (Dressler et al. 2010; Ritt & Kaźmierski 2015). As is evident from the above examples, this is not the case. Thus, coexistence phenomena like these need to be explained.

We suggest that the observable stable coexistence is grounded in asymmetric priming effects. Why is this plausible? A number of studies imply that morphonotactic consonant diphones are typically shorter than their lexical counterparts (Kemps et al. 2005; Plag et al. 2011; Leykum et al. 2015). If this is the case, then asymmetric priming should apply in such a way that morphonotactic diphones benefit from the presence of lexical diphones to a larger extent than the reverse. Hence, we can apply the model described in section 3 to the evolution of diphone lenght (we will use the terms 'length' and 'duration' interchangeably in this section) and check under which conditions two phonemically identical diphones, which merely differ in duration, can coexist.[10]

We specify the shape of the intrinsic growth rate r of diphones as a function length s. Kuperman et al. (2008) show that token frequency of Dutch, English, German and Italian diphone types exhibits the shape of an inverse 'U', respectively. Very short and very long diphones show relatively low token frequencies, while diphones in the middle of the duration spectrum are highly frequent in terms of tokens. Notably, this does not depend on the position of diphones within the word nor on whether or not diphones do belong to a language's phonotactics, although phonotactically illegal diphones are significantly longer than phonotactically legal ones (Kuperman et al. 2008: 3905). Importantly, this is orthogonal to the question of whether morphonotactic instances of a particular diphone type exhibit a shorter duration than their lexical counterparts that belong to the very same diphone type, as discussed above.

In their analysis, Kuperman et al. (2008) model this inverse-U shape as a result from a trade-off between articulatory and perceptual effort. Thus, the frequency distribution of diphones is shaped by pressures imposed both by the speaker and the listener. In contrast, Zipfian patterns such as the inverse relationship between length and token frequency are only determined by pressures imposed by the speaker. Similar to their model (Kuperman et al. 2008: 3902) we propose that the intrinsic growth rate r of a diphone as a function of length s is defined as

$$r(s) = Cs^{\alpha}(1 - s)^{\pi}$$

where C, α and π are strictly positive. In this function, α measures articulatory effort and π measures perceptual effort, while C simply bounds the height of the function from above. Note that these constants are assumed to be language specific and to apply to all items in a language's diphone inventory (Kuperman et al. 2008). The function above is locally concave (i.e. inverse-U shaped) at its maximum $s_{\max} = \alpha/(\alpha + \pi)$.[11] If $\alpha > \pi$, i.e. if articulatory effort outbalances perceptual effort (this is a listener friendly phonotactic system), then the peak of the function is

[10] Note that the durational differences between lexical and morphonotactic clusters are very small and thus probably do not classify as phonemic, but see Kemps et al. (2005) for a discussion about whether durational differences in phoneme sequences actually function as cues in word-decomposition. We would like to thank Martin Hilpert raising this issue.

[11] It is globally concave if $\alpha = \pi = 1$, and locally convex close to 0 and 1, if $\alpha > 1$ and $\pi > 1$, respectively.

shifted to the right. If $\pi > \alpha$ so that perceptual effort is larger than articulatory effort in diphone transmission (i.e. a speaker friendly phonotactics), then the peak is shifted to the left.

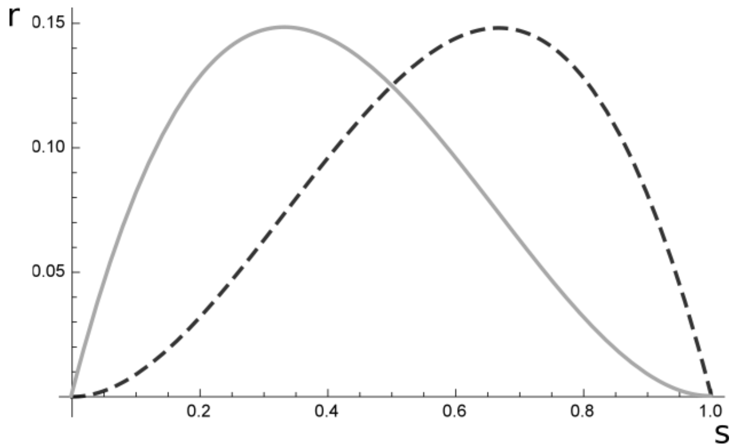


Figure 3. Intrinsic growth rate r as a function of s, where $r(s) = Cs^{\alpha}(1-s)^{\pi}$. Solid light gray curve: $\alpha = 1$, $\pi = 2$, i.e. perceptual effort dominates. Dashed dark gray curve: $\alpha = 2$, $\pi = 1$, i.e. articulatory effort dominates. In both cases, $C = 1$.

What can be said about the long-term evolution of acoustic duration? We show in Appendix A4 that the evolutionary dynamics of acoustic duration exhibit an evolutionary singularity which shall be labeled s^*. In the present scenario, s^* depends on articulatory effort α, perceptual effort π, the similarity threshold μ defining the scope of priming and strength of asymmetric priming τ (see Box 1 for a summary of the parameters involved).

In order to evaluate whether s^* is an evolutionary branching point (or indeed a CSS) we have to check if condition (4) is fulfilled. The computation is lengthy since the explicit expressions of s^*, intrinsic growth rate $r(s^*)$ and the derivatives it involves are a little cumbersome. Hence, we will not derive explicit conditions, but instead leave it at numerically plotting s^* as a function of α, π, μ and τ thereby distinguishing between the different types of evolutionary singularities. The results are shown in Fig. 4. It shows a 3-by-3 table consisting of nine bifurcation plots of the evolutionary singularity $s^*(\mu, \tau)$ (vertical axis) as a function of the parameters defining the impact of asymmetric priming μ and τ (horizontal axes). Across the single bifurcation plots, perceptual effort π increases from the left-most column to the right-most column, while articulatory effort α increases from the top row to the bottom row. In each plot, dark gray denotes singularities which are BPs, while light gray denotes singularities that are CSSs.[12] Also note that given the restrictions on the four parameters in this paper, s^* always exists and is non-negative.

[12] As can be seen, there are no repellors or Garden-of-Eden points for the admitted combinations of α, π, μ and τ. See appendix.

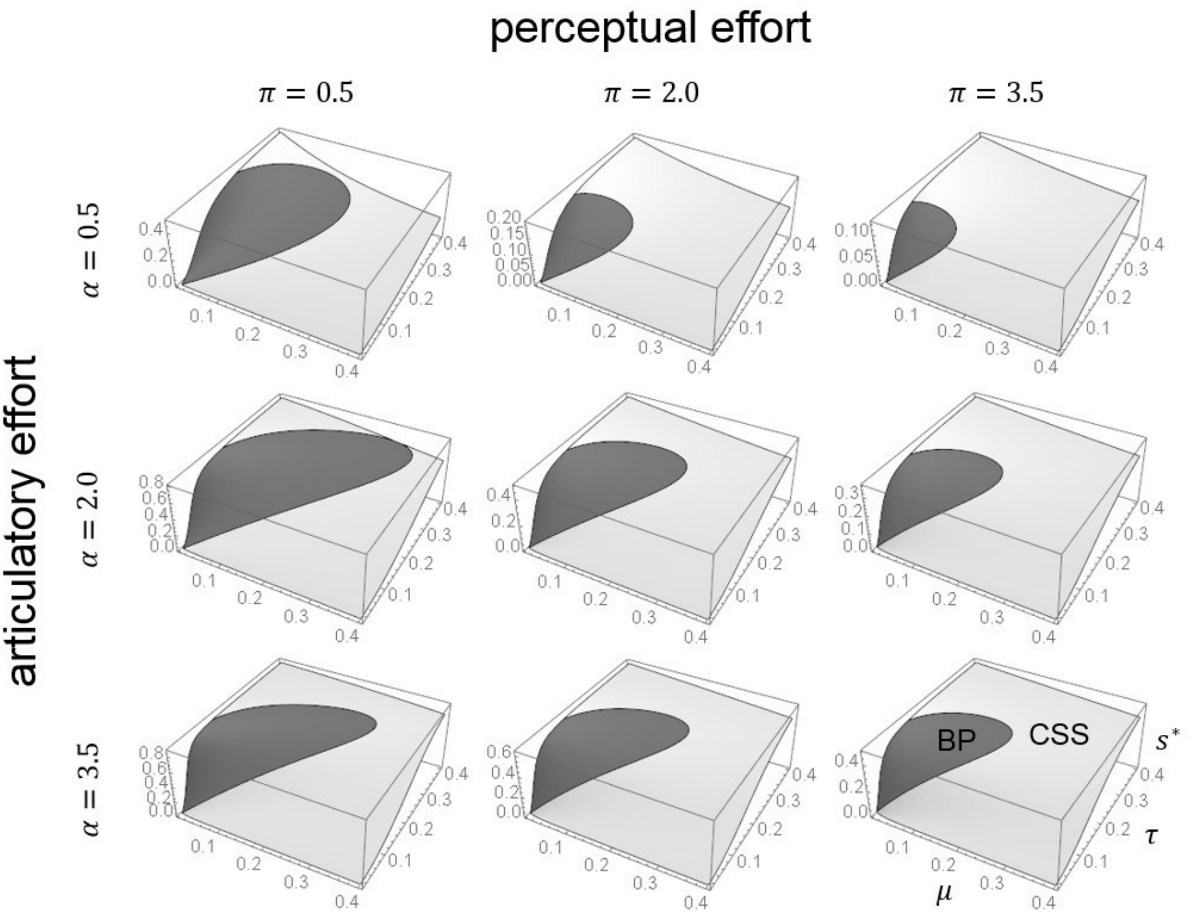


Figure 4. Bifurcation plots of the evolutionary singularity s^* depending on the similarity threshold μ and priming strength τ. Dark gray areas denote BPs, light gray areas denote CSSs. Plots are shown for different values of articulatory effort α (rows) and perceptual effort π (columns).

There are multiple observations to be discussed, the most relevant of which are summarized in Box 2 below. First, the evolutionary singularity s^* decreases in μ as can be seen from the decreasing values on the vertical axis. Since μ functions as a similarity threshold beyond which priming effects become weaker, this means that evolution drives length towards very small values, if asymmetric priming is relatively insensitive in the sense that it applies to pairs of items which are substantially different from another (large μ). In contrast, if asymmetric priming has a narrow scope (small μ), then formal reduction is hampered.

Second, s^* increases in τ, which determines the impact of asymmetric priming. If τ is small, then asymmetric priming has a strong impact. In that case, items tend to get shortened. If τ is large, so that asymmetric priming has relatively weak effects, then longer durations are maintained.

Third, the height of evolutionary singularity s^* is determined by articulatory and perceptual effort. While low perceptual effort supports long items, high perceptual effort drives reduction to shorter durations. This is plausible and consistent with what one would expect from the respective roles that speakers and listeners play in the evolution of diphone duration: speaker friendliness leads to reduction ('lenition') while listener friendliness supports long durations ('fortition'; see e.g. Dressler et al. 2001 and Dziubalska-Kolaczyk 2002 for some evidence in phonotactics).

Fourth, let us discuss the roles that the similarity threshold μ and strength of asymmetric priming τ play in evolutionary branching (dark gray region in Fig. 4). As can be seen in Fig. 4, μ must be relatively small in order to enable stable diversification. If μ is large so that the range of items that are subject to asymmetric priming is large then duration is simply optimized, i.e. approaches a CSS (light gray region in Fig. 4). Moreover, and consistent with the condition derived in 2.4, τ must be greater than μ, so that asymmetric-priming effects are relatively weak in order to accommodate BP. However, as can be seen from the elliptic shape of the dark gray region, τ must not be too large, and if τ is large then μ must not be too small. This illustrates that branching requires rather complicated conditions to occur, while optimization of duration is the default. Overall, stable coexistence of duration-wise substantially different diphone-type variants apparently is an exceptional phenomenon.

Finally, articulatory and perceptual effort have an impact on potential diversification. Looking at the size of the dark gray regions in Fig. 4 from left to right, i.e. increasing perceptual effort, we see that the dark gray area gets smaller making diversification less likely. However, when inspecting the size of the dark gray region from top to bottom, we see that it is maximal in the middle row, i.e. for intermediate values of articulatory effort. Interestingly, this means that speakers and listeners do not only exert differential impact on the extent of shortening, but that they also determine the potential for branching very differently. The more effort has to be allocated to the processing of a diphone in perception (i.e. the less listener friendly), the less likely it is that a language accommodates two variants of that diphone type. Conversely, if a language shows many coexisting diphones that differ in duration, then perceptual effort should be relatively small in that language (i.e. a more listener friendly configuration).[13] With respect to production, no such monotone relationship applies.

We can simulate the evolution of a diphone's duration s given articulatory effort α, perceptual effort π, similarity threshold μ and strength of asymmetric priming τ. Figure 5a shows the evolutionary trajectory of duration and the corresponding token frequency at population-dynamical equilibrium, i.e. $(s, \hat{x}(s))$, for $c_{\max} = 1, \mu = 0.1, \tau = 0.12, \pi = 1$ and $\alpha = 2$, i.e. articulatory effort being twice as large as perceptual effort. Note that the time axis measures the number of evolutionary steps rather than ecological time. Note that the diphone first undergoes durational reduction, i.e. pairwise competition of items in which the shorter item outcompetes the longer item. Reduction proceeds until an evolutionary singularity (at about $s^* \cong 0.25$) is reached. This singularity is an evolutionary branching point. Here, reorganization takes place, since from this point onwards, two variants of the diphone stably coexist. That is, the exemplar cloud (extension network) corresponding to the original item is split into two separate clouds (networks). As a consequence, the stored tokens from the set corresponding to the former prototype are divided among the two new sets. Consequently, the two new token frequencies are half as large as the former one. In Fig. 5a, this is represented by an abrupt drop in frequency displayed on the vertical axis.

[13] Coexisting diphones thus hint at increased listener friendliness, which seems contradictory given that the listener suffers most from ambiguous configurations. Note, however, that the model only captures the effect of duration and does not model the effect of complexity signaling in any way, apart from the assumption that lexical diphones are typically longer than their morphonotactic counterparts.

Box 2. Sublexical dynamics: key results

Assumptions	
Relationship between intrinsic growth r and substance s	Inverse U; governed by articulatory effort α and perceptual effort π
Directionality of asymmetric priming c	Long primes short more strongly than the reverse
Predictions	
Effect of strength of asymmetric priming τ	Relatively weak asymmetric priming promotes diversification; strong asymmetric priming leads to fierce reduction
Effect of scope of asymmetric priming μ	Narrow scope of priming promotes diversification; wide scope of priming promotes reduction towards optimal duration
Effect of articulatory effort α	High articulatory effort promotes reduction
Effect of perceptual effort π	High perceptual effort inhibits reduction and makes diversification less likely

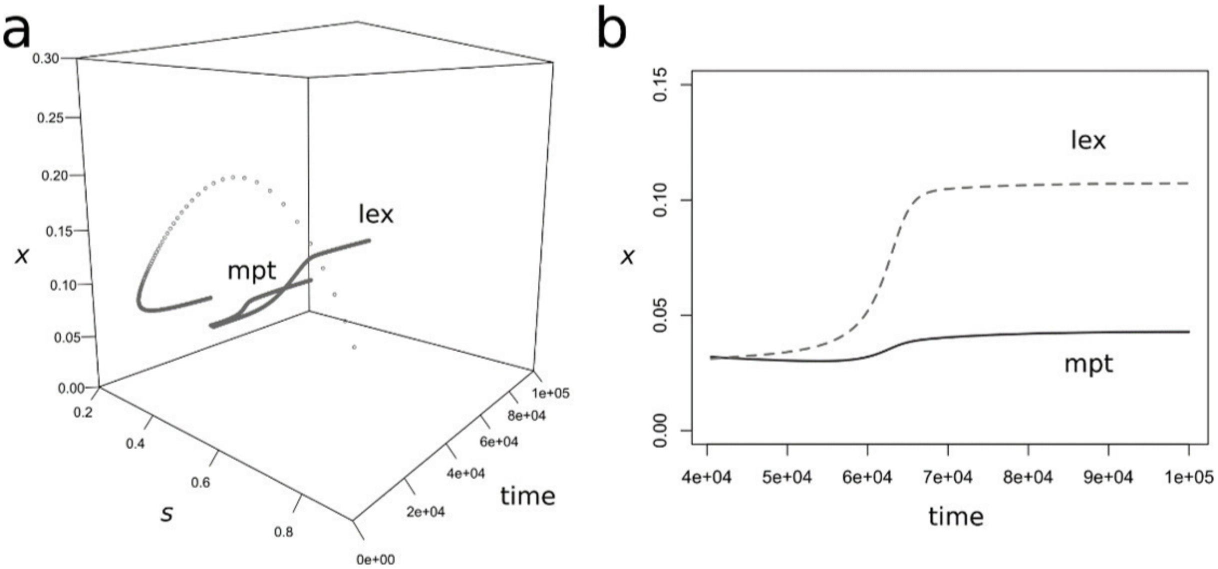


Figure 5. (a) Evolutionary trajectory of $(s, \hat{x}(s))$ before and after branching. Substance s proceeds towards a BP, subsequently followed by branching and coexistence of a shorter (morphonotactic, 'mpt') and a longer (lexical, 'lex') variant (only every 100th point displayed). (b) Frequency trajectories of both variants (dashed: lexical; solid: morphonotactic) after evolutionary branching ($c_{\max} = 1; \mu = 0.1; \tau = 0.12; \pi = 1; \alpha = 2$).

Beyond the branching point the dynamics support two subpopulations: the subpopulation of the reduced variant benefits from asymmetric priming while the subpopulation of the longer variant benefits from the listener friendliness assumed in the current scenario ($\alpha > \pi$). Figure 5b shows the development of the two token frequencies after the split. We argue that the more frequent variant represents lexical instances (dashed line) and the less frequent variant represents morphonotactic, i.e. boundary crossing, instances of the diphone type (solid line), since the former are longer than the latter. In this example, lexical diphones turn out to be roughly twice as frequent as their morphonotactic counterparts.

Although there is obviously no diachronic data that gives reliable information about diphone duration, we can at least compare the frequency development of morphonotactic diphones to that of their – apart from length – homophonous lexical counterparts by looking at diachronic corpus data. Overall, we would expect frequency trajectories of morphonotactic and lexical diphones to look roughly as the ones in Fig. 5b. In order to give empirically attested examples, we make use of the ECCE cluster database (cf. Baumann et al. 2016). It contains all word-final consonant diphones that occur in the Penn Helsinki corpora of Middle English and Early Modern English (Kroch et al. 2004; Kroch & Taylor 2000) together with weights that probabilistically account for the absence of word-final and inter-consonantal schwas. Most importantly, clusters are labeled as to whether they cross a morpheme boundary.

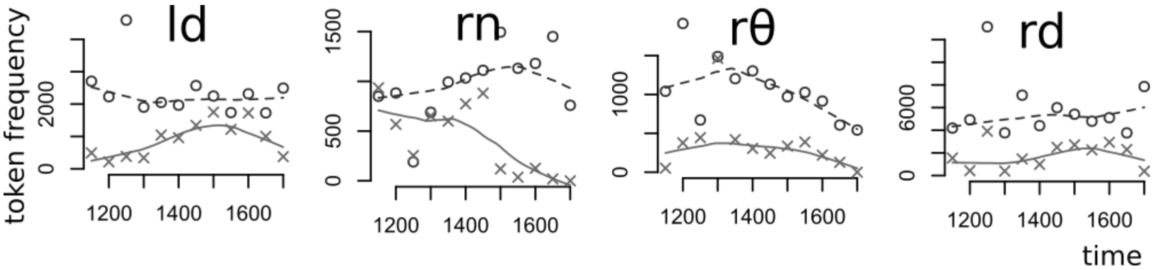


Figure 6. Empirical developments of four word-final consonant-diphone types retrieved from Middle and Early Modern English corpus data. Circles and crosses denote normalized frequencies (p.m.) of morpheme internal (lexcial) and boundary spanning (morphonotactic) diphones, while dashed and solid lines denote LOESS trajectories fitted to the lexical and morphonotactic data points, respectively.

For the purpose of this study, we only looked at a small set of ambiguous clusters, i.e. configurations in which morphonotactic and lexical instances of a diphone type co-occur in the data: /ld, rn, rθ, rd/ (which we assume to evolve independently from each other). We divided the observation period into sub-periods of 50 years each and computed the normalized token frequencies for each cluster type in each period, thereby differentiating between lexical and morphonotactic clusters. In this way, we computed a pair of frequency trajectories for each cluster type, which can be compared to trajectories resulting from the model, as the ones in Fig. 5b.

Figure 6 shows the resulting pairs of frequency trajectories for the four different ambiguous cluster types (lines denote fitted LOESS curves computed in R, R Development Core Team 2013). The respective trajectories of /ld, rn, rθ, rd/ roughly fit to the configuration predicted by the model in that morphonotactic and lexical clusters coexists so that the latter are consistently more frequent (cf. Fig. 5b).

4.2 Lexical: asymmetric priming in grammaticalization

When Jäger and Rosenbach (2008) brought forth their hypothesis of asymmetric priming they primarily had lexical items in mind: formally short and semantically bleached words are hypothesized to benefit more from their formally long and semantically rich counterparts than the reverse. We proceed in two steps. First, we apply our model to this problem and just consider asymmetric priming on the formal level. Second, we consider both form and meaning (by a unified degree of ‘grammaticality’ incorporating both dimensions) and define interaction among lexemes in such a way as suggested by Hilpert and Correia Saavedra (2016). As will be seen, stable lexical coexistence can only be predicted in the latter case.

In both steps, we assume an inverse relationship between reproductive success and length (Baayen 2001). For instance, we can define intrinsic growth rate in terms of a power law

$$r(s) = Cs^{-\kappa}$$

where κ and C are positive. Under these circumstances, diversification is not possible. Rather, formal substance unidirectionally evolves towards ever smaller values, as suggested by Jäger and Rosenbach (2008). Figure 7 shows an example of an evolutionary trajectory under the assumption of a Zipfian intrinsic growth rate. Mathematical details are shown in Appendix A5.

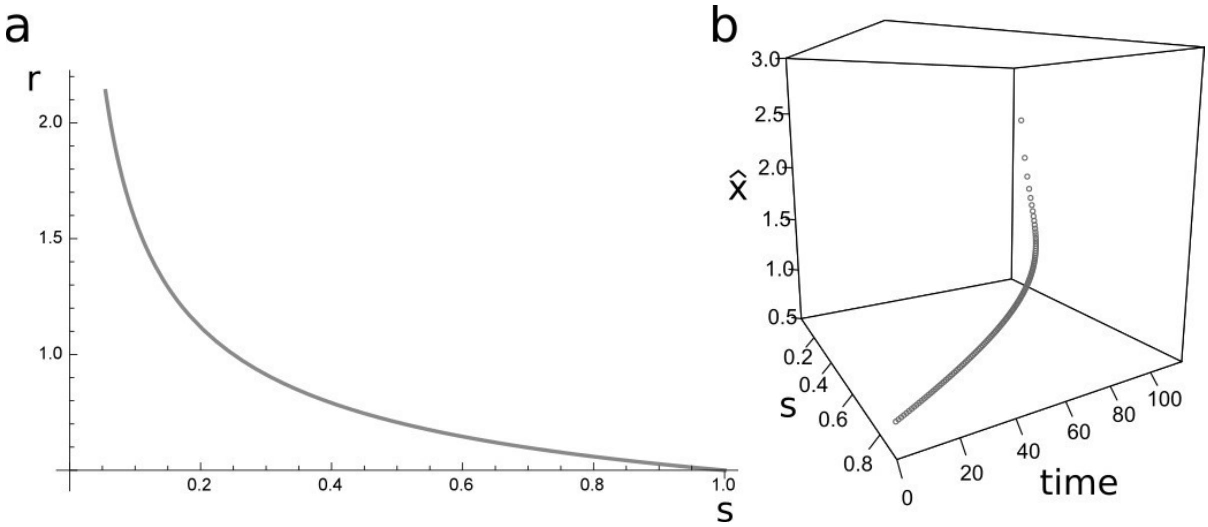


Figure 7. Evolution of formal substance s in grammaticalization under asymmetric formal priming and (a) Zipfian intrinsic growth. (b) Items undergo unidirectional reduction and become increasingly frequent (frequency $\hat{x}$ measured on the vertical axis; $C = 1, \kappa = 0.5, c_{\max} = 1, \mu = 0.1, \tau = 0.12$).

Although the model illustrates how unidirectional evolution of formal substance during grammaticalization might proceed and thereby formally supports Jäger and Rosenbach’s (2008) hypothesis that unidirectionality in grammaticalization is driven by asymmetric priming, the proposed scenario is not entirely convincing for at least two reasons. First, we see that according to the model, items get exponentially more frequent the more they are reduced rather than exhibiting a sigmoid frequency development as observed in many empirical grammaticalization studies (Hopper & Traugott 2003). What is more important, however, is that stable coexistence of related forms cannot be accounted for by the present model. This clearly speaks against what we see in the linguistic data.

The unrealistic behavior of the model might be grounded in the way in which asymmetric priming has been implemented, since in our model priming solely depends on formal differences between competing items ('more substance primes less substance'). Indeed, Hilpert and Correia Saavedra (2016) suggest asymmetric priming to work in the opposite direction if the semantic level is also taken into account (Hilpert & Correia Saavedra 2016). Lexical items are more inhibited less by grammaticalized variants than the reverse. If in the word domain, asymmetric semantic priming overrides the effects of asymmetric formal priming, then the roles of the two arguments in the asymmetric-competition term would be simply exchanged. As a result, stable diversification would be possible, provided the effect of asymmetric priming is sufficiently strong. Notably, this applies even if intrinsic growth rate is a decreasing function of formal substance.

For instance, let us define the 'degree of grammaticality', i.e. the degree to which a word is grammaticalized, as $g = 1 - s$ (because more grammatical words are typically shorter, cf. Hopper & Traugott 2003; Heine & Kuteva 2007).[14] We assume that, in the absence of competing variants, words benefit from higher degrees of grammaticality, for instance because of decreased effort in production, higher predictability, or higher syntactic productivity (Narrog & Heine 2011). Thus we let intrinsic growth rate increase in g, e.g. $g \mapsto C \cdot g^{\lambda}$, $\lambda, C > 0$ (see Fig. 8a). Then intrinsic growth rate, as a function of formal substance $r(s) = C \cdot (1 - s)^{\lambda}$, is decreasing. If we assume asymmetric priming on the word level to have exactly the opposite effects as defined in 2.2 so that 'grammaticalized primes lexical', we can set $c_{\text{word}}(\Delta) = c(-\Delta)$ (because $g_1 - g_2 = s_2 - s_1$), and replace $c(.)$ in the dynamical system by $c_{\text{word}}(.)$. Without going into detail about the evolutionary analysis of the adapted model, let us briefly consider Fig. 7 which shows evolution of the degree of grammaticality g, assuming $\mu = 0.2, \tau = 0.18, c_{\max} = C = 1$ and $\lambda = 2$.

As can be seen in Fig. 8b words become more grammatical and at the same time more frequent in terms of tokens until a branching point is reached. That is, lexical evolution unfolds as a sequence of invasion-substitution events in which variants compete without being able to coexist stably. At the branching point, the dynamics support the coexistence of two variants, one which is slightly more grammaticalized than the other one (as for instance seen in bridging contexts in the early stages of grammaticalization). At this point, both variants can coexist because the grammaticalized variant benefits from higher productivity and/or ease of production, while the lexical variant benefits from being asymmetrically primed by its more grammaticalized cousin. Subsequently, the subpopulations diverge until the two variants are sufficiently different from each other.[15] Notably, the more grammaticalized version also becomes more frequent than its more lexical counterpart and does so in a sigmoid way.

---

[14] Clearly, g is an abstract and simplified parameter in that it expresses multiple linguistic dimensions (formal substance, semantics, morphosyntax) associated with grammaticalization on a one-dimensional (gradual) scale. It lies in the qualitative nature of the model that we do not – even try to – give specific g values for particular words. What really matters is the ordering of lexical variants with respect to their degree of grammaticality.

[15] Note that in our simulation, evolution of g starts at a value close to 0, i.e. at the lexical end of the cline, because words usually enter the lexicon as open-class items. If we let evolution start close to 1, g would approach the BP from above. Thus, to be precise, the adapted model supports the unidirectionality hypothesis only in those cases, in which words enter a language as lexical items (which arguably holds true for the majority of all cases).

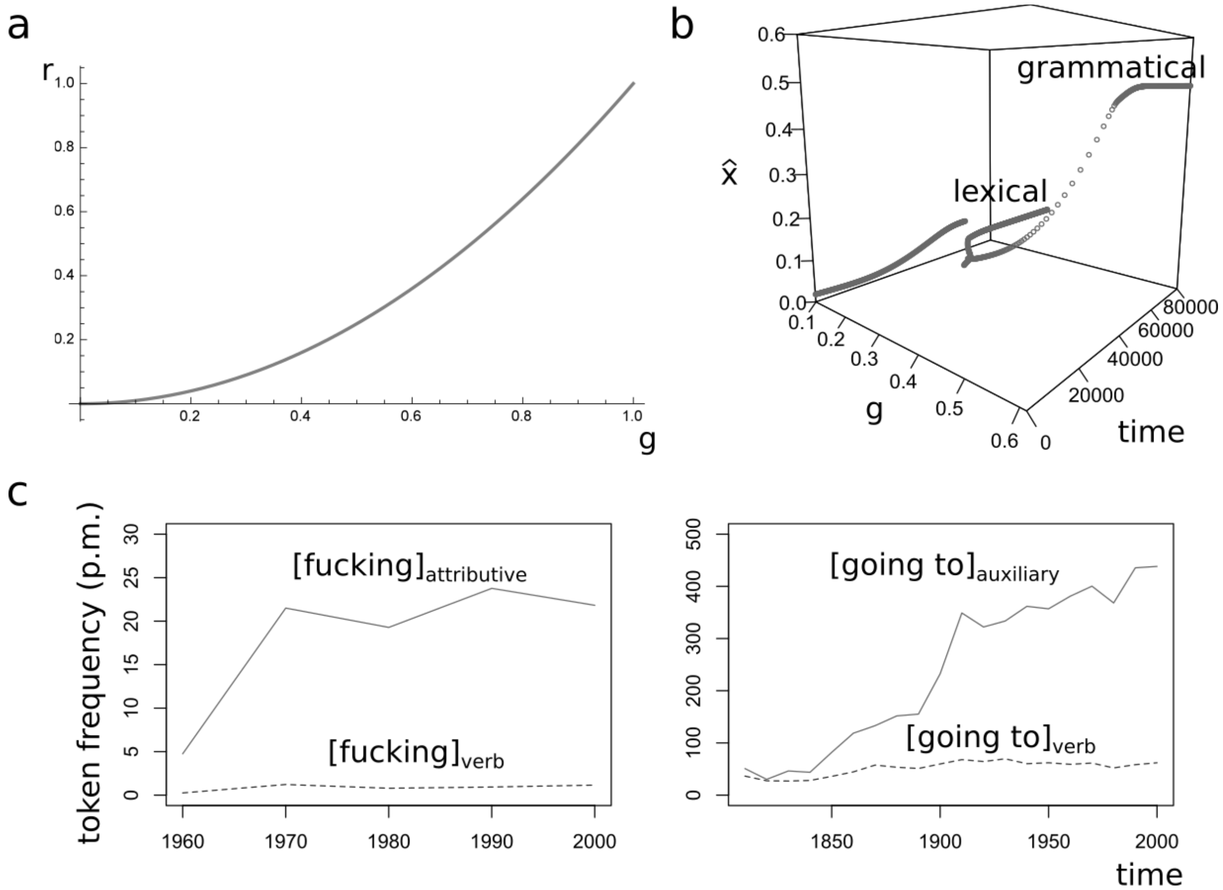


Figure 8. Evolution of the degree of grammaticality g in grammaticalization under asymmetric priming among words c_{word} and (a) a positive relationship between g and intrinsic growth rate: $r(g) = g^2$. (b) After a period of increasing grammaticality (and decreasing formal substance), the dynamics lead to stable coexistence of two words that differ with respect to their degree of grammaticality g and frequency $\hat{x}$. The more grammatical word is more frequent and more reduced than its more lexical cousin. Both trajectories exhibit sigmoid shapes ($c_{\max} = 1, \mu = 0.2, \tau = 0.18$; only every 100th point displayed). (c) Diachronic trajectories of grammaticalized (solid) and lexical (dashed) variants. On the left: attributive (grammaticalized) and verbal (lexical) instances of *fucking* (search queries: fucking _j* + fucking _nn* (attributive) vs. fucking_v* (verbal)). On the right: auxiliary (grammaticalized) and verbal (lexical) instances of *going to* (search queries: [going to _v?i*] vs. [going to]-[going to _v?i*]). The data was elicited from the *Corpus of Historical American English*.

The development shown in Fig. 8b strikingly converges with what is known from empirical research on grammaticalization phenomena (Narrog & Heine 2011). For instance, consider the development of the adverbial taboo intensifier 'fucking' (e.g. *fucking great*) and the *going to* future construction. The taboo intensifier developed out of the present participle form of the verb 'fuck' (with its meaning of sexual intercourse) which, in a first step, grammaticalized into an attributive adjective (*fucking losers*) and afterwards also took up the function of a taboo intensifier. During this grammaticalization process, the meaning of sexual intercourse bleached out and the form was also phonologically reduced (*fuckin'*; /ˈfʌkɪn/). On the other hand, the motion verb 'go' (*I am going to town*) grammaticalized into a future reference marker (*I am going to stay in town*). In both cases, the grammaticalized forms are much more frequent than the verbal source grams (Fig. 8c). This supports Hilpert and Correia Saavedra's (2016) observation that asymmetric priming on the lexical level works in precisely the opposite way than hypothesized by Jäger and Rosenbach (2008). The assumptions and predictions of both models are summarized in Box 3.

Box 3. Lexical dynamics: key results

	Substance only	Substance and meaning (degree of grammaticality g)
Assumptions		
Relationship between intrinsic growth r and substance s	Inverse	Inverse
Directionality of asymmetric priming c	Long primes short more strongly than the reverse	More grammatical (short) primes less grammatical (long) more strongly than the reverse
Predictions		
Effect of strength of asymmetric priming τ	Unidirectional reduction irrespective of τ	Diversification possible under weak asymmetric priming
Effect of scope of asymmetric priming μ	Unidirectional reduction irrespective of μ	Diversification possible if priming has a relatively small scope

5 Discussion and conclusion

Asymmetric priming among items that differ in formal substance has been argued to affect their long-term evolution. Although priming works on a very short time scale, multiple repeated production and perception processes affected by priming can lead to diachronic change of a linguistic item. One of these diachronic processes is formal reduction. Since items with more substance are supposed to prime less items with less substance rather than the reverse, this leads to unidirectional formal erosion (Jäger & Rosenbach 2008). Unfortunately, the premise of this hypothesis does not seem to hold if one investigates words rather than sublexical items. As Hilpert and Correia Saavedra (2016) demonstrate, it is the more lexical words which are inhibited less by their lexical counterparts than the other way round.

In this paper, we proposed a population-dynamical model that captures the effect of asymmetric priming among linguistic items to investigate the long-term diachronic effects of this short-term cognitive mechanism. Importantly, it also takes the relationship between formal substance and productivity into account. We applied the model to the sublexical domain (covering form only, more precisely strings of sounds) as well as to the lexical domain (covering words with form and meaning, and a corresponding degree of grammaticality). On both levels, we integrated empirically plausible functions that relate substance to reproductive success.

While we assumed that asymmetric priming works on the sublexical (phonotactic) level in the direction originally suggested by Jäger and Rosenbach (2008), we tested both directions on the lexical (word) level.

We could show that in all scenarios, reduction of full forms occurs as a combined effect of (negative) asymmetric priming, utterance frequency and formal substance. Crucially, in addition to the reducing tendencies that we find both lexically as well as sublexically, the model predicts diversification and coexistence of related forms that differ in formal substance under certain conditions. In particular, the effect of asymmetric priming must be relatively weak for diversification to occur. Diversification occurs on the lexical level only if interaction among lexemes acts in the way empirically attested by Hilpert and Correia Saavedra (2016). More grammatical items need to asymmetrically support their lexical counterparts, otherwise stable diversification is not supported. In fact, layering of related words is a common phenomenon, as exemplarily illustrated in 4.2 (Figure 7c). Thus, our model functions as a link between what we see on short time scales (within-utterance effects demonstrated by Hilpert & Correia Saavedra 2016) and in diachronic grammaticalization developments.

On the sublexical level, we integrated a function that accounts for the relative pressures imposed by the speaker and the listener (in order to relate duration to reproductive success), in addition to an asymmetric priming effect in which long items asymmetrically support short items. Several observations can be made: reduction is promoted (i) if asymmetric priming applies also to items which are very different from each other, (ii) if asymmetric priming has a strong effect, and (iii) if perceptual effort is high and if articulatory effort is low. The roles that perceptual and articulatory effort play in the likelihood of diversification are more complicated. Overall, diversification on the sublexical level seems to be the exception than the rule. Optimized durations are expected to be more dominant in sublexical inventories. But if it occurs, this points at pressures imposed by the listener, i.e. ease of perception. This seems contradictory, as ambiguous configurations, such as phonemically similar diphones, are expected to impute more effort to the listener. On the other hand, listeners benefit from an increased inventory of sublexical segments as this arguably allows for a larger number of contrastive (and thus listener friendly) configurations (albeit not larger contrasts; cf. de Boer 2000). We used the model to explain the semiotically dispreferred (ambiguous) configurations of coexisting lexical and boundary-spanning (morphonotactic) word-final consonant diphones (Hay & Baayen 2005; Dressler et al. 2010). In a nutshell, the model shows that stable coexistence among similar lexical (longer) and morphonotactic (shorter) diphones is possible because longer diphones are preferred by the listener and because shorter diphones benefit from the presence of their longer counterparts via priming.

Our model demonstrates that weak cognitive short-term effects can have major consequences on a larger time scale. It thus supports the notion that "weak inductive biases acting on learning can have strong effects in the cultural system as the effects of those biases accumulate" (Thompson et al. 2016: 4531) and that even weak biases can account for phenomena which are commonly seen as strong linguistic universals (Kirby et al. 2007; Evans & Levinson 2009). Indeed, phenomena like unidirectional reduction and unidirectional layering through grammaticalization have been conceptualized as "universals of language change" in the historical linguistic literature (Haspelmath 2004: 17; see also Greenberg 1966). In our account, 'weak biases' act on two different levels. The psychological process of (asymmetric) priming itself constitutes a weak process as it operates on a very short time scale. In addition to that, we show that within instances of that process it is only weak asymmetric effects as well as priming with a relatively narrow scope in terms of similarity which promotes an extremely common diachronic behavior, namely linguistic diversification. Diversification occurs on many linguistic levels, of which we only covered two in our study (evolution of lexical and phonotactic items). We leave applications to other linguistic diversification phenomena open for future research (examples are the split of phonemes into long and short variants, or constructional competition and diversification; for explicitly evolutionary accounts see Kaźmierski 2015 and Zehentner 2017, respectively).

Clearly, the complexity of the model is relatively restricted. Neither does it cover relationships between formally less related items, nor does it explicitly model semantic or complicated morphosyntactic relationships (let alone social or pragmatic factors). The only factors that are built into the model are asymmetric priming, utterance frequency and formal substance. However, as we have demonstrated, already a small set of interacting factors governing the production and perception of linguistic items can yield (perhaps) surprising reflexes in the long run. We take our study to demonstrate that (also relatively simple) mathematical models provide useful tools for systematically investigating interactions like this, testing linguistic hypotheses, and making sense of – in fact only seemingly – paradox empirical observations.

Appendix

A1 Stable ecological equilibria

In what follows, we discuss the equilibria of system (1) in the case of $N = 1$ and $N = 2$. The one-dimensional system can be shown to exhibit two population-dynamical equilibria where the rates of growth are zero: a trivial one at $\hat{x}_1 = 0$ and a non-trivial one at $\hat{x}_1 = r(s_1)/c(0) = K$, by substituting these two values into the equation. We will write $\hat{x}(s)$ to denote that equilibrium frequency is a function of substance s. A stability analysis of the trivial equilibrium reveals that it is unstable, i.e. that its stability modulus is positive, whenever $r(s_1) > 0$, so that the population of tokens approaches the non-trivial equilibrium (cf. e.g. Solé 2011: 168–171). According to our assumption about r this is always the case. In the absence of competitors, items remain in the language.

The situation becomes more complicated, when there are two competing items, i.e. $N = 2$. Then the system reads:

$$\frac{\mathrm{d}x_1}{\mathrm{d}t} = r(s_1)x_1 - c(0)x_1^2 - c(s_1 - s_2)x_1x_2$$
$$\frac{\mathrm{d}x_2}{\mathrm{d}t} = r(s_2)x_2 - c(0)x_2^2 - c(s_2 - s_1)x_1x_2$$

Let us assume that $s_1 < s_2$, that is item 1 has less formal substance (i.e. it is shorter) than item 2 does. Then, due to asymmetric priming, $c(s_1 - s_2) < c(s_2 - s_1)$. There are four equilibria at which no change occurs: (i) $(0,0)$, (ii) $(0, r(s_2)/c(0))$, (iii) $(r(s_1)/c(0), 0)$ and finally an internal equilibrium

$$\text{(iv)} \quad \hat{\boldsymbol{x}}_{\text{int}} = \left(\frac{c(0)r(s_1) - c(s_1 - s_2)r(s_2)}{c(0)^2 - c(s_1 - s_2)c(s_2 - s_1)}, \frac{c(0)r(s_2) - c(s_2 - s_1)r(s_1)}{c(0)^2 - c(s_1 - s_2)c(s_2 - s_1)}\right).$$

The latter is the case of stable coexistence. This equilibrium is stable if $1 > r(s_1)/r(s_2) > c(s_1 - s_2)/c(s_2 - s_1)$ (Hofbauer & Sigmund 1998: 26–27). Note in particular, that the intrinsic growth rate of a formally longer item is required to be larger than that of a formally shorter item. This will be important when we study diversification.

A2 Competition term

Let us inspect the competition term

$$c(\Delta) = c_{\max} \cdot e^{-\frac{(\Delta - \mu)^2}{2\tau^2}}$$

where $\Delta = s_j - s_i$ more closely. First, we see that it formally meets the requirements for c modeling asymmetric competition as outlined in 3.1. This is so, because $s_i < s_j$ implies $c(s_i - s_j) < c(s_j - s_i)$ as long as μ is positive (which is plausible because the effect of priming ultimately decreases with dissimilarity) and since $c(\Delta) > 0$ for all Δ. The parameter τ determines the steepness of the curve defined by c. If τ is small, then the effect of asymmetric priming is very strong. Conversely, if τ is large, then the curve is relatively flat so that asymmetric priming contributes less to the competition among the two items. At the same time τ defines the inflexion points of the function. If $\tau < \mu$ then the curve is locally convex in $c(0)$, as illustrated in Fig. 1, while it is locally concave if $\tau > \mu$. Also note that the first derivative fulfils $c'(s_i - s_j) > 0$ if $s_i \cong s_j$. That means, if j is only slightly longer than i then the strength of competition increases as the difference in substance between i and j increases. The latter observations will become important in the evolutionary analysis of the dynamical system (Appendix A3).

A3 Evolutionary diversification

We derive the conditions for evolutionary branching of formal substance, as a result of asymmetric priming. Let us denote invasion fitness, i.e. the expected growth rate of a rare item 2 exposed to an environment set by resident item 1 as $f(s_2, s_1)$. It is computed by taking the derivative of the right-hand side of equation (3a) with respect to x_2 and assuming that item 2 has frequency 0 (as it is rare) while item 1 rests at its population dynamical equilibrium $\hat{x}_1 = r(s_1)/c(0)$ (due to separation of time scales, see 3.3). We proceed as Kisdi (1999) and Law et al. (1997) (see also Doebeli 2011: 64–73 for a discussion of biological diversification driven by asymmetric competition). From the differential equation that defines the dynamics of item 1 (i.e. equation (3a)) we compute invasion fitness as

$$f(s_2, s_1) = r(s_2) - \frac{c(s_2, s_1)r(s_1)}{c(0)}.$$

Note that there is no term for self-regulation originating from item 2 (i.e. $c(0)$) since initially item 2 is supposed to be rare, so that self-regulation does not show any substantial effects. If $f(s_2, s_1)$ is positive, then item 2 can invade. If $f(s_2, s_1)$ is negative it will eventually go extinct so that the item 1, i.e. prototypical substance s_1, remains. Thus, if we want to know if items with slightly less or more substance can invade, we compute the partial derivative of $f(s_2, s_1)$ with respect to s_2 evaluated at s_1 This is the so-called ‘fitness gradient’:

$$D(s_2) \coloneqq \left[\frac{\partial f}{\partial s_1}\right]_{s_1 = s_2} = r'(s_2) - \frac{c'(0)r(s_1)}{c(0)}.$$

If the $D(s_2)$ is positive, variants with slightly more substance can invade, if $D(s_2)$ is negative, slightly shorter items can invade (Kisdi 1999: 152; Geritz et al. 1998: 37). As long as $D(s_2)$ is not close to zero, invasion implies that item 1 is replaced by item 2 (‘tube theorem’; see Geritz et al. 2002). The evolution of substance s unfolds as a stepwise sequence. Under the assumption of small and rare mutations, it can be shown (Dercole & Rinaldi 2008: 88–95) that evolution of s proceeds according to the differential equation

$$\dot{s} = k\hat{x}(s)D(s),$$

called the ‘canonical equation of adaptive dynamics’, where $k > 0$ denotes the ‘mutational rate’. It is proportional to the probability that an item is chosen to be a new prototype. In this paper, k is taken to be constant, although it is theoretically possible to let k depend on s. The equation operates on the evolutionary time scale measured in mutational steps. Since k is the rate of mutation, $1/k$ is the expected time between two substitution events, i.e. in our context between two events of adopting a new prototypical substance for some item.

Since $\hat{x}(s) > 0$, evolution goes either upwards if $D(s) > 0$ or downwards, i.e. representing successive formal reduction, if $D(s) < 0$. If, however, at some point s^* the fitness gradient vanishes, i.e. $D(s^*) = 0$, then evolution reaches an ‘evolutionary singularity’. In the present model this can be shown to be the case if

$$\frac{r'(s^*)}{r(s^*)} = \frac{c'(0)}{c(0)} = \frac{\mu}{\tau^2}.$$

If r is globally constant or decreasing, there is no such singularity, since r, μ and τ are positive by assumption.

In general there are four types of evolutionary singularities. First, evolution could have reached a local optimum at s^* which cannot be improved by changing s (‘continuously stable strategy’; CSS). Second, s^* could represent a local fitness-minimum so that evolution moves s away from s^* as soon as a mutant occurs (‘evolutionary repellor’). Third, s^* could represent an optimum, but if any perturbation occurs evolution drives s away from s^* ('Garden-of-Eden point'; GoE). Finally, and most relevant to our endeavor, s^* could represent an 'evolutionary branching point' (BP) at which the population splits into two coexisting variants. In biology, this is referred to as speciation; in linguistics this scenario represents synchronic coexistence of related linguistic variants.

Two formal criteria have been derived that have to be fulfilled for s^* to be an evolutionary branching point (Geritz et al. 1998: 38–40), namely that in the neighborhood of s^*

$$\text{(i)} \quad D'(s^*) < 0 \quad \text{and}$$

$$\text{(ii)} \quad \frac{\partial^2 f}{\partial s_2^2} > 0,$$

where condition (i) ensures that evolution proceeds towards s^*, since the fitness gradient is positive below s^* and negative above s^*, and condition (ii) ensures that s^* is not stable, since the fitness landscape in s^* is locally convex with respect to new variants. If both inequalities hold, then stable diversification is possible.

In order to evaluate the first condition the first derivative of the fitness gradient at the singular strategy has to be computed, which finally yields

$$\text{(i)} \quad r''(s^*) < r'(s^*) \underbrace{\frac{c'(0)}{c(0)}}_{>0},$$

where we know that $c'(0)/c(0) > 0$. Thus, (i) holds whenever r is strongly increasing at the singularity. If r is concave at the singularity ($r''(s^*) < 0$), and increasing ($r'(s^*) > 0$), then condition (i) follows immediately.

The second condition unfolds as

$$\text{(ii)} \quad r''(s^*) > c''(0) \underbrace{\frac{r(s^*)}{c(0)}}_{>0},$$

which holds if c is sufficiently concave around 0. If we explicitly compute $c'(0)$ and $c''(0)$ and substitute $c'(0)$ into $c''(0)$, we find that

$$c''(0) = \frac{c'(0)}{\tau^4} \cdot (\mu^2 - \tau^2)\,.$$

Furthermore we know that

$$\frac{c'(0)}{c(0)} = \frac{\mu}{\tau^2}$$

so that altogether, branching is possible if

$$\text{(i + ii)} \quad r'(s^*) \cdot \underbrace{\frac{\mu}{\tau^2}}_{>0} \underbrace{>}_{\text{(i)}} r''(s^*) \underbrace{>}_{\text{(ii)}} (\mu^2 - \tau^2) \cdot \underbrace{r(s^*) \cdot \frac{\mu}{\tau^6}}_{>0}\,.$$

A4 Sublexical evolutionary dynamics

We show that the evolutionary dynamics of the Lotka-Volterra system (1) where intrinsic growth is defined as

$$r(s) = Cs^{\alpha}(1-s)^{\pi}, r: [0,1] \to \mathbb{R}^{+},$$

exhibit an evolutionary singularity. To this end, we first have to derive the equilibrium of the system on the ecological time scale. In the case of a population consisting of a single type, i.e. a single exemplar/extension cloud whose prototypical diphone has length s, we find that at population-dynamical equilibrium frequency is given by $\hat{x} = Cs^{\alpha}(1-s)^{\pi}/c(0)$. Thus, the inverse-U shape of r is inherited by token frequency $\hat{x}$.[16] We know from Appendix A1 that two diphone variants of a specific diphone type with length s_1 and s_2, where $s_1 < s_2$, can coexist on the ecological time-scale if $1 > r(s_1)/r(s_2) > c(s_1 - s_2)/c(s_2 - s_1)$. This entails that coexistence is not possible if $s_1, s_2 > s_{\max} = \alpha/(\alpha + \pi)$. In that case, both lengths would be located in the decreasing region of r so that the first inequality would not be fulfilled. Thus, $s_{\max}$ provides a – necessary but not sufficient – upper bound for stable coexistence of two diphone variants of a single type that differ in duration. Put differently, two long variants of a diphone cannot coexist.

We know that an evolutionary singularity, if it exists, must fulfill $r'(s^*)/r(s^*) = \mu/\tau^2$ (see Appendix A3). After substituting r and the first derivative of r into this equation and solving it for s^* there are two solutions, only one of which is contained in the unit interval:

$$s^* = \frac{\mu + (\alpha+\pi)\tau^2 - \sqrt{-4\alpha\mu\tau + (\mu + (\alpha+\pi)\tau^2)^2}}{2\mu}.$$

A5 Lexical evolutionary dynamics

Here, we show that under the assumption of a Zipfian relationship between substance and utterance frequency, evolution of substance is unidirectional and that evolutionary branching is not possible. Let intrinsic growth be defined by a power law

$$r(s) = Cs^{-\kappa}, r: [0,1] \to \mathbb{R}^{+}$$

where $\kappa \geq 0$ and $C > 0$. From Appendix A1 we know that a single variant approaches a population dynamical equilibrium at $\hat{x} = Cs^{-\kappa}/c(0)$ so that the decreasing shape of the intrinsic growth rate is again inherited by token frequency at equilibrium as desired. However, since $r'(s) = -\kappa Cs^{-\kappa-1} < 0$ it follows that two variants which differ in length cannot stably coexist (see condition for the existence of an internal equilibrium in A1). If we compute the fitness gradient (Appendix A3) we see that

$$D(s) = -C\underbrace{\left(\kappa s^{-\kappa-1} + \frac{s^{-\kappa}\mu}{\tau^2}\right)}_{>0} < 0,$$

so that length evolves unidirectionally towards ever smaller values.

Since the fitness gradient never vanishes, there are no evolutionary singularities which immediately precludes evolutionary branching. Note, that this is even the case if $\kappa = 0$, i.e. if the intrinsic growth rate does not depend on formal substance. That is, if there is only

[16] It is worth pointing out that Kuperman et al.'s (2008) model in fact tracks logged token frequency as a function of duration rather than raw token frequency. We do not consider this a problem, since $e^{\hat{x}}$ as a function of s still displays an inverse-U shape.

asymmetric priming, then evolution of substance is unidirectional, as hypothesized by Jäger and Rosenbach (2008).

References

Altmann, G. 1983. Das Piotrowski-Gesetz und seine Verallgemeinerungen. In Karl-Heinz Best & Jörg Kohlhase (eds.), *Exakte Sprachwandelforschung*: *Theoretische Beiträge, statistische Analysen und Arbeitsberichte* (Göttinger Schriften zur Sprach- und Literaturwissenschaft 2), 54–90. Göttingen: Edition Herodot.

André Martinet. 1955. *Économie des changements phonétiques. Traité de phonologie diachronique*. Bern: A. Francke S. A.

Baayen, R. H. 2001. *Word frequency distributions*. Dordrecht: Kluwer Academic Publishers.

Baumann, Andreas, Nikolaus Ritt & Christina Prömer. 2016. Diachronic dynamics of Middle English phonotactics provide evidence for analogy effects among lexical and morphonotactic consonant clusters. *Papers in Historical Phonology* 1. 50–75.

Baumann, A., Kaźmierski, K.. 2016. A dynamical-systems approach to the evolution of morphonotactic and lexical consonant clusters in English and Polish. Yearbook of the Poznan Linguistic Meeting 2, 115-139.

Blythe, Richard A. & William Croft. 2012. S-curves and the mechanism of propagation in language change. *Language* 88(2). 269–304.

Bock, J. K. 1986. Syntactic persistence in language production. *Cognitive Psychology* 18. 355–387.

Bock, J. K. & Zenzi Griffin. 2000. The persistence of structural priming. Transient activation or implicit learning? *Journal of Experimental Psychology: General* 129(2). 177–192.

Bock, J. K. & Helga Loebell. 1990. Framing sentences. *Cognition* 35. 1–39.

de Boer, Bart. 2000. Self-organization in vowel systems. *Journal of Phonetics* 28(4). 441–465.

Breban, Tine, Jeroen Vanderbiesen & Kristin Davidse. 2012. Introduction: New reflections on the sources, outrcomes, defining features and motivations of grammaticalization. In Kristin Davidse, Tanja Mortelmans, Lieselotte Brems & Tine Breban (eds.), *Grammaticalization and language Change. New Reflections*, 1–35. Amsterdam: John Benjamins Pub. Co.

Brinton, Laurel J. & Elizabeth C. Traugott. 2005. *Lexicalization and Language Change*. Cambridge: CUP.

Bybee, Joan. 2010. *Language, Usage and Cognition*. Cambridge: Cambridge University Press.

Bybee, Joan, Revere D. Perkins & William Pagliuca. 1994. *The Evolution of Grammar. Tense, Aspect, and Modality in the Languages of the World*. Chicago: University of Chicago Press.

Chang, Franklin, Gary S. Dell & J. K. Bock. 2006. Becoming syntactic. *Psychological Review* 113(2). 234-272.

Croft, William. 2000. *Explaining language change*: *An evolutionary approach* (Longman linguistics library). Harlow, England, New York: Longman.

Cushing, Jim M. 1998. *An introduction to structured population dynamics*. Philadelphia: SIAM.

Daland, Rober & Janet Pierrehumbert. 2011. Learning diphone-based segmentation. *Cognitive Science* 35. 119–155.

Denison, David. 2003. Log(ist)ic and simplistic S-curves. In Raymond Hickey (ed.), *Motives for Language Change*, 54–70. Cambridge: Cambridge University Press.

Dercole, Fabio & Sergio Rinaldi. 2008. *Analysis of evolutionary processes*. Princeton, New York: Princeton University Press.

Dieckmann, Ulf & Richard Law. 1996. The dynamical theory of coevolution: a derivation from stochastic ecological processes. *Journal of Mathematical Biology* 34. 579–612.

Diessel, Holger. 2007. Frequency effects in language acquisition, language use, and diachronic change. *New ideas in psychology*(25). 108–127.
Diessel, Holger & Martin Hilpert. 2016. *Frequency Effects in Grammar*. Linguistics: Oxford Research Encyclopedias. Oxford: Oxford University Press.
Doebeli, M. 2011. *Adaptive diversification*. Princeton: Princeton University Press.
Dressler, Wolfgang U. 1990. The cognitive perspective of "naturalist" linguistic models. *Cognitive Linguistics* 1(1). 75–98.
Dressler, Wolfgang U. & Katarzyna Dziubalska-Kołaczyk. 2006. Proposing Morphonotactics. *Wiener Linguistische Gazette* 73. 69–87.
Dressler, Wolfgang U. Katarzyna Dziubalska-Kołaczyk & Lina Pestal. 2010. Change and variation in morphonotactics. *Folia Linguistica Historica* 31. 51–68.
Dressler, Wolfgang U. Katarzyna Dziubalska-Kołaczyk & Rosella Spina. 2001. Sources of markedness in language structures. *Folia Linguistica Historica* XXII(1). 103–135.
Dziubalska-Kolaczyk, Katarzyna. 2002. *Beats-and-binding phonology*. Frankfurt am Main, New York: Peter Lang.
Ernestus, Mirjam. 2014. Acoustic reduction and the roles of abstractions and exemplars in speech processing. *Lingua* 142. 27–41.
Evans, Nicholas & Stephen C. Levinson. 2009. The myth of language universals: Language diversity and its importance for cognitive science. *The Behavioral and brain sciences* 32(5). 429-48; discussion 448-494.
Fischer, Olga. 2007. *Morphosyntactic Change. Functional and Formal Perspectives*. Oxford: Oxford University Press.
Fowler, Carol A. & Jonathan Housom. 1987. Talkers‘ signaling of ‘new‘ and ‘old‘ words in speech and listeners‘ perception and use of the distinction. *Journal of Memory and Language* 25. 489–504.
Geritz, Stefan A. H. 2010. LV asymmetric competition.nb. Mathematica script. http://wiki.helsinki.fi/download/attachments/58368980/Competition.pdf?api=v2 (27 June, 2017).
Geritz, Stefan A. H. Mats Gyllenberg, Frans J. A. Jacobs & Kalle Parvinen. 2002. Invasion dynamics and attractor inheritance. *Journal of Mathematical Biology* 44. 548–560.
Geritz, Stefan A. H. Eva Kisdi, Géza Meszéna & Johan A. J. Metz. 1998. Evolutionarily Singular Strategies and the Adaptive Growth and Branching of the Evolutionary Tree. *Evolutionary Ecology* 12. 35–57.
Greenberg, Joseph. 1966. Synchronic and diachronic universals in phonology. *Language* 42(2). 508–517.
Haiman, John. 1994. Ritualization and the development of language. In William Pagliuca (ed.), *Perspectives on Grammaticalization*, 3–28. Amsterdam: John Benjamins Pub. Co.
Haspelmath, Martin. 1999. Why is grammaticalization irreversible? *LINGUISTICS* 37. 1043–1068.
Haspelmath, Martin. 2004. On directionality in language change with particular reference to grammaticalization. In Olga Fischer, Muriel Norde & Harry Perridon (eds.), *Up and down the Cline – The Nature of Grammaticalization* (Typological studies in language), 17–44. Amsterdam: John Benjamins Publishing Company.
Hawkins, John A. 2007. *Efficiency and complexity in grammars* (Oxford linguistics). Oxford: Oxford University Press.
Hay, Jennifer & Harald Baayen. 2003. Phonotactics, parsing and productivity. *Rivista di Linguistica* 15. 99–130.
Hay, Jennifer & Harald Baayen. 2005. Shifting paradigms: gradient structure in morphology. *Trends in cognitive sciences* 9. 342–348.
Heine, Bernd, Ulrike Claudi & Freiderike Hünnemeyer (eds.). *Grammaticalization: A Conceptual Framework*. Chicago: Chicago University Press.

Heine, Bernd & Tania Kuteva. 2002. *World Lexicon of Grammaticalization*. Cambridge: Cambridge Univ. Press.

Heine, Bernd & Tania Kuteva. 2007. *The Genesis of Grammar*. Oxford: Oxford University Press.

Heine, Bernd & Mechthild Reh. 1984. *Grammaticalization and Reanalysis in African languages*. Hamburg: Buske.

Hilpert, Martin & David Correia Saavedra. 2016. The unidirectionality of semantic changes in grammaticalization: An experimental approach to the asymmetric priming hypothesis. *English Language and Linguistics*. 1–24.

Himmelmann, Nikolaus P. 2004. Lexicalization and grammaticalization: opposite or orthogonal? In Walter Bisang, Nikolaus P. Himmelmann & Björn Wiemer (eds.), *What makes Grammaticalization? A Look from its Fringes and its Components*, 21–42. Berlin: Mouton de Gruyter.

Hofbauer, Josef & Karl Sigmund. 1998. *Evolutionary games and population dynamics*. Cambridge: Cambridge University Press.

Honeybone, Patrick. 2008. Lenition, weakening and consonantal strength: tracing concepts through the history of phonology. In J. Brandão de Carvalho, T. Scheer & P. Ségéral (eds.), *Lenition and Fortition*, 9–93. Berlin: Mouton de Gruyter.

Hopper, Paul J. & Elizabeth C. Traugott. 2003. *Grammaticalization*. Cambridge: Cambridge University Press.

Jäger, Gerhard. 2008a. Applications of game theory in linguistics. *Language and Linguistics compass* 2(3). 406–421.

Jäger, Gerhard. 2008b. Language evolution and George Price's ''General theory of selection. *Language in flux: Dialogue coordination, language variation, change and evolution. London: College Publications*. 53–82.

Jäger, Gerhard & Anette Rosenbach. 2008. Priming and unidirectional language change. *Theoretical Linguistics* 34(2). 85–113.

Jurafsky, Daniel, Alan Bell, Michelle L. Gregory & William D. Raymond. 2001. Probabilistic relations between words. Evidence from reduction in lexical production. In Joan Bybee & Paul J. Hopper (eds.), *Frequency and the Emergence of Linguistic structure*, 229–254. Amsterdam: John Benjamins Pub. Co.

Kaschak, Michael. 2007. Long-term structural priming affects subsequent patterns of language production. *Memory & Cognition* 35(5). 925–937.

Kaźmierski, Kamil. 2015. *Vowel-shifting in the English language*: *An evolutionary account* (Topics in English linguistics volume 88). Berlin: De Gruyter.

Kemps, Rachel, Lee Wurm, Mirjam Ernestus, robert Schreuder & Harald Baayen. 2005. Prosodic cues for morphological complexity in Dutch and English. *Language and Cognitive Processes* 20. 43–73.

Kirby, Simon, Mike Dowman & Thomas L. Griffiths. 2007. Innateness and culture in the evolution of language. *Proceedings of the National Academy of Sciences* 104(12). 5241–5245.

Kisdi, Eva. 1999. Evolutionary Branching under Asymmetric Competition. *Journal of Theoretical Biology* 197. 149–162.

Kooij. 1971. *Ambiguity in natural language: an investigation of certain problems in its linguistic description*. Amsterdam: North-Holland Publ. Co.

Koriat, Asher. 1981. Semantic facilitation in lexical decision as a function of prime-target association. *Memory & Cognition* 9. 587–598.

Kristiansen, Gitte. 2006. Towards a usage-based cognitive phonology. *International Journal of English Studies* 6(2). 107–140.

Kroch, Anthony. 1989. Reflexes of Grammar in Patterns of Language Change. *Language Variation and Change* 1. 199–244.

Kroch, Anthony, Beatrice Santorini & Lauren Delfs. 2004. *Penn-Helsinki Parsed Corpus of Early Modern English*. http://www.ling.upenn.edu/hist-corpora/.

Kroch, Anthony & Ann Taylor. 2000. *Penn-Helsinki Parsed Corpus of Middle English*. http://www.ling.upenn.edu/hist-corpora/.

Kuperman, Victor, Mirjam Ernestus & Harald Baayen. 2008. Frequency distributions of uniphones, diphones, and triphones in spontaneous speech. *The Journal of the Acoustical Society of America* 124. 3897–3908.

Langacker, Ronald W. 1977. Syntactic reanalysis. In Charles N. Li (ed.), *Mechanisms of Syntactic Change*, 57–139. Austin, TX: University of Texas Press.

Law, Richard, Paul Marrow & Ulf Dieckmann. 1997. On evolution under asymmetric competition. *Evolutionary Ecology* 11. 485–501.

Lehmann, Christian. 1985. Grammaticalization: synchronic variation and diachronic change. *Lingua e Stile* 20. 308–318.

Leykum, Hannah, Sylvia Moosmüller & Wolfgang U. Dressler. 2015. Word-final (mor-)phonotactic consonant clusters in Standard Austrian German. In The Scottish Consortium for ICPhS 2015 (ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow.

Loebell, Helga & J. K. Bock. 2003. Structural priming across languages. *LINGUISTICS* 41. 791–824.

Maynard Smith, John. 1982. *Evolution and the theory of games*. Cambridge, New York: Cambridge University Press.

McNamara, Timothy. 2005. *Semantic Priming. Perspectives from memory and word recognition*. New York: Psychology Press.

Mompeán-González, José. 2004. Category overlap and neutralization: The importance of speakers' classifications in phonology. *Cognitive Linguistics* 15. 429–469.

Narrog, Heiko & Bernd Heine. 2011. *The Oxford handbook of grammaticalization* (Oxford handbooks in linguistics). Oxford: Oxford University Press.

Nathan, Geoffrey S. 2006. Is the Phoneme Usage-Based? – Some Issues. *International Journal of English Studies* 6(2). 173–194.

Neely, James. H. 1991. Semantic priming effects in visual word recognition: A selective review of current findings and theories. In Derek Besner & Glyn W. Humphreys (eds.), *Basic processes in reading. Visual word recognition*, 264–336. Hillsdale, NJ: Erlbaum.

Niyogi, Partha. 2006. *The computational nature of language learning and evolution*. Cambridge, MA.

Norde, Muriel. 2009. *Degrammaticalization*. Oxford: Oxford University Press.

Nowak, Martin A. 2000. The basic reproductive ratio of a word, the maximum size of a lexicon. *Journal of theoretical biology* 204(2). 179–189.

Nowak, Martin A. 2006. *Evolutionary dynamics*: *Exploring the equations of life*. Cambridge, Mass: Belknap Press of Harvard University Press.

Nowak, Martin A. Joshua Plotkin & Vincent Jansen. 2000. The evolution of syntactic communication. *Nature* 404(6777). 495–498.

Page, Karen & Martin A. Nowak. 2002. Unifying evolutionary dynamics. *Journal of Theoretical Biology*(219). 93–98.

Pierrehumbert, Janet. 2001. Exemplar dynamics: Word frequency, lenition, and contrast. In Joan Bybee & Paul J. Hopper (eds.), *Frequency effects and the emergence of lexical structure*, 137–157. Amsterdam [a.e.]: Benjamins.

Pierrehumbert, Janet. 2016. Phonological Representation: Beyond Abstract Versus Episodic. *Annual Review of Linguistics* 2. 33–52.

Pietarinen, Ahti-Veikko. 2003. Games as formal tools versus games as explanations in logic and science. *Foundations of Science* 8. 317–364.

Plag, Ingo, Julia Homann & Gero Kunter. 2015. Homophony and morphology: The acoustics of word-final S in English. *Journal of Linguistics*. 1–36.

Plag, Ingo, Gero Kunter & Mareile Schramm. 2011. Acoustic correlates of primary and secondary stress in North American English. *Journal of Phonetics* 39(3). 362–374.

R Development Core Team. 2013. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
Ritt, Nikolaus. 2004. *Selfish sounds and linguistic evolution*: *A Darwinian approach to language change*. Cambridge: Cambridge University Press.
Ritt, Nikolaus & Kamil Kaźmierski. 2015. How rarities like gold came to exist: on co-evolutionary interactions between morphology and lexical phonotactics. *English Language and Linguistics*(19). 1–29.
Roberts, Ian. 2007. *Diachronic syntax*. Oxford: Oxford University Press.
Roberts, Ian G. & Alexiadou Roussou. 2003. *Syntactic Change. A Minimalist Approach to Grammaticalization*. Cambridge: Cambridge Univ. Press.
Rueckl, Jay G. 1990. Similarity effects in word and pseudoword repetition priming. *Journal of Experimental Psychology: Learning, Memory and Cognition* 16(3). 374–391.
Schmid, Hans-Jörg (ed.) (2016). *Entrenchment and the psychology of language learning*. Berlin: Mouton de Gruyter.
Shields, Lynne W. & David A. Balota. 1991. Repetition and associative context effects in speech production. *Language and Speech* 34(1).
Smet, Hendrik de. 2013. *Spreading patterns: Diffusional change in the English system of complementation*. Oxford: Oxford University Press.
Snider, Neal. 2009. Similarity and Structural Priming. *Proceedings of the Cognitive Science Society* 31.
Solé, Ricard V. 2011. *Phase transitions*. Princeton: Princeton University Press.
Solé, Ricard V. Bernat Corominas-Murtra & Jordi Fortuny. 2010. Diversity, competition, extinction: the ecophysics of language change. *Journal of The Royal Society Interface* 7(53). 1647–1664.
Sommerer, Lotte. 2015. The influence of constructions in grammaticalization: Revisiting category emergence and the development of the definite article in English. In Jóhanna Barðdal, Elena Smirnova, Lotte Sommerer & Spike Gildea (eds.), *Diachronic Construction Grammar*, 107–133. Amsterdam: John Benjamins Pub. Co.
Thompson, Bill, Simon Kirby & Kenny Smith. 2016. Culture shapes the evolution of cognition. *Proceedings of the National Academy of Sciences of the United States of America* 113(16). 4530–4535.
Thompson-Schill, Sharon L. Kenneth J. Kurz & Gabrieli, John D. E. 1998. Effects of semantic and associative relatedness on automatic priming. *Journal of Memory and Language* 38. 440–458.
Tooley, Kristen M. & Matthew J. Traxler. 2010. Syntactic Priming Effects in Comprehension: A Critical Review. *Language and Linguistics compass* 4(10). 925–937.
van Gelderen, Elly. 2004. *Grammaticalization as Economy*. Amsterdam: John Benjamins Pub. Co.
Wang, William & James Minett. 2005. The invasion of language: Emergence, change and death. *Trends in Ecology and Evolution* 20. 263–269.
Wedel, Andrew. 2006. Exemplar models, evolution and language change. *The Linguistic Review* 23(3). 247–274.
Wolfram Research, Inc. 2016. *Mathematica*. Champaign, Illinois.
Yang, Charles. 2000. Internal and external forces in language change. *Language Variation and Change* 12. 231–250.
Zehentner, Eva. 2017. Ditransitives in Middle English: On semantic specialisation and the rise of the dative alternation. *English Language and Linguistics* 11. 1–27.
Zipf, Georg K. 1949. *Human Behaviour and the Principle of Least Effort: An Introduction to Human Ecology*. Cambridge, MA.: Addison-Wesley Press.

DE GRUYTER OPEN
Yearbook of the Poznań Linguistic Meeting 2 (2016), pp. 115–139
DOI: 10.1515/yplm-2016-0006

A dynamical-systems approach to the evolution of morphonotactic and lexical consonant clusters in English and Polish

Andreas Baumann[1] and Kamil Kaźmierski[2]
[1] *University of Vienna, andreas.baumann@univie.ac.at*
[2] *Adam Mickiewicz University, Poznań, kamil.kazmierski@wa.amu.edu.pl*

Abstract
Consonant clusters appear either lexically within morphemes or morphonotactically across morpheme boundaries. According to extant theories, their diachronic dynamics are suggested to be determined by analogical effects on the one hand as well as by their morphological signaling function on the other hand. This paper presents a mathematical model which allows for an investigation of the interaction of these two forces and the resulting diachronic dynamics. The model is tested against synchronic and diachronic language data. It is shown that the evolutionary dynamics of the cluster inventory crucially depend on how the signaling function of morphonotactic clusters is compromised by the presence of lexical items containing their morpheme internal counterparts.

Keywords: morphonotactics; mathematical modeling; evolutionary linguistics; morphological complexity; analogy.

1. Introduction

There is strong evidence that consonant clusters in general, and final consonant clusters in particular, are dispreferred. First, they are much rarer typologically than languages allowing at most one coda consonant. For example, in the World Phonotactics Database (Donohue et al. 2013), which contains data on 2,378 languages, as many as 79.3% of the languages allow no more than one coda consonant. Second, consonant clusters are reduced in casual speech (Dziubalska-Kołaczyk and Zydorowicz 2014; Madelska 2005; Shockey 2003). Third, they are acquired late in first language acquisition (Jarosz et al., submitted). Despite this evidence of bias against them, however, consonant clusters are well-established in a number of languages, including Polish and English.

Language specific phonotactics often restrict the set of clusters that are allowed in morphologically simple words. For example, English does not allow its words to end in /-md/ and Polish does not allow its simplex words to end in /-gw/. However, such phonotactic restrictions are considerably relaxed when it comes to morphologically complex forms. And so /md/ can surface in English *seemed* /siːmd/ and /gw/ in Polish *pomógł* /ˈpɔmugw/ 'he helped' when a past tense suffix is attached. The difference in the range of sound sequences allowed to surface lexically (i.e. phonotactically) on the one hand, and those allowed to surface in morphologically complex words (i.e. "morphonotactically"), is accounted for in the theory of "morphonotactics", as proposed by Dressler and Dziubalska-Kołaczyk (2006). One of the claims they make is that markedness restrictions are less tight on morphonotactic clusters than on lexical ones, because the markedness of morphonotactic clusters might help to signal morphological operations. In fact, it is the very ill-formedness of morphonotactic clusters that makes them semiotically conspicuous, and therefore well-designed for signaling morphological complexity.[1]

Some clusters are purely morphonotactic, i.e. in a given language they can be found only in morphologically complex words. This happens to be the case with /md/ in English (e.g. *seemed*, *doomed*) which never occurs in any morphologically simple form. Other clusters, such as /mp/ in English (e.g. *lamp*, *damp*) are purely lexical, since they can be found only in morphologically simple forms, and never arise from morphological operations. These two cases can be seen as located at the opposite ends of a scale indicating the degree to which a cluster is morphonotactic (Dressler and Dziubalska-Kołaczyk 2006). Between the two endpoints, intermediate cases can be found. Moving one step away from purely morphonotactic clusters, clusters which are morphonotactic by "strong default" can be encountered. A good example is English /ts/, which in the vast majority of cases shows up across word boundaries, as in *cats*, *sits* but occasionally inhabits morphologically simple words, as in *waltz*. The next step on the scale is occupied by clusters that are morphonotactic by "weak default", such as /ks/ in English *lacks*, *sacks* but also *tax*, *box* along with a number of other simple forms. These are more often morphonotactic than lexical, just like the strong default clusters are, but they show up as lexical more often than the

---

[1] On a related note, as one of the Reviewers points out, statistical information about phoneme sequences can also be used to help segment speech into words. For example, English phoneme pairs have been found to follow a strongly bimodal distribution in that each phoneme pair tends to appear either within words or across word boundaries (Hockema 2006).

strong default clusters do. And finally, there are clusters which are roughly as likely to appear in morphologically complex and morphologically simple forms, and these can be termed ambiguous. Such is the status of /nd/ in English, frequent in past tense and past participle forms such as *banned*, *sinned* but also in numerous simple forms such as *hand* or *kind*. In sum, the cluster types that occur in any language can always be arranged on a scale reflecting the different proportions to which they are lexical or morphonotactic. Now, as Dressler and Dziubalska-Kołaczyk (2006: 83) very plausibly suggest, the position of a given cluster on this scale influences the adequacy of that cluster to signal morphological complexity. The more lexical it is, the less well does it work as a signal of morphological complexity.

Crucially, this claim has interesting implications regarding the historical evolution of languages (cf. Dressler et al. 2010): although the factors it refers to are grounded in cognition and physiology, their reflexes should be observable in diachronic data.[2] Thus, if speakers seek to enhance the signaling function of clusters they should prefer (and select for) clusters to be more frequent near the two ends of the lexicality scale than near the middle (see also 2.1 below).

At the same time, however, there are also arguments for predicting the opposite. In particular, analogy effects have been posited (e.g. in Hogg and McCully 1987) by which morphonotactic cluster models support the emergence and stabilizes lexical counterparts – and *vice versa*. This clearly predicts that clusters should be more frequent near the middle of the lexicality scale than near its ends (see 2.2. below)

As for how such preferences shape the cluster inventory of a language, there is a number of mechanisms which can alter the status of a cluster as morphonotactic/lexical, as well as its prevalence diachronically. These include the reinterpretation of a morphonotactic cluster as lexical through loss of transparency (/l+θ/ → /lθ/ as in *health*), borrowing of loan words with a particular cluster (e.g. /ns/ as in *commence*, and a wealth of other Romance loan words); sound change such as cluster simplification (/mb/ → /m/ as in *bomb*), devoicing (/ld/ → /lt/ as in *spelt*, in certain dialects), schwa loss /nəd/ → /nd/ as in *sinned*), and schwa epenthesis (e.g. /lm/ → /ləm/ as in *film*, in certain dialects). The issue of the implementation of the change in status and prevalence at the level of an individual cluster, i.e. why a particular mechanism is applied to a particular cluster

[2] Thus, morphonotactics could provide tools for explaining long-term developments which otherwise defy a unified account, such as long-term reduction of consonantal complexity in English. This might be particularly true in cases where sociolinguistic explanations fail.

at a particular point in time, however, is beyond the scope of our investigation. We are taking a broader perspective and asking why such mechanisms are applied to particular clusters and not to others.

More specifically, we address the question of how morphological signaling and analogy interact and determine the diachronic evolution of consonant clusters by means of a mathematical dynamical-systems model. In a nutshell, the model describes the growth (or decline) of a population of cluster items that belong to a specific cluster type (like e.g. word-final /nd/). More precisely, this population of cluster items is divided into two subpopulations: morphonotactic instances and lexical instances of that cluster type, respectively. The dynamics of this population of cluster items is determined by various factors such as the likelihoods of transmitting or memorizing a cluster. Most importantly, the respective effects that morphonotactic signaling and analogy have on the growth of the cluster population are built into the model. The model then allows us to investigate the diachronic evolution of a specific cluster type with respect to its distribution of morphonotactic and lexical instances. If this is done for a large number of cluster types, one can simulate the diachrony of entire cluster inventories. These simulated developments finally are compared against historical data, in order to check the validity of the model.

Our research questions are the following:

(1) How do the opposed forces deriving from the semiotic signaling function of clusters on the one hand and analogy effects on the other interact and contribute to the diachronic evolution and synchronic distribution of consonant clusters?

(2) How are the clusters which can act as both lexical and morphonotactic spread over the lexicon in written/spoken, synchronic/diachronic data?

(3) How do the results of mathematical modeling and simulations of the evolutionary dynamics of consonant clusters compare to empirically observable distributions?

The paper is structured as follows. Section 2 introduces the Semiotic Utility Function, which is a formal expression of the relationship between the degree to which a cluster is morphonotactic and its ability to signal morphological complexity. It formalizes the predictions of morphonotactics (Dressler and Dziubalska-Kołaczyk 2006: 83) for the shape of this function (2.1), discusses the role of analogy (2.2), and compares and contrasts the predictions of morphonotactics and of analogy for the distribution of ambiguous and unambiguous clusters in the lexicon (2.3). It ends with the description of the Polish and English data used in the study (2.4). Section 3 introduces our modeling approach, while Section 4 discusses the applications of the model to diachronic English data (4.1.) and synchronic Polish data (4.2). Section 5 summarizes our conclusions.

2. Opposing forces in morphonotactics

2.1. Morphonotactics and predictions about semiotic utility

Dressler and Dziubalska-Kołaczyk (2006: 83) state that "[p]rototypical morphonotactic clusters [...] have the function of co-signaling the existence of a morphological rule, morphonotactic default clusters [...] fulfill this function less adequately, while phonotactic clusters [...] cannot fulfill this function [...]". As a formalization of this claim, we propose the Semiotic Utility Function. Semiotic utility is the degree to which a given cluster is capable of signaling morphological complexity. That the Semiotic Utility Function must be a decreasing function follows from the statement cited above. However, Dressler and Dziubalska-Kołaczyk's claim is agnostic about the shape of the function (linear, concave or convex). The various possible shapes of the function are presented in Figure 1. Consider a consonant cluster that occurs only across morpheme boundaries, i.e. purely morphonotactically. While a convex shape indicates that by adding items to the lexicon which include the same cluster morpheme internally, the complexity-signaling ability of the morphonotactic cluster is substantially compromised, a concave shape means that adding the same items to the lexicon would not have a huge impact on the cluster's semiotic utility.

We suggest that the curvature of the Semiotic Utility Function may be linked to the extent to which a specific language makes use of inflectional morphology. Languages clearly differ in the extent to which they rely on morphological coding versus word order. Polish relies on morphology much more than English does, as is clear from Sadeniemi et al.'s (2008) study employing both a Kolmogorov complexity-based approach and a morpheme-level comparison to measure the morphological complexity of languages.

Arguably, languages that tend to express grammatical functions by morphemes are expected to be more sensitive to lexical items that feature a usually morphonotactic cluster morpheme internally, than languages that express the same grammatical functions analytically are, because the latter do not have to rely that much on proper recognition of morphemes. That Polish relies on clear boundaries between morphemes more than English does is suggested by the observation relating to language acquisition stating that "the segmentation of bound morphemes may be important much earlier for learners of highly inflecting languages [...]" (Peters 1997). Indeed, the hypothesis that "the richer the morphology of the language they are acquiring is, the faster children will develop morphology" (Dressler 2007: 3) is supported by the studies on the development on morphology in typologically disparate languages collected in Laaha and Gillis (2007), where morphology has been found to develop faster in children acquiring Slavic languages (Croatian and Russian) than Germanic languages (German and Dutch). This is in line with the idea that recognizing morphological boundaries is more important for speakers of languages with more synthetic morphology, such as Polish, than for speakers of more analytic languages, such as English. Thus, we will assume that synthetic languages exhibit more convex Semiotic Utility Functions, while analytic languages feature more concave ones.[3] We will come back to this in Section 4, when we apply our approach to English (representing a more analytic language) and Polish (being a more synthetic one). Indeed, as we shall see, the shape of this function will play a crucial role in the diachronic development of consonant clusters.

2.2. Analogy and mutual support

There are analogy-based lines of argumentation for a mutually beneficial relationship between morphonotactic and lexical clusters. The general idea is that the clusters that occur morphonotactically support lexical occurrences, or that vice versa, the presence of a certain cluster in a given language lexically supports its morphonotactic occurrences. Hogg and McCully (1987: 47) point to a possible interaction of the former type, saying that "the type of syllable structure found in a simplex word such as *wind* (/waind/) has been protected through analogy with inflected forms such as *weaned*". In another study that corrobo rates analogical support from morphonotactic to lexical consonant clusters, Bau-

[3] Crucially, this assumption is made *a priori* and independently from other phonological features of a language. Rather, the shape of this function is assumed to be conditioned by morphosyntactic properties of that language.

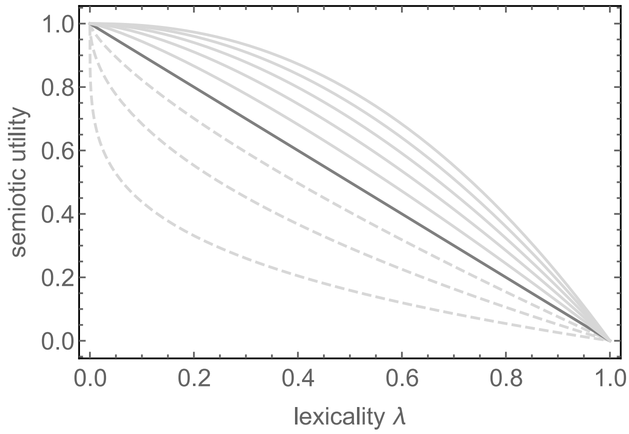


Figure 1. Semiotic utility as a decreasing function of lexicality λ, where λ is a measure of the number of simple word types in the lexicon a cluster occurs in. It can be linear (solid dark gray; each additional item decreases the cluster's semiotic signaling function to an equal extent), convex (dashed light gray; additional lexical items substantially decrease the signaling function of a primarily morphonotactic cluster) or concave (solid light gray; additional lexical items decrease the signaling function of a primarily morphonotactic cluster not to a large extent).

mann et al. (2015) suggest that the diachronic development of the English consonant cluster inventory reveals frequency effects among lexical and morphonotactic clusters.

On the other hand, based on a data driven study of English noun-noun compounds, Martin (2007: 99) claims that "the categorial phonotactic restrictions that hold within morphemes also hold gradiently across morpheme boundaries", and thus indirectly supports the latter direction, in that morphonotactic consonant clusters benefit from being lexically licensed by their phonotactic counterparts. In total, this gives us a bidirectional mutual relationship between morphonotactic and lexical clusters.

2.3. Comparing the predictions

The two approaches to the relationship between morphonotactic and lexical clusters, the morphonotactic and the analogy-based approach, make contradictory predictions regarding the frequency distribution of cluster types. Following the logic of morphonotactics, clusters either specialize for the lexical domain or for the morphonotactic domain. Since the ambiguous clusters are not apt to signal morphological complexity, they should be susceptible to loss through language change. The surviving lexicon is expected to be populated by high proportions of purely morphological clusters, which signal morphological complexity reliably, and of purely lexical clusters, which do not signal morphological complexity at all. Thus, morphonotactics predicts a V-shaped distribution of morphonotactic and lexical clusters throughout the lexicon.

If we follow the analogy-based view of the relationship between morphonotactic and lexical clusters, on the other hand, we would predict a Λ-shaped distribution. If the two kinds of clusters are assumed to reinforce one another, then ambiguous clusters should be favored in language change. Such conditions would leave the lexicon populated with high rates of clusters which are morphonotactic and lexical at the same time. The two predictions are represented schematically in Figure 2.

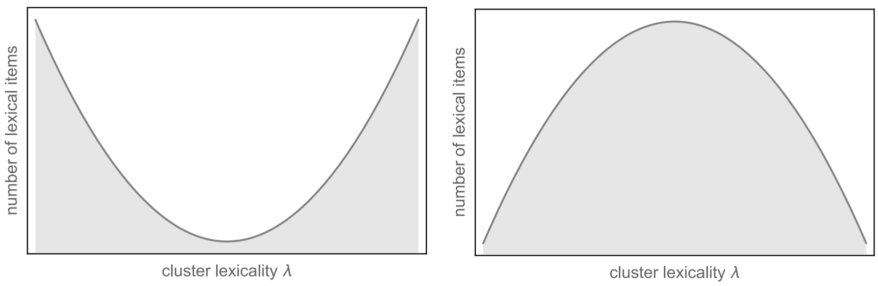


Figure 2. Schematic representation of the frequency distributions of cluster types with respect to the degree of lexicality λ which is a measure of the number of lexical items containing a given cluster type. On the left: Under the morphonotactic approach, a lack of ambiguous cluster types is expected, while most clusters are either primarily lexical or primarily morphonotactic. On the right: In the presence of mutually supporting analogical effects, most cluster types are expected to be ambiguous, i.e. medially lexical.

2.4. Data

For the present study we used three data sets. The first set comprises English diachronic written data. It is the current version of the database created by the

Evolution of Consonant Clusters in English (ecce) research project (www.ecce.univie.ac.at). It is a morphologically coded database of all final consonant clusters in the Penn-Helsinki Parsed Corpus of Middle (and Early Modern) English (Kroch and Taylor 2000; Kroch et al. 2004), containing approximately 240,000 tokens, spanning the 12th to the 17th century. The second set comprises Polish synchronic dictionary data, namely all initial clusters, yielding 232 cluster types spread over 2,137 words (Zydorowicz et al., forthcoming). The third set comprises Polish synchronic spoken data. It comprises all consonant clusters appearing in a corpus of spoken Polish, containing 375 cluster types distributed over 10,051 tokens (Kaźmierski 2015).

A preliminary inspection of the data allows for some conclusions regarding the predictions following from morphonotactics and from analogy for the distribution of cluster types. To begin with, the Polish data follow a V-shaped distribution. This is the case both for spoken and for dictionary data, with no appreciable difference in distribution between the two sets (comparison of 5-bin distributions; distribution-difference effect size estimated at Cramer's $V = .18 \pm .08$ (95% CI); $\chi^2 = 19.35$). Thus, we find no evidence of substantial analogy effects on the stability of consonant clusters in Polish. Additionally, the lack of difference between spoken and written data supports the validity of using written evidence for phonological research. This is welcome news not least in view of our diachronic English data, which by necessity, are written only. English diachronic data form a suitable testing ground for comparing the effects of morphonotactics and analogy. In Middle through early Modern English, large numbers of consonant clusters were created due to schwa loss. Schwa loss itself is assumed to have been phonologically conditioned, and not sensitive to morphology. However, if morphonotactics and/or analogy are indeed relevant for the diachronic survival of clusters, the reflexes of these forces should be seen in the English data. Indeed, for the English data set, we observe diachronic change in the distribution of morphonotactic and lexical clusters. Comparing the frequency distribution of clusters with increasing ratio of lexical tokens for the earliest stage (EME) with that of a later stage (EmodE), we notice a transition from a V-shaped bimodal to a W-shaped trimodal distribution (see Figure 3). This indicates a combined influence of morphology and analogy. While analogy not playing a role would give a V-shaped bimodal distribution and analogy playing a dominating role would give a Λ-shaped unimodal distribution, a W-shaped distribution is a combination of the two, with analogy exerting some, but not all-powerful, influence.

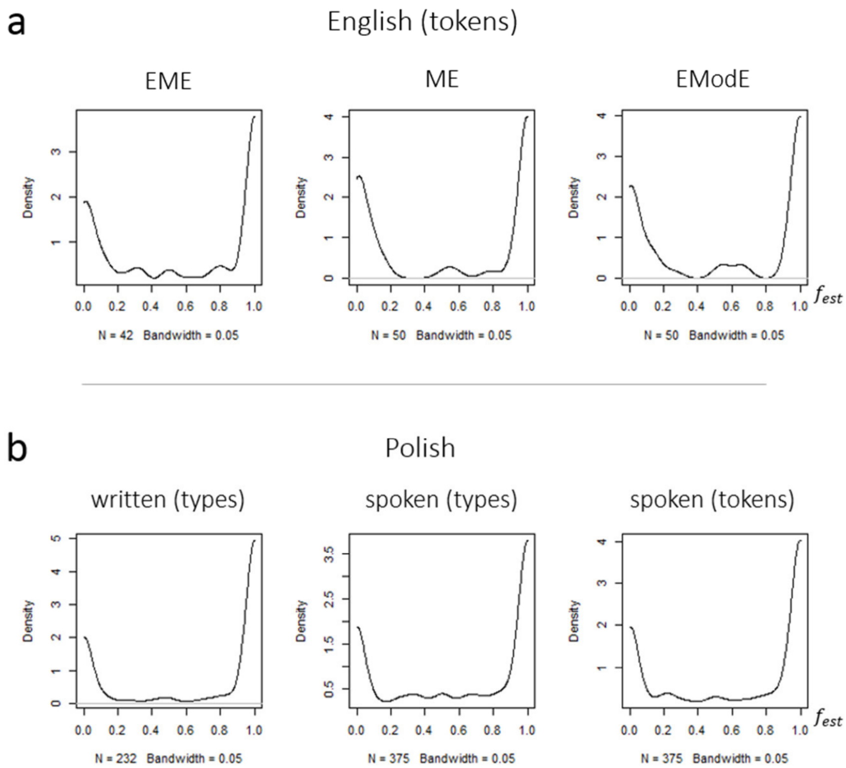


Figure 3. Estimated density distributions for the English and Polish consonant cluster inventories (Gaussian kernel; fixed bandwidth $h = .05$). (a) Diachronic development of the distribution of fractions of lexical cluster tokens $f_{est} = L_{est}/(M_{est} + L_{est})$ from Early Middle English (EME) to Early Modern English (EModE), where L_{est} and M_{est} are the estimated token frequencies of lexical and morphonotactic clusters, respectively. Diachronically, the distribution establishes a trimodal W-shape. (b) Distributions of f_{est}-values in terms of type frequency (written and spoken) and token frequency (spoken only). All distributions are V-shaped and bimodal. The difference between the former two distributions (comparison of 5 equally sized bins) is small at an effect size of $V = .18 \pm .08$ (95% CI); $\chi^2 = 19.35$.

Thus, the distributional shape for Polish is consistent with the predictions of morphonotactics, and the distributional shape for English indicates a joined influence of both morphonotactics and analogy. The modeling presented in the following section tackles the issue of how such distributional shapes come about.

3. Modelling approach

3.1. Modelling the interaction of consonant clusters

As outlined above, analogy effects will, in the long run, lead to configurations with large numbers of ambiguous consonant cluster types, with respect morphonotacticity, while morphonotactics predicts distributions characterized by large numbers of purely morphonotactic cluster types on the one hand and purely lexical cluster types on the other hand. However, in general it is difficult to assess what the development looks like when we face an interaction between these two pressures. The long-term dynamics that result from such a complex interaction are by no means trivial.

Thus, mathematical modeling suggests itself. It allows for a formalization of the interacting forces, and, at least, qualitative predictions. There might be other factors, like the precise social structure of the speaker community, that influence the dynamics quantitatively; however, they would probably not change the overall behavior of the dynamics. The predictions of a mathematical model then are – not only but also – as qualitative in nature as the diachronic predictions that are derived from hypotheses about analogy and/or morphonotactics. The model does nevertheless provide a more elaborate and complete picture.

In the present study we opt for a Lotka-Volterra type structured-population dynamical system. This model family is chosen not because of its quantitative accuracy, but due to its flexibility and simplicity. It is well studied and allows for a straightforward analysis of different types of interactions, such as competition, parasite-host configurations, or – as in this paper – mutualism, i.e. bidirectionally supporting effects (Hofbauer and Sigmund 1998: 28–29).

The term *structured population* refers to the fact that the population of instances of a given consonant cluster type, e.g. /nd/, is composed of two different subpopulations, namely morphonotactic instances and lexical instances of that cluster type. What the mathematical model does is that it simply tracks the frequencies of morphonotactic and lexical instances in time. We do not model the dynamics of uttered tokens on the one hand and speakers knowing a certain cluster on the other hand explicitly. Rather, we assume for simplicity that the abundances of uttered instances (that can be retrieved from a corpus) reflect the cognitive entrenchment of linguistic constituents in the speaker population (see e.g. Pierrehumbert 2001 and Wedel 2006 for exemplar based accounts; but also Yang 2000 for an approach in a generative framework and Nowak 2000 and Solé 2011 for related models of word dynamics). The two subpopulations interact according to the forces introduced in 2.1 and 2.2, and the parameters that specify the growth of the subpopulations reflect these interactions. In the following, these parameters will be described and motivated. For more formal details and a rigorous analysis of the model see Authors (####) as well as the Appendix.

First, the model incorporates the most fundamental parameters responsible for the spread of linguistic items: on the one hand reproduction, i.e. the combination of perception, learning and production, and on the other hand disappearance due to restricted memory or speaker death. We furthermore assume that the growth of linguistic items is bounded from above, since the number of linguistic items within memorized utterances, the number of utterances per speaker, and the number of speakers are obviously limited. These factors are cognitive, physiological, biological or simply physical in nature. Social biases are not explicitly incorporated, although this could easily be done (e.g. by increasing the reproduction rate of a certain class of items). This is, however, not the focus of this paper. In the mathematical model, reproduction is represented by a strictly positive growth rate r, meaning that per time unit r items are produced per already existing item. That is, if for example a consonant cluster occurs in an utterance, it will on average lead to the utterance of r new instances of the same type. In the absence of other populations of linguistic items, growth is bounded by a so-called carrying capacity K, which represents the maximal number of slots in memorized utterances, in which linguistic items of interest could be found. It deserves a more elaborate discussion. As suggested above, the amount of items should be limited for multiple reasons. We assume that the reproduction rate decreases linearly as the population size approaches K. This means that the growth of the population depends (a) on the number of its reproducing members and (b) on the number of yet available slots in which its members can be placed. This type of regulating population growth is referred to as logistic growth and regarded as the null model of density regulation (Otto and Day 2007: 75). Logistic behavior corresponds to S-shaped diachronic trajectories, which have been frequently observed in diachronic linguistics (Kroch 1989; Denison 2003; Blythe and Croft 2012) and applied to models of language or word spread (Solé 2011: 169–170; Nowak 2000). Finally, disappearance is formalized by a strictly positive parameter d so that per time unit, d items are removed per existing item (cf.

Pierrehumbert 2001). This is plausible, since memorized utterances can be forgotten or simply disappear due to speaker death.

In our model, the parameters characterizing the growth of morphonotactic consonant clusters can differ from that of their lexical counterparts. In particular, we assume that the morphonotactic growth rate (r_M) is larger than the lexical growth rate (r_L), and that the lexical growth rate depends on the number of lexical items (i.e. morphemes and words) which a consonant cluster can occur in. This amount shall be expressed by a parameter λ, which is – for convenience – normalized to the unit interval. Note in particular that λ is a normalized measure of type frequency rather than token frequency. We furthermore assume that – in line with morphonotactics – the reproduction rate of the morphonotactic subpopulation is supposed to be decreasing in λ, which represents the decreasing semiotic utility for signaling morphological complexity of the cluster. This will play a key role in the evolutionary dynamics of the system.

Finally, we build mutually supporting analogy into the model. We assume that analogical transfer between morphonotactic and lexical consonant clusters is proportional to the product of the two population sizes. In the theory of dynamical systems this is referred to as the law of mass action (Heesterbeek 2005). If one population has a positive impact on another population, it is more likely that the latter benefits from the former if both populations are large than if one of them is rare. This is plausible also from a linguistic point of view. Analogical transfer between two populations of sufficiently similar linguistic items takes place if they co-occur in a language at a given point in time. We formalize this as a strictly positive interaction parameter a, meaning that per co-occurrence event of morphonotactic and lexical consonant clusters a new clusters of each type are produced. Note that this allows the populations to exceed K. This is reasonable, since via analogy the chance of memorizing an item due to the presence of another item can be larger than in cases where the latter is absent.

3.2. Dynamics of consonant clusters

Under the assumption of the analogical-interaction parameter a and the disappearance parameter d being sufficiently small one can show that the model outlined in the previous subsection converges to a stable mix of morphonotactic and lexical consonant clusters (Hofbauer and Sigmund 1998: 28–29; see also Figure A1). The question now is how this mix changes if the number of lexical items containing a cluster represented by the parameter λ varies and, even more importantly, whether λ gets larger or smaller, i.e. whether a cluster type evolves in such a way that it occurs in more and more lexical items (e.g. by interconsonantal schwa deletion as in ME *godes* to ModE *gods*), or whether its occurrence in lexical items gets reduced (e.g. by cluster reduction as in *womb* from ME to ModE).

The model allows us to answer this question. Given a cluster-type specific λ used in the language and a new and different value λ' that comes into usage, one can determine if the new variant – which stands for a configuration in which the cluster occurs in more or less lexical items – replaces the old variant, or if no change occurs. The latter is the case, if an innovation like a hypothetical cluster reduction of *st* to, say, *s* in *vast* or *gist* is not supported. By this logic, we can determine the evolutionary trajectory of λ by means of a sequence of innovation-invasion-substitution events (cf Appendix A.1).[4]

As described in the previous section, the dynamics of the system depend on the parameter λ in that the reproduction rate of the morphonotactic subpopulation is assumed to decrease in this parameter due to decreasing semiotic utility. This is a direct consequence of how semiotic utility is operationalized (as described in 2.1). Since languages differ in the extent to which they make use of morphology, we suppose that the precise shape of the Semiotic Utility Function is language specific. As outlined in 2.1, our conjecture is that languages with elaborate morphological systems are more sensitive in the sense of morphonotactics, i.e. the semiotic utility of a morphonotactic cluster will already be to a large extent decreased if it also appears non-morphonotactically in a small number of lexical items. Thus the Semiotic Utility Function is assumed to be convex (dashed curves in Figure 1). On the other hand, languages making less use of inflectional or derivational morphology are supposed to be more tolerant to ambiguous clusters, so that their dynamics are better described by employing concave Semiotic Utility Functions (solid light-gray curves in Figure 1; see also Appendix A.2).[5]

---

[4] Note that this differs from the sequence of innovation/actuation and propagation proposed by Croft (2000: 37–38, 98) in that in the present approach, innovation is not necessarily functionally driven while propagation is not exclusively socially but rather also functionally driven (cf. Ritt 2004: 74–78).

[5] It should not be unmentioned that, as a Reviewer of this paper points out correctly, one could think of more complex shapes than just uniformly convex, linear, or concave functions, such as sigmoid or piecewise sigmoid shapes. While this is undoubtedly true (and psycholinguistically indeed likely), implementing more complicated shapes would at the same time increase the complexity of the model significantly: the functions which we use for modelling decreasing semiotic utility

Thus, on the evolutionary time-scale our model predicts the diachronic development of the number of lexical items a given cluster type occurs in specified by λ which in turn determines the numbers of morphonotactic and lexical cluster instances, which shall be labelled M_λ and L_λ, respectively. From these values the fraction of lexical instances can be calculated according to $f_\lambda = L_\lambda/(M_\lambda + L_\lambda)$ for each point in time, where $f_\lambda = 0, 0.5,$ and 1 correspond to purely morphonotactic, maximally ambiguous, and purely lexical configurations, respectively (see Authors ####). In this way, the model shows the trajectory of the relative frequency of lexical instances of a given cluster type. In the subsequent section it will be shown by means of simulations that the long term dynamics of this relative frequency depends on the language specific shape of the Semiotic Utility Function. We will see that, depending on this shape, clusters can either evolve towards the boundaries (i.e. so that they become either purely morphonotactic or purely lexical) or towards a fair mix of morphonotactic and lexical cluster instances. Note that the system solely models the evolution of single cluster types on the scale from purely morphonotactic to purely lexical. It does not account for the complete loss of consonant clusters once they have reached either of the two boundaries.

4. Applications of the model

4.1. Diachrony of English

The model described above needs to be tested against actual language data. We will start with an analysis of the diachronic development of the English inventory of word final consonant clusters from the beginning of the ME period (12th cent.) up to the EModE period (17th cent.). The respective distributions have been described in 2.4 and Figure 3a. We expect the predictions of the proposed model to be consistent with the distributions found in the 12th-century and 17th-century data.

We proceed as follows. First, the fractions of lexical instances $f_{est} = L_{est}/(M_{est} + L_{est})$ are estimated for each consonant cluster in the 12th-century

are power functions of the form $\lambda \to 1 - \lambda^c$ (see Appendix A.1), while sigmoid functions usually involve exponential terms. The function family chosen in this approach provides a reasonably cheap way of accounting for a large number of functional relationships. Thus, the present approach strikes a balance between formal complexity and explanatory power.

cluster inventory in order to determine the distribution of f_{est} values. From this empirically estimated initial distribution a random sample of $N = 80$ values[6] is drawn (empirical initial distribution in Figure 4a). This set serves as a starting point for the simulation (simulated initial distribution in Figure 4a). Since English is regarded as moderately inflectional and derivational as compared to other languages such as German or Slavic languages (Szmrecsanyi 2012), we assume the Semiotic Utility Function to be non-convex.[7] For each initial value in the simulated initial distribution, a sequence of repeated innovation-invasion-substitution events is simulated according to the structured-population dynamical system outlined in the previous section. In order to make the model more realistic, we allow for uniformly random fluctuations. The respective trajectories are shown in the middle panel in Figure 4a. In general, we see that trajectories that start close to the boundaries evolve towards even more extreme f_λ values, while trajectories initiating in more or less ambiguous configurations evolve towards a fair mix of morphonotactic and lexical instances. After several simulation steps, the resulting distribution of f_λ values is determined and compared to the empirical distribution of f_{est} from the 17th century data. It can be seen that the simulated final distribution and the empirical final distribution both display a considerable number of ambiguous clusters (histograms in the right panel of Figure 4a).

That is, under the assumption that an English speaker's ability to process morphonotactic consonant clusters is not to a large extent reduced by the occurrence of the same cluster type within a small number of lexical items, the model predicts that the English consonant-cluster inventory evolves in such a way that there are (a) a large number of purely morphonotactic consonant clusters, (b) a large number of purely lexical consonant clusters, and (c) a substantial number of ambiguous consonant clusters that show a fair share of morphonotactic and lexical instances (but see Appendix A.2 for more formal information about the diachronic reflex of the Semiotic Utility Function's curvature on the set of possible evolutionary outcomes). This is consistent with the trimodal distribution we find in the English data.

---

[6] The sample size is motivated by a pre-simulation power analysis which revealed that – if actually present – significant medium differences in terms of Cramer's V between the simulated and the empirical final 5-bin distributions, respectively, could be detected at a power of .80.

[7] In particular, note that according to quantitative analyses conducted by Szmrecsanyi (2012: Fig. 3) no monotonous trend of the syntheticity and analyticity of English is identifiable in the ME period. This supports our notion that the curvature of the English semiotic utility function stayed constant from the 12th to the 17th century.

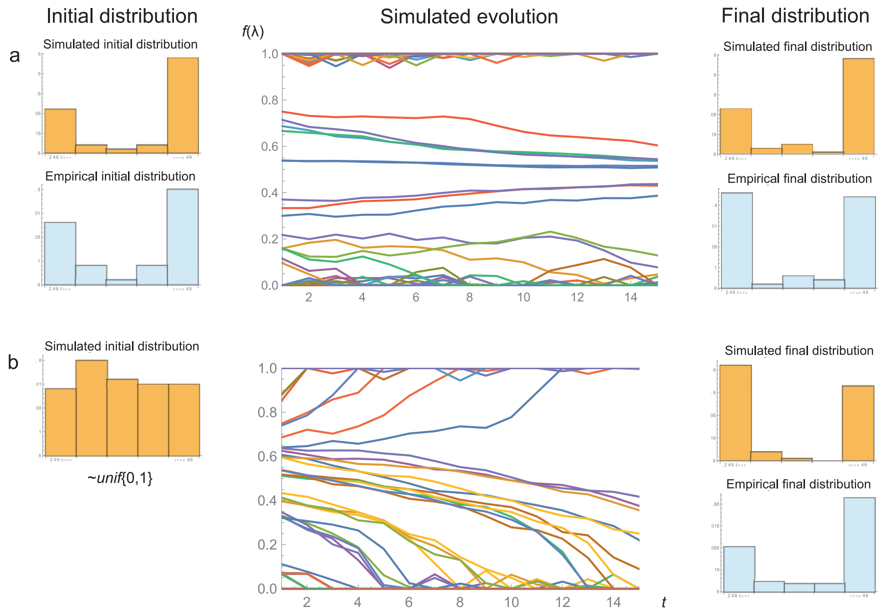


Figure 4. Simulating the evolution of the (a) English and (b) the consonant-cluster inventory, respectively. Orange histograms represent simulated distributions while blue histograms represent empirically measured distributions. The repeated invasion-analysis simulations ($N = 80$) are based on (a) a model with a non-convex Semiotic Utility Function in the case of English and (b) a model with a convex Semiotic Utility Function in the case of Polish ($a = 0.01$, $d = 2$, $K = 50$, $r_M = r_L = 10$, random drift $\sim unif\{\pm 0.01\}$).

4.2. Synchronic analysis of Polish data

In contrast to English, the Polish consonant-cluster inventory exhibits a bimodal distribution. Cluster types tend to be either unambiguously morphonotactic or unambiguously lexical, while ambiguous consonant clusters are relatively rare (see 2.4 and Figure 3b). While we have collected solely synchronic Polish data, which renders the direct comparison of simulated and empirically observed diachronic trajectories impossible, we can still test the proposed model against the Polish data.

Instead of using a historical Polish cluster-inventory distribution as initial state for the simulation, we are even stricter and test whether the Present Day

Polish distribution would be approximated when starting with a completely random distribution. Thus, the initial values f_λ for the simulation are randomly drawn from a uniform distribution on the unit interval $unif(0,1)$. Note, that this does not and should not in any way reflect the distribution of the Polish consonant-cluster inventory at any point in time. Rather, in the lack of historical data we stay completely agnostic about the history of Polish and test our model against the least informed initial state.

As in 4.1, the precise shape of the Semiotic Utility Function must be specified. In contrast to what was assumed for the English case, a convex semiotic signaling function was chosen for Polish. This means that already small numbers of lexical items containing a consonant cluster substantially decrease the semiotic signaling function of its morphonotactic version. This can be motivated by the inflectional and derivational richness of Polish.

The evolution from the assumed initial state towards the Polish-like simulated final distribution is shown in Figure 4b. After $t = 15$ simulation steps, the resulting distribution of f_λ values is bimodal, i.e. the cluster-inventory evolves in such a way that clusters are predominantly either purely morphonotactic or purely lexical, while ambiguous consonant clusters have almost disappeared (cf. Appendix A.2). This is as expected and coincides with the configuration found in the synchronic Polish data.

5. Conclusion

The complexity-signaling ability of morphonotactic consonant clusters and analogy effects between morphonotactic and lexical consonant-cluster instances constitute two cognitive forces which are crucial to the replication of consonant clusters. They are counteracting in the sense that the complexity-signaling ability – in other words: the semiotic utility – of morphonotactic clusters is handicapped by the appearance of lexical clusters, while analogy constitutes a mutual relationship between the two cluster categories. The interaction of these two forces is complex and its diachronic reflexes are nontrivial. However, it can be addressed by means of a mathematical model which tracks the respective abundances of the morphonotactic and lexical versions of a given cluster type, and allows for an analysis of the diachronic development of its degree of morphonotacticity. We have shown in this paper that the way in which the semiotic utility of consonant clusters is operationalized is crucial for the diachronic long-term development.

It has been shown that the development of the inventory of English word-final consonant clusters from the 12th to the 17th century is best captured by the model, if a non-convex Semiotic Utility Function is assumed (cf. Appendix A.2). This means that the semiotic utility of a consonant cluster type is not substantially diminished by the presence of lexical instantiations of that cluster. Crucially, the implementation of analogy is necessary for obtaining a trimodal W-shaped distribution. In the absence of analogical effects between morphonotactic and lexical consonant clusters, the stable amount of ambiguous English clusters could not be accounted for. We take this to support the hypotheses about the relevance of analogy in the reproduction of morphonotactic consonant clusters, which have been outlined in 2.2 (see Appendix A.3 for a slightly more formal argument).

In contrast, Polish synchronic data exhibit a bimodal V-shaped distribution in which ambiguous consonant clusters are considerably less prominent. In the model, this can be explained by a very convex Semiotic Utility Function, which in turn means that the reproduction of morphonotactic consonant clusters is impeded significantly by the appearance of lexical items that contain a lexical consonant cluster. It is worth pointing out that the model shows that even if there are analogical effects, this can be compensated by a very convex Semiotic Utility Function, so that finally the amount of ambiguous consonant clusters gets reduced.

While the theory of morphonotactics suggests that the Semiotic Utility Function is decreasing, this paper stresses – in addition – the relevance of its curvature. The shape of the Semiotic Utility Function thus is modeled as to be language specific. As a motivation, we hypothesize that it is related to the extent to which a language makes use of inflectional or derivational morphology, i.e. its syntheticity. Speakers who are exposed to elaborated morphological systems are supposed to be challenged by slightly ambiguous clusters to a larger extent than speakers of less synthetic languages. We take this as a starting point for further (i) experimental research which tries to evaluate the language specific shapes of the Semiotic Utility Function and (ii) quantitative and comparative corpus research on the respective (mor)phonotactic consonant-cluster inventories.

Appendix

A.1 Mathematical outline of the model and its evolutionary analysis

In the following, the model which underlies the simulation described in Section 3 and 4 shall be outlined in a bit more mathematical detail. The dynamics of the densities of morphonotactic consonant clusters M and lexical or phonotactic consonant clusters L are modelled by means of a two-dimensional dynamical system in continuous time, which belongs to the Lotka-Volterra model family. The changes in the respective densities, i.e. the first derivatives of M and L with respect to time, are given by the system

$$\begin{pmatrix} \dot{M} \\ \dot{L} \end{pmatrix} = \mathrm{A}(\lambda) \bullet \begin{pmatrix} M \\ L \end{pmatrix},$$

where

$$\mathrm{A}(\lambda) \coloneqq \begin{pmatrix} r_M \, u(\lambda) \left(1 - \frac{M}{K}\right) - d & a \\ a & r_L \lambda \left(1 - \frac{L}{K}\right) - d \end{pmatrix}$$

is a 2 by 2 matrix depending on $\lambda \in [0,1]$. The parameters r_M, r_L (i.e. the reproduction rates r specific to M and L, respectively), K, d and a are assumed to be real and strictly positive. They are motivated in Section 3. Semiotic utility is formalized by a strictly decreasing function $u \colon \lambda \to 1 - \lambda^c$ on the unit interval, where the exponent $c > 0$ determines the curvature of u. For $c > 1$ the function is concave and for $c < 1$ it is convex (Figure 1). If d and a are sufficiently small, the system exhibits a single stable population-dynamical attractor $\left(\widehat{M}_\lambda, \widehat{L}_\lambda\right)$ (Hofbauer and Sigmund 1998: 29). For most functions u an explicit analytical representation of this equilibrium would be too large and complicated, so that it cannot be shown in this paper. The ecological dynamics shall, however, be illustrated in Figure A1.

The successful invasion of a new configuration of consonant clusters characterized by λ' can be shown to depend on

$$s(\lambda, \lambda') \coloneqq \mu(\mathrm{A}(\lambda')|_{(M_{\lambda'}, L_{\lambda'}) = (\widehat{M}_\lambda, \widehat{L}_\lambda)}),$$

where μ is the stability modulus (i.e. the largest Eigenvalue of the Jacobian matrix of the system) and where (M_λ, L_λ) shall denote the pair of the densities of morphonotactic and lexical consonant clusters depending on the number of lexi-

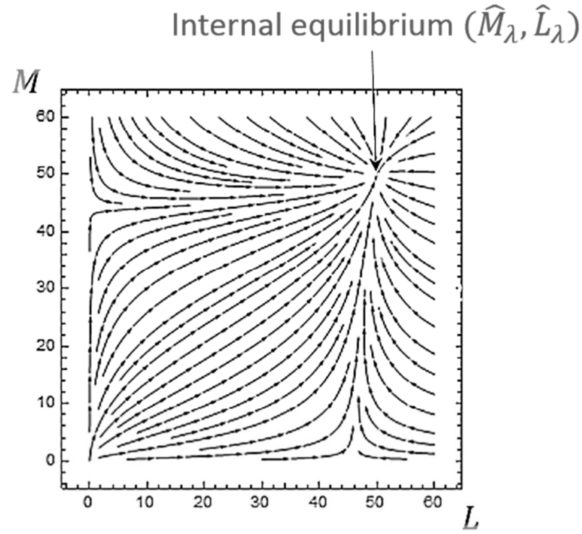


Figure A1. Phase portrait of the ecological dynamics of the structured population. The horizontal axis measures the number of lexical instances while the vertical axis measures the number of morphonotactic items of a specific cluster type. If death rates and analogy rates are sufficiently small, the population always converges to a stable ecological equilibrium consisting of a mixed population of morphonotactic and lexical clusters. In this example there are as many lexical items as morphonotactic ones.

cal items represented by λ. See Geritz et al. (1998), Dercole and Rinaldi (2008: ch. 2.8), and Hoyle and Bowers (2008) for more details. If the difference between λ and λ' is sufficiently small,

$$s(\lambda, \lambda') > 0$$

implies that the new configuration replaces the old one (Geritz et al. 2002) and converges to a new ecological equilibrium $(M_{\lambda'}, L_{\lambda'})$ (as in Figure A1). Otherwise λ stays the same. Thus, by checking the latter inequality condition for each pair of successive values λ and λ', one can keep track of the change in the number of lexical items a cluster occurs in. This is what is done at each step of the simulation. Furthermore, we allow for random fluctuations drawn from a uniform distribution at each simulation step. The sequence of all simulation steps finally constitutes the long-term evolutionary trajectory for each initial value of λ.

A.2 Analytical results: possible evolutionary outcomes

Certain qualitative properties of the evolutionary dynamics of λ can, in fact, be predicted analytically. In particular, it can be shown that the set of possible evolutionary outcomes differ depending on the curvature of the Semiotic Utility Function. For concave, linear, and mildly convex Semiotic Utility Functions, λ will either approach an intermediate value (a so-called mixed evolutionarily stable strategy, ESS) or one of the two boundaries of the unit interval. In contrast, if the Semiotic Utility Function is sufficiently convex, then the mixed ESS disappears so that λ always approaches either 0 or 1, depending on its initial position. A rigorous derivation of these results is beyond the scope of this paper, but see Hoyle & Bowers (2008) and Rüffler et al. (2004) for some discussion on the relationship between trade-off curvatures and possible evolutionary outcomes.

For the applications discussed in Section 4, this has the following consequence. If the Semiotic Utility Function is not strongly convex (as in the English case), then clusters with intermediate λ values will always move towards the middle of the unit interval so that an accumulation of ambiguous clusters get visible after a sufficient number of simulation steps. In contrast, if the Semiotic Utility Function is strongly convex, the final distribution will always end up being bimodal with a large numbers of non-ambiguous clusters, as long as not all clusters enter the simulation in the same basin of attraction of either 0 or 1.

A.3 Evolutionary optimization in non-interacting populations

The above described model is based on the assumption that morphonotactic and lexical cluster items interact via analogy. In the model, this is represented by the condition that a is strictly positive. Indeed, if it were zero, i.e. if there was no interaction between the two populations of clusters, then the evolutionary dynamics of λ would be much less complicated. If $a = 1$, then the matrix A becomes reducible so that the corresponding stability modulus only depends on one of the two diagonal entries (Horn & Johnson 1985). Evolutionarily, this inevitably leads to an optimization of λ towards either 0 or 1 (Metz and Diekmann 2008). In this scenario, mixed strategies, i.e. ambiguous clusters, will always disappear in the long run.

Acknowledgements

The research reported on in this paper was supported by Österreichischer Austauschdienst grant No. PL8/2014 and by Austrian Science Fund (FWF) grant No. P27592-G18.

References

Baumann, A., C. Prömer and N. Ritt. 2015. "Diachronic reflexes of frequency effects among word final (mor)phonotactic consonant clusters in Middle and Early Modern English". Paper presented at the 3rd International Workshop on Phonotactics and Phonotactic Modeling.

Blythe, R.A. and W. Croft. 2012. "S-curves and the mechanism of propagation in language change". *Language* 88(2). 269–304.

Croft, W. 2000. *Explaining language change. An evolutionary approach*. Harlow, New York: Longman.

Denison, D. 2003. "Log(ist)ic and simplistic S-curves". In: Hickey, R. (ed.), *Motives for language change*. Cambridge: Cambridge University Press. 54–70.

Donohue, M., R. Hetherington, J. McElvenny and V. Dawson. 2013. "World phonotactics database". Department of Linguistics, The Australian National University. <http://phonotactics.anu.edu.au>. Last accessed: 16 Sep 2015.

Dressler, W.U. 2007. "Early development of nominal and verbal morphology from a typological perspective". In: Laaha, S. and S. Gillis (eds.), *Typological perspectives on the acquisition of noun and verb morphology*. (*Antwerp Papers in Linguistics* 112.). 3–9.

Dressler, W.U., K. Dziubalska-Kołaczyk and L. Pestal. 2010. "Change and variation in morphonotactics". *Folia Linguistica Historica* 31. 51–67.

Dressler, W.U. and K. Dziubalska-Kołaczyk. 2006. "Proposing morphonotactics". *Wiener Linguistische Gazette* 73. 1–19.

Dziubalska-Kołaczyk, K. and P. Zydorowicz. 2014. "The production of high-frequency clusters by native and non-native users of Polish". *Concordia Working Papers in Applied Linguistics* 5. 130–144.

Geritz, S.A.H., E. Kisdi, G. Meszéna and J.A.J. Metz. 1998. "Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree". *Evolutionary Ecology* 12. 35–57.

Geritz, S.A.H., M. Gyllenberg, F.J. Jacobs and K. Parvinen. 2002. "Invasion dynamics and attractor inheritance". *Journal of Mathematical Biology* 44. 548–560.

Heesterbeek, J. 2005. "The law of mass-action in epidemiology: A historical perspective". In: Cuddington, K. and B. Beisner (eds.), *Ecological paradigms lost: Routes of theory change*. Amsterdam: Academic Press. 81–104.

Hockema, S.A. 2006. "Finding words in speech: An investigation of American English". *Language Learning and Development* 2(2). 119–146.

Hofbauer, J. and K. Sigmund. 1998. *Evolutionary games and population dynamics*. Cambridge: Cambridge University Press.

Hogg, R. and C.B. McCully. 1987. *Metrical Phonology: A course book*. Cambridge: Cambridge University Press.

Horn, R. and C. Johnson. 1985. *Matrix analysis*. Cambridge: Cambridge University Press.

Hoyle, A. and R.G. Bowers. 2008. "Can possible evolutionary outcomes be determined directly from the population dynamics?" *Theoretical Population Biology* 74. 311–323.

Jarosz, G., S. Calamaro and J. Zentz. Submitted. "Input frequency and the acquisition of syllable structure in Polish". *Language Acquisition.* <http://people.umass.edu/jarosz/jarosz_etal_submitted.pdf>

Kaźmierski, K. 2015. "Consonant clusters in casual speech in Polish: Testing the predictions of NAD and Beats-and-Binding phonology". Paper presented at the 23rd Manchester Phonology Meeting.

Kroch, A. 1989. "Reflexes of grammar in patterns of language change". *Language Variation and Change* 1. 199–244.

Kroch, A. and A. Taylor. 2000. "Penn-Helsinki Parsed Corpus of Middle English". <http://www.ling.upenn.edu/hist-corpora/>.

Kroch, A., B. Santorini and L. Delfs.2004. "Penn-Helsinki Parsed Corpus of Early Modern English". <http://www.ling.upenn.edu/hist-corpora/>.

Laaha, S. and S. Gillis (eds.). 2007. *Typological perspectives on the acquisition of noun and verb morphology*. (*Antwerp Papers in Linguistics* 112.)

Madelska, L 2005. *Słownik wariantywności fonetycznej współczesnej polszczyzny* [A dictionary of phonetic variability in contemporary Polish]. Kraków: Collegium Columbinum.

Martin, A. 2007. The evolving lexicon. (PhD dissertation, UCLA, Los Angeles.)

Metz, J.A.J., S.D. Mylius and O. Diekmann. 2008. "When does evolution optimize?" *Evolutionary Ecology Research* 10. 629–654.

Nowak, M.A. 2000. "The basic reproductive ratio of a word, the maximum size of a lexicon". *Journal of Theoretical Biology* 204 (2). 179–189.

Otto, S. and T. Day. 2007. *A biologist's guide to mathematical modeling in ecology and evolution*. Princeton: Princeton University Press.

Peters, A.M. 1997. "Language typology, prosody and the acquisition of grammatical morphemes". In: Slobin, D.I. (ed.), *The crosslinguistic study of language acquisition* (vol. 5). Hillsdale, NJ: Lawrence Erlbaum Associates. 136–197.

Pierrehumbert, J. 2001: "Exemplar dynamics: Word frequency, lenition, and contrast". In: Bybee, J. and P.J. Hopper (eds.), *Frequency effects and the emergence of lexical structure*. Amsterdam: John Benjamins. 137–157.

Rueffler, C., T. Van Dooren and J.A.J. Metz. 2004. "Adaptive walks on changing landscapes: Levins' approach extended". *Theoretical Population Biology* 65. 165–178.

Sadeniemi, M., K. Kettunen, T. Lindh-Knuutila and T. Honkela. 2008. "Complexity of European Union languages: A comparative approach". *Journal of Quantitative Linguistics* 15(2). 185–211.

Shockey, L. 2003. *Sound patterns of spoken English*. Malden, MA: Blackwell.

Solé, R. 2011. *Phase transitions*. Princeton: Princeton University Press.

Szmrecsanyi, B. 2012. "Analyticity and syntheticity in the history of English". In: Nevalainen, T. and E. Closs Traugott (eds.), *The Oxford handbook of the history of English*. Oxford: Oxford University Press. 654–665.

Wedel, A. 2006. "Exemplar models, evolution and language change". *The Linguistic Review* 23. 247–274.

Yang, C. 2000. "Internal and external forces in language change". *Language Variation and Change* 12. 231–250.

Zydorowicz, P., P. Orzechowska, M. Jankowski, K. Dziubalska-Kołaczyk, P. Wierzchoń and D. Pietrala. Forthcoming. *Phonotactics and morphonotactics of Polish and English. Theory, decription, tools and applications.*

Address for correspondence:
Andreas Baumann
Institut für Anglistik und Amerikanistik
Universität Wien
Campus d. Universität Wien
Spitalgasse 2-4/Hof 8.3
1090 Wien
Austria
andreas.baumann@univie.ac.at

Papers in Historical Phonology
http://journals.ed.ac.uk/pihph
ISSN 2399-6714
Volume 1, 50–75
DOI: 10.2218/pihph.1.2016.1693

Diachronic dynamics of Middle English phonotactics provide evidence for analogy effects among lexical and morphonotactic consonant clusters

ANDREAS BAUMANN
University of Vienna

NIKOLAUS RITT
University of Vienna

CHRISTINA PRÖMER
University of Vienna

Abstract

Consonant clusters that rarely occur lexically (i.e. within morphemes) may function as complexity markers when they span a morpheme boundary, i.e. when they occur morphonotactically. In this study we observe patterns in the diachronic dynamics of Middle English which hint at mutually beneficial effects between morphonotactic and lexical clusters. We suggest that the patterns revealed can be explained by frequency-based analogy effects in language acquisition.

1 Introduction

On the basis of diachronic corpus evidence from Middle and Early Modern English, this paper studies interactions between word-final consonant clusters that occur within morphemes, like /nd/ in *hand*, /lb/ in *bulb*, or /st/ in *fast+est,* and those that span morpheme boundaries, like /nd/ in *quicken+ed* or /mz/ in *seem+s*. The former are by definition phonotactically licensed, and, following in this respect Dressler and Dziubalska-Kołaczyk (2006), we refer to them as 'phonotactic' or 'lexical'. The latter are referred to as 'morphonotactic clusters'. They may be phonotactically licensed as well, but often they are not. For instance, the /nd/ in *quicken+ed* is, while the /mz/ in *seem+s* is not, because the latter does not occur morpheme-internally.

In English, as in many languages, the sets of lexical and morphonotactic clusters are not identical. This is to be expected, since phonotactic constraints are known to be tightest at the stem level, i.e. morpheme-internally (Kiparsky 1982; Giegerich 1999; McMahon 2002), and since consonant clusters in general count as phonologically marked, or dispreferred, they are rare within morphemes (Shockey; Berent et al. 2007; Dziubalska-Kołaczyk & Zydorowicz 2014). Even when they are not permitted within morphemes, however, they may be produced through morphological or syntactic concatenation, and by virtue of being ruled out morpheme-internally, such clusters then have the potential of signaling syntactic (McQueen 1998) or morphological boundaries (Post et al. 2008; Dressler et al. 2010). Thereby, they serve an important function in the decomposition of speech into meaningful units, and it may be for this reason that they have become stably established at the word or phrase level.

Although the sets of lexical and morphonotactic clusters are not identical they often overlap. This is the case, for example, in Polish, French, German, and also in English. Thus, final /nd/, which represents a morphonotactic cluster in *quicken+ed*, occurs also morpheme-internally in words like *hand* or *wind*. It is these clusters that our study focuses on.

Specifically, we ask whether—and under what conditions—morphonotactic clusters inhibit or promote the emergence of homophonous lexical counterparts. The question is motivated by the following considerations. On the one hand, and as argued by Dressler & Dziubalska-Kołaczyk (2006), the signaling function of morphonotactic clusters is clearly diminished when there are lexical homophones, so that clusters that span, and thereby indicate, morpheme boundaries should inhibit the emergence of lexical clusters. On the other hand, however, children may acquire highly frequent morphonotactic clusters before they recognize the morphological boundaries they signal (Jusczyk et al. 2002), which may loosen the constraints that prohibit such clusters within morphemes and thereby promote, rather than inhibit, the establishment of lexical homophones.

We address this question through a quantitative corpus study, in which we chart the development of word-final lexical and morphonotactic clusters in Middle English (ME) and Early Modern English (EModE). We show that the pattern in the diachronic dynamics in ME provides evidence of analogy effects by which morphonotactic clusters promote rather than inhibit the establishment of lexical homophones.

The paper is structured as follows. Section 2 reviews the relevant aspects of morphonotactic theory (2.1), focusing on the diverse relationships that may be established between morphonotactic and lexical clusters, and elaborating the research question (2.2). Section 3 introduces the data (3.1), presents an outline of the quantitative approach (3.2), and introduces the analysis and our findings (3.3). Finally, the results are discussed and summarized in the concluding section (4).

2 Consonant clusters, morphonotactics, and analogy

2.1 Phonotactic and morphonotactic consonant clusters

2.1.1 Inhibitory effects among consonant clusters

As outlined above, morphonotactic clusters, which span morpheme boundaries, can signal these boundaries by virtue of their markedness. Clearly, this works best when clusters do not at the same time also occur within morphemes. A good example is ModE /md/. It occurs only when the suffix *-ed* is added to a stem ending in /m/, as in *seem-ed*, thereby facilitating the decomposition of past tense verbs or past participles. When morphonotactic clusters have lexical homophones, however, their facilitating effect is diminished (Dressler & Dziubalska-Kołaczyk 2006; Dressler et al. 2010; Calderone et al. 2014). This is the case in English /nd/, which occurs not only in past-tense verbs or past participles, but also in numerous lexical base forms such as *hand*, *band*, *demand*, as well as in highly frequent function words like *behind* or *and*. Thus, clusters like /nd/ are very weak indicators of the morphological structure of the words in which they occur. Hence, lexical clusters would be expected to inhibit the establishment of morphonotactic homophones and *vice versa*. Assuming that the inhibitory pressure a cluster exerts on its homophonous counterpart correlates with its frequency, lexical /nd/ should greatly inhibit morphonotactic /nd/.

The general prediction that this hypothesis proposes is that cluster types should diachronically tend to become either purely morphonotactic or purely phonotactic (Dressler et al. 2010). Such a scenario could come about via selective repair processes such as cluster reduction (cf. Labov 1989, who reports that final coronal deletion more frequently affects /nd/ clusters in simple items such as *find* than in complex forms such as *fine+d*), schwa epenthesis (rare in English, cf. the lexicalized adjective *learned*, /ləːnɪd/, but see Schlüter 2005), selective devoicing of /nd/ in past tense or participle forms (e.g. *learn+t*, *burn+t* < *learn+ed*, *burn+ed*), or theoretically also by the avoidance (and eventually the loss) of ambiguous word forms.

2.1.2 Supporting effects among consonant clusters

In addition, and to a certain degree in contrast, to the inhibiting effects outlined in the previous section, mutually supporting effects between morphonotactic and lexical clusters have also been suggested. On the one hand, Martin (2007: 99) investigated consonant clusters that occur at the boundary of English noun-noun compounds and concludes that "the categorical phonotactic restrictions that hold within morphemes also hold gradiently across morpheme boundaries". This provides evidence for a mutually supporting relationship between morphonotactic and lexical clusters, since lexical clusters license the presence of their boundary-spanning counterparts. Note that Martin's argument concerns both lexical clusters and morphonotactic clusters at constituent boundaries in compounds; the same relationship can be assumed to hold, even more so, between lexical clusters and morphonotactic clusters in the prosodically weaker word-final position.

On the other hand, Hogg and McCully (1987: 47) investigated VVCC rhymes and state that "the type of syllable structure found in a word such as *wind* (/waind/) has been protected through analogy with inflected forms such as *weaned*". Hence, they claim that morphologically produced word-final VVCC rhymes stabilize their lexical counterparts via analogy, thus providing support for the hypothesis that lexical clusters may also benefit from the presence of morphonotactic clusters. Hogg and McCully (1987) focus on coda clusters following a long vowel, but in the remainder of this paper, their claim will be extended to coda clusters in general.

2.2 Elaborating the analogy hypothesis

We want to test whether morphonotactically produced consonant clusters support their morpheme-internal counterparts. The hypothesized mechanism at work is frequency-based analogy, and the diagnostic method for detecting these analogy effects involves the analysis of the diachronic development of the consonant clusters in question.

2.2.1 Word-internal phonotactics and analogy

We suggest two reasons for analogical transfer from morphonotactics to morpheme-internal phonotactics. First, morphonotactic and lexical instances of a cluster type obviously share properties such as place or manner of articulation, voicing, or sonority of the respective consonants involved, although in certain articulatory or acoustic features, morphonotactic and lexical clusters of a certain type might exhibit slight differences. For instance, Plag et al. (2015) show that the acoustic duration of word-final /s/ in English is significantly longer if it is non-morphemic (i.e. if it does not represent an inflectional suffix or clitic). Thus, if /s/ is part of a word-final /Cs/ cluster, morphonotactic instances of that cluster are supposed to be shorter than their lexical counterparts. Nevertheless, structural similarity between instances of the two cluster categories should be substantial, so that on the discourse level the production and perception of tokens of one category is supposed to have a priming effect on the other category, along the lines of structural priming theory in syntax (Ferreira & Bock 2006; Pickering & Ferreira 2008). These effects on the level of discourse then facilitate the establishment of certain—in our case phonologically primed—patterns in grammatical knowledge (Gries 2005; Fehér et al. 2016). Clearly, when restricting oneself to phonological structure, this argument in principle goes in both directions: morphonotactic clusters have facilitating effects on lexical clusters, and *vice versa*. However, referring to a study by Shields and Balota (1991) about priming and duration, Jäger and Rosenbach (2008: 97) argue that phonetic priming is asymmetric in that "a phonetic full form has a stronger priming effect on the corresponding reduced form than the other way round". This suggests that via priming, lexical clusters support their morphonotactic counterparts to a larger extent than the converse, since the former are phonologically less reduced (Plag et al. 2015).

We hypothesize a stronger version of the opposite direction, i.e. that morphonotactic clusters support lexical ones, and that this can be accounted for in terms of language acquisition. It is known that during the first two years of first-language acquisition, learners acquire highly frequent inflected word forms as lexical chunks (cf. Brown 1973; Rumelhart & McClelland 1986). This entails that highly token-frequent, and specifically morphonotactic clusters are acquired before the morphological operations that actually produce them in adult speech (the very same mechanism has been suggested to drive the lexicalization of words (Brinton & Traugott 2005: 91–95)). Crucially, during this first stage these items that were originally produced as morphonotactic clusters are processed as lexical clusters by the learner, which would logically facilitate the acquisition of words containing actual lexical clusters of the same cluster type. These acquired words would, in all likelihood, not be 'unlearned' after the onset of the acquisition of morphology, resulting in lexical clusters surfacing more frequently in simplex words. This would entail that morphonotactic clusters promote the acquisition of their lexical counterparts, and this supporting effect is expected to be larger, the more token-frequent the morphonotactic clusters are. We summarize these thoughts in the following hypothesis:

(1) **Analogy among consonant clusters.** Morphonotactic clusters and lexical clusters of the same cluster type mutually support each other via analogy.

2.2.2 Diachronic reflexes of frequency-based analogy effects

If highly token-frequent morphonotactic clusters promote the acquisition of words containing lexical clusters, then diachronically the number of instances of the lexical cluster should obviously increase. Impressionistically, this is evident from the developments of the word-final clusters /nd/ and /md/. Through morphological operations, the first one comes about roughly four times as often as the second one in ME. Lexically, /nd/ surfaces in many lexical items such as ME/PDE *and*, *fiend*, *behind*, *wind*, or OE/ME *kalend* ('(first day of a) month') and *healend* ('savior'), whereas /md/ occurs sporadically in items such as ME *fremd* ('foreign'). The crucial point is that while /nd/ even appears in more recently imported loans such as *defend* or *command*, thereby increasing in frequency, /md/ gradually lost its lexical use.[2]

In the following, this phenomenon will be investigated more systematically. The hypothesis to be tested in this paper thus reads as follows:

(2) **Diachronic reflexes of analogy.** If the first stages of language acquisition feature analogical transfer from morphonotactic to lexical clusters, then, as a diachronic reflex, the lexical counterparts of highly token-frequent morphonotactic clusters should in the long run appear in many lexical items and hence become more lexical. Similarly, token-infrequent clusters are expected to become less lexical.

Conversely, albeit strictly speaking not logically accurate, we will see the appearance of such a diachronic reflex as indirect evidence for frequency-based analogy effects between morphonotactic and lexical consonant clusters, i.e. hypothesis 1. According to the above hypothesis,

[2] Clearly, importing loans or disfavoring particular words is not the only way in which a cluster can become more or less acceptable in lexical items. Similarly, fusion (e.g. *whence* < *whenne+s*) or phonological change (e.g. cluster reduction in *bomb*, *knight* or *damn*) provide other sources of variability.

analogy may very well interfere with the pressure of decreasing cluster ambiguity as outlined above. This is the case if a primarily morphonotactic cluster which is also very frequent in terms of morphonotactic tokens by analogy increases in lexical items, and thus becomes more ambiguous.

3 Detecting diachronic reflexes of analogy effects

In this section, we will explain how we tested the previously stated hypothesis that morphonotactically token-frequent clusters should become more lexical by investigating the diachronic development of the ME inventory of word-final consonant clusters. It is structured as follows: first, the ME data are introduced (3.1), then the hypothesis is operationalized in order to investigate it statistically (0), and finally, the data are analyzed and interpreted by means of two modeling approaches (1.1).

3.1 Data description

The dataset used for this study consists of word-final sequences extracted from the Penn Helsinki Corpora of Middle English (PPCME2, Kroch & Taylor 2000) and Early Modern English (PPCEME, Kroch et al. 2004). The compilation dates of the texts included range from 1138 to 1698. All words ending in graphemic C(V)C(V) sequences were extracted with the exception of words labeled as foreign (i.e. cases of code-switching); however, for the present study, only those words which end in a consonant cluster are of interest. Hence, excluded from the data set were sequences for which there is evidence that at least one of the two vowels did not get reduced (such as e.g. *plenty*) and words that are already monosyllabic (e.g. *for*). All other potential clusters were labeled as 'morphonotactic' (e.g. *bann+ed*), 'lexical' (e.g. *hand*) or 'weakly morphonotactic'. The latter intermediate category consists of cases like *concept*, which are not morphonotactically transparent (for etymological reasons, for example), but which might feature an inflectional or derivational operation. Cases labeled as weakly morphonotactic were excluded from the dataset for the following analyses.

In total, 314,158 potential final consonant cluster tokens were included in the dataset, of which 206,427 are lexical and 82,384 are morphonotactic. For each token, the corresponding date (depending on the text it was extracted from) was recorded, and the data cover roughly six centuries. Due to the unequal distribution of texts across this time span, the whole range was divided into sub-periods of 50 years, starting with the period from 1100 to 1150. The short-hand notation '1200' represents the period from 1200 to 1250, '1250' for the period from 1250 to 1300, etc. Table 1 shows the numbers of potential morphonotactic and lexical consonant clusters for each period as well as the sizes of the respective sub-corpora. Due to the small number of word-final consonant clusters in the Early Middle English period and the fact that schwa-loss began to spread no earlier than the 12th century (Brunner 1984; Fisiak 1968; see also Section 3.2.1 below), the first three half-centuries (1100, 1150, 1200) were excluded from the analysis.

period	morphonotactic	lexical	total count		sub-corpus size
1250	137 (26.3%)	384 (73.7%)	521	(2.7 pm)	192,086,758
1300	1,892 (25.4%)	5,559 (74.6%)	7451	(32.8 pm)	226,997,791
1350	10,786 (29.0%)	26,378 (71.0%)	37,164	(151.5 pm)	245,362,411
1400	15,409 (29.4%)	36,934 (70.6%)	52,343	(118.6 pm)	441,525,895
1450	13,889 (26.2%)	39,206 (73.8%)	53,095	(80.9 pm)	656,369,953
1500	6,356 (27.5%)	16,742 (72.5%)	23,098	(22.9 pm)	1,009,235,900
1550	6,963 (29.1%)	16,950 (70.9%)	23,913	(14.6 pm)	1,642,395,212
1600	6,525 (28.7%)	16,213 (71.3%)	22,738	(10.9 pm)	2,091,129,356
1650	9,153 (30.3%)	21,088 (69.7%)	30,241	(9.9 pm)	3,065,964,242

Table 1. Frequencies of potential word-final morphonotactic and lexical consonant clusters in the half-centuries from 1250 to 1700 together with the sizes of the corresponding sub-corpora. Figures in brackets denote fractions of morphonotactic and lexical sequences among all C(V)C(V) sequences (as %), and fractions of the latter sequences among the total number of words in the respective sub-corpora (per million words).

3.2 Operationalization of the hypothesis and its parameters

As described above, we are investigating the impact that the token frequency of morphonotactic instances of a given cluster type has on the number of lexical items its phonotactic counterpart occurs in, i.e. the cluster's lexicality, and we address the question of whether this impact changes diachronically. To this end, three variables have to be operationalized: (a) time, (b) morphonotactic token frequency, and (c) lexicality. These variables will be described in the following section, before presenting a more formalized version of our hypothesis.

3.2.1 Dramatis personae: time, frequency and lexicality

The first variable, time, simply measures the discrete 50-year periods from 1200 to 1700. The second variable to be covered is morphonotactic token frequency. The goal is to obtain an estimate of the number of morphonotactic tokens of a particular cluster type that a listener was exposed to. To this end, the raw number of morphonotactic tokens ending in a sequence /C(ə)C(ə)/ was determined for each cluster type and for each text, where /ə/ could be represented by any vowel grapheme. Since, particularly due to schwa-loss, the graphemic representation does not necessarily provide a reliable estimate of its phonological counterpart (think of the graphemic representation of the past tense suffix *-ed*, to name an obvious example), the frequency of the sequences with the above structure was adjusted probabilistically in order to attain more reliably frequencies of actual occurrence.

To explain this step in more detail, the process of schwa-loss in English is actually a combination of two deletion processes, one of which accounts for the loss of word-final schwa while the other deletes inter-consonantal checked schwa. The first process is believed to have initiated at the latest at around 1200 (Fisiak 1968: 36; Minkova 1991; Brunner 1984: 348) and finished no later than sometime in the 15th century (Dobson 1957: 879). The second process, i.e. the loss of checked schwa, started slightly later in the 14th century (Mossé 1991: 35) and was completed in nominal and verbal inflections at around 1600 (Dobson 1957: 883). Accordingly, for the purpose of this analysis, $t_0 = 1200$ and 1300 were taken as rough onset times and $t_1 = 1500$ and 1600 as rough offset times for the respective processes.

Phonological processes, such as schwa-loss, which act on a population of linguistic items often exhibit a sigmoid trajectory (Denison 2003; Wang & Minett 2005; Blythe & Croft 2012; cf. also Kroch 1989). Such a shape arises, in particular, if the process proceeds logistically, i.e. if its growth rate depends (a) on the amount of items that have already been affected by a change and (b) on the amount of items that have not yet been affected. Mathematically, logistic growth is modeled by the logistic function $p(t) = 1/(1 + e^{-c-rt})$, where $p(t)$ measures the proportion of items affected by the change at time t. Given the values of pairs $(t, p(t))$ for two different times t, the values of the constants c and r can be determined by making use of the logit transform of the above equation. Once c and r are known, the proportion p of affected items can be determined for any time t. For a single randomly drawn token at time t, $p(t)$ can now be interpreted as the probability of being affected by the change.

A separate logistic-spread process for each of the two sub-processes of schwa-loss was implemented. The respective proportions are $p_{\text{final}}(t)$ and $p_{\text{checked}}(t)$. In order to determine the respective constants c_i and r_i (i standing for 'final' or 'checked'), we made use of the above mentioned onset and offset times t_0 and t_1 and defined the onset proportion as $p_i(t_0) = .01$ and the offset proportion as $p_i(t_1) = .99$, i.e. 1% and 99% affected items, respectively. We assume that losing final schwa has no effect on the likelihood of losing interconsonantal schwa, and *vice versa*. In other words, the two sub-processes are regarded as independent. Thus, the probability of a word-final /C(ə)C(ə)/ sequence being a consonant cluster is at least $p_{\text{final}}(t) \times p_{\text{checked}}(t)$ at time t (Figure 1).

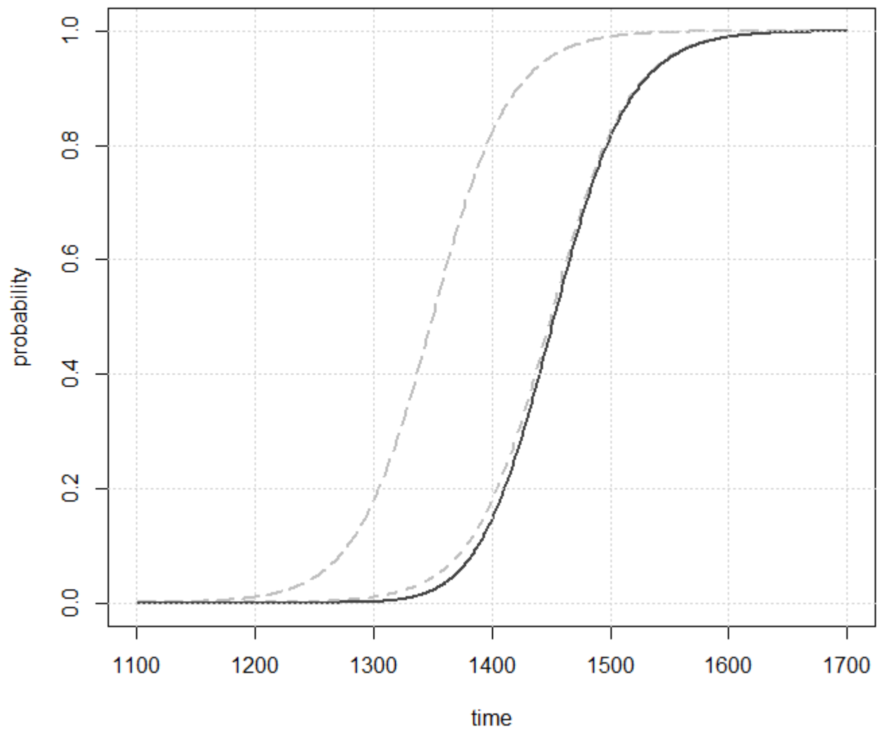


Figure 1. Spread of schwa-loss (dark gray) as an interacting process of the loss of final (light gray, long dashes) and checked schwas (light gray, short dashes). The vertical axis measures the probability of a final /CəCə/ sequence being affected by the change.

We take this product as a conservative estimate of the probability that an item /C(ə)C(ə)/ actually is a cluster, $p_{\text{cluster}}(t)$. It is crucial to note that this estimate constitutes a lower bound in the sense that in a /C(ə)C(ə)/ sequence, one of the two schwas might have been already lost or indeed may have never been present before the onset of schwa-loss. Thus, it ensures that the token frequencies in each period are not underestimated. We assume that this provides us with a more reliable estimate than just resorting to the problematic graphemic representations, which would result in much lower frequencies of cluster tokens.

The period-wise frequencies of morphonotactic word-final consonant-cluster tokens were then calculated according to the following procedure. The raw token frequencies of the sequences /C(ə)C(ə)/ corresponding to a cluster type CC, were determined for each text. These raw frequencies were multiplicatively adjusted by the above-described probability $p_{\text{cluster}}(t)$, where t is the estimated date of the text (see Kroch et al. 2004; Kroch & Taylor 2000). For each cluster type and each period, these adjusted frequencies were summated and subsequently normalized with respect to the period-specific sample size (i.e. the total number of words in all texts that belong to the half-century period). The base of normalization was set at 1 million. The same adjustment and normalization procedure was applied to the lexical instances of the respective cluster types, and the resulting frequencies are denoted as φ_{mpt} (or *mpt.frequency*) and φ_{lex} (or *lex.frequency*), respectively.

Finally, the third variable, lexicality (denoted as λ), is intended to measure in how many instances a cluster type occurs lexically rather than morphonotactically. We simply define it as the fraction of lexical tokens among all instances of that cluster type in a given period, thus, $\lambda = \varphi_{\text{mpt}}/(\varphi_{\text{lex}} + \varphi_{\text{mpt}})$.[3] Lexicality, computed in this way, has a straightforward and theoretically relevant interpretation. If for a cluster type the score is close to 0, it is morphonotactic, if it is close to 1, it is lexical, and if it is in-between, the cluster type is ambiguous with respect to its complexity-signaling function. Note that the lexicality scores of a given cluster type in a particular period can range from 0 to 1. Obviously, this score is always strictly smaller than 1, since the present analysis is restricted to cluster types with $\varphi_{\text{mpt}} > 0$ only. According to morphonotactic theory, cluster types are expected to disambiguate, hence approaching either 0 or 1 on the lexicality scale (cf. Section 2 and Dressler et al. 2010). Selecting proportional frequencies allows for a direct application of our findings to morphonotactic theory,

[3] Note that the fact that φ_{mpt} appears in the definition of λ is unproblematic, since for large frequencies, the strength of a monotone relationship between morphonotactic frequency and lexicality provides a lower bound for the strength of a monotone relationship between φ_{mpt} and φ_{lex}.

so that we can test whether clusters, under certain circumstances, show the opposite dynamics.

The choice of (adjusted) token frequency over type frequency in the definition of lexicality can be attributed to the lack of lemmatized data in the corpora used (which is obviously a consequence of the spelling variation and inconsistency in ME and, to a lesser extent, EModE).

3.2.2 The changing morphonotactic space and how it should evolve

In each period, the inventory of final morphonotactic consonant clusters can now be conceptualized by means of the Cartesian product of φ_{mpt} and λ, which we will refer to as the *morphonotactic space*. That is, for each cluster type, such as /nd/, /ns/ or /rn/, we determined, first, morphonotactic frequency and, second, lexicality. The scores on these two variables determine the cluster's location in the morphonotactic space. If this is done for all potentially morphonotactic cluster types, the cloud of resulting points in this space constitutes the morphonotactic cluster inventory in that period. Figure 2 below shows the cluster inventories in all the half-centuries from 1250 to 1700. In each scatterplot, the horizontal axis measures the morphonotactic frequency while the vertical axis measures lexicality. The locations of the cluster types are represented by points in the plot.

For our hypothesis we want to show that those cluster types which are morphonotactically frequent, in the long run appear in more and more lexical items, i.e. they should become more lexical. Thus, clusters that are located on the right of the scale should evolve in such a way that they also score high on the lexicality scale. In other words, we hypothesize that diachronically the cluster inventory establishes a positive monotone relationship between φ_{mpt} and λ.

The dynamic component in this hypothesis is crucial. It is not sufficient to show that at a given point in time there is such a monotone relationship. Rather, it has to be shown that frequent clusters evolve in such a way that they become integrated into lexical items via analogy effects. A synchronic view alone fails to shed light on this matter, since the development could just as well move in the opposite direction. In the present analysis, we exploit the fact that in Old English, word-final morphonotactic clusters were quite rare, since inflectional endings were typically syllabic. Thus, schwa-loss has given rise to a completely novel inventory of clusters, in which changes, such as those analogy-driven ones found within lexical items, should be clearly observable.[4] In the subsequent section, it will be confronted with the historical data.

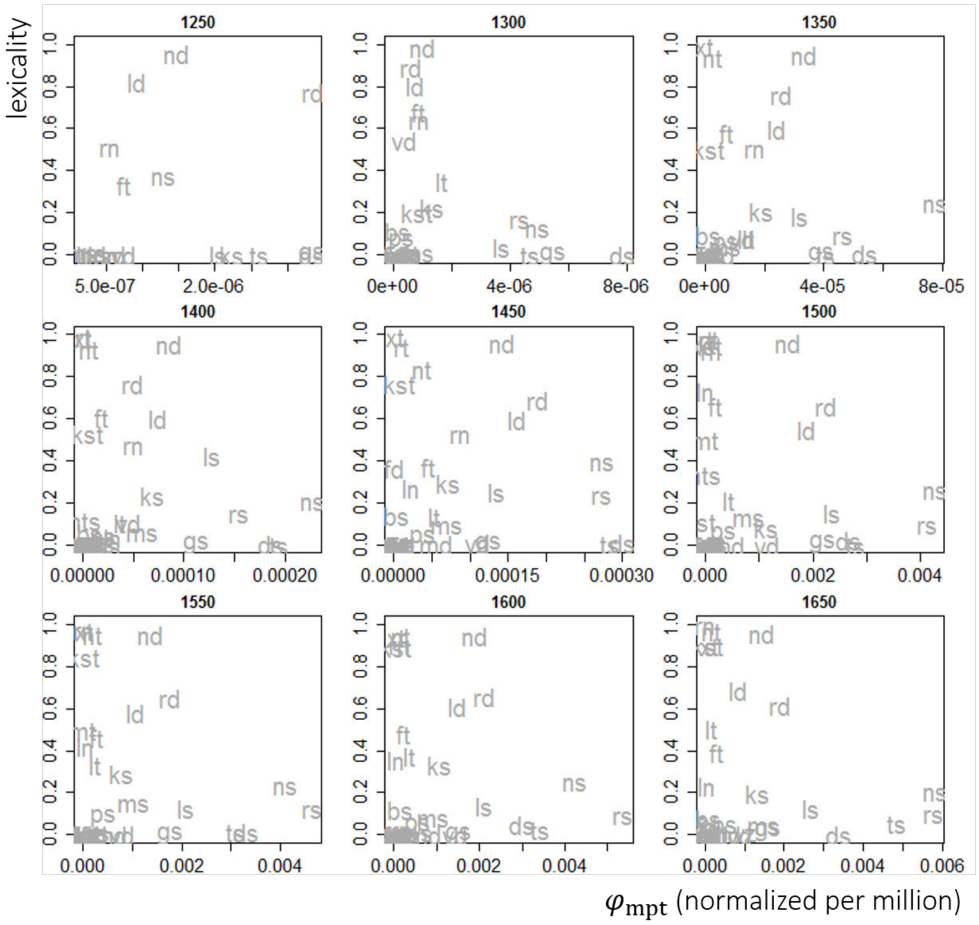


Figure 2. Labelled scatterplots showing the morphonotactic space in the nine semi-centuries from 1250 to 1700. Frequency scores were adjusted probablistically according to the spread of schwa-loss and normalized per million with respect to the period-wise subcorpora.

3.3 Data analysis

This section describes two approaches which are intended to answer the questions of whether token frequency has an effect on a cluster's lexicality and, furthermore, whether or not this effect varies over time. First, a generalized additive model will be fitted to the complete data set (3.3.1). Second, we will investigate in which way the

[4] It is worthwhile pointing out that due to this, the history of English provides an ideal testing ground for the hypothesis outlined above.

correlation (or more precisely: the period-specific correlation coefficients) between frequency and lexicality evolve diachronically (3.3.2).

3.3.1 Fitting a generalized additive model

We are interested in how the interaction between time and the frequency of morphonotactic clusters affects the lexicality of a cluster: thus, multidimensional modelling of the dependence of lexicality on the other two variables is required. In this analysis, a generalized additive model (GAM) was selected. In conventional linear regression models, interactions between predictor variables result in multiplicative linear terms. This means, that if one predictor variable—say time, as in the present case—is held constant, the dependent variable (λ) is a linear function of the second variable (φ_{mpt}). However, we are not exclusively interested in linear relationships between the latter two variables. Instead, any monotone (decreasing or increasing) relationship between frequency and lexicality would be of interest according to our hypothesis. Hence, a more flexible modeling technique not restricted to linear dependencies is required, and GAMs fulfill these requirements. In a nutshell, GAMs are models which are composed of linear and nonlinear components (so-called 'splines'), thus yielding smoothly curved (or 'wiggly') surfaces that fit to the data in a statistically satisfying way. GAMs have been used extensively in ecology and evolution, and more recently in linguistics (Wieling et al. 2011; Baayen 2013; Fruehwald 2015).

Before feeding the data into the model, some adjustments had to be made. The φ_{mpt} scores were first normalized with respect to the period-wise maximal scores. This was necessary, since due to the S-shaped spread of schwa-loss, frequency scores were concentrated close to 0 in the earlier periods. By normalizing the data with respect to the maximal scores, a more appropriate model of the relationship between frequency and lexicality was achieved. Second, as the φ_{mpt} scores were strongly skewed to the right (see Figure 2), they were Box-Cox transformed, i.e. put into a shape that resembles normally distributed data (Box & Cox 1964). Figure 3a displays the morphonotactic space changing in time. Overall, the dynamics look rather complicated, so that in order to detect diachronic patterns, fitting a model to the data indeed might provide more insights.

In the GAM, lexicality is modeled as being related to the interaction of time and morphonotactic frequency. The `mgcv` package in R (Wood 2006a; R Development Core Team 2013) enables us to include so-called *tensor product smooths* into a GAM. While a detailed explanation is not relevant here, suffice it to note that tensor products provide a simple way of modeling interactions between predictor variables in a GAM (Wood 2006a, 2006b). The GAM computed from the data yields a significant intercept (at 0.224; $p < .0001$) and a tensor-product term ($p = .0079$; estimated $df = 7.597$), which means that the morphonotactic space indeed changes significantly over time rather than staying roughly the same.

In order to interpret the model, it has to be visualized. Figure 3b shows the surface defined by the GAM, in which the following three patterns can be observed. (a) Very rare clusters evolve from medially lexical to predominantly morphonotactic because the surface defined by the GAM heads downwards for low φ_{mpt} values. This fits the frequency-driven analogy effects part of our hypothesis. (b) In accordance with the same hypothesis, medially frequent clusters become more lexical. However, (c) morphonotactically highly frequent clusters evolve from a primarily lexical state to a slightly less lexical one. This contradicts our predictions, since these clusters would be expected to become even more lexical. Looking at the period-wise one-dimensional curves, which depict the dependence of lexicality on frequency by fixing a point on the time axis and moving along the grid on the smooth wiggly surface in the direction of φ_{mpt}, it can be seen that the one which comes closest to an increasing monotone relationship is attained somewhere around the 1450 period. After this date, intermediately frequent clusters (b) overtake highly frequent ones in terms of lexicality (c).

In summary, the data in the first part of the observed time span, which corresponds to a large share of the ME period, seems to provide evidence for frequency effects among morphonotactic and lexical consonant clusters (see hypothesis 2), while the later data do not. This contrast is too interesting to be ignored, and therefore, the following section outlines another approach, which allows for a systematic analysis of this antithetic behavior.

3.3.2 Analysis of the correlation-coefficient trajectory

In the second approach, all periods are dealt with separately. The aim is to investigate whether the monotone relationship between frequency and lexicality increased diachronically. To this end, the corresponding correlation coefficients were determined for each period. This allows the investigation of the trajectory of correlation coefficients, which should increase according to the hypothesis being tested.

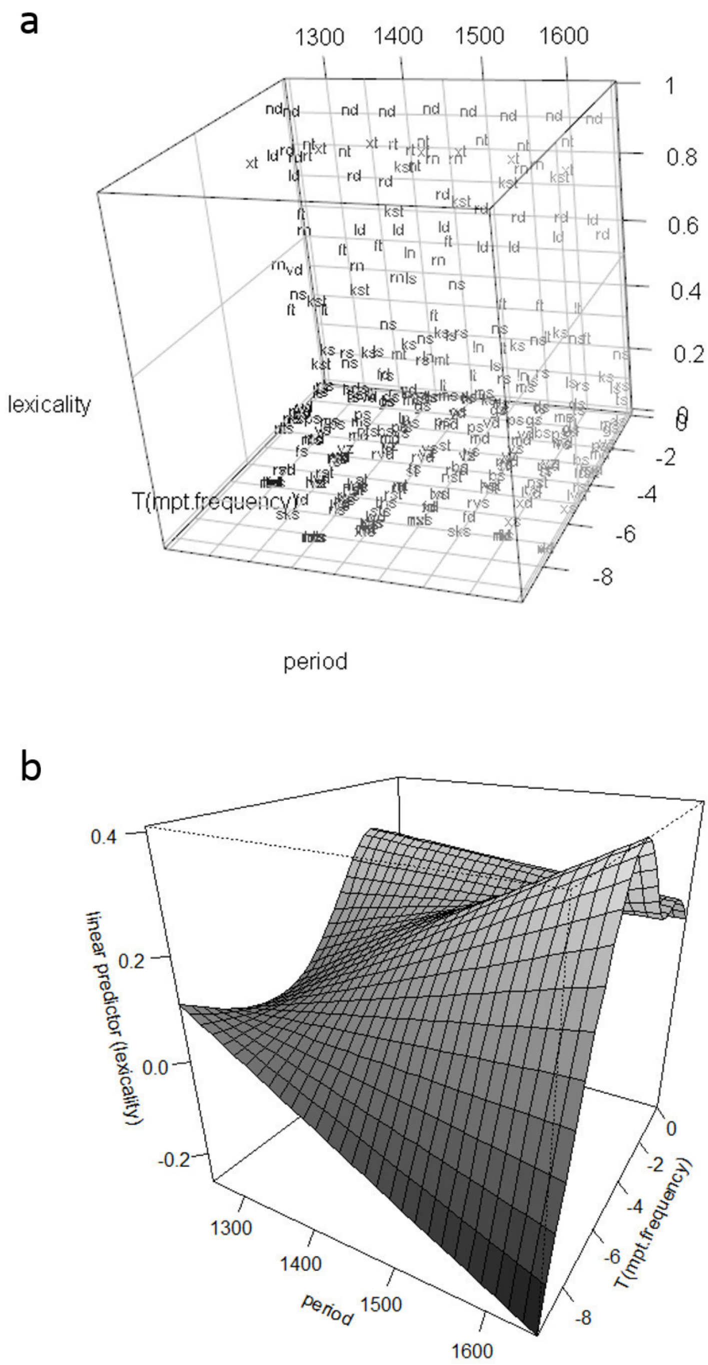


Figure 3. (a) Scatterplot showing the diachronic development of the morphonotactic space, where φ_{mpt} is normalized with respect to the period-wise maximal scores and Box-Cox transformed to $T(\varphi_{\text{mpt}})$. (b) Plot of a generalized additive model fitted to the transformed data. The impact of the morphonotactic token frequency on cluster lexicality is evidently changing over time.

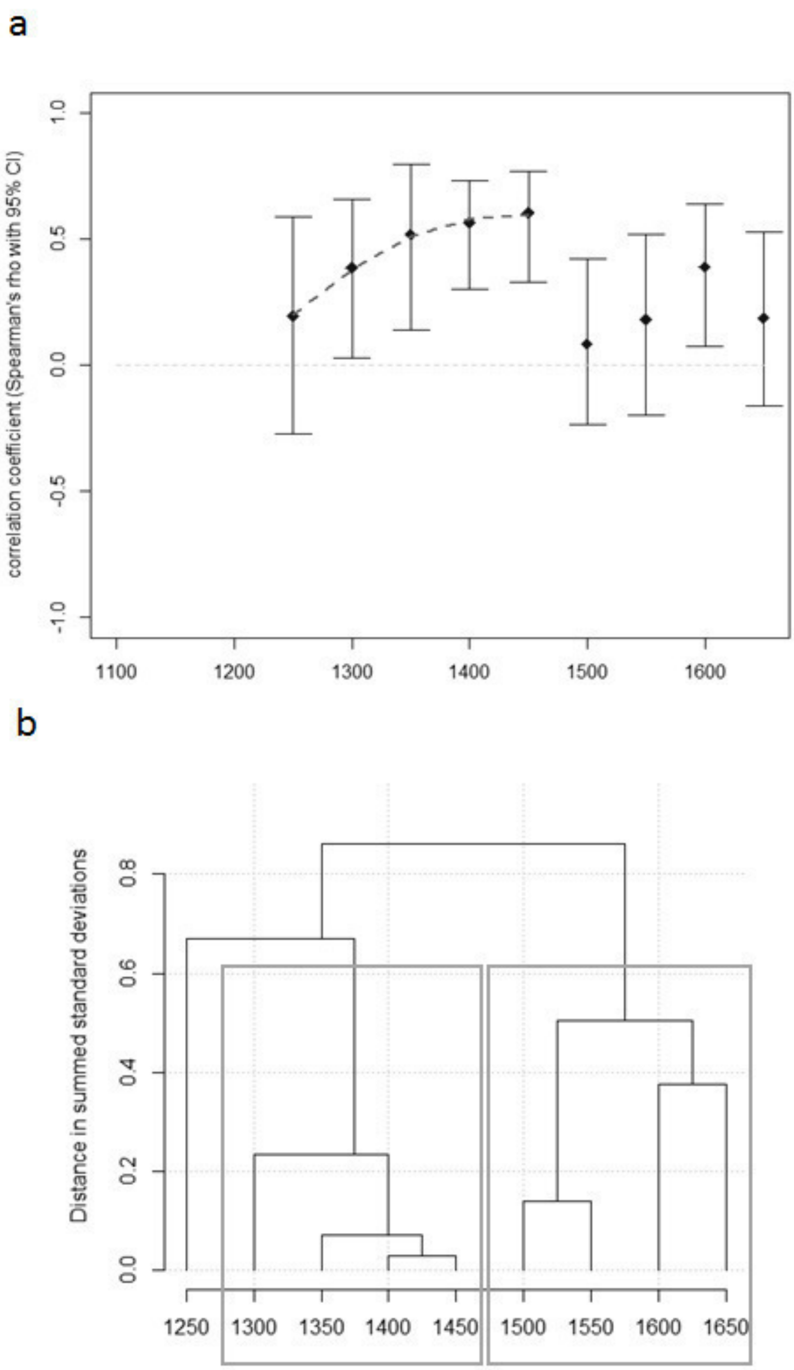


Figure 4. (a) Trajectory of estimated correlation coefficients (Spearman's ρ) together with 95% confidence intervals for all half-centuries from 1250 to 1700. The dashed line corresponds to a fitted quadratic model (adjusted $R^2 = .99$, $p = .004$). From 1250 to 1500 a significant positive correlation is established. (b) VNC based dendrogram of the successive correlation coefficients. Between 1450 and 1500 a break in the diachronic development is clearly observable.

For the present analysis, we selected Spearman's ρ as the correlation coefficient of φ_{mpt} and λ. This is motivated by the skewed distribution of the data and, more importantly, by the fact that we want the correlation measure to be sensitive to any monotone relationship. Due to the non-parametric nature of Spearman's ρ, the data were not transformed. Nine correlation coefficients ρ_i $(i = 1,2,\dots,9)$ were determined $(t_i = 1250\,, 1300,\dots,1650$; cluster-inventory sizes N_i ranging from 18 to 44). The trajectory of correlation coefficients together with the corresponding 95% confidence intervals (computed with the `RVAideMemoire` package in R) is shown in Figure 4a.

An inspection of Figure 4a clearly shows that, as expected from the results of the previous analysis, the correlation between φ_{mpt} and λ increases until the end of the 15th century, not significantly at first, but then reaching significance and approximating a strong correlation of $\rho \cong 0.5$ (Cohen 1992). However, in the 1500 period, the correlation drops close to zero and becomes non-significant again. Apart from the 1600 period, which exhibits a significant relationship again, this trend stays the same from then on.

Looking at the confidence intervals alone, which, crucially, do overlap when looking at the 1450 and 1500 periods, we cannot confidently claim that the data show the existence of two substantially different periods (before and after 1500). Hence, a clustering technique, variability-based neighbor-clustering, was employed which allows for the identification of stages in sequential data.

Variability-based neighbor clustering (VNC, Gries & Hilpert 2008) is a hierarchical clustering method which has the advantage of keeping a fixed ordering of the leaves of the hierarchy tree, because the clustering proceeds in such a way that only the direct neighbors in a previously defined sequence—here, successive time periods—are eligible for clustering on the various hierarchical levels. Hence VNC provides an excellent method to detect sets of similarly behaving periods in diachronic developments.

Figure 4b shows the dendrogram[5] which results from the application of VNC to the trajectory of correlation coefficients ρ_i. It clearly divides the observation period into two stages, one corresponding to the periods before 1500 and one to the periods after 1500. Hence, it can be concluded that from 1250 to 1500, the cluster

[5] VNC computations were done in R (version 3.0.2) with a script written by Stefan Gries and Martin Hilpert (see http://global.oup.com/us/companion.websites/fdscontent/uscompanion/us/static/companion.websites/nevalainen/Gries-Hilpert_web_final/vnc.individual.html; accessed 16.02.2016).

inventory evolved as hypothesized. After the break in 1500, the development does not show a clear pattern.[6]

3.3.3 Interpretation of the break: the transition from LME to EModE

Both analyses, the GAM as well as the VNC analysis of the correlation coefficients, suggest that at the end of the ME period, the dynamics of the cluster inventory showed a substantial change. Figure 5 shows the cluster inventories from these two periods overlaid in the same plot. Morphonotactic frequency was normalized with respect to the maximum score in each respective period. In order to identify the clusters that behave differently in the two periods, the cluster locations are labeled by the respective phonological representation (light gray indicates 1450 data, and dark gray represents the 1500 data). For the sake of illustration, linear regression lines were included, although due to the distributional properties of the data, they should be treated with caution. The regression lines are nevertheless helpful for identifying the cluster types which are responsible for the change in the correlation coefficient.

Two sets of cluster types seem to be particularly involved in the changing behavior: /Cs/ clusters, which, contrary to our predictions, become more frequent and less lexical, and /Cn/ clusters, which become slightly less frequent and more lexical (locations indicated by circles).

In ME, instances of the /Cs/ group occur as verbal present tense inflections (Northern dialects, Horobin & Smith 2002: 117), as well as nominal plural and genitive forms (all dialects). At the end of the 15th century, the inflectional competitors of the -(*e*)*s* ending (-*eth* and contracted -*t*) were ousted, partially due to migration and, in the end, standardization processes which took place at this time (cf. Horobin & Smith 2002), so that -(*e*)*s* became the default choice for expressing 3rd person singular in verbs and plural in nouns. Thus, morphonotactic final /Cs/ clusters became more frequent.

---

[6] One might wonder, at this point, why a configuration of morphonotactic and lexical clusters which is so dispreferred that it needs to be repaired by processes operating during first language acquisition, as we intended to demonstrate by the development in the pre-1500 era, came about at all. We propose the following quite straightforward answer: only when schwa-loss had produced a reasonably large number of consonant clusters was it possible for analogy to function in language acquisition. We assume that this must have been at around 1200.

In contrast, word-final /Cn/ clusters went through a completely different development. The ME -*e*(*n*) suffix played a substantial role in the inflectional morphology of nouns (as a plural suffix) and verbs (as a subjunctive and partially indicative plural suffix, and as an infinitival suffix). In the transition from the ME to the EModE period, this suffix began to become less productive, and eventually dropped out of inflectional morphology. As a consequence, morphonotactic word-final /Cn/ clusters became less frequent and more lexical.

In summary, the above observations show that both sets of clusters exhibit a development that is not in line with the proposed monotonously increasing relationship between morphonotactic frequency and lexicality.

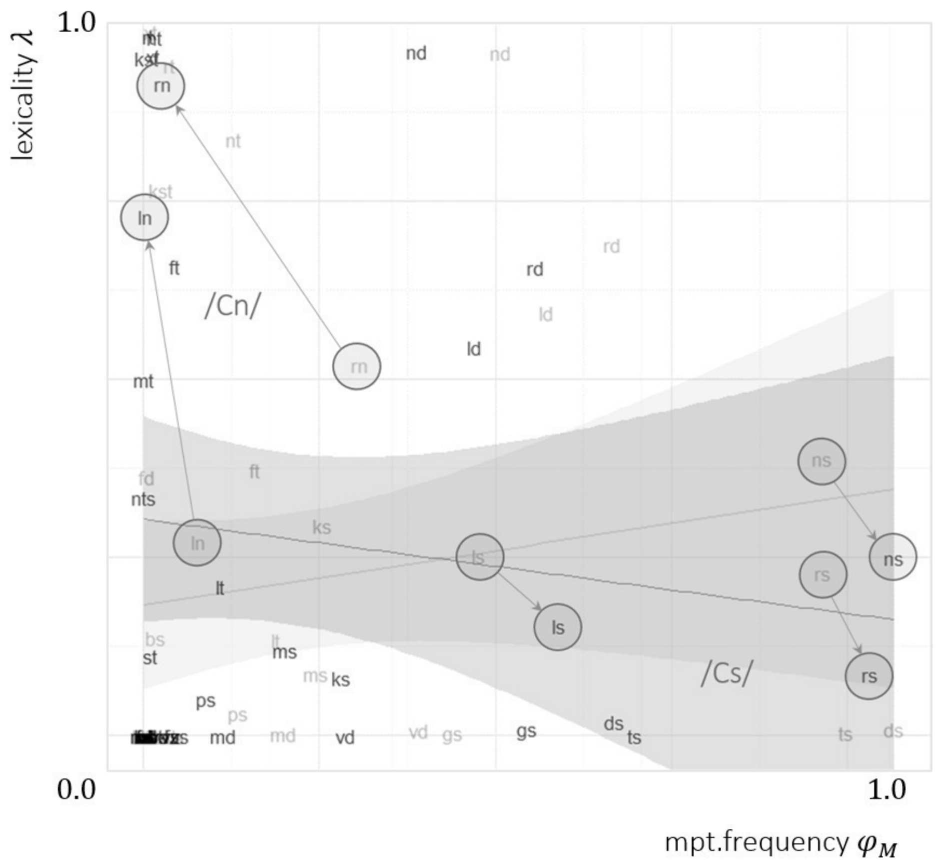


Figure 5. Plot of the superimposed morphonotactic spaces of the cluster inventories in the successive periods 1450–1500 (light gray labels) and 1500–1550 (dark gray labels), respectively. The horizontal axis was scaled in such a way that the respective maximal and minimal scores in both periods coincide. Linear regression models (solid lines) together with 95% confidence areas (gray) were added and illustrate a positive correlation in the 1450-1500 period in contrast to a (not significant) negative correlation in the 1500-1550 period. Circles indicate morphosyntactically relevant dynamics corresponding to a set of /Cn/ and /Cs/ clusters.

The crucial point is that these developments are driven by morphosyntactic and sociolinguistic factors, and hence are not phonologically or phonotactically conditioned. We conclude that the period before 1500 provides corroborating evidence for the hypothesis about analogy effects. However, the period after 1500 does not yield a clear picture. Indeed, more recent data (i.e. ModE after 1700) is needed in order to satisfactorily address the question at hand.

4 Conclusion

At the outset of this paper we put forth the question of whether morphonotactic consonant clusters provide supporting effects for lexical instances of the same cluster type via analogy. This hypothesis was motivated by observations and conjectures found in the (mor)phonotactic literature about the inhibiting and supporting effects among the two cluster categories. We hypothesized that frequency effects in the first stages of language acquisition could give rise to these supporting effects, and that as a diachronic reflex of these effects, characteristic diachronic patterns were proposed to be observable. More specifically, we expected morphonotactically token-frequent clusters to become more lexically present (see hypothesis 2).

The latter claim was formalized in the following fashion: diachronically, a positive monotone relationship between the morphonotactic token frequency and the lexicality of the consonant clusters in the inventory of the language was expected to establish itself. Data from the ME period were used to test this hypothesis quantitatively by means of two different modeling approaches. Using ME data for addressing the research question at hand suggested itself, since through schwa-loss a completely new set of consonant clusters was created, so that the diachronic reflexes of the hypothesized analogy effects should be clearly observable.

In the first modeling approach, a generalized additive model (GAM) was fit to the data. It showed that indeed in the first part of the ME period, the relationship between morphonotactic frequency and lexicality evolved as expected, but that later on and contrary to expectation, intermediately frequent clusters became more lexical than highly frequent clusters. The second modeling approach provided a detailed look at when the change in the behavior of the cluster inventory took place. It was shown that before 1500 the inventory behaved as expected under the assumption of frequency-driven analogy in language acquisition, while after 1500 no particular pattern could be observed. Looking at the clusters involved in this change, we showed that the shifts in frequency or lexicality in a number of cluster types can be attributed to morphosyntactic or sociolinguistic, and thus phonology-external, changes. Although an investigation of the development after 1500 would naturally be interesting, a systematic survey of this period exceeds the scope of our data. Hence, we can conclude that at least before the onset of inflectional reduction and standardization, the diachronic dynamics of the ME coda-cluster inventory suggest a supporting relationship between morphonotactic and lexical clusters (see hypothesis 1).

This has interesting implications for morphonotactic theory. One of the major claims about morphonotactic consonant clusters is that their functionality in terms of signaling morpheme boundaries is diminished by the presence of structurally similar lexical clusters, so that there is an inhibitory relationship between the two cluster categories, as outlined in Section 2. A corollary of this is that cluster types should disambiguate so that they become either purely morphonotactic or purely lexical. The findings from the present study add two novel aspects to the expected diachronic dynamics of consonant clusters.

First, it can be specified which clusters should become more lexical and which ones should become more morphonotactic. According to our findings it should be—somewhat counterintuitively—the morphonotactically *highly* token-frequent clusters that evolve towards the lexical boundary, while their low-frequency counterparts are expected to evolve into the less lexical, i.e. morphonotactic direction.

Second, the findings in this study bolster the evidence for the supporting rather than inhibiting effects between morphonotactic and lexical clusters, so that a diachronic development towards more ambiguous configurations would be expected. In this sense, two opposing forces are at work in the diachronic dynamics of consonant clusters, one which favors unambiguous clusters and a second one which favors ambiguous ones. The—doubtless language-specific—nature of the interaction between these two forces, however, still remains to be explored in further studies.

Comments invited

PiHPh relies on post-publication review of the papers that it publishes. If you have any comments on this piece, please add them to its comments site. You are encouraged to consult this site after reading the paper, as there may be comments from other readers there, and replies from the author. This paper's site is here:

http://dx.doi.org/10.2218/pihph.1.2016.1693

Acknowledgements

We would like to thank the participants of the 3rd International Workshop on Phonotactics and Phonotactic Modeling (Vienna, November 2015), in particular Basilio Calderone, for many helpful comments on the present project, Kamil Kaźmierski, Ines Nitsche, Paula Orzechowsa, and Paulina Zydrowicz for helping us with analyzing the historical data, as well as Elnora ten Wolde and Klaus Hofmann for their support. This research was supported by the Austrian Science Fund (FWF, grant No. P27592-G18) and the Österreichische Austauschdienst (OeAD, grant No. PL8/2014).

First Author contact details

Andreas Baumann
Department of English and American Studies
University of Vienna
Spitalgasse 2-4, Hof 8.3
A–1090 Vienna, Austria

andreas.baumann@univie.ac.at

References

Baayen, Harald. 2013. Multivariate Statistics. In Robert Podesva & Devyani Sharma (eds.), *Research Methods in Linguistics*, 337–372. Cambridge: Cambridge University Press.

Berent, Iris, Donca Steriade, Tracy Lennertz & Vered Vaknin. 2007. What we know about what we have never heard: Evidence from perceptual illusions. *Cognition* 104(3). 591–630.

Blythe, Richard A. & William Croft. 2012. S-curves and the mechanism of propagation in language change. *Language* 88(2). 269–304.

Box, George & David Cox. 1964. An analysis of transformations. *Journal of the Royal Statistical Society Series B* 26(2). 211–252.

Brinton, Laurel J. & Elizabeth C. Traugott. 2005. *Lexicalization and language change* (Research surveys in linguistics). Cambridge, UK, New York: Cambridge University Press.

Brown, Roger. 1973. *A first language: The early stages*. Harvard: Harvard University Press.

Brunner, Karl. 1984. *Die englische Sprache*: *Ihre geschichtliche Entwicklung* (Sammlung kurzer Grammatiken germanischer Dialekte. B, Ergänzungsreihe Nr. 6), 2nd edn. Tübingen: Niemeyer.

Calderone, Basilio, Chiara. Celata, Katharina Korecky-Kröll & Wolfgang U. Dressler. 2014. A computational approach to morphonotactics: evidence from German. *Language Sciences* 46. 59–70.

Cohen, Jacob. 1992. A power primer. *Psychological Bulletin* 112. 155–159.

Denison, David. 2003. Log(ist)ic and simplistic S-curves. In Raymond Hickey (ed.), *Motives for Language Change*, 54–70. Cambridge: Cambridge University Press.

Dobson, Eric J. 1957. *English Pronunciation 1500-1700*: *Volume II: Phonology*. Oxford: Clarendon Press.

Dressler, Wolfgang U. & Katarzyna Dziubalska-Kołaczyk. 2006. Proposing Morphonotactics. *Wiener Linguistische Gazette* 73. 69–87.

Dressler, Wolfgang U. Katarzyna Dziubalska-Kołaczyk & Lina Pestal. 2010. Change and variation in morphonotactics. *Folia Linguistica Historica* 31. 51–68.

Dziubalska-Kołaczyk, Katarzyna & Paulina Zydorowicz. 2014. The production of high-frequency clusters by native and non-native users of Polish. *Concordia Working Papers in Applied Linguistics* 5. 130–144.

Fehér, Olga, Elizabeth Wonnacott & Kenny Smith. 2016. Structural priming in artificial languages and the regularisation of unpredictable variation. *Journal of Memory and Language*.

Ferreira, Victor S. & Kathryn Bock. 2006. The functions of structural priming. *Language and Cognitive Processes* 21(7-8). 1011–1029.

Fisiak, Jacek. 1968. *A Short Grammar of Middle English*: *Part One: Graphemics, Phonemics and Morphemics*. Warsaw: Panstwowe Wydawnictwo Naukowe.

Fruehwald, Josef. 2015. *Generations, lifespans, and the zeitgeist*. The Second Edinburgh Symposium on Historical Phonology. Edinburgh.

Giegerich, Heinz. 1999. *Lexical strata in English*: *Morphological causes, phonological effects*. Cambridge: Cambridge University Press.

Gries, Stefan T. 2005. Syntactic priming: A corpus-based approach. *Journal of Psycholinguistic Research* 34(4). 365–399.

Gries, Stefan T. & Martin Hilpert. 2008. The identification of stages in diachronic data:variability-based neighbour clustering. *Corpora* 3(1). 59–81.

Hogg, Richard M. & Christopher B. McCully. 1987. *Metrical Phonology*: *A Course Book*. Cambridge: Cambridge University Press.

Horobin, Simon & Jeremy J. Smith. 2002. *An Introduction to Middle English*. Oxford: Oxford University Press.

Jäger, Gerhard & Anette Rosenbach. 2008. Priming and unidirectional language change. *Theoretical Linguistics* 34(2). 85–113.

Jusczyk, Peter W. Paul Smolensky & Theresa Allocco. 2002. How English-Learning Infants Respond to Markedness and Faithfulness Constraints. *Language Acquisition* 10. 31–73.

Kiparsky, Paul. 1982. Lexical phonology and morphology. In Ik-Hwan Lee (ed.), *Linguistics in the morning calm*, 1st edn. 3–91. Seoul: Hanshin Publ.

Kroch, Anthony. 1989. Reflexes of Grammar in Patterns of Language Change. *Language Variation and Change* 1. 199–244.

Kroch, Anthony, Beatrice Santorini & Lauren Delfs. 2004. *Penn-Helsinki Parsed Corpus of Early Modern English*. http://www.ling.upenn.edu/hist-corpora/.

Kroch, Anthony & Ann Taylor. 2000. *Penn-Helsinki Parsed Corpus of Middle English*. http://www.ling.upenn.edu/hist-corpora/.

Labov, William. 1989. The child as linguistic historian. *Language Variation and Change* 1. 85–97.

Martin, Andrew. 2007. *The evolving lexicon.* Los Angeles: UCLA PhD dissertation.

McMahon, April M. S. 2002. *An Introduction to English Phonology* : Edinburgh University Press.

McQueen, James. 1998. Segmentation of Continuous Speech Using Phonotactics. *Journal of Memory and Language* 39. 21–46.

Minkova, Donka. 1991. *The history of final vowels in English*: *The sound of muting* (Topics in English linguistics 4). Berlin, New York: M. de Gruyter.

Mossé, Fernand. 1991. *A handbook of Middle English*, 10th edn. Baltimore, Md.: Johns Hopkins University Press.

Pickering, Martin J. & Victor S. Ferreira. 2008. Structural priming: a critical review. *Psychological Bulletin* 134(3). 427–459.

Plag, Ingo, Julia Homann & Gero Kunter. 2015. Homophony and morphology: The acoustics of word-final S in English. *Journal of Linguistics*. 1–36.

Post, Brechtje, William D. Marslen-Wilson, Billi Randall & Lorraine K. Tyler. 2008. The processing of English regular inflections: Phonological cues to morphological structure. *Cognition* 109(1). 1–17.

R Development Core Team. 2013. *R: A Language and Environment for Statistical Computing*: *A language and environment for statistical*. Vienna, Austria: R Foundation for Statistical Computing.

Rumelhart, David E. & James L. McClelland. 1986. On learning the past tenses of English verbs. In PDP Research Group (ed.), *Parallel distributed processing: explorations in the microstructures of cognition*, 216–271. Cambridge, MA: MIT Press.

Schlüter, Julia. 2005. *Rhythmic grammar*: *The influence of rhythm on grammatical variation and change in English* (Topics in English linguistics 46). Berlin, New York: Mouton de Gruyter.

Shields, Lynne W. & David A. Balota. 1991. Repetition and associative context effects in speech production. *Language and Speech* 34(1).

Shockey.

Wang, William & James Minett. 2005. The invasion of language: emergence, change and death. *Trends in Ecology and Evolution* 20. 263–269.

Wieling, Martijn, John Nerbonne & Harald Baayen. 2011. Quantitative Social Dialectology: Explaining Linguistic Variation Geographically and Socially. *PloS one*.

Wood, Simon. 2006a. *Generalized Additive Models: an introduction with R*. Boca Raton, Florida: Chapman & Hall/CRC.

Wood, Simon. 2006b. Low-Rank Scale-Invariant Tensor Product Smooths for Generalized Additive Mixed Models. *Biometrics* 62. 1025–1036.

Assessing the effect of ambiguity in compositionality signaling on the processing of diphones

Andreas Baumann & Kamil Kaźmierski

Abstract: Consonantal diphones differ as to their *ambiguity* (whether or not they indicate morphological complexity reliably by occurring exclusively either within or across morphemes) and *lexicality* (how frequently they occur within morphemes rather than across morpheme boundaries). This study empirically investigates the influence of ambiguity and lexicality on the processing speed of consonantal diphones in speech perception. More specifically, its goal is to test the predictions of the Strong Morphonotactic Hypothesis, which asserts that phonotactic processing is influenced by morphological structure, and to clarify the two conceptions thereof present in extant research. In two discrimination task experiments, it is found that the processing of cross-morpheme diphones decreases with their ambiguity, but there is no processing difference between primarily cross-morphemic and morpheme-internal diphones. We conclude that the predictions of the Strong Morphonotactic Hypothesis are borne out only partially, and we discuss the discrepancies.

Highlights:

- ★ Ambiguity in signaling morphological complexity affects diphone processing
- ★ Speakers have probabilistic knowledge of how often diphone types span morpheme boundaries
- ★ Diphones that occur prototypically within morphemes are processed as fast as prototypically cross-morphemic diphones
- ★ Processing of cross-morphemic diphones is slow if they are ambiguous
- ★ Participants can be primed for analyzing diphones in nonce words as spanning a morpheme boundary

Keywords: morphonotactics, compositionality signaling, ambiguity, perception

1 Introduction

The processing of sound sequences, and that of word-internal consonant sequences in particular, have been argued to depend, among other factors, on the morphology of words they are embedded in: some diphones, such as /ld/ or /nd/, occur across morpheme boundaries (*call+ed, wan+ed*) as well as morpheme internally (*cold, wand*), while others are restricted to a single morphological environment (/md/ as in *seem+ed*, and /mp/ as in *lamp*, respectively). This has been suggested in turn to affect their acquisition and diachronic development (Dressler et al., 2010; Korecky-Kröll et al., 2014; Leykum et al., 2015a; Zydorowicz, 2007).

This work aims at assessing the influence of morphological status of consonantal sequences on the ease of their processing in speech perception. We address this aim by means of two related experiments conducted with speakers of Polish. Our experimental setup will, more specifically, address two divergent propositions that have been drawn from - and sometimes equated with - a central hypothesis in the research focusing on the interaction of phonotactics and morphology, i.e. the morphonotactic research paradigm (see

Table 1 for terminological clarification). This hypothesis in a nutshell asserts that sound sequences may have the function of signaling morpheme boundaries and triggering the decomposition of a complex word. Above the morphological level, it is well known that phonotactic knowledge helps listeners in the decomposition of the speech stream into words (McQueen, 1998; Mattys et al., 1999; Mattys and Jusczyk, 2001; Daland and Pierrehumbert, 2011; van der Lugt, 2001). Thus, sound sequences which rarely occur within words function as boundary signals and thus speed up the parsing process. In morphonotactics, this principle is transferred to the word-internal domain, i.e. the decomposition of words into morphemes.

Put into semiotic terms, sound sequences are hypothesized to function as *signifiants* for the *signifié* ‘morphological boundary’ (Dressler and Dziubalska-Kołaczyk, 2006). If sound sequences indeed fulfill this semiotic function, then the reliability of this function and by consequence the ease of processing of boundary spanning sequences should be diminished as soon as signaling becomes ambiguous (in the sense that the same consonant sequence can be additionally used within morphemes). The question is whether the latter condition holds true. This is what we test in our first experiment. In the second experiment we consider the question of whether the ambiguity of a sequence in general affects its processing. The subtle difference between these two questions, which have both been addressed but not always clearly distinguished from each other in morphonotactic research, is this: The former is about the effect of ambiguity on the quality of a sign, which as a consequence is expected to affect the processing of a sequence (the sign’s *signifiant*; ‘Is the boundary-signaling sequence /md/ in *seem+ed* processed faster than /nd/ in *wan+ed*?’). The latter considers the effect of ambiguity on the processing of a sequence without being restricted to denoting a morpheme boundary (‘Is /md/ generally processed faster than /ld/, irrespective of whether /ld/ occurs in *call+ed* or in *cold*).

We will show that the central hypothesis is confirmed by our experiments, albeit only partially: boundary-spanning instances of diphone types (such as /ld/ in *called*) are processed most slowly if the type occurs across morpheme boundaries and within morphemes at roughly equal frequencies (e.g. /ld/). Thus, speakers have probabilistic knowledge of the morphological environment of diphones. We argue that this suggests a cognitive model of phonotactics in which memories of instances of sound strings are stored together with morphological information (Plag et al., 2017). We do, however, not find a general advantage of non-ambiguous (/md/) over ambiguous (/ld/) diphone types if cross-morpheme instances are not explicitly tested, nor did we detect a general advantage of primarily boundary spanning (/md/) over primarily morpheme-internal (/rl/) diphone types (or the reverse).

In our analysis, we employ different ways of measuring ambiguity of signaling morpheme boundaries, in particular differentiating between type and token frequencies. In order to detect potentially nonlinear effects of ambiguity, we use generalized additive models (Wood, 2006), a modeling technique which recently gained momentum in linguistic research (e.g. Wieling et al., 2011; Baayen, 2013; Fruehwald, 2017). Thus, in addition to providing results on the processing of sequences of sounds, this study, on a more theoretical level, seeks to highlight and clarify some of the argumentative vagueness that seems to be present in the morphonotactic literature, while at the same time featuring relatively novel analytical methods.

The cornerstones of morphonotactics shall be described together with our specific research questions in the remainder of this section and in Section 2. Afterwards, the two experiments together with their analyses (Section 3 and 4) shall be presented and finally discussed (Section 5).

Table 1. Phonotactic and morphonotactic terminology

term	meaning	example
diphone	sequence of two single sound segments	/hæ/, /æn/, and /nd/ in *hand* /hænd/
consonant cluster	sequence of consonants; sometimes restricted to sequences within syllables (not in this study)	/nd/ in *hand* /hænd/
morphonotactic instance of a cluster	token of a cluster which spans a morpheme boundary; sometimes referred to as *morphotactic*, *boundary spanning* or *cross-morphemic* cluster	/nd/ in *bann+ed* /bænd/
lexical instance of a cluster	token of a cluster which is morpheme internal; also referred to as *phonotactic*	/nd/ in *hand* /hænd/
primarily morphonotactic cluster	cluster type which has exclusively or almost exclusively morphonotactic instances; sometimes measured in type frequency rather than token frequency; also referred to as *morphonotactic strong default*, *prototypically morphonotactic*, or if token frequency is used *low probability*	word final /ts/ as in *bit+s* or *cut+s* (but also in a few items like *blitz*)
primarily lexical cluster	cluster type which has exclusively or almost exclusively lexical instances; sometimes measured in type frequency rather than token frequency; also referred to as *lexical strong default*, *prototypically lexical*, or if token frequency is used *high probability*	word final /lk/ as in *bulk* or *milk*
ambiguous cluster	cluster type which many morphonotactic as well as many lexical instances; also referred to as *mid probability* if token frequency is used	word final /ld/ in *call+ed* or *cold* or /nd/ in *bann+ed* or *bind*
lexicality of a cluster	Fraction of lexical instances of a cluster type; also *probability* of a cluster type if token frequency is used	Close to 0 if primarily morphonotactic (English /ts/); close to 1 if primarily lexical (English /lk/); close to 1/2 for a perfectly ambiguous cluster (English /ld/)
ambiguity of a cluster	Similarity of a cluster distribution with a 1:1 distribution of morphonotactic and lexical instances	Close to 1 for a perfectly ambiguous cluster (English /ld/); close to 0 for primarily lexical or morphonotactic clusters (English /lk/ or /ts/)

1.1 Phonotactics

The phonotactics of a given language consists in imposing limitations on or expressing preferences with regard to sound sequences in that language. In this article, we limit the scope of our investigation to consonantal diphones, that is, to sequences of exactly two consecutive consonants. One approach to phonotactics is to look for universal rules (or constraints), whose ordering (or ranking) accounts for cross-linguistic differences as to which sound sequences are licit in particular languages. These rules (or constraints) can be formulated with regard to the syllable as the domain of their application (Kahn, 1976), with regard to strings (Steriade, 1999), or with regard to both strings and syllables (Albright, 2015). A different approach (e.g. Vennemann, 1988; Dziubalska-Kołaczyk, 2014), whose predictions for processing in speech perception will be tested here, is not to determine the legal sound sequences for a language, but to formulate preferences for particular sound sequences. The point of departure is the assumption that all consonantal diphones are in general 'dispreferred'. Various observations are used to support this notion, including typology (e.g. consonantal diphones in syllable codas are allowed in fewer than 21% of the world's languages; Donohue et al., 2013), casual speech phenomena (diphones are reduced in fast speech; Dziubalska-Kołaczyk and Zydorowicz, 2014) and language acquisition (they are acquired late; Levelt and Vijver, 1998, 2004; Jarosz et al., 2016).

While all consonantal diphones are dispreferred compared to singletons, they are said to differ as to the degree to which they are so. This approach, then, does not categorize sound sequences as licit or illicit, but instead ranks the observed sequences with respect to their 'preferability'. This is done with regard to proposed universal preferences regarding the distance (in place and manner of articulation) between the members of the consonantal sequence and the neighboring vowel or vowels. For example, for medial diphones, which are the focus of this paper, a diphone is preferred if the distance between the two consonants is less than or equal to the distance between each consonant and its neighboring vowel (Dziubalska-Kołaczyk, Pietrala, & Aperliński, 2014). Crucially, while preferability and frequency are related, they cannot be equated. Preferred diphones are expected to be, or become, frequent, but there are are other, e.g. lexico-grammatical and pragmatic factors influencing a diphone's frequency, besides its preferability.

A considerable amount of research has been devoted to the influence of phonotactics on speech segmentation, i.e. on spotting words in the speech stream. Diphones with a high frequency of occurrence between words and a low frequency of occurrence within words have been repeatedly found to help segmentation, both when listeners are infants (Mattys et al., 1999; Mattys and Jusczyk, 2001) and adults (van der Lugt, 2001). Daland and Pierrehumbert (2011), having tested learning models on speech corpora, show that phonotactic knowledge (here: phrase-medial word boundaries) is learnable given the input that infants typically receive. In contrast to the research on segmentation, however, we are looking at word-internal diphones only, and taking into account the probability with which they occur within or across morphemes rather than within or across words.

1.2 Phonotactics vs. morphonotactics

It has been observed that the phonotactics of a language interacts with its morphology. For example, final consonantal sequences in English words allowing four members are exclusively non-monomorphemic, e.g. *six+th+s, glimpse+d* (Cruttenden, 2014, p. 262). The interaction of morphology and phonotactics has been the focus of a proposed theory of ‘morphonotactics’ (Dressler and Dziubalska-Kołaczyk, 2006). Here, the very fact that consonantal diphones spanning morphemic boundaries are ranked low on the preferability scale is actually argued to be their strength. Dispreferred diphones are claimed to signal morphological complexity better through their status as ‘dispreferred’. A diphone which is dispreferred stands out in semiotic terms, the argument goes, and thus may be an indication of a morphological operation having taken place.

1.3 The Strong Morphonotactic Hypothesis

While some consonantal diphones - ‘purely morphonotactic’ ones - always occur across morpheme boundaries in a given language (e.g. ENG /md/ as in *seem+ed*) and others - ‘purely lexical’ ones - always occur within morphemes (e.g. ENG /mp/ as in *lamp*), yet others might occur both across and within morphemes, and so can be seen as ambiguous. Ambiguous diphones differ as to the degree of their ambiguity, i.e. the relative frequency with which they occur across and within morphemes. And so clusters such as ENG /ts/ are morphonotactic by strong default, as they usually (e.g. *cat+s*) occur across morpheme boundaries, though not always (e.g. *waltz*). There are diphones which act as morphonotactic and lexical roughly equally frequently (e.g. ENG /ld/ as in *call+ed* and *cold*). Finally, there are clusters such as ENG /nd/, which are lexical by strong default as they usually (e.g. *hand*) occur within morphemes, though also occur across morpheme boundaries (e.g. *bann+ed*). Table 1 provides an overview of the morphonotactic terminology adopted in this paper.

Thus, going from purely morphonotactic diphones, through the three categories of ambiguous diphones all the way to purely lexical diphones, a lexicality scale can be formed - see Table 2. We will consider diphones in category 3 as maximally ambiguous, diphones in category 2 or 4 as less ambiguous, and diphones in category 1 or 5 as least ambiguous. Note that the frequency measure used to determine ambiguity is not a priori clear. Indeed, Dressler and Dziubalska-Kołaczyk (2006) left the question of whether type frequencies (number of word types a diphone occurs in) or token frequencies (number of diphone instances) should be employed as an open question.[1] We will account for both frequency measures in our analysis.

Table 2. The lexicality scale (morphonotactic - ambiguous - lexical)

1	2	3	4	5
Morphonotactic	Strong default	Equally frequent	Strong default	Lexical
/md/ *seemed*	/ts/ *cats, waltz*	/ld/ *called, cold*	/nd/ *banned, hand*	/mp/ *lamp*

Dressler and Dziubalska-Kołaczyk (2006: 83) postulate the following hypothesis with regard to the relationship between the position of a given diphone[2] on the lexicality scale and its

[1] We would like to thank an anonymous reviewer for raising this issue.
[2] Consonantal diphones are, as in (1) above, often referred to as ‘consonant clusters’, or simply ‘clusters’. Since the question of whether clusters are restricted to being contained within syllables is under debate, we prefer the more neutral term ‘diphone’. Whenever we use the term cluster in this paper it simply denotes ‘consonantal diphone’.

ability to signal morphological complexity, which has come to be known as the Strong Morphonotactic Hypothesis (SMH; cf. Korecky-Kröll et al., 2014; Calderone et al., 2014; Ritt & Kaźmierski, 2015), although Dressler and Dziubalska-Kołaczyk (2006) did not actually coin this term in this very paper:

(1) Strong morphonotactic hypothesis (SMH):
   a) "Prototypical morphonotactic clusters [...] have the function of co-signaling the existence of a morphological rule [i.e. presence of a morphological operation],"[3]
   b) "morphonotactic default clusters [...] fulfill this [signaling] function less adequately,"
   c) "while phonotactic clusters [...] cannot fulfill this [signaling] function [...]" (Dressler and Dziubalska-Kołaczyk 2006: p. 83)

The SMH figures centrally in morphonotactic research, and numerous attempts have been made to test it drawing on data from language acquisition (Freiberger et al., 2011), diachronic linguistics (Dressler et al., 2010), experimental research (Korecky-Kröll et al., 2014; Leykum et al., 2015a), or by means of computational modeling (Calderone et al., 2014). The authors of these studies, however, have not always tested the same hypothesis, as it seems.

This deserves to be elaborated on in more detail. The SMH as phrased in the quote above implies (a) that clusters that span a morpheme boundary (i.e. morphonotactic clusters) have the semiotic function of signaling that boundary and (b) that the success at which morphonotactic clusters signal morpheme boundaries decreases in their degree of lexicality as shown in Table 2 (cf. solid line in Figure 1a below). Morpheme-internal clusters (phonotactic or lexical clusters) - trivially - lack this function, i.e. (c).

Notably, part (a) and (c) of the SMH do not directly assert anything about whether or not this signaling function exhibits some beneficial effect on the processing of morphonotactic clusters as opposed to their lexical counterparts (nor does (b), obviously). This is interesting, because previous studies such as Korecky-Kröll et al. (2014: p. 57) have experimentally[4] investigated the following operational hypothesis:

(2) Operational hypothesis associated with SMH:
"[I]f a certain sequence occurs only over a morpheme boundary and is thus a prototypical morphonotactic sequence, it should be processed more easily than a purely phonotactic sequence"

According to Korecky-Kröll et al. (2014: 57) this operational hypothesis is meant to shed light on "[t]he Strong Morphonotactic Hypothesis, which assumes that phonotactics helps in the decomposition of words into morphemes". Arguments to the same effect can also be found elsewhere. Leykum et al. (2015b: p. 1) who propose that "as an extension of the Strong Morphonotactic Hypothesis [...] morphonotactic clusters are more robust and more

[3] In this definition, morphonotactic clusters are those which arise from any morphological operation rather than just morphological concatenation (e.g. morphologically induced vowel drop between two consonants). In this paper, however, we restrict ourselves to morphological concatenation, so that morphonotactic clusters are equated with clusters spanning a morpheme boundary.
[4] In a series of experiments, participants were asked to find a particular substring of triconsonantal clusters. Response times were significantly lower if a morpheme boundary was present.

highlighted in speech production than phonotactic clusters”.[5] Similarly, Calderone et al. (2014: pp. 59-60) state that

> “[a]ccording to the strong morphonotactic hypothesis [...], speakers use morphonotactic consonant clusters as morphological boundary signals. Morphonotactic clusters are thereby assigned a morphological function in processing [...], which is assumed to facilitate processing and acquisition of complex consonantal structures.”

The underlying rationale is this: morphonotactic diphones have the burden of signaling morpheme boundaries in a confident way in order to be of any help in morphological decomposition. To this end, they must be easily detected (cf. 1.1), and hence they are required to have properties that make them being easily processed in perception (such as beneficial perceptual contrast between segments or longer duration). Notably, these properties must outweigh any cognitive costs that are imposed by the process of morphological decomposition (otherwise there is no reason to expect hypothesis (2) to hold). Consequently, as we infer, ease of processing is expected to be a decreasing function of a diphones lexicality (see solid line in Figure 1b, below).

The argument contrasts with findings from the research on phonotactic signaling of word boundaries outlined in 1.1. It has been shown that diphones which occur word internally (i.e. which are ‘lexically licensed’ and thus belong to the phonotactic inventory) are perceived more easily, produced more accurately, and less likely subject to repair processes than diphones which occur only across word boundaries (e.g. Moreton, 2002, and Berent et al., 2007). If the same mechanisms also apply to morpheme boundaries within words, ease of processing must be an increasing function of a diphones lexicality (dotted line in Figure 1b).

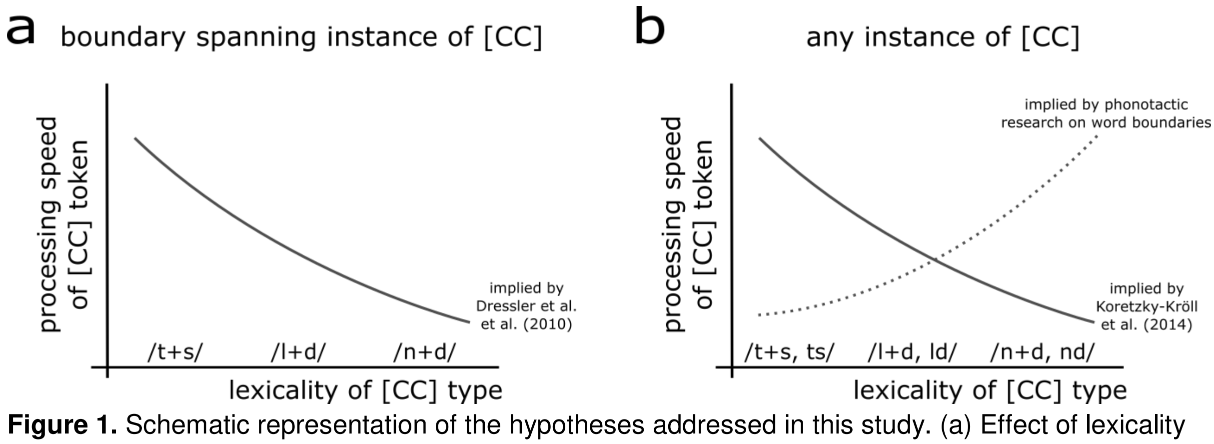


Figure 1. Schematic representation of the hypotheses addressed in this study. (a) Effect of lexicality on the processing of morphonotactic instances of a diphone type. Under one interpretation of SMH (1a), ease of processing is a decreasing function of lexicality (solid line), because diphones signal boundaries less reliably if they also occur morpheme internally. The question is addressed by our first Experiment 1. (b) Under a second interpretation of SMH (2), ease of processing of consonant diphones in general is a decreasing function of lexicality (solid line), because morpheme-boundary signaling diphones need to stand out to be detected easily. Phonotactic research on word boundaries suggests the opposite (dotted line). The question is subject to Experiment 2.

[5] They fail to show that this is the case in an experimental reading task assessing duration and intensity of word final diphones. Note that in their study, the respective articulation of morphonotactic versus phonotactic instances of the very same cluster type is compared against each other.

Clearly, interpretation (2) of the SMH lacks part (1b) in the formulation of Dressler and Dziubalska-Kołaczyk (2006). Ambiguity with respect to signaling morphological complexity is not relevant under this interpretation; it is only lexicality that seems to play a role.

In contrast, although not explicitly referring to the SMH, Freiberger et al. (2011) investigate in a series of visual-target experiments if prototypical morphonotactic clusters are processed faster than morphonotactic default clusters, thus explicitly covering (1b). Likewise, diachronic studies such as Dressler et al. (2010) and Ritt and Kaźmierski (2015) focus on the relevance of ambiguity with respect to signaling morphological complexity to the diachronic stability (i.e. resistance against deletion processes) of clusters. In a neuro-computational simulation study, Calderone et al. (2014) find differences between the representational setup of purely morphonotactic and ambiguous clusters. In doing so, these studies tackle (1b) but they do not test whether exclusively boundary-spanning clusters are processed more easily (or acquired earlier or diachronically more stable), than exclusively morpheme-internal clusters, i.e. (2).

Indeed, the logical relationship between (1a), i.e. that clusters facilitate morphological decomposition, and the operational hypothesis (2) tested in many of the above mentioned studies - as relevant and interesting as it may be in itself - is not entirely clear. For instance, it can be argued that, even though morphonotactic clusters fulfill their function of signaling morpheme boundaries, they are not acquired earlier (cf. Freiberger, 2007) or produced more accurately (cf. Leykum et al., 2015b) than their morpheme internal counterparts, just because morphological processing takes its cognitive toll. Likewise, the very fact that morphonotactic clusters are generally less preferred than lexical clusters from an articulatory and perceptual perspective (cf. Marecka and Dziubalska-Kołaczyk, 2014 and section 1.2) could mean that the processing of morphonotactic clusters is hampered, so that any advantage due to boundary signaling is immediately overridden. Conversely, showing that morphonotactic clusters are detected faster than phonotactic clusters (cf. Korecky-Kröll et al., 2014), i.e. (2), does not immediately entail that the former assist the speaker/hearer in recognizing a morpheme boundary, because it could in principle be the case that morphonotactic clusters are detected earlier just because they are located at very prominent positions (in that sense, facilitated processing of a cluster would be an epiphenomenal consequence of morphological parsing rather than the reverse).

We conclude that comparing the processing of morphonotactic against that of phonotactic clusters does not *a priori* allow for immediate conclusions about the presence of a signaling function in morphonotactic clusters, as proposed in (1a). Rather, we suggest that the ambiguity of morphonotactic clusters in signaling morpheme boundaries as originally proposed in (1b) should be taken as a more reliable diagnostic tool for testing the existence of their signaling effects, i.e. (1a). Clearly, if morphonotactic clusters exhibit a signaling function, then this function is expected to be diminished by ambiguity, so that clusters with differential degrees of ambiguity should also show differential degrees of ease of processing. This follows from basic principles of semiotics (Peirce, 1965). By contraposition, the absence of differences in processing among clusters with differential degrees of ambiguity renders compositionality signaling as dominant factor in the processing of morphonotactic clusters unlikely.

This stresses the relevance of an approach that explicitly incorporates ambiguity with respect to signaling morpheme boundaries as an explanatory factor. We do so in two slightly different experiments. Importantly, the differential design of these experiments allows for addressing both the hypothesis that the processing of morphonotactic clusters is diminished by ambiguity (Figure 1a), as well as the hypothesis that the processing of clusters in general is influenced by their degree of lexicality (Figure 1b). In the analysis of our experiments, we will employ both type and token frequencies of diphones to assess their degree of ambiguity.

2 Research questions

We will make use of the following terminology (see also Table 1). For our study, we conceptualize the 'degree of lexicality' of a cluster type as the amount of phonotactic instances among all instances of that type (i.e. phonotactic instances plus morphonotactic instances). The higher the degree of lexicality, the more phonotactic instances there are for a given cluster type. Note that we focus on the investigation of morphonotactic clusters with variable degrees of lexicality. This is why we obviously do not include purely phonotactic clusters, although our data set includes cluster types that surface morphonotactically extremely rarely (cf. Table A1). As a consequence, 'ambiguity' is largest if there are as many lexical as morphonotactic instances of a cluster type. We investigate two slightly different research questions:

(3) Research questions:
   a) Are there differences in how quickly cross-morphemic consonant-diphone instances with variable degrees of lexicality are processed in speech perception?
   b) Are there differences in how quickly instances of consonant-diphone types with variable degrees of lexicality are processed in speech perception?

The difference between these two research questions is very subtle. Question (3a) is about the processing of morphonotactic clusters (cf. Figure 1a), while (3b) is about the processing of clusters in general (Figure 1b). While the items tested in the latter question are cluster types that could be classified as primarily morphonotactic clusters (e.g. ENG /md/), ambiguous clusters (/ld/) or primarily lexical clusters (/nd/), the former question is about the processing of cross-morphemic instances of primarily morphonotactic clusters (/md/ in *seem+ed*), ambiguous clusters (/ld/ in *call+ed*) and of primarily lexical clusters (/nd/ in *bann+ed*).

The reason why we have chosen this set of research questions is twofold. First and foremost, answering (3b) will allow us to evaluate whether the experiment that addresses (3a) really captures the processing of morphonotactic instances of cluster types. In this sense, (3b) functions as a control hypothesis. If there is no difference between the respective outcomes of the experiments, then our experiment obviously failed to address morphonotactic clusters specifically. The second reason is that while (3a) directly refers to (1b) and is thus of major relevance to our study, (3b) relates to the operational hypothesis (2) mentioned in 1.3, namely that morphonotactic cluster tokens should on average be processed faster than phonotactic ones. Clearly, (3b) is not exactly the same since by design we do not actually consider purely phonotactic cluster types. However, we think that (3b) is nevertheless an interesting extension of what has been tested frequently in morphonotactic research (Korecky-Kröll et al., 2014). Moreover it relates to within-type comparison studies of the processing of morphonotactic and phonotactic clusters (cf. Freiberger, 2007; Leykum et al., 2015b) because it can be argued that if morphonotactic instances of some cluster type are on average processed faster than phonotactic instances of that type (i.e. hypothesis (2)), then, everything else being equal, (a random representative of) a less lexical cluster type should be on average processed faster than (a random representative of) a more lexical cluster type (i.e. the scope of (3b)). Whether or not (2) or (3b) are actually related to the SMH (1) is a different question, although not an uninteresting one, as we have pointed out in the previous section.

We are able to compare these two research questions with the differing experimental setups in Experiment 1 and Experiment 2 below, the former addressing (3a) and the latter addressing (3b). They shall be described in more detail in the following.

3 Experiment 1

3.1 Methods

3.1.1 Materials

A set of token-wise equally frequent Polish consonant diphones differing in their variability of occurrences within morphemes and across morpheme boundaries have been selected. These were the following medial sequences: /ʂk, lk, ɕm, vn, ʐn, kw, ɕɲ, lɲ, ɕl, zn/. Variability was operationalized by determining the fraction of morpheme internal occurrences for each diphone type (in terms of token frequency, see 3.1). Frequency counts were taken from a corpus of spoken Polish collected by one of the authors (Kaźmierski 2015). The particular set of consonantal diphones was chosen for several reasons:

a) They are existing Polish consonantal diphones in order to ensure the familiarity of our Polish participants with them;
b) They differ in their ambiguity (both, in terms of types and tokens), as it is the influence of ambiguity on processing speed that we set out to test;
c) They are roughly equally frequent in order to avoid effects of variable entrenchment (frequency effects);
d) They are all reasonably frequent to ensure that the participants are familiar with both their morphological and lexical instances;
e) The set is sufficiently large in order to include a range of different consonants so as to exclude articulatory bias;
f) We wanted to exclude the length of the sequence as an additional variable, hence only diphones were considered.

The only category of consonantal diphones fulfilling all of the above criteria are word-medial consonantal diphones listed above (Table 1 below lists the 9 cluster types together with their lexicality scores, frequency counts, and other properties).

It is worth noting that the comparably large size and diversity of the set of consonant diphones used in this study could only be achieved in the first place since Polish in general features a huge number of consonant-diphone types (138 initial, 382 medial, and 34 final consonant-diphone types in our underlying corpus of spoken Polish, of which the above 9 types fit the criteria (a-f) above; see also Zydorowicz et al., 2016).

The diphones were embedded in nonce words in order to prevent the token frequency of actual lexical items, as well as the relationship between the frequency of the base and the derived word (cf. Hay, 2001) from affecting the results. The stimuli were recorded by a native speaker of Polish (one of the authors) in the anechoic chamber of the

Centre for Speech and Language Processing at Adam Mickiewicz University in Poznań through a head-mounted condenser microphone (Sennheiser HSP2) plugged into a Roland Duo Capture USB interface. The audio interface was connected to a laptop computer running the Speech Recorder program (Draxler and Jänsch, 2015), used to display the stimuli and automatically save the recorded words to individual, and uniquely named sound files. All recordings were normalized with respect to duration: 600ms.

Table 1. Cluster types together with measures considered in the experiments. For a given cluster, lexical type ratio is the fraction of word types featuring the cluster morpheme internally in medial position among all word types featuring that cluster medially. Lexical probability is the fraction of medial and morpheme internal tokens of that cluster among all medial instances. Numeric measures are derived from Kaźmierski (2015). Note that some diphones feature fractions of 1.00 based on this corpus; however, they do in fact rarely occur across morpheme boundaries. MoA and PoA denote manner and place of articulation, and 1 and 2 the first and second segment of the diphone, respectively. NAD denotes whether the cluster is preferred according to the net auditory distance metric. Preferred items show high articulatory intersegmental contrast (see 3.2.1 for details).

cluster	Lexical type ratio	Lexical probability	Token frequency	MoA1	MoA2	PoA1	PoA2	NAD
ɕɲ	1.00	1.00	84	fricative	nasal	coronal	coronal	yes
ɕl	1.00	1.00	71	fricative	liquid	coronal	coronal	no
ʂk	0.95	0.90	100	fricative	stop	coronal	dorsal	yes
kw	0.88	0.96	85	stop	glide	dorsal	dorsal	no
zn	0.86	0.96	71	fricative	nasal	coronal	coronal	yes
lɲ	0.56	0.94	87	fricative	nasal	coronal	coronal	yes
vn	0.54	0.57	93	fricative	nasal	labial	coronal	yes
lɲ	0.45	0.24	84	liquid	nasal	coronal	coronal	yes
ɕm	0.02	0.01	95	fricative	nasal	coronal	labial	no

3.1.2 Participants

Twenty-two participants took part in Experiment 1. They were native speakers of Polish, undergraduate students at Adam Mickiewicz University in Poznań. They signed consent forms and filled out personal questionnaires. None of the participants reported any speech disorders.

3.1.3 Design

The materials were used for an AX discrimination task. Instructions were presented on a computer screen, auditory stimuli were presented through headphones, and participants responded by pressing keys on a keyboard.[6] The test phase was preceded by a training phase with additional, unrelated items which were superficially similar to the test items that followed. The experiment was implemented in PsychoPy (Peirce, 2007), an open-source Python-based software.

Altogether, the test phase consisted of 110 trials: 90 with test items and 20 with distractors. Each trial began with a pair of actual Polish words in which the respective diphone spans a morpheme boundary (‘priming pair’), and the participant had to make a decision as to whether the two items were the same or different. These responses were not recorded, as the sole purpose of the priming pairs was to induce the processing of the diphone as morphonotactic: the participants were primed with words in which the test diphone spans a morpheme boundary. This was meant to ensure that the processing of morphonotactic items is evaluated, as formulated in research question (3a). Afterwards, a pair of nonce words with the same diphone (‘test pair’) was presented, and the participant had to make a decision as to whether the two items were the same or different. Accuracy and reaction times of correct responses to the nonce word pairs in the test phase were recorded. The reaction time clock was started at the onset of X of AX in each test pair. Thus, participants were exposed to 220 word-pair stimuli including primes. Table 2 illustrates the procedure for one token.

Table 2. An illustration of the experimental procedure for one token of one diphone: /ɕm/ in Experiment 1. The primes were meant to induce the treatment of the diphone in the nonce word as spanning a morpheme boundary, but we do not want to prejudge the issue of whether the diphone really was processed as morphonotactic. In this example, the correct response is ‘different’.

<table>
<tr><td colspan="3">1. Priming pair</td><td colspan="3">2. Test pair</td></tr>
<tr><td colspan="2">Exposure</td><td>Decision</td><td colspan="2">Exposure</td><td>Decision</td></tr>
<tr><td>/ʂliɕ+mɨ/</td><td>/ʂliɕ+mɨ/</td><td rowspan="2">Same or different?</td><td>/iɕmi/</td><td>/ɛɕmi/</td><td rowspan="2">Same or different?</td></tr>
<tr><td>‘we went’</td><td>‘we went’</td><td>-</td><td>-</td></tr>
</table>

3.2 Analysis

Overall, N=1980 responses were collected. A single data point was deleted as it showed a reaction time of almost zero. Of the remaining responses, 1906 (96%) were correct. Mean reaction time was 1.153s (SD=0.349). See Figure 2a for the distribution of reaction times. In the following we describe the statistical analysis of the collected data.

[6] To control for possible influence of participants’ handedness, it was included as a regressor in the analysis. See 3.2.1.

3.2.1 Variables

Two outcome variables were considered in our analysis: reaction time (RT) and accuracy. Response time was measured in seconds and therefore implemented as a continuous variable, thereby considering only those word pairs featuring the same diphones that have been identified correctly, whereas accuracy was measured as a binary variable assuming the values 1 ('similarity correctly identified') or 0 ('similarity not correctly identified', defined as baseline). We are interested in the way in which the ambiguity of a cluster with respect to signaling a morpheme boundary affects processing. As discussed before, there are several options of how ambiguity can be operationalized.

First, we compute `lexical probability` of a diphone type by calculating the fraction $tokens_{lex}/(tokens_{lex}+tokens_{mpt})$, where $tokens_{lex}$ and $tokens_{mpt}$ are the token frequencies of lexical and morphonotactic (i.e. boundary spanning) occurrences of that diphone, respectively. Token frequencies were taken from Kaźmierski (2015). Thus, `lexical probability` assumes scores in the unit interval. Cluster types closer to 1 are high probability clusters, while cluster types closer to 0 are low probability clusters. If the production, perception and processing of a cluster primarily depends on the number of lexemes it occurs in rather than on its utterance frequency (cf. Pierrehumbert, 2016) type frequencies should be considered. Thus we compute the `lexical type ratio` as $types_{lex}/(types_{lex}+types_{mpt})$, where $types_{lex}$ and $types_{mpt}$ are the respective morpheme internal (lexical) and boundary spanning (morphonotactic) type frequencies.

Both measures range from primarily lexical or high probability (score close to 1) to primarily morphonotactic or low probability (score close to 0) with perfectly ambiguous diphones in the middle at 0.5. That is, if a listener is exposed to a perfectly ambiguous cluster she has a chance of 50% to correctly predict the presence of a morpheme boundary. Thus, we operationalize the ambiguity of a diphone in the narrow sense (`ambivalence`) by means of how close the diphone type is to being perfectly ambiguous, i.e. `ambivalence` is defined as $1-|p - 1/2|/2$ where p is either `lexical probability` or `lexical type ratio` (which shall be denoted as `token ambivalence` and `type ambivalence` respectively). A score close to 1 means that a diphone type is very ambiguous, while a score close to 0 means that it is not (i.e. either almost exclusively lexical or morphonotactic, respectively). We will refer to `lexical probability`, `lexical type ratio`, `token ambivalence` and `type ambivalence` as primary predictors. Table 3 gives an overview of these four measures.

Due to the experimental design (AX), the binary variable `condition` (`same/different`) was included as an additional categorical predictor. A number of (potentially) phonologically relevant factors entered our analysis as secondary predictor variables. First, preferability classification based on Net Auditory Distance (`NAD`) was included as a measure of well-formedness of a cluster (Dziubalska-Kołaczyk, 2014). In a nutshell, NAD measures the articulatory difference between segments involved in the composition of a cluster in terms of manner and place of articulation. If this difference is larger than the contrast between the consonantal segments and their neighboring vowels, then a cluster is assumed to be preferred, and dispreferred otherwise. Binary values (`preferable: yes/no`) were computed for all clusters with the 'NAD Phonotactic Calculator' (Dziubalska-Kołaczyk et al., 2014), also considering consonant voicing. Second, token frequencies (`frequency`) were retrieved from Kaźmierski (2015). Third, articulatory features (manner of articulation and place of articulation) of the first (`MoA1`, `PoA1`) and the second consonant (MoA2, PoA2) of the diphone were determined. We opted for a rather rough articulatory classification due to the relatively small number of diphone types in our study: fricative (baseline), liquid, nasal, stop, and coronal (baseline), dorsal, labial. Since phonological proximity of nonce words to actual Polish words can influence reaction times, we included edit distance (edit) between both nonce words in each trial and their closest neighbor as additional covariate. Due to the way in which responses were recorded, handedness (left, right, n/a) was included as an additional factor. Finally, participant was included as a cluster variable (random effect) in our analysis. There were no repeated measures per test item (nonce-word pair) per participant. In hierarchical models, random effects are assumed to be nested (Baayen et al., 2008: 391; West et al., 2015). Consequently no additional random effect was considered.

Table 3. Four different ways of measuring ambiguity in signaling morpheme boundaries.

Measure	Involved frequency measure	Computation	Terminology (0 vs. 1)	Maximally ambiguous score
lexical probability	tokens	$tokens_{lex}/(tokens_{lex}+tokens_{mpt})$	Low probability vs. high probability	1/2
lexical type ratio	types	$types_{lex}/(types_{lex}+types_{mpt})$	Primarily morphonotactic vs. primarily lexical	1/2
token ambivalence	tokens	1-|*p* - 1/2|/2; *p* = lexical probability	Non-ambiguous vs. ambiguous	1
type ambivalence	types	1-|*p* - 1/2|/2; *p* = lexical type ratio	Non-ambiguous vs. ambiguous	1

3.2.2 Calculation

In order to assess the effect of ambiguity in compositionality signaling on reaction time, a generalized additive mixed model (GAMM, Wood, 2006) was fitted to the data. The choice of GAMMs as opposed to (generalized) linear models was crucial, since we did not want to limit our analysis to linear or, more generally, monotone dependencies between lexical probability and the ease of processing consonantal diphones. In GAMMs, continuous variables can be integrated as so-called smooth terms, i.e. curves, allowing for more complicated functional relationships (Wood, 2006).

In a nutshell, smooth terms are composed of several relatively simple functions (so-called ‘basis functions’) which are added up in order to yield a more complicated curved shape which fits well to a given set of data points (hence ‘additive’ model). The composed function is then fit to the data so that its deviation from the data points (i.e. residuals as in conventional regression models) and at the same time the overall curvature (‘wiggliness’) is minimized. The family the basis functions belong to can be specified by the modeler. In our case we selected so-called ‘thin-plate regression splines’ which have the advantage that the modeler does not have to bother about where to place the basis functions (the computational cost incurred by this function family can be neglected given the relatively small sample size in our case). In addition, we allowed smooth terms to vanish ('shrinkage smoother') so that they effectively drop out of a model. In our model selection procedure (to be described below) this is particularly useful since we deliberately kept smooth terms of the respective ambiguity measure in all models. The number of basis functions is then determined automatically during the modeling procedure based on an initial value which can be specified by the modeler. The selected initial number of basis functions was checked with the `gam.check` function in order to avoid overspecification (`mcgv` package; Wood 2006).

As in generalized regression modeling, various link functions and distributional families can be implemented into GAMMs (hence 'generalized'). In the present analysis, we opted for an exponential model with inverse link. First, this transformation (i.e. 1/RT) accounts for the slightly positively skewed distribution of reaction times (see Figure 2a), second, and more importantly, reciprocal reaction time 1/RT can be interpreted as reaction or processing speed (see also Kliegl et al., 2010; Balota et al., 2013; Lo & Andrews, 2015).

Finally, random effects can be implemented as well (hence 'mixed') in order to capture hierarchical data structure, i.e. clustered data such as multiple data points belonging to a single participant. GAMMs allow for complex mixed effects (smoothing over every single cluster in the data). In our case we opted for the GAMM analogue of random intercepts to model participant random effects. All calculations were done in R (R Development Core Team, 2013). GAMMs were computed with the `mcgv` package (Wood 2006, see appendix for details on the R code used).

We employed the following bottom-up nesting procedure, in order to derive the most parsimonious and at the same time the most informative model for each constellation (West et al., 2015), starting with a minimal model in which reaction time only depends on `lexical probability`.[7] Pairs of nested models differing in exactly one predictor variable were compared with the `compareML` function from the `itsadug` package (van Rij et al. 2015). If the larger model was preferred to the smaller model, the latter model was rejected, and retained otherwise. In case of multiple models scoring better, the one with the lowest AIC (also provided by `compareML`) was selected. This procedure was applied iteratively until the model could not be further improved by adding fixed and/or random effects (Model 1.1). With the same procedure, three additional models were computed, one in which `lexical probability` was replaced by `lexical type ratio` (Model 1.2), two in which `lexical probability` was replaced by `token ambivalence` computed via token frequencies (i.e. `lexical probability`; Model 1.3) and `type ambivalence` computed via type frequencies (i.e. `lexicality ratio`; Model 1.4), respectively.

Concerning `accuracy`, ceiling effects could be observed (the number of incorrect responses was extremely low at 4%), which rendered any statistical analyses of this variable unfeasible. Consequently, we will neglect `accuracy` scores for the remainder of this paper.

[7] We are aware that optimal models determined through model-optimization procedures are in general inferior to averaged models generated by multimodel-inference techniques (Burnham & Anderson 2002). However, averaging of generalized additive mixed models is still subject to ongoing research (Grueber et al. 2011). We thus stick with more traditional step-wise model optimization to identify the best model. Nevertheless, we will employ certain methods from the multimodel-inference paradigm in the post-hoc analyses of our results (see 3.2.3)

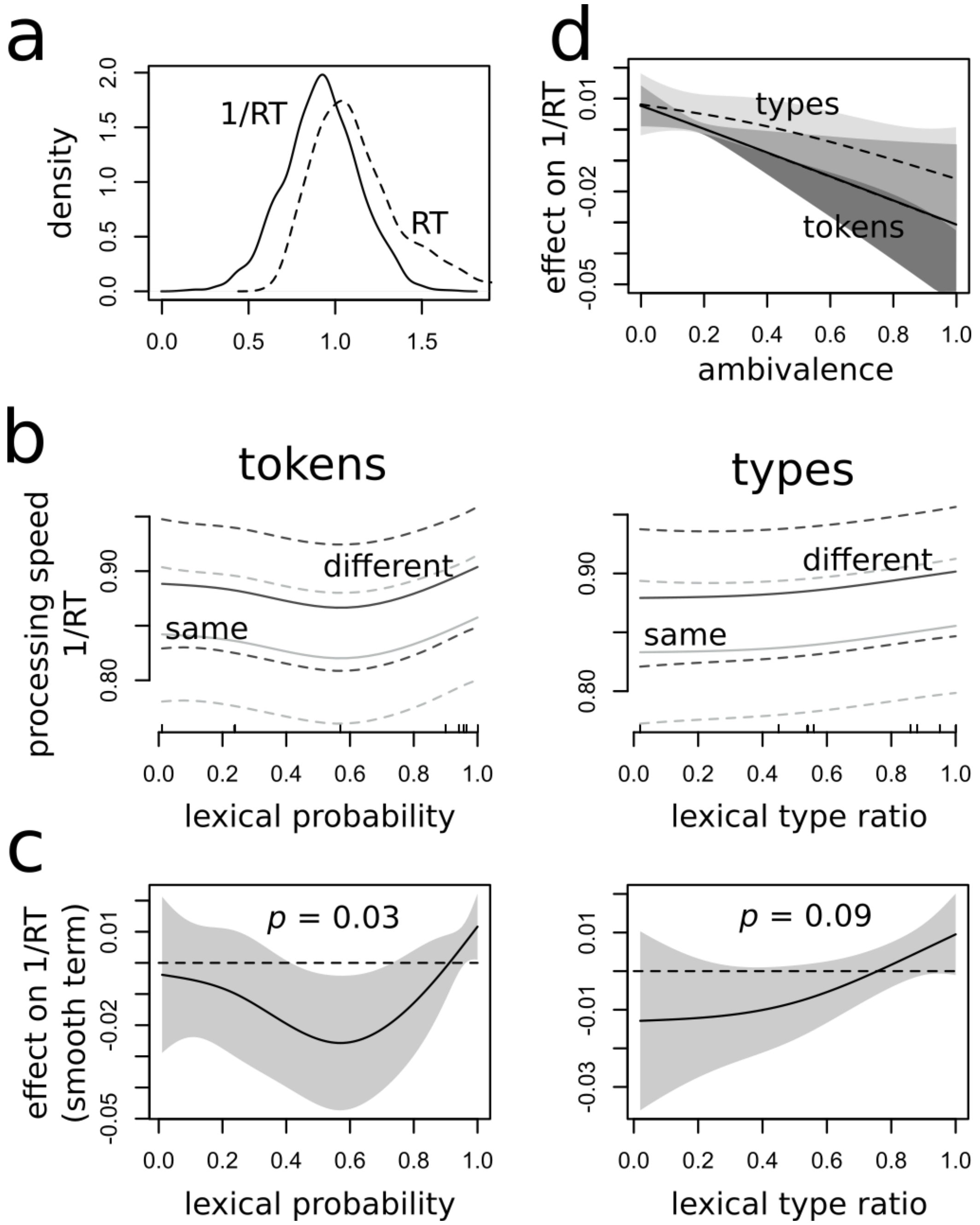


Figure 2. (a) Distributions of reaction time RT and processing speed 1/RT, respectively, the latter being less skewed. (b) Predicted GAMMs of processing depending on lexical probability (Model 1.1 on the left) and lexical probability (Model 1.2 on the right), respectively, as well as condition (same or different nonce words in the stimulus). Stimuli comprised of different nonce words are processed faster. (c) Smooth terms of the effect of lexical probability (left) and lexical type ratio (right) on 1/RT (Model 1.1-2). The shape of the effect of lexical probability resembles a U with items in the mid range being processed significantly slower. (d) Smooth terms of two different ambiguity measures: `token ambivalence` (dashed, dark gray; Model 1.3) and `type ambivalence` (solid, light gray; Model 1.4) affecting 1/RT. Reaction speed decreases significantly in both measures.

3.2.3 Results

Each model resulting from the optimization procedure described in the previous section only contains its respective primary predictor from the list presented in Table 4 (as smooth term), as well as condition (as expected, same word pairs throughout led to significantly faster reaction times than different word pairs, see Figure 2b) and participant (both reaching statistical significance in all cases). All remaining variables turned out not to contribute to the predictive strength of the models and were excluded by the model-optimization procedure.

Let us focus on what is most interesting, namely the primary predictors. Table 4 gives an overview of the most relevant features of Models 1.1-1.4 (see appendix for more details). We find that the effect that lexical probability exerts on 1/RT (significantly non-trivial smooth term at p=0.039; Model 1) exhibits the shape of a U which is significantly different from the null-assumption (i.e. 0 baseline). Diphone types in the middle of the spectrum take longer to be processed than those that surface either within morphemes or across morpheme boundaries (Figure 2c, left). As can be seen from the confidence region, the difference between diphones on the lexical end and those in the middle of the lexical probability spectrum can be classified as more substantial than the difference between the latter and low probability diphones. In contrast, looking at type frequencies (i.e. Model 1.2, lexical type ratio) we find that processing speed mildly increases the more lexical diphones are (marginally significant smooth term at p=0.092, Figure 2c, right) with the steepest slope on the lexical end of the spectrum.

The models in which type/token ambivalence figures as primary predictor show even clearer results. In both cases, ambiguity decreases processing speed. In the case of we token ambivalence we find a significant decreasing linear effect (p=0.010, Figure 1d, solid line), while type ambivalence only yields a marginally significant linear to mildly concave effect on processing speed (p=0.056, Figure 2d, dashed line). Note that the latter curve is persistently less steep than the effect imposed by token ambivalence.

Given that all resulting models predict the same outcome variable (namely processing speed), we can apply post-hoc model-comparison techniques in order to assess which model, and in turn which primary predictor, accounts best for the differences in processing speed. Thus, we derive Akaike weights (Burnham & Anderson, 2002; Burnham et al., 2011) from the respective model AICs, assuming that the set of candidate models consists of Model 1.1 to 1.4. A model's Akaike weight can be interpreted as the probability of the model given the data and all other competing candidate models. Akaike weights for the four models are shown in Table 4 (see brackets in AIC column).

It can be seen that Model 1.3 scores highest and that its probability is more than twice as large as that of the second-best Model 1.1. Models 1.4 and finally 1.4 show a much lower probability. This further corroborates what we have pointed out above: type frequencies are less relevant than token frequencies, and within each way of measuring frequency ambivalence is a better predictor than the fraction of boundary spanning items. Overall, it seems to be token ambivalence which captures ambiguity in signaling morpheme boundaries best.

Table 4. Model overview for Experiment 1. For further details see appendix. Significance code: '*': $p < 0.05$; '°': $p < 0.1$.

	Primary predictor	Significance of primary predictor	Shape of primary predictor	AIC (Akaike weight)	Visualization
Model 1.1	`lexical probability`	p = 0.039 *	U shaped	176.12 (0.27)	Fig. 1b-c (left)
Model 1.2	`lexical type ratio`	p = 0.092 °	Convexly increasing	178.79 (0.07)	Fig. 1b-c (right)
Model 1.3	`token ambivalence`	p = 0.010 *	Linearly decreasing	174.71 (0.56)	Fig. 1d (dashed)
Model 1.4	`type ambivalence`	p = 0.056 °	Slightly concavely decreasing	178.19 (0.18)	Fig. 1d (solid)

3.3 Discussion

In this experiment, we were testing the effect of ambiguity in signaling morpheme boundaries on the processing of diphones that are in fact spanning a boundary. In order to encourage participants to analyze a diphone surfacing in a stimulus (nonce word) as spanning a boundary, primes were first presented to the participants in which diphones signal a morpheme boundary (we will see in the discussion of the second experiment lacking primes that the primes in the first experiment indeed have an effect).

There are two main findings to be discussed. First, ambiguity measures based on token frequency show larger effects on processing than ambiguity measures based on type frequency do. In fact, the latter effects were only marginally statistically significant. As a consequence, this means that the heuristic that listeners rely on in order to analyze whether or not a boundary is present is based on previously encountered utterance frequencies rather than on the number of word types a diphone occurs in. We will come back to this in the conclusion section.

Second, the ease of processing of boundary spanning diphones is a decreasing function of ambiguity rather than a decreasing function of lexical probability. This is evident from the U shape of the effect that lexical probability exerts on processing speed and becomes even clearer when the (linearly) decreasing effect of ambivalence on processing speed is considered. The result is surprising under the assumption that participants actually analyzed diphones as boundary-spanning instances. This is so because low-probability diphone types are expected to provide much worse boundary-signaling cues than those that signal a boundary more often (e.g. 50% of the time).

There are at least two possible explanations to this. First, it could be the case that positive effects on processing imposed by lexical licensing (i.e. the abundant presence of diphones within morphemes) overshadows the negative effects that result from deficient boundary-signaling properties of low probability diphones. We will see in the next section that this possibility can be ruled out because high probability diphones are not processed faster in the absence of boundary-signaling primes.

Second, it is possible that the participants in fact did not always analyze the diphones as spanning a boundary. That is, the prime did perhaps not trigger the decomposition of the subsequently presented nonce word into two parts. Under that interpretation, primes can be assumed to be most successful in low probability diphones. During perception, diphones are either categorized as signaling a boundary or occurring morpheme internally (in spite of the presented primes) but this categorization process is inhibited if the boundary-detection heuristic available to the speaker is not reliable. Consequently, reaction speed is lowest in maximally ambivalent diphones.

Finally, the robustness of our analysis is supported by the fact that neither phonological factors (manner and place of articulation), nor handedness or the proximity of nonce words to existing Polish lexemes (edit distance) contributed to the quality of our models. This indicates that the set of diphone types considered in this study is relatively balanced. Interestingly, the wellformedness metric NAD (Dziubalska-Kołaczyk, 2014) did not contribute to the explanation of differences in RT either. One would expect diphones which are preferred according to the NAD principle to show higher processing speed in our experiment than dispreferred diphones because NAD-preferred items are postulated to have advantages during perception. The possibility remains that the effects of NAD are obscured by that of morphological signaling (or frequency, see Experiment 2).

4 Experiment 2

4.1 Methods

4.1.1 Materials

The stimuli in Experiment 2 were the same as those in Experiment 1 with the sole exception that no existing Polish words were included.

4.1.2 Participants

Thirteen new participants took part in Experiment 2, all of them being native speakers of Polish. Again, none of the participants reported any speech disorders. The number of participants was determined in such a way that there are approximately equally many data points in both experiments. This helped to exclude sample size as a potential explanatory factor of the differences between Experiment 1 and Experiment 2.

4.1.3 Design

The design of Experiment 2 closely matches that of Experiment 1, with one major difference being that Experiment 2 does not include primes (the procedure is presented in Table 5). This control experiment is meant to show whether the priming implemented in Experiment 1 addressing research question (3a) had an effect, and to address research question (3b). Thirteen (new) participants took part in Experiment 2. Alltogether, Experiment 2 consisted of 192 word pairs, among them 150 test pairs and 42 distractor pairs. Recall that Experiment 1 featured 220 word pairs, hence both experiments took roughly the same time in total. Accuracy and reaction times of correct responses to the test pairs were recorded.

Table 5. An illustration of the experimental procedure for one token of one diphone, [ɐm], in Experiment 2.

Test phase		
Exposure		Decision
[iɐmi]	[ɛɐmi]	Same or different?

4.2 Analysis

In total 1950 responses were collected, 1889 (97%) of which were correct. Thus, sample sizes are roughly equal in both experiments (1889 vs. 1906). Mean reaction time was 1.05s (SD=0.344s). The distribution of reaction times in shown in Figure 2a.

4.2.1 Variables

Variables were defined and analyzed precisely as for Experiment 1 (3.2.1).

4.2.2 Calculation

The statistical modeling procedure matches the one presented before in 3.2.2. That is, four models were computed, one for each primary predictor (`lexical probability`, `lexical type ratio`, `token ambivalence`, `type ambivalence`). The analysis of accuracy scores was omitted again due to clear ceiling effects.

4.2.3 Results

As in the analysis of Experiment 1, `condition` and the respective primary predictor survived the optimization procedure in all of the four models. However, none of the effects of the primary predictors reached statistical significance. Table 6 shows the main characteristics of the computed models.

Table 6. Model overview for Experiment 2. For further details see appendix ('n.s.' denotes 'not significant')

	Primary predictor	Significance of primary predictor	Shape of primary predictor	AIC (Akaike weight)	Visualization
Model 2.1	`lexical probability`	*p* = 0.940 (n.s.)	Flat	29.85 (0.30)	Fig. 2b (left)
Model 2.2	`lexical type ratio`	*p* = 0.219 (n.s.)	Flat	31.67 (0.12)	Fig. 2b (right)
Model 2.3	`token ambivalence`	*p* = 0.613 (n.s.)	Flat	29.92 (0.29)	-
Model 2.4	`type ambivalence`	*p* = 0.935 (n.s.)	Flat	29.92 (0.29)	-

As can be seen from Figure 3b-c, none of the ambiguity measures shows a significant impact on processing speed in the absence of boundary-signaling primes. In addition, token frequency turned out to contribute significantly to the quality of all models (in contrast to Experiment 1). That is, in the absence of primes, the effect of morphological structure is overshadowed by that of token frequency although diphone types with roughly equal frequency were selected for the experiments (cf. 3.1.1; notably none of the ambiguity measures reached statistical significance even in the absence of token frequency as a predictor in the model). Interestingly, the effect of frequency on reaction speed turned out to be non-monotonous (mid-frequency items scoring lower reaction speed) rather than strictly increasing. This is exemplarily shown for Model 2.1 in Figure 3c (the effect of frequency displays a similar shape in all other models, 2.2-2.4). A comparison of the respective Akaike weights (see 3.2.3) reveals that there is no clear single best model. Model 2.2 (featuring `lexical type ratio`) shows the lowest probability given the data and the set of four candidate models.

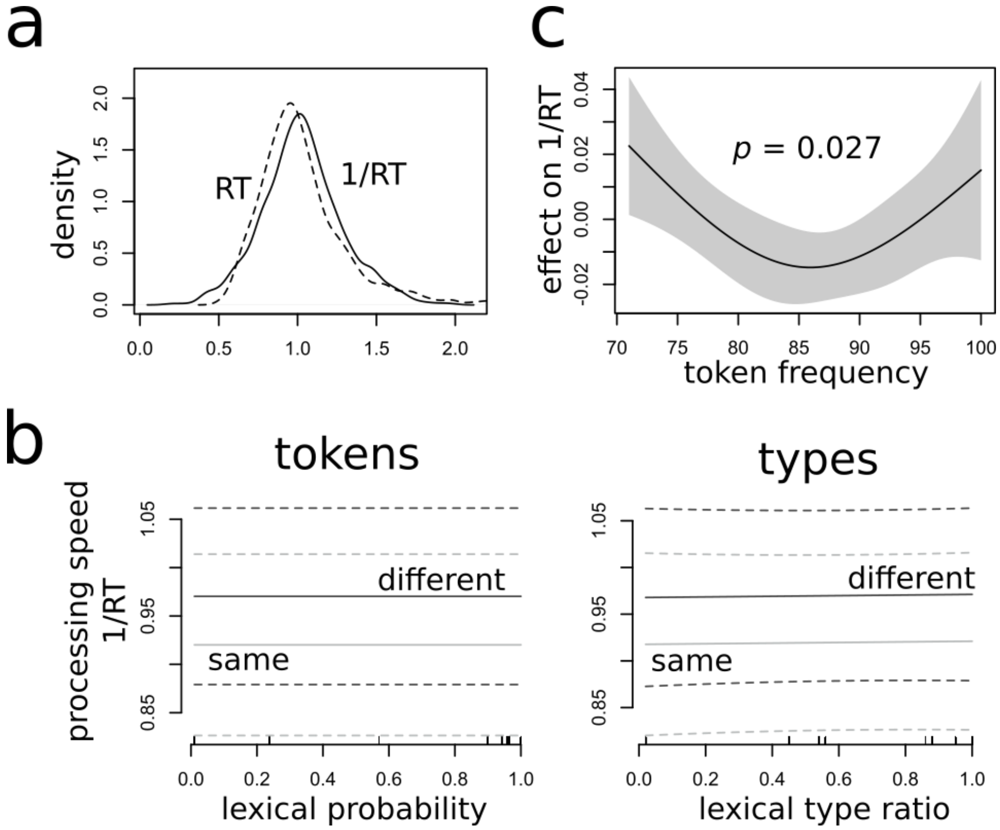


Figure 3. (a) Distributions of reaction time RT and processing speed 1/RT, the latter being slightly less skewed. (b) GAMMs of processing speed depending on ambiguity (lexical probability, Model 2.1, on the left; lexical type ratio, Model 2.2, on the right) and condition. Again, stimuli with different words are processed faster. Neither lexical probability nor lexical type ratio have a significant impact on 1/RT. Models 2.3-4 are not displayed as there is no statistically robust effect either. (c) Significant smooth term of the effect of token frequency on 1/RT in Model 2.1. It has a similar shape in Models 2.2-4.

4.3 Discussion

Our failure to detect any significant effects of ambiguity on processing speed in Experiment 2 can have different causes. First, the collected sample might have been simply too small (although we highlight that sample sizes in both experiments were roughly equal). Second, and more interestingly, the lack of an effect of ambiguity on processing speed might result from the missing primes. Note that since we did not prime participants for detecting a morpheme boundary, we can assume that participants were free to analyze the diphone in the way they preferred. That is, diphones were analyzed more generally as sequences that may or may not span a morpheme boundary. If this is true, then it is not at all surprising that ambiguity does not affect phonotactic processing. Participants simply choose the diphone category (boundary vs. morpheme internal) which suggests itself based on distributional grounds. What the result of Experiment 2 then suggests is that in Experiment 1 participants indeed were encouraged to analyze diphones as spanning a boundary. Thus, Experiment 2 functions as a control experiment, which supports our assumption that the primes in Experiment 1 worked in the way they were meant to.

The results of Experiment 2 do have another consequence, as they help to assess research question (3b). As discussed in Section 2, it has been hypothesized that morphonotactic diphones are processed faster than lexical diphones (Korecky-Kröll et al., 2014). The claim is, that since morphonotactic diphones must confidently signal morpheme boundaries they have the necessity to show properties that facilitate processing in order to be easily detected. Thus, we would expect low-probability/primarily morphonotactic diphone types to be processed faster in our experiment than high-probability/primarily lexical diphone types. Based on our results, this cannot be confirmed.

Any potential effects of boundary signaling are overshadowed by token frequency which turns out to predict reaction speed in a U-shaped manner. Mid-range items are processed more slowly than rare or frequent items. This is interesting, as the effect of frequency on processing speed, if there is one, is rather expected to be strictly positive. The effect of frequency on phonotactic processing, however, is not the focus of our study. Diphone types from a broader frequency range are needed to investigate this matter more thoroughly.

Finally, note that response times in Experiment 2 where on average shorter than those in Experiment 1, which could be seen as evidence against the hypothesis that morphonotactic instances of cluster types are processed faster than their homophonous lexical counterparts (Celata et al,. 2015; Leykum et al., 2014a). However, the differences in design between the two experiments render a direct comparison of response times difficult. We thus remain agnostic with respect to this question.

5 Conclusion

Two propositions that are related to or indeed part of what is generally referred to as the Strong Morphonotactic Hypothesis are present in the morphonotactic literature. The first one is that consonant diphones are processed faster the less lexical they are, and in particular that consonantal diphones which span a morpheme boundary (i.e. morphonotactic diphones) are processed faster than morpheme internal (i.e. lexical) consonant diphones (operational hypothesis (2), cf. 1.3). The second hypothesis is that the compositionality-signaling function of consonant-diphone types decreases the more frequently it is also used morpheme internally as this decreases the reliability at which a diphone signals morphological structure (hypothesis (1b)). In the morphonotactic literature, both hypotheses have been suggested to be linked with the hypothesis that clusters have the function of signaling morpheme boundaries (hypothesis (1a)). In this study, we experimentally addressed both hypotheses ((2) and (1b)).

In order to do so, it was necessary to operationalize ambiguity - and at the same time reliability - of signaling morphological structure. We proposed four different ways of doing so: lexical probability (the fraction of boundary-spanning diphone tokens); lexical type ratio (the fraction of word-types in which a diphone spans a boundary); token ambivalence (the extent to which lexical probability deviates from the most ambiguous configuration); type ambivalence (the extent to which the lexical type ratio deviates from the most ambiguous configuration). This allowed us to assess (a) which type of frequencies based on previous exposure and (b) which corresponding heuristic for measuring ambiguity most relevant to morphonotactic processing.

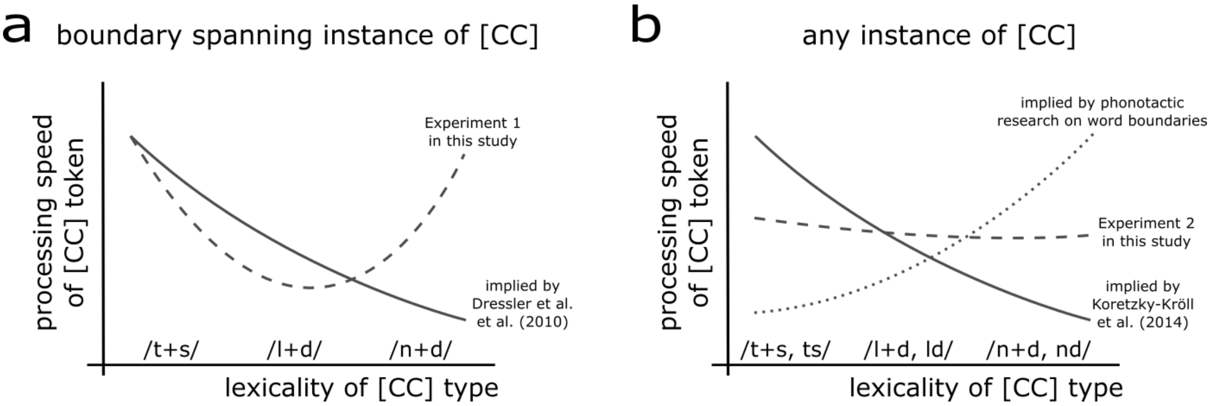


Figure 4. Schematic representation shown in Figure 1 extended by our results (dashed). (a) Research question (3a): from interpretation (1a) of the SMH follows that ease of processing of morphonotactic instances is a decreasing function of lexicality (solid line). Experiment 1 shows that the functional relationship is U shaped, maximally ambiguous clusters scoring the lowest processing speed (dashed line). (b) Research question (3b): interpretation (2) of the SMH implies, ease of processing of consonant diphones to decrease with lexicality (solid line), while phonotactic research on word boundaries suggests the reverse (dotted line). Experiment 2 does not reveal any clear non-trivial relationship (dashed line).

Hypothesis (2) (Figure 4b, solid line) cannot be confirmed by our results. In Experiment 2 we did not reveal any significant effects of a diphone's ambiguity in signaling boundaries on diphone processing (Figure 4b, dashed line). This result is independent of how ambiguity is operationalized. Neither the fraction of boundary spanning word types nor the fraction of boundary spanning tokens showed an effect on reaction speed in our experiment. Thus, as long as there is no morphological processing involved, speakers do not differentiate between diphones which occur always, sometimes, rarely, or never across morpheme boundaries. This contrasts with the findings of Korecky-Kröll et al. (2014). The property of being prone to signaling a boundary alone does not significantly promote a diphone's processing during perception. This goes in line with reported differential effects of morphological structure on the acquisition of consonant diphones (Zydorowicz, 2007; Freiberger et al., 2011). At the same time, our results do not support the hypothesis that low-probability diphones are generally less preferred (i.e. processed more slowly) than their high-probability counterparts if the morphological level is taken into account. If, in contrast, lexical probability is defined as the fraction of word internal items (vs. crossing a word boundary) different pressures seem to apply. On the lexical level, words composed of high-probability diphones are less likely subject to repair processes and hence assumed to be processed faster than words composed of low-probability diphones (Moreton, 2002; see also Vitevich & Luce, 1998, for similar results in nonce words; Figure 4b, dotted line). Thus, it seems that there is a subtle difference between phonotactically guided decomposition of the speech stream into words, and that of words into morphemes, respectively.

Hypothesis (1b) shown in Figure 4a (solid line), and as a consequence likely also (1a), has been partially corroborated. When participants were primed for analyzing a diphone as morphonotactic, they took longer to identify the difference between two words containing that diphone, if it is commonly used ambiguously in speech (rather than either predominantly morpheme internal or spanning a boundary, respectively; Figure 4a, dashed line). Semiotically, this is plausible as signs, and consequently also markers of compositionality, in general tend to avoid ambiguity. In our study, it was token ambivalence which turned out to be the most reliable predictor. Our result has multiple consequences.

First and most fundamentally, it indicates that listeners rely on previously encountered diphone instances when parsing diphones. That is, individuals have detailed memories of diphone instances which also include morphological information, i.e. information about whether or not an encountered diphone has spanned a morpheme boundary. Otherwise, the differential effects of ambiguity on diphone processing observed in our study are hard to explain. Thus, sensitivity to phonotactic probability applies not only on the lexical level as demonstrated previously (Saffran et al. 1998; Vitevich & Luce 1998; McQueen 1998; Adriaans & Kager 2010) but also on the morphological level. This goes very much in line with an exemplar based approach to language perception and production (Pierrehumbert 2001; Wedel 2006; Bybee 2013; Ernestus 2014). Since the diphones in our experiments were embedded into nonce-words this suggests that speakers indeed store exemplars of sublexical units (pace Välimaa-Blum 2009) which carry morphological information, namely whether or not a morpheme boundary is present. This converges with Plag et al. (2017) who demonstrate that properties of sounds (such as duration) indicate the morphological structure of a word. This, as they argue, requires a phonological model which builds on detailed memories of sublexical items, such as those provided by exemplar theory. Finally, this notion also conforms with Calderone et al. (2014) and Celata et al. (2015) who argue that morphonotactic and lexical diphones show differential cognitive representations.

Second, we demonstrated that it is token frequency rather than type frequency which determines whether a diphone is ambiguous. This is interesting as it means that individuals do not necessarily differentiate between multiple lexical types in phonotactic processing. Rather it is overall exposure to a certain sound string which is more relevant. Again, this speaks for relatively self-contained sets of exemplars of sublexical items. However, the relationship between sublexical exemplars and (strings of) lexical exemplars is complicated. On the one hand, the former draw on information about the distributional properties of the latter (because phonotactic knowledge is based on which diphones occur within lexical items). On the other hand, word types are abstracted away during phonotactic processing in favor of a more general classification (boundary vs. no boundary). This entails that mental representations of phonotactic items must be subject to both abstractionist and episodic effects (Adriaans & Kager 2010; Ernestus 2014; Pierrehumbert 2016).

Third, coming back to our primary research question, it is the extent to which the distributional pattern of a diphone type deviates from the most ambiguous configuration (which amounts to tossing a coin) which matters for processing morpheme boundaries. This is interesting, because one would actually expect the probability of spanning a boundary to correlate with processing speed in the presence of boundary spanning primes (cf. solid line in Figure 4a). Why might that be the case? We suggest that the boundary spanning primes in our experiment indeed increased the likelihood of analyzing a subsequently perceived diphone as spanning a boundary as well. However, this likelihood decreases if an encountered item (a boundary spanning diphone) only vaguely fits the previously experienced instances of that diphone type. In the case of a high probability diphone (i.e. a prototypically morpheme internal diphone) this set of previously experienced instances largely consists of morpheme internal items. The consequence of this conflict must be that the listener classifies the encountered item as a member of the more prototypical category, i.e. as being morpheme internal. It is plausible then, that the analysis of diphone types which are maximally ambiguous by representing a nearly equal amount of boundary spanning and morpheme internal instances incurs the highest cost.

Fourth, it is interesting to see that processing speed decreases linearly with ambivalence (i.e. the degree to which an item is ambiguous). This indicates that every additional instance of a diphone type which does not fit to the prototypical usage of that type decreases the quality of the type's signaling function to the same degree. We have demonstrated elsewhere (Baumann & Kaźmierski 2016) by means of computational simulations that the exact shape of the functional relationship between ambiguity and quality of the signaling function of a diphone type can have consequences for the diachronic development of diphone inventories. Strongly convex functional relationships lead to diphone inventories almost lacking any ambiguous diphones while strongly concave relationships promote the stable establishment of some ambiguous diphones. Polish rather belongs to the former category. The hypothesis is that strongly convex relationships characterize languages that make much use of morphology while strongly concave relationships belong to less synthetic languages. The results of the present experiment at least do not contradict this hypothesis in that we did not detect a strongly concave relationship between ambivalence and processing speed in speakers of Polish. Running a similar experiment with speakers of languages that accommodate a larger amount ambiguous diphones, e.g. English (which is at the same time less synthetic than Polish), would be interesting (the conjecture being that having English speakers in the experiment leads to a more concave shape; cf Baumann & Kaźmierski 2016).

One question we cannot address is if there is a general bias towards analyzing ambiguous diphones as morpheme internal. Neither does our second experiment shed any light on this issue (because we cannot be certain as to whether participants classified an encountered item as morpheme internal or boundary spanning), nor does our first experiment do so (because we only provided boundary spanning primes). What would be needed in order to assess whether there is such a bias, is an experimental setup similar to our first one but with morpheme internal instead of boundary spanning primes. We leave this question open for further experimental research on the interaction between phonotactics and morphology.

Appendix

Supplementary data: All collected data together with phonological characteristics and ambiguity measures can be found in the supplementary data files `Experiment_1.csv` and `Experiment_2.csv`.

Notation: In all models reported below, the following significance code applies: '***': $p < 0.001$; '**': $p < 0.01$; '*': $p < 0.05$; '°': $p < 0.1$. Here, the null-hypothesis always corresponds to a trivial (zero) effect (or zero intercept). Relevant abbreviations: '*edf*': estimated degrees of freedom (see below); '*SE*': standard error; '*AIC*': Akaike information criterion; '*N*': sample size (correct instances only). A description of the variables involved can be found in 3.2.1 and 4.2.1.

Remark on GAMM code: Smooth terms are coded as `s()` in the `mgcv` package (Wood 2006), and the specification `bs="ts"` refers to thin-plate spline modeling of terms which may vanish (i.e. shrunk to zero). The parameter `k` specifies the initial number of basis functions ('knots'). For modeling random intercepts in `mgcv`, the procedure of implementing random effects as penalized regression terms with `s(...,bs="re")` in `gam()` was chosen (for that reason, `participant` is listed as smooth term, although this variable is clearly categorical). For details on this see Wood (2015). The abbreviation '*edf*' in the `gam()` output refers to estimated degrees of freedom. If a smooth term corresponding to a one-dimensional continuous predictor shows high overall curvature (i.e. if it is highly nonlinear) then *edf* is high, while *edf* is close to 1 if the term is effectively linear. See 3.2.2 for more information on GAMMs.

Model 1.1: Exponential GAMM with inverse link of RT depending on `lexical probability` and `condition`; `participant` as random effect. $R^2 = 0.228$; $AIC = 176.12$; $N = 1907$. R code: `frm = RT ~ s(lexical_probability, k = 5, bs = "ts") + condition + s(participant, bs = "re"); gam(frm, data = Experiment_1, family = Gamma(link=inverse)).`

Parametric terms			
`intercept`	0.89±0.03*SE*	$t = 32.4$	$p < 0.001$ ***
`condition (same)`	-0.05±0.01*SE*	$t = -4.2$	$p < 0.001$ ***
Smooth terms			
`lexical probability`	*edf* = 2.4	$F = 1.7$	$p = 0.039$ *
`participant`	*edf* = 20.3	$F = 30.4$	$p < 0.001$ ***

Model 1.2: Exponential GAMM with inverse link of RT depending on `lexical type ratio` and `condition`; `participant` as random effect. $R^2 = 0.226$; $AIC = 178.79$; $N = 1907$. R code: `frm = RT ~ s(lexical_type_ratio, k = 5, bs = "ts") + condition + s(participant, bs = "re"); gam(frm, data = Experiment_1, family = Gamma(link = inverse)).`

Parametric terms			
intercept	0.89±0.03*SE*	$t = 32.5$	$p < 0.001$ ***
condition (same)	-0.05±0.01*SE*	$t = -4.2$	$p < 0.001$ ***
Smooth terms			
lexical type ratio	$edf = 1.4$	$F = 1.8$	$p = 0.09$ °
participant	$edf = 20.3$	$F = 30.4$	$p < 0.001$ ***

Model 1.3: Exponential GAMM with inverse link of RT depending on `token ambivalence` and `condition`; `participant` as random effect. $R^2 = 0.228$; $AIC = 174.71$; $N = 1907$. R code: `frm = RT ~ s(token_ambivalence, k = 5, bs = "ts") + condition + s(participant, bs = "re"); gam(formula, data = Experiment_1, family = Gamma(link = inverse)).`

Parametric terms			
intercept	0.89±0.03*SE*	$t = 32.4$	$p < 0.001$ ***
condition (same)	-0.05±0.01*SE*	$t = -4.2$	$p < 0.001$ ***
Smooth terms			
token ambivalence	$edf = 0.88$	$F = 1.4$	$p = 0.010$ *
participant	$edf = 20.3$	$F = 30.4$	$p < 0.001$ ***

Model 1.4: Exponential GAMM with inverse link of RT depending on `type ambivalence` and `condition`; `participant` as random effect. $R^2 = 0.226$; $AIC = 178.19$; $N = 1907$. R code: `frm = RT ~ s(type_ambivalence, k = 5, bs = "ts") + condition + s(participant, bs = "re"); gam(formula, data = Experiment_1, family = Gamma(link = inverse)).`

Parametric terms			
intercept	0.89±0.03*SE*	$t = 32.4$	$p < 0.001$ ***
condition (same)	-0.05±0.01*SE*	$t = -4.2$	$p < 0.001$ ***
Smooth terms			
type ambivalence	$edf = 1.2$	$F = 1.0$	$p = 0.055$ °
participant	$edf = 20.3$	$F = 30.3$	$p < 0.001$ ***

Model 2.1: Exponential GAMM with inverse link of RT depending on `lexical probability`, `frequency` and `condition`; `participant` as random effect. $R^2 = 0.196$; $AIC = 29.85$; $N = 1889$.
R code: `frm = RT ~ s(lexical_probability, k = 5, bs = "ts") + condition + s(frequency, k = 5) + s(participant, bs = "re"); gam(frm, data = Experiment_2, family = Gamma(link = inverse)).`

Parametric terms			
`intercept`	0.98±0.05*SE*	$t = 21.3$	$p < 0.001$ ***
`condition (same)`	-0.05±0.01*SE*	$t = -3.5$	$p < 0.001$ ***
Smooth terms			
`lexical probability`	*edf* < 0.1	$F = 0.0$	$p = 0.94$
`participant`	*edf* = 11.8	$F = 43.7$	$p < 0.001$ ***
`frequency`	*edf* = 2.0	$F = 3.4$	$p = 0.027$ *

Model 2.2: Exponential GAMM with inverse link of RT depending on `lexical type ratio`, `frequency` and `condition`; `participant` as random effect. $R^2 = 0.195$; $AIC = 31.67$; $N = 1889$.
R code: `frm = RT ~ s(lexical_type_ratio, k = 5, bs = "ts") + condition + s(frequency, k = 5) + s(participant, bs = "re"); gam(frm, data = Experiment_2, family = Gamma(link = inverse)).`

Parametric terms			
`intercept`	0.98±0.05*SE*	$t = 21.3$	$p < 0.001$ ***
`condition (same)`	-0.05±0.01*SE*	$t = -3.5$	$p < 0.001$ ***
Smooth terms			
`lexical type ratio`	*edf* = 1.0	$F = 0.0$	$p = 0.88$
`participant`	*edf* = 11.8	$F = 43.7$	$p < 0.001$ ***
`frequency`	*edf* = 2.0	$F = 2.4$	$p = 0.022$ *

Model 2.3: Exponential GAMM with inverse link of RT depending on `token ambivalence`, `frequency` and `condition`; `participant` as random effect. $R^2 = 0.196$; $AIC = 19.92$; $N = 1889$.
R code: `frm = RT ~ s(token_ambivalence, k = 5, bs = "ts") + condition + s(frequency, k = 5, bs = "ts") + s(participant, bs = "re"); gam(frm, data = Experiment_2, family = Gamma(link = inverse)).`

Parametric terms			
`intercept`	0.98±0.05*SE*	$t = 21.3$	$p < 0.001$ ***
`condition (same)`	-0.05±0.01*SE*	$t = -3.5$	$p < 0.001$ ***

Smooth terms			
`token ambivalence`	*edf* < 0.1	$F = 0.0$	$p = 0.61$
`participant`	*edf* = 11.8	$F = 43.7$	$p < 0.001$ ***
`frequency`	*edf* = 2.0	$F = 2.0$	$p = 0.016$ *

Model 2.4: Exponential GAMM with inverse link of RT depending on `type ambivalence`, `frequency` and `condition`; `participant` as random effect. $R^2 = 0.196$; $AIC = 29.92$; $N = 1889$.
R code: `frm = RT ~ s(type_ambivalence, k = 5, bs = "ts") + condition + s(frequency, k = 5, bs = "ts") + s(participant, bs = "re"); gam(frm, data = Experiment_2, family = Gamma(link = inverse)).`

Parametric terms			
`intercept`	0.98±0.05*SE*	$t = 21.3$	$p < 0.001$ ***
`condition (same)`	-0.05±0.01*SE*	$t = -3.5$	$p < 0.001$ ***
Smooth terms			
`type ambivalence`	*edf* < 0.1	$F = 0.0$	$p = 0.93$
`participant`	*edf* = 11.8	$F = 43.7$	$p < 0.001$ ***
`frequency`	*edf* = 2.0	$F = 2.0$	$p = 0.016$ *

References

Adriaans, F., Kager, R., 2010. Adding generalization to statistical learning: The induction of phonotactics from continuous speech. Journal of Memory and Language 62 (3), 311–331. 10.1016/j.jml.2009.11.007.

Albright, A., 2015. Cluster phonotactics require both syllables and strings: Evidence from Lakhota, Manchester.

Baayen, H., 2013. Multivariate Statistics. In: Podesva, R., Sharma, D. (Eds.) Research Methods in Linguistics. Cambridge University Press, Cambridge, pp. 337–372.

Balota, D.A., Aschenbrenner, A.J., Yap, M.J., 2013. Additive effects of word frequency and stimulus quality: The influence of trial history and data transformations. Journal of experimental psychology. Learning, memory, and cognition 39 (5), 1563–1571. 10.1037/a0032186.

Baumann, A., Kaźmierski, K., 2016. A dynamical-systems approach to the evolution of morphonotactic and lexical consonant clusters in English and Polish. Yearbook of the Poznan Linguistic Meeting 2, 115–139.

Berent, I., Steriade, D., Lennertz, T., Vaknin, V., 2007. What we know about what we have never heard: Evidence from perceptual illusions. Cognition 104 (3), 591–630. 10.1016/j.cognition.2006.05.015.

Burnham, K.P., Anderson, D.R., 2002. Model selection and multimodel inference: a practical information-theoretic approach. Springer, New York.

Burnham, K.P., Anderson, D.R., Huyvaert, K., 2011. AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. Behav Ecol Sociobiol 65, 23–35.

Bybee, J.L., 2013. Usage-based Theory and Exemplar Representations of Constructions. In: Hoffmann, T., Trousdale, G. (Eds.) The Oxford Handbook of Construction Grammar. Oxford University Press.

Calderone, B., Celata, C., Korecky-Kröll, K., Dressler, W.U., 2014. A computational approach to morphonotactics: evidence from German. Language Sciences 46, 59–70.

Celata, C., Korecky-Kröll, K., Ricci, I., Dressler, W.U., 2015. Phonotactic processing and morpheme boundaries: wordfinal /Cst/ clusters in German. Italian Journal of Linguistics 27 (1), 85–110.

Cohen, J., 1992. A power primer. Psychological Bulletin 112, 155–159.

Cruttenden, A., 2014. Gimson’s pronunciation of English, 8th ed. Routledge, London.

Daland, R., Pierrehumbert, J., 2011. Learning diphone-based segmentation. Cognitive Science 35, 119-155.

Donohue, M., Hetherington, R., McElvenny, J., Dawson, V., 2013. World phonotactics database. http://phonotactics.anu.edu.au. Accessed September 16th, 2015.

Dressler, W.U., Dziubalska-Kołaczyk, K., 2006. Proposing Morphonotactics. Wiener Linguistische Gazette 73, 69–87.

Dressler, W.U., Dziubalska-Kołaczyk, K., Pestal, L., 2010. Change and variation in morphonotactics. Folia Linguistica Historica 31, 51–68. 10.1515/flih.2010.003.

Dziubalska-Kołaczyk, K., 2014. Explaining phonotactics using NAD. Language Sciences 46, 6–17.

Dziubalska-Kołaczyk, K., Pietrala, D., Aperliński, G., 2014. The NAD Phonotactic Calculator – an online tool to calculate cluster preference in English, Polish and other languages. http://wa.amu.edu.pl/nadcalc.

Dziubalska-Kołaczyk, K., Zydorowicz, P., 2014. The production of high-frequency clusters by native and non-native users of Polish. Concordia Working Papers in Applied Linguistics 5, 130–144.

Ernestus, M., 2014. Acoustic reduction and the roles of abstractions and exemplars in speech processing. Lingua 142, 27–41.

Freiberger, E.M., Abbrederis, C., Luckabauer, F., Stammler, M., 2011. Morphonotaktik in der Sprachverarbeitung: Online- und Offline-Tests mit deutschsprachigen Erwachsenen. Wiener Linguistische Gazette 75, 33–52.

Fruehwald, J., 2017. Generations, lifespans, and the zeitgeist. Lang. Var. Change 29 (01), 1–27. 10.1017/S0954394517000060.

Grueber, C.E., Nakagawa, S., Laws, R.J., Jamieson, I.G., 2011. Multimodel inference in ecology and evolution: Challenges and solutions. Journal of evolutionary biology 24 (4), 699–711. 10.1111/j.1420-9101.2010.02210.x.

Hay, J. 2001. Lexical frequency in morphology: is everything relative? Linguistics 39 (6), 1041-1070.

Jarosz, G., Calamaro, S., Zentz, J., 2016. Input Frequency and the Acquisition of Syllable Structure in Polish. Language Acquisition. 10.1080/10489223.2016.1179743.

Kahn, D., 1976. Syllable-based generalizations in English phonology. Dissertation.

Kliegl, R., Masson, M.E.J., Richter, E.M., 2010. A linear mixed model analysis of masked repetition priming. Visual Cognition 18 (5), 655-681. 10.1080/13506280902986058.

Kaźmierski, K., 2015. Consonant clusters in casual speech in Polish: Testing the predictions of NAD and Beats-and-Binding phonology. Paper presented at the 23rd Manchester Phonology Meeting.

Korecky-Kröll, K., Dressler, W.U., Freiberger, E.M., Reinisch, E., Mörth, K., Libben, G., 2014. Morphonotactic and phonotactic processing in German-speaking adults. Language Sciences 46, 48–58.

Levelt, C., van de Vijver, R. 1998/2004. Syllable types in cross-linguistic and developmental grammars. In: Kager, R., Pater, J., Zonneveld, W. (Eds.) Constraints in phonological acquisition. Cambridge: Cambridge University Press, pp. 204–218.

Leykum, H., Moosmüller, S., Dressler, W.U., 2015a. Homophonous phonotactic and morphonotactic consonant clusters in word-final position. In: Proceedings of Interspeech 2015, 1685-1689.

Leykum, H., Moosmüller, S., Dressler, W.U., 2015b. Word-final (mor-)phonotactic consonant clusters in Standard Austrian German, in: The Scottish Consortium for ICPhS 2015 (Ed.), Proceedings of the 18th International Congress of Phonetic Sciences. Glasgow, UK: the University of Glasgow. ISBN 978-0-85261-941-4. Retrieved from https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0701.pdf

van der Lugt, A. H. 2001. The use of sequential probabilities in the segmentation of speech. Perception and Psychophysics 63 (5), 811–823.

Marecka, M., Dziubalska-Kołaczyk, K., 2014. Evaluating models of phonotactic constraints on the basis of sC cluster acquisition data. Language Sciences 46, 37–47.

Martin, A., 2007. The evolving lexicon. PhD dissertation, Los Angeles.

Mattys, S. L., Jusczyk, P. W. 2001. Phonotactic cues for segmentation of fluent speech by infants. Cognition 78, 91–121.

Mattys, S. L., Jusczyk, P. W., Luce, P. A., Morgan, J. L. 1999. Phonotactic and prosodic effects on word segmentation in infants. Cognitive Psychology 38, 465–494.

McQueen, J., 1998. Segmentation of Continuous Speech Using Phonotactics. Journal of Memory and Language 39, 21-46.

Moreton, E., 2002. Structural constraints in the perception of English stop-sonorant clusters. Cognition 84 (1), 55–71. 10.1016/S0010-0277(02)00014-8.

Nakagawa, S., Cuthill, I., 2007. Effect size, confidence interval and statistical significance: a practical guide for biologists. Biological Reviews 82 (4), 591–605. 10.1111/j.1469-185X.2007.00027.x.

Peirce, C. S. 1965. Collected papers. Cambridge: Harvard University Press.

Peirce, J., 2007. PsychoPy: Psychophysics software in Python. Journal of Neuroscience Methods (162), 8–13.

Pierrehumbert, J., 2001. Exemplar dynamics: Word frequency, lenition, and contrast. In: Bybee, J., Hopper, P.J. (Eds.) Frequency effects and the emergence of lexical structure. Benjamins, Amsterdam [a.e.], pp. 137-157.

Pierrehumbert, J., 2016. Phonological representation: Beyond abstract versus episodic. Annual Review of Linguistics 2, 33-52.

Pitt, M.A., 1998. Phonological processes and the perception of phonotactically illegal consonant clusters. Perception & Psychophysics 60 (6), 941–951.

Plag, I., Homann, J., Kunter, G., 2017. Homophony and morphology: The acoustics of word-final S in English. Journal of Linguistics 53, 181–216.

R Development Core Team, 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Ritt, N., Kaźmierski, K., 2015. How rarities like gold came to exist: on co-evolutionary interactions between morphology and lexical phonotactics. English Language and Linguistics (19), 1-29. 10.1017/S1360674315000040.

Saffran, J.R., Aslin, R.N., Newport, E.L., 1996. Statistical learning by 8-month-old infants. Science (New York, N.Y.) 274 (5294), 1926–1928.

Steriade, D., 1999. Alternatives to the syllabic interpretation of consonantal phonotactics, in: Proceedings of the 1998 Linguistics and Phonetics Conference. The Karolinum Press, pp. 205–242.

Välimaa-Blum, R., 2009. The phoneme in cognitive phonology: episodic memories of both meaningful and meaningless units? CogniTextes 2. 10.4000/cognitextes.211.

van Rij, J., Wieling, M., Baayen, H., van Rijn, H., 2015. itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs: R package version 1.0.1.

Vennemann, Theo. 1988. Preference laws for syllable structure and the explanation of sound change: With special reference to German, Germanic, Italian, and Latin. Berlin: Mouton de Gruyter.

Vitevitch, M.S., Luce, P.A., 1998. When Words Compete: Levels of Processing in Perception of Spoken Words. Psychological Science 9 (4), 325-329.

Wedel, A., 2006. Exemplar models, evolution and language change. The Linguistic Review 23, 247-274.

West, B., Welch, K., Gałecki, A., 2015. Linear mixed models. CRC Press, Boca Raton, Florida.

Wieling, M., Nerbonne, J., Baayen, H., 2011. Quantitative Social Dialectology: Explaining Linguistic Variation Geographically and Socially. PloS one. 10.1371/journal.pone.0023613.

Wood, S., 2006. Generalized Additive Models: an introduction with R. Chapman & Hall/CRC, Boca Raton, Florida.

Wood, S., 2015. Random effects in GAMs. https://stat.ethz.ch/R-manual/R-devel/library/mgcv/html/random.effects.html

Zydorowicz, P., 2007. The acquisition of Polish morphonotactics. Wiener Linguistische Gazette Online 74, 24–44.

Stabilizing determinants in the transmission of phonotactic systems: diachrony and acquisition of coda clusters in Dutch and Afrikaans

Andreas Baumann & Daan Wissing

Abstract
The phonotactic system of Afrikaans underwent multiple changes in its diachronic development. While some consonant clusters got lost, others still surface in contemporary Afrikaans. In this paper we investigate to what extent articulatory difference between the segments of a cluster contribute to its successful transmission. We proceed in two steps. First, we analyze the respective effects of differences in manner of articulation, place of articulation and voicing on the age at which a cluster is acquired by analyzing Dutch acquisition data. Second, we investigate the role that these articulatory differences play in the diachronic frequency development from Dutch to Afrikaans. We demonstrate that large differences in manner of articulation between segments contribute to a cluster's success in acquisition and diachrony. In contrast, large differences in place of articulation have impeding effects, while voicing difference shows a more complicated behavior.

Keywords: Dutch/Afrikaans phonotactics, articulatory difference, first-language acquisition, diachronic change

1 Introduction

In its history, the sound system of Afrikaans underwent a number of changes. A considerable amount of these changes is related with the way in which sounds are combined to strings in speech, i.e. with the language's phonotactics. Of course, not all combinations of sounds – or more precisely: phonemes – are permitted in speech. For instance, no Afrikaans word ends in /ʃn/, a sequence that is for instance perfectly fine in (Austrian) German, as in /d̥e:tʃn/ (*Tetschn*, 'slap in the face'), nor is there an Afrikaans word ending in /rm/. The latter was not always the case. For instance, words like *arm* ('arm'), *skerm* ('screen') or *wurm* ('worm') were (and still are) articulated with a final /rm/ sequence in Dutch while the two phonemes are separated in contemporary Afrikaans by a neutral vowel so that they end in /rəm/. Similarly, some consonant sequences have been reduced in the history of Afrikaans, so that words that originally ended in /xt/ like (historical and contemporary) Dutch *nacht* ('night') or *specht* ('woodpecker') are pronounced with a single word-final consonant in contemporary Afrikaans (i.e. /nɑx/, *nag*, and /spɛx/, *speg*, although *houtkapper* is admittedly more widely used in Afrikaans; see Figure 1).

However, although obviously some phonotactic restructuring took place in the diachrony of the Afrikaans phonotactic system, not all sequences of consonants (or 'consonant clusters') have been affected by deletion processes. Still a large number of consonant pairs surface in Afrikaans speech, /rx/ (in *berg*, 'mountain'), /nt/ (in *kind*, 'child'), /rt/ (in *boord*, 'edge') and /ls/ (in *dikwels*, 'often') being just a small selection of examples offered by the consonant-cluster inventory of Afrikaans.

Figure 1. Irrespective of whether or not these animals are fine with it, woodpeckers and flatworms underwent cluster-deletion processes in the history of Afrikaans: *specht* > *speg* ('woodpecker') by consonant elimination from /xt/ to /x/ and *worm* > *wurm* ('worm') by schwa epenthesis from /rm/ to /rəm/. But why?[1]

The question of why some clusters have been deleted in Afrikaans diachrony, while others are still abundantly used – and thus diachronically more stable – figures centrally in this paper. More precisely, we consider which articulatory factors might determine the diachronic stability of word-final consonant clusters in Afrikaans. Consonants are conventionally described by three primary articulatory features:

1. Place of articulation (PoA): where does the obstruction occur in the vocal tract?
2. Manner of articulation (MoA): how tight is the obstruction?
3. Voicing: do the vocal folds vibrate?

These features, naturally, play a crucial role when consonants are paired to form strings, i.e. phonotactic items. There are reasonable arguments for the assumption that clusters whose segments differ with respect to some of these articulatory features are more easily processed and, as a consequence over multiple production-and-perception cycles, diachronically more stable. Likewise, there are reasons for assuming on the contrary that phonotactic items are particularly successful if the segments they are composed of mesh with each other with respect to manner, place or voicing (cf. discussion in 1.2).

The aim of this paper is to shed light on the differential effects that manner-of-articulation difference, place-of-articulation difference and difference in voicing exert on the diachronic stability of phonotactic items in Afrikaans. Thus, we seek to explore which of these factors represent relevant determinants in the emergence of the Afrikaans inventory of sequences of two consonants. We address this aim in two separate studies. First, the differential impact of articulatory distances in the acquisition of Dutch is assessed. In a second study, we investigate the effects of articulatory distances on the diachronic success of consonant clusters in the transition from Dutch to Afrikaans. The underlying argument goes like this. If a certain articulatory difference, say, difference in voicing between the segments of a cluster has beneficial effects in phonotactic acquisition, then similar effects should apply diachronically,

[1] Photographs taken from Wikimedia Commons (2005); Wikimedia Commons (2012) and modified (phonological transciptions).

so that differentially voiced clusters are on average diachronically more stable than those clusters in which the segments share the same voicing pattern.[2] Altogether, we show that clusters benefit from large intersegmental manner-of-articulation differences and are restrained by large place-of-articulation differences, while voicing place a more complicated role in the acquisition and change of word-final phonotactics in Afrikaans.

The paper is structured as follows. First, the history of Afrikaans phonotactics is recapitulated (1.1), followed by a more detailed discussion of articulatory differences in phonotactic production and perception (1.2). Subsequently, the link between language acquisition and change is discussed, thereby particularly focusing on the phonological domain. In the end of the first section, the research hypotheses to be investigated in this paper are brought forth (1.4). The introductory section is followed by a detailed description of the two studies which assess the above sketched research questions. Section 2 describes the acquisition study of Dutch phonotactics, while section 3 presents the (historically) comparative study meant to evaluate the diachronic stability of phonotactic items. Finally, the respective results are compared to each other and discussed (4).

1.1 Processes in phonotactic change: from Dutch to Afrikaans

We first discuss the processes that shaped Afrikaans phonotactics in its history originating in Dutch. We focus on word-final phonotactics, which implies that we restrict ourselves to dynamics in the coda. As all Germanic languages, Dutch features (and has already featured in the past) complex consonant phonotactics in the coda, mostly clusters consisting of two consonants such as /sp/ in *wesp* 'wasp' or /xt/ in *nacht* 'night'. In its evolution from Dutch, starting in the 17th century, Afrikaans coda phonotactics underwent a number of processes that lead to the loss but also to the emergence of cluster types.

Leaving changes of the quality of a consonant aside, we can distinguish between two major types of processes that are relevant to coda phonotactics: (a) deletion and (b) insertion processes. In word-final V(C)CC structures, both of them can apply either (i) in the end of the cluster or (ii) cluster medially. In the diachronic development of Afrikaans, all of the four logically possible changes have occurred as shown in Table 1:

Table 1. Coda processes in the diachrony of Afrikaans (Conradie 2017; Donaldson 1993; Roberge 2002)

	Process	Example
(ai)	final t-deletion	/xt/ > /x/ in *nacht* > *nag* ('night')
(aii)	medial t-deletion	/xts/ > /xs/ in *slechts* > *slegs* ('only')
(bi)	final t-insertion	/n/ > /nt/ in *oven* > *oond* ('oven')

[2] This is expected, because the same cognitive mechanisms apply to language learners and proficient speakers and because linguistic knowledge of the latter clearly depends on the process of language acquisition, if only by the differently strong entrenchment of early and late acquired linguistic items. We argue that the question if children or adults contribute more to linguistic change is secondary in this respect, but see 1.3 for a discussion.

(bii)	medial p-insertion; medial schwa-insertion	/mt/ > /mpt/ in *hemt* > *hempt* > *hemp* ('shirt'); /rm/ > /rəm/ in *worm* > *wurm* ('worm')

Let us go through (ai) to (bii) in more detail. Final t-deletion (ai), also t-apokope, is a process that, according to van Veen (1964), is known since the Middle Ages. It first occurred in and around Utrecht in the Netherlands and later also in the vernacular of Dutch cities in the west of the Netherlands. It is generally assumed to have been transferred to the Cape, where it developed further, while since then being reversed in Standard Dutch (Schouten 1982). In Afrikaans, the process is context sensitive and only applies to word-final VCC structures in which the first consonant is an obstruent (Conradie 2017). This is illustrated in Table 2, which lists all obstruent-obstruent clusters in Dutch and Afrikaans. The process does not apply when the first consonant is a sonorant, e.g. /lt, rt, nt/ in Dutch and Afrikaans *asfalt* ('asphalt'), *hart* ('heart') or *land* ('land'; due to final devoicing), respectively [Author], although Ponelis (1989) argues that general stop-deletion after sonorants was productive in early Afrikaans due to Malay influence: e.g. /mp/ > /m/ in *lamp* 'lamp', /lt/ > /l/ in *wild* 'wild' or /rt/ > /t/ in *boord* 'board'. According to Kotzé (1984), however, this contact induced process is restricted to clusters in which the first segment is either a nasal or /l/ (e.g. in /mɔnt/, *mond* 'mouth'; /xəstamp/, *gestamp* 'bumped'; /xɛlt/, *geld* 'money'). The latter process still occurs in the Cape region. Beyond that it is seen sporadically in unstressed syllables, e.g. /rt/ > /t/ in *mosterd* 'mustard' (Ponelis 1989: 4). Note that the dental stop still surfaces in plural forms like *nagte* /naxtə/ ('nights') when the /t/ occurs in the onset of the final syllable (Watermeyer 1996).

Table 2. Word-final obstruent-obstruent clusters in Dutch and Afrikaans. Dutch examples are taken from Linke (2017). Sporadically unreduced types are indicated with a dagger (†).

	Dutch		Afrikaans	
	example	translation	example	translation
/pt/	*concept*	'concept'	—	—
/ps/	*rups*	'caterpillar'	*raps*	'slightly'
/ts/	*muts*	'hat'	*flits*	'torch'
/tʃ/	*kitsch*	'kitsch, junk'	*kitsch*	'kitsch, junk'
/kt/	*pact*	'pact'	*pakt*†	'pact'
/ks/	*heks*	'witch'	*heks*	'witch'
/ft/	*kaft*	'cover'	—	—
/fs/	*vergeefs*	'vainly'	*vergeefs*	'vainly'
/sp/	*wesp*	'wasp'	*wesp*	'wasp'
/st/	*beest*	'beast'	—	—
/sk/	*kiosk*	'stall'	*grotesk*	'grotesque'
/xt/	*macht*	'power'	*agt*†	'eight'
/xs/	—	—	*slegs*	'only'

In contrast to final position, t-deletion within a coda cluster (aii), necessarily VCCC, to our knowledge only occurs in /xts/ > /xs/ (*slechts* > *slegs*, 'only') and it is debatable if this process merely represents final t-deletion together with *-s* suffixation (the latter likely being an adverbial derivational suffix). This boils down to the question of whether or not *sleg+s* (derived from *sleg*, 'bad'; originally in Dutch *slecht* 'bad' but formerly also 'simple') is still morphologically transparent.

Final t-insertion (bi), or t-paragoge, applies sporadically in forms like *oond* from *oven* ('oven'), *reent* from *regen* ('rain') or *behoort* from *behooren* ('belong') (Conradie 2017). It seems to apply only before sonorants, which is consistent with the context sensitive t-deletion process discussed before.

Finally, there are some interesting processes of epenthesis (bii) involving consonants and schwas. On the one hand there is p-insertion in final /mt/ clusters to yield /mpt/, e.g. in *hemt* > *hempt* > *hemp* ('shirt'), a process which likely is articulatorily motivated, the opening of the lips in the transition from bilabial /m/ to /t/ functioning as a release of the bilabial stop /p/ (Conradie 2017). Similar processes can be observed in (Austrian) German in *Hemd*, /hɛmpt/, 'shirt' or *Samt*, /sɑmpt/, 'velvet'. More commonly, however, is schwa-epenthesis in clusters in which the first segment is a liquid /r, l/ and the second segment a nasal /n, m/, e.g. in *worm* /vɔrm/ > *wurm* /vʌrəm/ ('worm') or *psalm* /psalm/ > / pəsaləm/ ('psalm', here also in the onset) (Donaldson 1993). Donaldson (1993) points out that schwa-epenthesis does not apply to liquid-stop clusters, a process which occurs sporadically in Dutch (e.g. /rk/ > /rək/ in *kerk*, 'church'). We suspect that the latter represents a more recent development in Dutch (Kuijpers, van Donselaar and Cutler 2000), so that a process of checked schwa deletion /rək/ > /rk/ should not be assumed for Afrikaans.

From what has been covered so far, it becomes clear that the phonotactics of Afrikaans underwent numerous changes. That is, some phonotactic items, i.e. strings of consonants in our case, got lost while others have emerged during the past couple of centuries. In the subsequent section, we discuss the articulatory and perceptual factors that influence the historical stability of a phonotactic item, in order to better understand the motivation behind the phonotactic processes (ai) to (bii) discussed above.

1.2 Articulatory and perceptual determinants in phonotactic stability

A phonotactic item, e.g. a consonant cluster, that remains within a language's phonotactic inventory through multiple generations of speakers is considered as diachronically stable. Diachronic stability has at least two different aspects. First, phonotactic items can be stable because they are frequently uttered and used. Trivially, a certain sound sequence that is not produced any more although speakers might have mental representations of that sequence stored in their memories, will cease to exist and drop out of a language's phonotactic inventory. Sound sequences that are, in contrast, produced frequently have a higher chance of being perceived and memorized by other speakers, which leads to enforced cognitive entrenchment of that sequence. This, in turn, entails higher utterance frequencies, as many cognitive linguists argue (Bybee 2007; Bybee 2010; Croft 2000; Langacker 2008).

The second aspect relevant to phonotactic stability is that of accuracy in production and perception. If it is extremely difficult to produce a certain sequence of sounds, for instance because the articulation of that sequence involves inconvenient and laborious movements of the tongue, it is likely that speakers will facilitate that sequence by, for instance, switching places of articulation. Notably, this might be the case no matter how frequently the sequence should surface in speech, e.g. caused by very productive morphological operations (e.g. plural *-s* in Afrikaans as in *kinders* 'children'). Indeed, it has been argued that utterance frequency even decreases accuracy in sound production (Bybee 2007; Diessel 2007), a notion also referred to as frequency-driven phonological or phonetic erosion in diachronic change (Heine and Kuteva 2007).

We conclude that utterance frequency and accuracy (in more general terms also referred to as 'fecundity' and 'copying fidelity'; Ritt 2004: 123) are two orthogonal aspects of diachronic stability in phonotactics, which may or may not be linked to each other by a trade-off, where accuracy is diminished by high utterance frequency and *vice versa*. The focus of this paper, however, is not on phonotactic utterance frequency but rather on articulatory and perceptual determinants in phonotactics, that is, on factors that determine the accurate transmission of phonotactic items. Frequency may well interact with these determinants in one way or another, but this is not of primary concern to our endeavor (see Pagel, Atkinson and Meade 2007, Diessel 2007 and works cited above for more discussion). Nevertheless and crucially independent thereof, we will, in this paper, operationalize diachronic stability of phonotactic sequences by means of diachronic increase or decrease in frequency. This is not contradictory, since inaccurate production and perception of a phonotactic sequence yield, everything else being equal, a diachronic decrease in frequency irrespective of whether or not that sequence was initially highly frequent. Synchronic frequency and diachronic growth (or decline) in frequency are clearly conceptually different from each other.

The question now is which articulatory factors determine accuracy and hence diachronic stability of phonotactic items. We focus on the three most prominent articulatory features in phonological research throughout various theoretical approaches: manner of articulation, place of articulation and phonation, i.e. voicing (Chomsky and Halle 1968; Dressler 1989; Dziubalska-Kołaczyk 2014; Hogg and McCully 1987; Prince and Smolensky 2002). Although the cognitive implementation of these articulatory features is by no means clear (Bybee 1994; Daland and Pierrehumbert 2011; Välimaa-Blum 2005; Wedel 2006), they have clear physiological interpretations and without doubt serve as valuable models in phonological theory.

Three phonotactic concepts related to these articulatory features shall be discussed exemplarily: 'net auditory distance', 'voicing harmony', and 'sonority sequencing'. Although we focus on pairs of consonants in this paper, these principles can be extended to larger sequences of phonemes.

First, the framework of beats-and-binding phonology (Dziubalska-Kołaczyk 2014; Marecka and Dziubalska-Kołaczyk 2014) predicts that phonotactic sequences in general and sequences of consonants in particular are preferred if the articulatory difference between the sounds they are composed of is large. In order to operationalize articulatory difference, Dziubalska-Kołaczyk and colleagues derive the so-called 'net auditory distance' (NAD) for a sequence of consonants. This metric combines differences between the respective manners of articulation, places of articulation of the consonants and vowels involved. Via NAD it is determined whether or not a consonant sequence is preferred (or 'well-formed'); generally, the larger NAD between the consonantal segments (and the smaller NAD between neighboring V and C) the better the cluster. For instance, NAD would predict that word-final /mʃ/ is more preferred than word final /mf/, because the latter cluster does not exhibit a sufficiently large difference between PoA of its segments, while the former does so (Dziubalska-Kołaczyk, Pietrala and Aperliński 2014). Although defined in terms of articulatory features, the NAD principle is motivated on perceptual grounds. It is argued that large articulatory differences facilitate the perception of a phonotactic sequence and its decomposition into segments. However, the dominance of assimilation as opposed to dissimiliation processes in casual – and crucially speaker-friendly – speech suggests that speakers favor small articulatory differences in phonotactic sequences (Dziubalska-Kołaczyk 2014: 17).

A second important strand of phonotactic research involving articulatory differences is that of 'voicing harmony' (Blevins 2004; Coetzee 2014; Hansson 2004). Regressive and progressive voicing agreement among pairs of consonants occurs in various languages across long distances (i.e. crossing intervening segments; Cho 1991; Hansson 2004) as well as within consonant clusters (Grijzenhout and Krämer 2000).

Voicing agreement in general has been argued to be motivated in multiple ways (see Coetzee 2014: 696–700 for an excellent discussion). On the one hand, voicing agreement serves the speaker since changing voicing in the transition from one consonant to another incurs increased articulatory effort to the speaker (although Dziubalska-Kołaczyk 2014: 17 challenges that notion by posing the question of whether actually retention or modulation is physiologically more costly). Similarly, regressive voicing agreement could be driven by 'anticipatory activation' in production (Coetzee 2014; Hansson 2001), in which an articulatory feature of a consonant that is about to be produced is activated already before the consonant is actually produced and mapped onto a preceding consonant. On the other hand, progressive voicing agreement might be driven by perception errors, in which the listener maps formant values characterizing voicing in a consonant to a nearby and originally voiceless consonant (Coetzee 2014: 697).[3]

It is important to note that voicing agreement crucially depends on the position of the consonants involved, or more precisely, on other processes that apply to consonants in a certain position. For instance, voicing agreement among word-final consonant clusters coincides with the process of word-final devoicing which could – perhaps accidentally – produce voicing harmony (in the sense that both consonants share the same voicing feature) if the penultimate consonant is voiceless (as e.g. in the German pronunciation of the acronym *OMFG* of *oh my fucking god*, /ɔmfk/). Likewise, devoicing could disturb voicing agreement if the penultimate consonant is voiced (e.g. Afrikaans *hond* /hont/, 'dog', vs. *honde* /hondə/, 'dogs'). In some languages such as English, however, word-final devoicing is restricted to cases in which it assists voicing harmony, as illustrated by the differential voicing of consonantal suffixes attached to base forms ending in voiceless and voiced consonants (e.g. *picked* /pɪkt/ vs. rigged /rɪgd/).

[3] Coetzee (2014) mentions yet a third plausible mechanism namely that of 'lexical accumulation' which led to voicing agreement in Afrikaans CVC sequences as a consequence of multiple unrelated sound changes.

Finally, let us consider the principle of 'sonority sequencing' (SSP) which in a nutshell asserts that consonant sequences in the onset position must rise, while consonant sequences in the coda position must successively fall in sonority (Clements 1990). For the present discussion, the SSP is relevant, because sonority is tightly linked to manner of articulation and – to a lesser extent – to voicing. In general, consonants are the more sonorous the less the airstream is obstructed by the articulators, so that glides and liquids are more sonorous than nasals followed by fricatives, affricates, and stops in decreasing order. Within these categories, voiced sounds are considered more sonorous than voiceless sounds (Burquest and Payne 1993: 101).
Crucially, apart from directionality the SSP entails a strict ordering with respect to sonority and prefers large sonority differences to small ones. Cross-linguistically, large sonority rises (in onset position) or drops (in coda position) are more common than small rises or drops, which in turn are more widely attested than sonority plateaus (Berent et al. 2007: 294; Clements 1990). More specifically, Berent et al. (2007) have shown experimentally, that onset-clusters with rising sonority are processed faster than onset-clusters remaining on the same sonority level by speakers of English. They do, however, point out that this behavior is language specific and might depend on linguistic experience. Russian speakers did not show differential processing in their experiment, which is in turn reflected in the generosity – or laxness – of the Russian language with respect to the SSP.

Sonority sequencing interferes with the above described principles in an interesting way. Although it is trivially consistent with the principle of maximizing NAD, since manner of articulation corresponds to the sonority scale, it counteracts the principle of voicing harmony in those cases in which progressive voicing assimilation applies to coda clusters, e.g. in English plural *bells* /bɛlz/ or in the loan *Pils* /pɪlz/ imported from German into English. Sonority sequencing would prefer /ls/ to /lz/ while voicing harmony would prefer /lz/ to /ls/ (*mutatis mutandis* the same holds for regressive assimilation in onset clusters).

What we can infer from this discussion is, at least, that articulatory differences matter in phonotactic production and perception. More specifically, manner of articulation seems to play a particularly important role as it is a defining feature in the principle of sonority sequencing, which is arguably quite prominent in phonotactic research. Interestingly, the dominance of manner of articulation over place of articulation and voicing in the formation of phonotactic sequences is supported by recent neurological research. Mesgarani et al. (2014) have found that during processing in the superior temporal gyrus in the brains of speakers of English, consonants sharing the same manner of articulation are locally patched together, while in each of these patches consonants with different places of articulation and voicing patterns are mixed and processed closely together. This implies that on the neurological level, manner of articulation has higher discriminatory power than place of articulation, while voicing is shown to have discriminating effects within the subset of fricatives and plosives (Mesgarani et al. 2014: 1009). As a consequence, consonant clusters exhibiting large manner-of-articulation differences should be less confusable and hence more stable than clusters with large place-of-articulation differences. Whether this neurological setup reflects universal properties related to articulatory organs or is exclusively shaped by linguistic experience, and is thus language specific, is – as far as we know – a matter of future research.

1.3 The link between phonotactic acquisition and change

In this paper, observations from language acquisition data and diachronic data are contrasted with each other, the underlying hypothesis being that what is acquired early tends to be diachronically more successful. This deserves some discussion, since the exact relationship between language acquisition and language change is under debate.

While it is generally acknowledged in generative research that language acquisition and change are inherently linked to each other since in this framework the grammatical output of the language-acquisition process ultimately determines the next generation's grammar (Roberts 2007; Yang 2000), the respective roles that children and adults play in language change are much more contested in more functional and usage-based paradigms (Bybee 2010). Trivially, only what is acquired can be passed on to the next generation, but the question remains if it is really the acquisition process which constitutes the source of linguistic variation.

Indeed, arguments have been brought forth that adults are to a larger extent responsible for linguistic change (Bybee 2010: 114–119). First, changes do occur at the adult stage and the phonological domain particularly seems to be subject to changes in linguistic behavior, as known from research on articulatory loss and phonological attrition (Ballard et al. 2001; Seliger and Vago 1991). Furthermore it is argued that – particularly phonological – errors occurring during the process of language acquisition do not resemble attested diachronic developments (Diessel 2012) and that these errors do not persist at later ages (Bybee 2010).

Nevertheless, it has been shown empirically, that age of acquisition (which we are focusing on in this paper) indeed correlates with diachronic stability. Words that are acquired early tend to be more resistant against phonological change than later acquired words (Monaghan 2014), and early used phonotactic strings have shown to be abundant in historically old word forms (MacNeilage and Davis 2000). Thus, a certain link between language acquisition and language change cannot be denied. Indeed, from a cognitive perspective it makes sense that items which are acquired late are to a larger extent prone to change as a consequence of being cognitively less entrenched (Bybee 2007; Croft 2000; Diessel 2012; Rosenbach 2008). This is so because items that are acquired late have less opportunities to be processed and produced – and consequently less entrenched – than words which are acquired early. Thus, it may well be the case that it is variation in adult speech, which drives linguistic change, but the underlying mechanism that enables this variation might to some extent be originally grounded in language acquisition.[4]

1.4 Summary and research hypotheses

We conclude that contrasting phonotactic acquisition with phonotactic change is worthy of investigation and moreover that we expect the same articulatory and perceptual pressures to apply and be visible in both domains. More specifically, we suppose that if a certain articulatory

[4] In that sense, the generative model of diachrony discussed above is an abstraction of what might happen in language change, in that it does not precisely show what happens in an individual through his or her lifetime, but from a less fine-grained perspective it fits as it predicts a similar outcome.

distance in phonotactic items, say manner-of-articulation difference between consonants, is shown to be positively correlated with ease of acquisition, so that consonant clusters featuring large manner-of-articulation differences are acquired early, then this articulatory difference is expected to positively correlate with diachronic stability. Hence, clusters featuring large differences should diachronically become more established and clusters featuring small – in this example, manner-of-articulation – differences are expected to become less frequently used on the diachronic time scale. The same reasoning applies, *mutatis mutandis*, to place of articulation and voicing.

By the same token, we expect that the relative ordering of the respective strength of effect of manner, place and voicing to language acquisition is preserved in diachrony. That is, if for instance voicing difference is shown to exert a larger facilitating effect on the acquisition of consonant clusters than place of articulation does, then place of articulation is expected to contribute less to the diachronic stability (or more to diachronic change) of a cluster than voicing does.

Thus, given the discussion in 1.2 on the relevance of differences in manner, place and voicing to processes and phenomena in phonotactics, we propose the following two research hypotheses:

(1) **Articulatory differences in phonotactic acquisition.** Large manner-of-articulation difference between the segments of a cluster decrease its age of acquisition. Facilitating effects should be smaller for difference in voicing and smallest for place-of-articulation difference.

(2) **Articulatory differences in phonotactic diachrony.** Large manner-of-articulation difference between the segments of a cluster increase its diachronic stability. These stabilizing effects should be smaller for difference in voicing and smallest for place-of-articulation difference.

It is possible that these two hypotheses causally hang together if acquisition is indeed a driving force of diachronic change, as outlined in the previous section. If (1) is confirmed, then (2) is expected to hold as well.[5] They will be approached by two separate studies drawing on Dutch and Afrikaans data: Study 1 (Section 2) tackles research hypothesis (1) by investigating phonotactic acquisition in Dutch while Study 2 (Section 3) addresses hypothesis (2) by means of a (historic) comparison of Dutch and Afrikaans.

Since both studies are based on empirical data, it is clear that the causal relationship between (1) and (2) can only be partially accounted for, the reason for this being the obvious lack of acquisition data of historical Dutch or Afrikaans. In other words, the present project rests on the simplifying assumption that Dutch did not change much phonotactically (or at least less than Afrikaans) through the past two to three centuries (but see for instance Szemerényi 1996

[5] Clearly, it is logically impossible to verify an implicational relationship between (1) and (2) based on a single study, even if both hypotheses turn out to be correct. For instance, and quite plausibly, there might be factors independently supporting both hypotheses. Diessel (2012) suggests a number of psychological and cognitive factors that would do so.

for a justification of this assumption). Being fully aware of this, we assume for the sake of argument that the results from Study 1 about the acquisition of Dutch phonotactics can be transferred to historical Afrikaans (see 3.1 below for more comments). Both studies will be described in detail in the subsequent sections.

2 Study 1: articulatory effects on phonotactic acquisition in Dutch

We tackle the question of whether large differences are beneficial to the acquisition of Dutch consonant clusters. More specifically, hypothesis (1) in the previous section is addressed, which states that the facilitating effects of manner of articulation and voicing should be larger than those of place of articulation. This hypothesis can be corroborated. It can be shown that large place-of-articulation differences delay cluster acquisition as opposed to voicing and manner-of-articulation differences. However, it is voicing difference (and not MoA) which excels in enhancing phonotactic acquisition. Only in highly frequent consonant clusters do both voicing and manner show similarly strong facilitating effects (cf. 2.4). Data, variable operationalization, statistical modeling procedure, analysis and results are discussed in more detail below.

2.1 Data

We combined two data sets in order to assess the effect of articulation on the acquisition of phonotactic sequences in Dutch. Age-of-acquisition (AoA) data were taken from Brysbaert et al. (2014) who provide AoA ratings for 30,000 Dutch lemmas. These ratings were collected in a large study in which participants were asked to estimate the age at which a given stimulus word was acquired, a methodology which has been shown to yield high correlations with estimates obtained under laboratory conditions (Kuperman, Stadthagen-Gonzalez and Brysbaert 2012). Phonological transcriptions were taken from the CELEX database (Baayen, Piepenbrock and Gulikers 1995) and added to the lemmas. Additionally, word-length in terms of the number of phonemes and token frequency were retrieved from CELEX. Only a core vocabulary of the 5,000 most frequent words was considered (based on INL token-frequency scores in CELEX) in order to exclude non-prototypical low-frequency items.[6] Foreign words were excluded as well.

Since we are analyzing consonant clusters in word-final position in this study, only words ending in a sequence of two consonants were considered. In total, 828 Dutch lemmas met these requirements, featuring a range of 33 word-final consonant-cluster types. Articulatory features of the two final consonants were assigned based on the IPA chart.

2.2 Variables

The aim of this study is to assess the effect of articulatory differences on the AoA of consonant clusters. Thus, a number of variables were considered in the statistical analysis. First and foremost, for each cluster type, age-of-acquisition (`AoA`) was determined by calculating the first decile of all AoA ratings given by Brysbaert et al. (2014) of the words showing that cluster

[6] Inspecting CELEX (INL frequency), it can be easily shown that the 5,000 most frequent lemmas cover about three fourths of all Dutch tokens.

type word finally ($\mu_{\text{age}} = 6.7\,yr; sd_{\text{age}} = 1.5\,yr; rg_{\text{age}} = (4.3,10.7)$). The first decile was chosen as it provides a more robust, albeit at the same time more conservative, estimate of the minimum of cluster-wise AoA ratings than the actual minimum does. This is so, since the actual minimum is obviously sensitive to outliers at the lower range of AoA. In other words, this procedure ensures that cluster `AoA` is not just determined by the knowledge of a single word. As `AoA` featured a slightly skewed distributional pattern the data were Box-Cox transformed (Box and Cox 1964) in order to fit the statistical modeling requirements (after transformation: $\mu_{\text{AoA}} = 0.84;\ sd_{\text{AoA}} = 0.03; rg_{\text{AoA}} = (0.77,0.91)$).[7]
The primary predictors we are interested in are difference in manner of articulation, difference in place of articulation and difference in voicing of the two final consonants. In order to parametrize these differences, manner of articulation as well as place of articulation were first translated into ordinal scores as in Dziubalska-Kołaczyk (2014) and Dziubalska-Kołaczyk et al. (2014). While manner-of-articulation scores (`MoA1`, `MoA2`) depend on sonority (from 1 to 7; higher sonority yielding a higher score), place-of-articulation scores (`PoA1`, `PoA2`) were defined by the phoneme's place of articulation in the vocal tract (from 1 to 7; the closer to the front, the lower the score). Voicing was simply considered as a binary variable (1/0; voiced/unvoiced). See Table A1 in the appendix.

Subsequently, difference in manner of articulation (`ΔMoA`) was operationalized as the absolute value of the difference between manner-of-articulation scores, normalized to the unit interval with respect to the maximal absolute difference, thus restricting values to scores going from 0 to 1, where 1 is the maximal difference of all cluster types considered and 0 denotes identity with respect to manner of articulation. Taking the absolute value ensures that the arbitrarily assigned directionality of the ordinal manner-of-articulation scores does not skew the results, while the normalization procedure facilitates interpretation of and comparisons between the respective effect of manner, place and voicing without the need of computing normalized regression coefficients in the statistical analysis (Nakawaga and Cuthill 2007), and at the same time retains any relevant information. Difference in place of articulation (`ΔPoA`) was determined, *mutatis mutandis*, as above. Difference in voicing (`ΔVoice`) was simply operationalized as a binary variable ('`different`' if consonants are voiced differently and '`same`' else).

A side note is in order on the operationalization of `ΔMoA` of consonant clusters in word-final position by means of sonority. One might wonder whether relying on absolute values as opposed to actual differences makes sense from a theoretical perspective, since the principle of sonority sequencing (see 1.2) would predict sonority to decline word finally. However, it was found in our data that absolute and actual differences in `MoA` deviate from each other in only 3 out of 33 cluster types, which as a matter of fact illustrates that the sonority-sequencing

[7] This procedure of using AoA ratings corresponding to word types to estimate AoA of a cluster type might potentially strike one as odd, since it tends to suggest words to be acquired, say, late *because* they contain a particularly ill-formed cluster. However, although the latter suggestion may – all other things being equal – be correct, we will remain agnostic about it. Rather, the decision for parametrizing cluster-wise AoA as done above is a primarily pragmatic one. Since phonotactic sequences hardly occur in isolation there is simply no better way of estimating their AoA than via the linguistic items they are embedded in.

principle is fulfilled anyway to a large extent in Dutch. Thus, taking actual differences would not substantially change the results reported below.

Finally, two additional controlling variables entered the analysis. For each cluster type, token frequency (`frequency`) was determined by computing the median INL token frequency (see section 2.1) of all words featuring that cluster. We preferred the median of all words to, e.g., the frequency of only the subset of initially acquired words featuring that cluster as it better represents the average overall exposure of the language learner to the cluster type. Due to expected distributional properties, `frequency` was Box-Cox transformed as well, before it entered the statistical analysis (after transformation: $\mu_{\text{frequency}} = 2.73$; $sd_{\text{frequency}} = 0.11$). The remaining controlling variable, phonological length (`length`) was simply operationalized as the median number of phonemes of all words featuring that cluster ($\mu_{\text{length}} = 5.42$; $sd_{\text{length}} = 1.29$). All scores are summarized in Table A2.

2.3 Modeling procedure

The goal of the statistical analysis is to assess the differential effects of `ΔMoA`, `ΔPoA` and `ΔVoice` on `AoA`. We opted for linear models (LM, Baayen 2013; West, Welch and Gałecki 2015) as this model family provides a flexible way of combining numerical as well as categorical variables and at the same time allows for including controlling factors. Thus, `AoA` was implemented as a dependent variable into a LM in which `ΔMoA`, `ΔPoA` and `ΔVoice` function as predictor variables. In addition, six interacting terms were included in which `ΔMoA`, `ΔPoA` and `ΔVoice` are controlled by `length` and `frequency`, respectively. All distributional requirements were met (see previous subsection). Computations were done in R (R Development Core Team 2013).

This resulted in a model featuring nine regression terms which harbors the risk of being overspecified and hence insufficient fitting properties, rendering conclusions drawn from the estimated coefficients unreliable. In order to find the most informative and at the same time most parsimonious model of `AoA` with the best fit, AICc-driven model selection was employed. This requires some elaboration. AICc ('corrected Akaike Information Criterion'; Johnson and Omland 2004) is a measure of information (or more precisely, of information loss relative to the data) of a given model which balances goodness-of-fit and model complexity, and which is in addition corrected for applications to small samples. The smaller AICc the better the model. AICc is superior to plain goodness-of-fit measures such as (adjusted) R^2 in that the latter automatically increases, the more predictors are added to a model. Thus, AICc accounts for model overspecification.

In the model-selection procedure, linear models for all theoretically interesting subsets of predictor regression terms together with their AICc are computed. In the present analysis, nine predictor terms were considered, three for the isolated variables and six controlling terms, as described above. We assumed that controlling interaction terms always co-occur with their corresponding controlled predictor in isolation. For instance, if `ΔMoA` is controlled by `frequency` in a model it includes the configuration `ΔMoA+ΔMoA:frequency`. This restriction ensured, that the controlling variables (`frequency` and `length`), which we are actually not interested in in this study, do not occur in isolation in the analysis. This resulted in a set of 124 candidate models. The optimal – or 'AICc-best' – model then is the one model with the lowest AICc score, i.e. the least loss of information.

This information-theoretic model selection procedure allows for yet deeper investigations, namely 'multimodel inference' (Burnham and Anderson 2002; Burnham, Anderson and Huyvaert 2011). To begin with, an important observation is, that although there is always a single best model, that model does not necessarily have to be much better than other candidate models. There might be some other relevant information contained in some of the remaining candidate models, which would be lost if one only considers the single best model. By comparing a candidate model's AICc score with that of the best model, the relative strength of evidence of that candidate model, the so-called Akaike weight w, can be computed. It can be interpreted as the probability of that model given the data and the set of all competing candidates. Thus, the Akaike weight measures how much evidence there is in the data for a candidate model (Johnson and Omland 2004: 104). Clearly, the best model has the largest Akaike weight.

In multimodel inference one can exploit Akaike weights in order to combine all candidate models. A whole set of models obviously contains more information than a single best model. From the model set and the corresponding set of Akaike weights, average regression coefficients $\bar{c}_i$ can be computed. These regression coefficients can then be used to calculate average predictor effects, under the assumption of average token frequency and phonological length. For instance, the average effect $\bar{e}$ of ΔPoA can be computed as $\bar{e}_{\Delta\mathrm{PoA}} = \bar{c}_{\Delta\mathrm{PoA}} + \mu_{\mathrm{frequency}} \cdot \bar{c}_{\Delta\mathrm{PoA:frequency}} + \mu_{\mathrm{length}} \cdot \bar{c}_{\Delta\mathrm{PoA:length}}$, where μ_j denotes the mean.

Moreover, using the Akaike weights it is possible to compute 'relative variable importance' (RVI), a measure of how often a predictor appears in the models contained in the candidate set (Burnham and Anderson 2002: 168). This measure is very informative: not only can one determine the average strength of the effect of some predictor variable, it is also possible to make assertions about how important that predictor is for obtaining information about the output variable, relative to the other predictors. Note that large importance of a variable does not necessarily imply a large effect and *vice versa*. Finally, predictor variables can be ranked by their RVI. This procedure allows for an in-depth analysis of the effects that articulatory differences have on acquisition as well as of their respective informational importance.

2.4 Results

The AICc-best model of AoA obtained by the procedure described above features six terms[8], as can be seen in Table 3. Most notably, ΔMoA only contributes to AoA in isolation, having a slight enhancing effect on acquisition, so that no significant interactions surface in this model. Translating transformed AoA back into age of acquisition measured in years (based on the respective ranges), the effect on AoA corresponds to roughly 3 years if ΔMoA is maximal (note

[8] Note that the Akaike weight of the best model equals about 0.123 while that of the maximal model containing all terms equals 0.001. Thus, evidence for the best model is roughly 123 times stronger than for the maximal and least parsimonious model. This illustrates the necessity of careful model building in quantitative research.

at this point that the intercept in the model conveniently coincides – approximately – with μ_{AoA}). A similar acquisition-enhancing effect can be seen in the averaged model. Taking average length and frequency into account, maximal ΔMoA reduces AoA by about $\bar{e}_{\Delta\text{MoA}} = -0.06$ (Table 4). Also note that of all three variables manner-of-articulation difference scores highest on RVI, suggesting that ΔMoA plays an important role in phonotactic acquisition.

In the best model, ΔPoA shows three significant effects on acquisition (Figure 2a). In isolation, ΔPoA considerably increases AoA, an effect which becomes even significantly larger in long words, but which is diminished significantly the more frequent clusters are in terms of tokens (Table 3). Inspecting the average model, and assuming average length and frequency, it can be seen that ΔPoA has an inhibiting effect on phonotactic acquisition of about $\bar{e}_{\Delta\text{PoA}} = 0.04$ (Table 4). From all three primary predictors, place-of-articulation difference seems to be least relevant to explaining age of acquisition and to be the most dependent on the controlling variables `frequency` and `length`, as can be judged from the respective RVI scores in Table 4 (see Figure 2b).

Finally, ΔVoice shows significant enhancing effects on acquisition (Table 1). Crucially, though, these effects are diminished considerably in frequent clusters, as shown by the significant interaction term in the best model. This is confirmed, by the average effect computed from the averaged model, $\bar{e}_{\Delta\text{Voice}} = -0.11$. In the upper part of the frequency spectrum, however, this effect becomes weaker.

We conclude that, everything else being equal, differences in manner of articulation as well as in voicing have a facilitating effect on the acquisition of word-final consonant clusters in Dutch, while clusters are acquired later if the phonemes they are composed of differ with respect to their place of articulation (e.g. /pt/, /lp/ or /kt/). In the upper part of the frequency spectrum, and thus in cognitively more entrenched clusters, manner becomes more enhancing while voicing becomes less enhancing ($\bar{e}_{\Delta\text{Voice}} = -0.10$ vs. $\bar{e}_{\Delta\text{MoA}} = -0.07$ at maximum `frequency` and average `length`). For making predictions about acquisition, manner is most important, followed by voicing and finally place of articulation (cf. RVI in Table 4 and Figure 2b). This corroborates hypothesis (1) presented in section 1.4, albeit only partially, as voicing difference turns out to enhance acquisition more than expected.

Table 3. AICc-best model: $R^2_{\text{Adj}} = 0.42$; AICc $= -140$; $w = 0.12$.

Variables	Estimate (c_i)	*SE*	*t*	*p*
Intercept	0.86	0.01	82.9	<0.001
ΔMoA	-0.08	0.02	-3.29	<0.01
ΔPoA	0.43	0.15	2.83	<0.01
ΔVoice(different)	-0.38	0.19	-2.05	0.05
ΔPoA:length	0.01	0.00	3.41	<0.01
ΔPoA:frequency	-0.18	0.06	-3.32	<0.01
ΔVoice(different):frequency	0.15	0.07	2.21	0.04

Table 4. Predictor coefficients in the combined model resulting from model averaging procedure over 124 candidate models. Average coefficients together with standard errors, relative variable importance and averaged overall effects are shown.

Variables	Average estimate ($\bar{c}_i$)	*SE*	*RVI*	Average effect ($\bar{e}_i$)
`ΔPoA`	0.16	0.21	0.73	
`ΔPoA:length`	0.01	0.01	0.46	0.04
`ΔPoA:frequency`	-0.06	0.07	0.39	
`ΔMoA`	-0.01	0.37	0.96	
`ΔMoA:length`	0.01	0.01	0.45	-0.06
`ΔMoA:frequency`	-0.03	0.13	0.19	
`ΔVoice(different)`	-0.27	0.27	0.83	
`ΔVoice(different):length`	0.00	—	0.16	-0.11
`ΔVoice(different):frequency`	0.05	0.09	0.32	

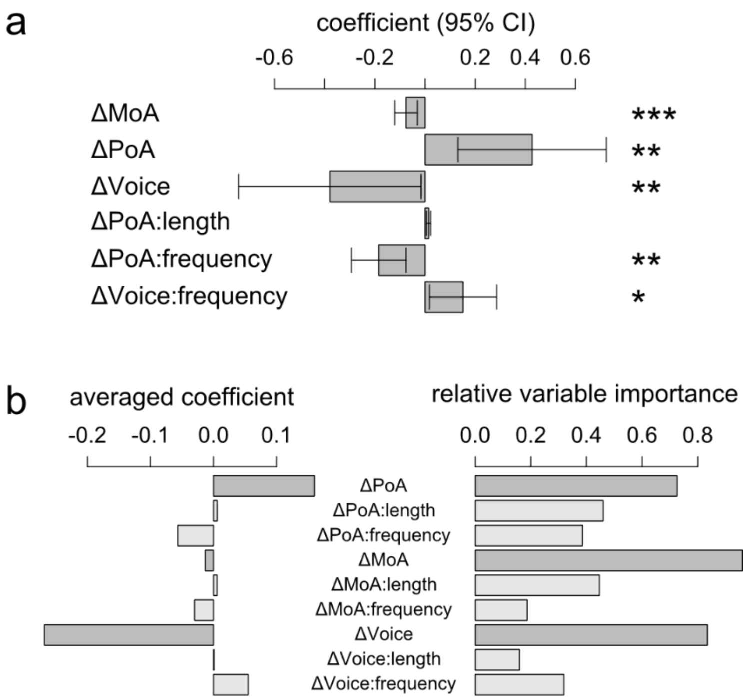


Figure 2. (a) Plot of the coefficients in the optimal model resulting from model-selection procedure. Vertical axis measures effect on `AoA` (transformed age-of-acquisition). Error bars denote 95% confidence intervals. Significance code: '***' $p < 0.001$, '**' $p < 0.01$, '*' $p < 0.05$ with respect to zero-effect hypothesis. (b) Averaged model resulting from multi-model inference. Dark gray and light gray bars correspond to isolated variables and interacting terms, respectively. On the left: averaged coefficients measuring the effect on `AoA`. On the right: RVI scores corresponding to variables and interaction terms.

Interestingly, this goes in line with results from neurological and cognitive research on the organization of phonemes, as discussed in Section 1.2. The question is, whether the differential behavior of manner of articulation, place of articulation and voicing in phonotactic acquisition also yields diachronic long-term reflexes. This question will be dealt with in the subsequently presented study.

3 Study 2: articulatory effects on stability in the history of Afrikaans

We assess the second research hypothesis – i.e. (2) section 1.4 – in this study and expect clusters featuring large manner-of-articulation differences to be diachronically more stable than clusters with differentially voiced segments followed by clusters that show large differences with respect to place of articulation. Indeed, as presented in what follows this hypothesis is largely supported by data from (historical) Dutch and contemporary Afrikaans. While large place-of-articulation difference and difference in voicing diminishes diachronic stability, clusters showing large differences in terms of manner of articulation are more successful in contemporary Afrikaans than what would be expected under the null-hypothesis that no change occurred in the development of the Afrikaans phonotactic system. The hypothesis and data considered in this study require an analytic approach which differs from that in the first study. It will be described in the following subsections together with its outcome.

3.1 Data

Two additional corpora were used to address the research question rephrased above. First, Afrikaans final clusters were retrieved from the NCHLT corpus (Eiselen and Puttkammer 2014). The corpus consists of 58,096 annotated word tokens (distributed among 6,464 word types) retrieved from written Afrikaans, mostly from government websites. The fact that its tokens are lemmatized as well as morphologically decomposed was crucial in order to extract all tokens ending in a cluster, excluding those tokens in which the final cluster involves a morphological operation and thus spans a morpheme boundary. This was necessary, as the Dutch data did not include word-forms containing boundary-spanning clusters either. Phonological information was taken from Coetzee (1969). In total, 445 word types ending in a consonant cluster were retrieved, featuring 26 different cluster types. Subsequently, type frequencies were obtained for each cluster type, i.e. the number of word-types a cluster type surfaces in word-finally.

Second, in order to obtain slightly more representative historical Dutch data, the 5,000 most frequent word types (cf. Section 2.1) were retrieved from the pre-1900 subset of the ‘De Gids’ corpus (henceforth DGC; see van de Velde 2009). The data were matched with the CELEX lemma list in order to obtain phonological transcriptions. After excluding those items not ending in a consonant cluster, 819 word types featuring 33 cluster types remained. Cluster specific type frequencies were also based on DGC. It is worth to mention that the 33 cluster types found in historical Dutch represent a proper superset of the 26 Afrikaans cluster types (only double clusters considered).

We emphasize that the procedure of retrieving frequency data from historical corpora but phonological data from contemporary language sources implies that the present approach is, at best, pseudo-historical. Also, 19th century Dutch data can obviously not be equated with historical Afrikaans which – at least phonologically – emerged from Early Modern Dutch vernacular spoken in the 17th or 18th century (Roberge 1993). Nevertheless, historical arguments are often based on comparative grounds solely considering contemporary data (cf. e.g. Szemerényi 1996). In that sense, the present approach finds itself somewhere halfway in between a synchronic comparative study and a purely diachronic one.

Having clarified this, the reader might be relieved to see that the distributional difference between the coda phonotactics in contemporary and 19th-century Dutch is not substantially large anyway (phi-coefficient based on chi-squared test of independence: $\varphi = 0.12$, 95%-CI: $(0.11, 0.12)$; $\chi^2 = 67{,}662.94; N = 5{,}042{,}350; df = 32$; cluster-wise token frequencies in CELEX and DGC considered).

3.2 Variables

We address the following question: do articulatory differences determine the historical stability of word-final consonant clusters in the diachrony of Afrikaans? We do so by measuring the productivity of word-final clusters in historical Dutch and in contemporary Afrikaans and subsequently testing whether articulatory differences have an effect on the relationship between both productivity scores. We straight-forwardly define the productivity of a cluster as the number of word types it occurs in. Note, that these scores will substantially differ when considering raw figures in historical Dutch and Afrikaans due to the different corpus sizes of DGC and NCHLT, respectively. This does not pose any problem, since relationships among these scores rather than raw differences between them will be considered (but see 3.3 below). Cluster productivity in historical Dutch and contemporary Afrikaans shall be denoted as `DutProd` and `AfrProd`, respectively. Whenever a cluster did only occur in DGC but not in NCHLT, its `AfrProd` score was set to 0, in order to maintain as many data points as possible. We decided to leave both variables untransformed (e.g. log, BoxCox) and to resort to nonparametric methods instead. `DutProd` and `AfrProd` scores for the set of 33 clusters are illustrated in Figure 3a. Articulatory differences (`ΔMoA`, `ΔPoA` and `ΔVoice`) were defined exactly as in 2.2. All scores are summarized in Table A3.

3.3 Modeling procedure

In a preliminary analysis, the strength of the relationship between productivity in historical Dutch (`DutProd`) and productivity in contemporary Afrikaans (`AfrProd`) was assessed. This was done by means of Spearman's ρ (rank correlation) due to the nonparametric nature of the productivity scores. The relationship between `DutProd` and `AfrProd` can be shown to be relatively strong at $\rho = 0.68$ (95% CI: $(0.44, 0.83)$; $t = 5.2, df = 31$). This was to be expected: on average, clusters which have been productive in historical Dutch are also productive in contemporary Afrikaans.

Subsequently, the influence that `ΔMoA`, `ΔPoA` and `ΔVoice` exert on the relationship between `DutProd` and `AfrProd` was analyzed in three separate generalized additive models (GAMs, Wood 2006a). GAMs are statistical models which do not only involve linear terms (as (generalized) linear models do) but combine linear, quadratic and even more complicated components. Hence, in contrast to straight lines or piece-wise linear surfaces they potentially yield curves and 'wiggly' surfaces, of course depending on the underlying data. They provide an efficient way of detecting non-linear, or more generally non-monotone, interactions among variables (Baayen 2013). Indeed, the relationship between productivity in Dutch and Afrikaans might be for instance curved rather than linear, and this holds even more so for the relationship between productivity and articulatory difference. Most importantly, however, interactions among two predictor variables that influence a third (dependent) variable can be easily modeled using GAMs by implementing so-called tensor-product terms (Wood 2006b), especially if predictor variables are located on different scales (e.g. here: articulatory differences on the one hand and type frequencies on the other hand). Details are not relevant at this point, it is sufficient to note that this modeling tool-kit provides a convenient way of analyzing the interaction of articulatory difference on the relationship between productivity in Dutch and Afrikaans. Finally, GAMs have the advantage of being innately non-parametric, thus not imposing any particular distributional requirements on the data to be analyzed. This is particularly convenient given the skewed distribution of the productivity scores (Figure 3a, see also previous section). We opted for three separate GAMs as opposed to a single GAM featuring interactions among all variables due to the relatively small number of cluster types. Moreover, by comparing R^2 scores of the three separate models, the relative explanatory power of each of the three articulatory differences can be assessed (somewhat similar to – albeit not identical with – the information theoretic measure RVI in Study 1).

The question to be asked is the following: in which way does the relationship between `DutProd` and `AfrProd` change, if we consider articulatory differences at different degrees? For instance, it could be the case that the relationship between `DutProd` and `AfrProd` is positive and increasing for small articulatory differences, say, in manner of articulation, but decreasing for large articulatory differences. This would indicate that `ΔMoA` would have a negative effect on cluster stability, since in this scenario clusters with large manner-of-articulation differences are less frequent in Afrikaans than what would be expected based on Dutch productivity scores. Thus, in the first GAM, `AfrProd` is implemented as an outcome variable depending on the interaction of `DutProd` and `ΔMoA` (integrated as a tensor-product term). In the second GAM, `ΔMoA` is simply replaced by `ΔPoA`, in order to assess the effect of place-of-articulation. In the third GAM, finally, `AfrProd` again functions as outcome variable predicted by `DutProd` which is controlled (or 'smoothed'; Wood 2006a) by the binary variable `ΔVoice`. All computations were done in R, in particular using the `mgcv` package for computing GAMs (R Development Core Team 2013; Wood 2006a).

3.4 Results

The models reveal differential results about the impact that articulatory differences have on the relationship between cluster productivity in Dutch and Afrikaans, and hence on the diachronic stability of clusters. Let us begin with manner of articulation. The first GAM shows a significantly non-zero intercept at 13.49 ($SE = 1.59; t = 8.50; p < 0.001$) and a significant tensor-product term modeling the interaction of `DutProd` and `ΔMoA` ($edf = 7.66; F = 23.37; p < 0.001$) as well as remarkable fitting properties ($R^2 = 0.87; N = 33$). From the significant interaction term we see that `DutProd` and `ΔMoA` indeed affect `AfrProd`. In order to better understand the interacting behavior, however, the model has to be visualized.

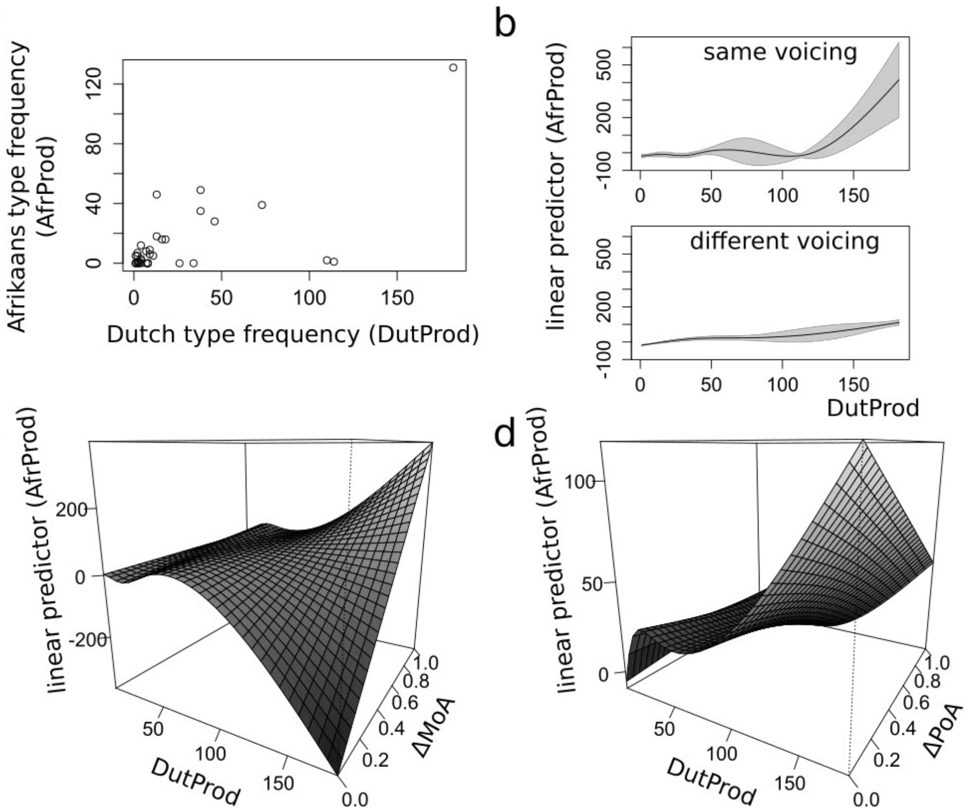


Figure 3. (a) Distribution of Dutch and Afrikaans type frequency in word-final consonant clusters. There is a significant correlation between both productivity scores. (b-d) Effect of voicing difference, manner-of-articulation difference and place-of-articulation difference on the relationship between productivity in Dutch and Afrikaans, respectively.

Figure 3c shows `AfrProd` as a two-dimensional function of `DutProd` and `ΔMoA`, illustrated by a curved surface. If `ΔMoA` is held constant at a certain value, say `ΔMoA` $= 0.2$, one can inspect the relationship between `DutProd` (horizontal) and `AfrProd` (vertical), just as in Figure 3a, represented by one of the solid black lines in the grid superimposed on the curved surface. For low `ΔMoA` scores, this functional relationship is decreasing. Clusters which do not or only slightly differ with respect to manner of articulation are less productive in Afrikaans than expected based on historical Dutch, i.e. under the null-hypothesis that there was no change between historical Dutch and contemporary Afrikaans. In contrast, for high `ΔMoA` scores, the relationship is increasing. Clusters exhibiting large manner-of-articulation differences are more productive in Afrikaans than expected.

The second GAM, which analyzes place of articulation, also shows a significantly non-zero intercept at 13.49 ($SE = 2.49; t = 5.42; p < 0.001$), a significant tensor-product term of the interaction of interaction of `DutProd` and `ΔPoA` ($edf = 6.68; F = 10.11; p < 0.001$) and slightly reduced explanatory power ($R^2 = 0.68; N = 33$).

In contrast to the first GAM, however, the relationship between `DutProd` (horizontal in Figure 3d) and `AfrProd` (vertical) is roughly increasing for all `ΔPoA` scores (with a tiny dip in the lower-mid-frequency range). Additionally, the strength of this relationship seems to be strongest (i.e. showing the steepest slope) for small differences in place of articulation and comparably weak if consonants differ in their place of articulation. Thus, increasing `ΔPoA` weakens the relationship between both productivity scores. The larger the difference in place of articulation, the smaller the productivity of Afrikaans clusters than what would be expected by the Dutch data. It can be concluded that, overall, `ΔPoA` exerts a diachronically destabilizing influence on word-final clusters.

Finally, the third GAM yields two separate one-dimensional curves (smooth terms) for the relationship between `DutProd` (horizontal in Figure 3b) and `AfrProd` (vertical), one for `same` voicing and one for `different` voicing (upper and lower graph in Figure 3b, respectively). Overall, the model shows an intercept of 19.30 ($SE = 2.49; t = 5.42; p < 0.001$) and again good fitting properties ($R^2 = 0.89; N = 33$). Both smooth terms show a significantly non-trivial behavior (`same`: $edf = 5.18; F = 4.78; p = 0.002$; `different`: $edf = 3.17; F = 63.36; p < 0.001$), the latter staying more constant than the former, as can be seen from the two graphs in Figure 3b. In the low and mid frequency range, voicing does not seem to differentially determine cluster productivity. However, in the high frequency range, clusters that have no voicing difference are more productive than their differently voiced counterparts. This shows that at least highly frequent clusters do not benefit from voicing contrasts, diachronically speaking. The opposite seems to be the case. This goes in line with various voicing-assimilation processes discovered in historical language research (e.g. Horobin and Smith 2002; Colantoni and Steele 2007). Infrequent clusters, however do not seem to show this differential behavior.

We conclude that, if diachronic stability of word-final clusters is assessed by comparing productivity in historical Dutch and in contemporary Afrikaans, increasing manner-of-articulation differences between the building blocks of a cluster seem to have a promoting (or stabilizing) effect, while increasing place-of-articulation differences and voicing differences exhibit demoting (or destabilizing) effects on word-final consonant clusters. This effect is enhanced the more productive and consequently the more frequent a cluster is. Thus, it is clusters like /rt/, /lt/ and /mp/ which are diachronically most stable, and clusters such as /pt/, /kt/ or /lm/ which are expected to undergo diachronic deletion processes. In summary, this goes in line with research hypothesis (2) in 1.4.

4 Discussion and conclusion

Based on considerations about the differential relevance of manner of articulation, place of articulation and voicing in phonotactic production and perception, we addressed the question of whether articulatory difference between the segments of a consonant sequence exerts a promoting effect in phonotactic acquisition and change. More specifically, we hypothesized difference in manner of articulation to have a stronger promoting effect than difference in voicing and difference in place of articulation, as suggested by research on the neuro-cognitive organization of phonemes (Mesgarani et al. 2014). We tested this set of hypotheses against Dutch acquisition data (word-final consonant clusters) and by means of a (pseudo-historical) comparative study of Dutch and Afrikaans. That is, we sought to identify the articulatory factors in the acquisition of Dutch coda phonotactics and the diachronically stabilizing determinants in the history of the Afrikaans phonotactic system.

It was found that indeed manner-of-articulation difference incurs a stronger promoting effect on phonotactic acquisition and diachronic stability than place-of-articulation difference does. That is, it is the set of clusters in the lower right corner in Figure 4 below (dashed box), which is most successful in acquisition and change. Consequently, developments such as t-deletion in *specht* (/xt/ > /x/) or schwa-epenthesis in *worm* (/rm/ > /rəm/) as shown in Figure 1 follow a systematic and articulatorily as well as perceptually motivated trend rather than occurring just by coincidence.

The behavior of voicing difference is slightly more complicated. While voicing difference has turned out to yield the strongest enhancing effects in acquisition of Dutch, it shows demoting effects on the diachronic time scale in the history of Afrikaans. Interestingly, voicing difference seems to suffer from utterance frequency in the sense that in clusters in the upper part of the frequency spectrum face the strongest demoting effects – both in acquisition and in diachrony – if their constituents differ in voicing. With respect to manner and place, the effect of frequency goes in precisely the opposite direction. This is surprising, given that sonority (which is linked to manner of articulation) by definition at least slightly correlates with voicing.

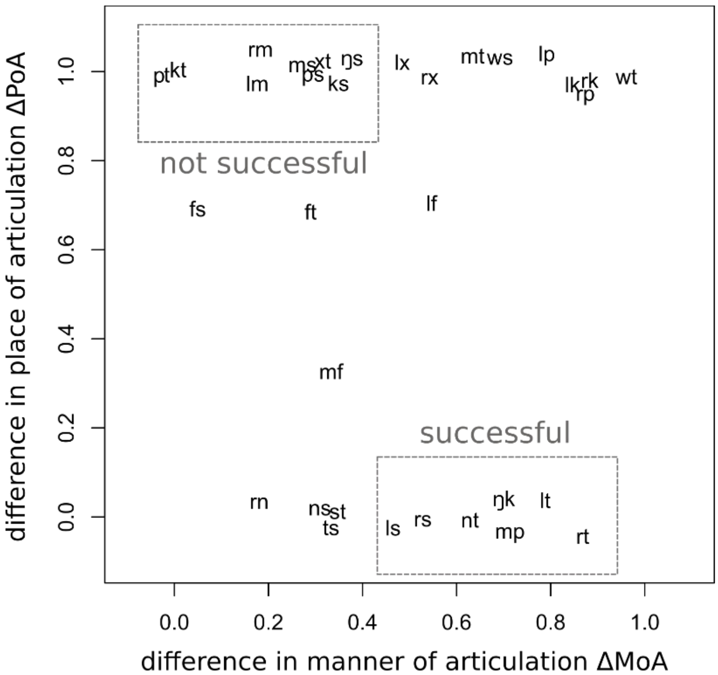


Figure 4. Dutch/Afrikaans word-final cluster types in the space defined by ΔMoA (horizontal axis) and ΔPoA (vertical axis). Data points are randomly jittered to make the labels readable. The dashed box in the lower right corner denotes clusters which are particularly successful in acquisition and change, i.e. clusters with segments showing similar place and different manner of articulation. Less successful types are located in the dashed box in the upper left corner.

Our results seem to converge with established phonotactic principles (cf. 1.2). The importance of manner-of-articulation difference in phonotactic acquisition and – more evidently – in phonotactic change goes in line with the principle of sonority sequencing which favors decreasing sonority in coda position (Clements 1990). The diverging effect of voicing difference in acquisition and change, respectively, is admittedly more puzzling and probably hints at a conflict between voicing agreement and final devoicing in Afrikaans, especially in adult speech. Finally, the fact that large place-of-articulation differences are neither particularly beneficial to phonotactic acquisition nor to the diachronic stability of word-final clusters (in fact the opposite seems to be the case) questions the validity of net auditory distance (Dziubalska-Kołaczyk 2014) as an overall measure of phonotactic well-formedness. Rather, a more differentiated approach is required to assess whether or not consonant clusters are preferred, if only in the case of Afrikaans and Dutch coda phonotactics.

It is important to remark, that our results are restricted to word-final sequences of two consonants in the coda. Neither did we specifically address longer sequences of consonants (although it is reasonable to assume that similar restrictions with respect to manner, place and voicing hold in longer phonotactic items), nor did we account for dynamics in the onset position.

Apart from the empirical results about the differential roles that manner, place and voicing play in the acquisition and change of Dutch and Afrikaans phonotactics, the present study more generally contributes to the discussion about the link between language acquisition and change (Bybee 2010; Diessel 2012; MacNeilage and Davis 2000; Monaghan 2014). Whether or not language change primarily happens during first-language acquisition cannot be clearly answered by our results. What can be confirmed is that in the domain of phonotactics similar articulatory determinants influence acquisition and change.

References

Baayen, H. 2013. Multivariate Statistics. Podesva, R., D. Sharma (eds.). *Research Methods in Linguistics*. Cambridge, 337–372.

Baayen, H., R. Piepenbrock and L. Gulikers. 1995. CELEX2 LDC96L14. Web download.

Ballard, K.J., D. Robin, G. Woodworth and L. Zimba. 2001. Age-related changes in motor control during articulator visuomotor tracking. *Journal of Speech, Language, and Hearing Research* 44, 763–777.

Berent, I., D. Steriade, T. Lennertz and V. Vaknin. 2007. What we know about what we have never heard: Evidence from perceptual illusions. *Cognition* 104 (3), 591–630.

Blevins, J. (ed.). 2004. *Evolutionary phonology: The emergence of sound patterns*. Cambridge, New York.

Box, G. and D. Cox. 1964. An analysis of transformations. *Journal of the Royal Statistical Society Series B* 26 (2), 211–252.

Brysbaert, M., M. Stevens, S. de Deyne, W. Voorspoels and G. Storms. 2014. Norms of age of acquisition and concreteness for 30,000 Dutch words. *Acta psychologica* 150, 80–84.

Burnham, K.P. and D.R. Anderson. 2002. *Model selection and multimodel inference: a practical infromation-theoretic approach*. New York.

Burnham, K.P., D.R. Anderson and K. Huyvaert. 2011. AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behav Ecol Sociobiol* 65, 23–35.

Burquest, D.A. and D.L. Payne. 1993. *Phonological analysis: A functional approach*. Dallas, TX.

Bybee, J. 2007. *Frequency of use and the organization of language*. Oxford.

Bybee, J. 2010. *Language, Usage and Cognition*. Cambridge.

Bybee, J.L. 1994. A view of phonology from a cognitive and functional perspective. *Cognitive Linguistics* 5, 285–306.

Cho, Y.-m. 1991. A typology of voicing assimilation. Halpern, A. (ed.). *Proceedings of the 9th West Coast Conference on Formal Linguistics*, 141–155.

Chomsky, N. and M. Halle. 1968. *The sound pattern of English*. New York.

Clements, G.N. 1990. The role of the sonority cycle in core syllabification. Kingston, J., M. Beckman (eds.). *Papers in Laboratory Phonology I: Between the grammar and the physics of speech*. Cambridge, 282–333.

Coetzee, A. 1969. *Dictionary: Afrikaans–English, English–Afrikaans. Ponouncing dictionary containing more than 39,000 entries.* Johannesburg.

Coetzee, A. 2014. Grammatical change through lexical accumulation: voicing cooccurrence restrictions in Afrikaans. *Language* 90 (3), 693–721.

Colantoni, L. and J. Steele. 2007. Voicing-dependent cluster simplification asymmetries in Spanish and French. Prieto, P., J. Mascaró, M.-J. Solé (eds.). *Segmental and prosodic issues in Romance phonology*, 109–130.

Conradie, C. 2017. Taalverandering in Afrikaans. Carstens, W., N. Bosman (eds.). *Kontemporêre Afrikaanse taalkunde*. Pretoria.

Croft, W. 2000. *Explaining language change: An evolutionary approach*. Harlow, England, New York (Longman linguistics library).

Daland, R. and J. Pierrehumbert. 2011. Learning diphone-based segmentation. *Cognitive Science* 35, 119–155.

Diessel, H. 2007. Frequency effects in language acquisition, language use, and diachronic change. *New ideas in psychology* (25), 108–127.

Diessel, H. 2012. New Perspectives, Theories and Methods: Diachronic change and language acquisition. Bergs, A., L.J. Brinton (eds.). *Historical linguistics of English*. Berlin [a.e.], 1599–1613.

Donaldson, B.C. 1993. *A Grammar of Afrikaans*. Berlin, Boston (Mouton Grammar Library [MGL], 8).

Dressler, W.U. 1989. Markedness and naturalness in phonology: The case of natural phonology. Tomić, O.M. (ed.). *Markedness in synchrony and diachrony*, Vol. 39. Berlin, New York, 111–120.

Dziubalska-Kołaczyk, K. 2014. Explaining phonotactics using NAD. *Language Sciences* 46, 6–17.

Dziubalska-Kołaczyk, K., D. Pietrala and G. Aperliński. 2014. *The NAD Phonotactic Calculator – an online tool to calculate cluster preference in English, Polish and other languages.*

Eiselen, E.R. and M.J. Puttkammer. 2014. Developing text resources for ten South African languages. *Proc. of the 9th International Conference on Language Resources and Evaluation*, 3698–3703.

Grijzenhout, J. and M. Krämer. 2000. Final devoicing and voicing assimilation in Dutch. Stiebels, B., D. Wunderlich (eds.). *Lexicon in Focus*. Berlin, 55–83.

Hansson, G.O. 2001. The phonologization of production constraints: Evidence from consonant harmony. *Chicago Linguistic Society* 37 (1), 187–200.

Hansson, G.O. 2004. Long-distance voicing agreement: An evolutionary perspective. *Berkeley Linguistics Society* 30, 130–141.

Heine, B. and T. Kuteva. 2007. *The Genesis of Grammar*. Oxford.

Hogg, R.M. and C.B. McCully. 1987. *Metrical Phonology: A Course Book*. Cambridge.

Horobin, S. and J.J. Smith. 2002. *An Introduction to Middle English*. Oxford.

Johnson, J. and K. Omland. 2004. Model selection in ecology and evolution. *Trends in Ecology and Evolution* 19 (2), 101–108.

Kotzé, E.F. 1984. Afrikaans in die Maleierbuurt: 'n Diachroniese perspektief. *Tydskrif vir Geesteswetenskappe* 24 (1), 41–73.

Kuijpers, C., W. van Donselaar and A. Cutler. 2000. Phonological variation: epenthesis and deletion of schwa in Dutch. *ICSLP 96: Proceedings, Fourth International Conference on Spoken Language Processing*. [Piscataway, N.J.], 149–152.

Kuperman, V., H. Stadthagen-Gonzalez and M. Brysbaert. 2012. Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods* 44, 978–990.

Langacker, R.W. 2008. *Cognitive grammar: A basic introduction*. Oxford, New York.

Linke, K. 2017. *Coda*. Taalportal. http://www.taalportaal.org/taalportaal/topic/pid/topic-14008431983824103. Accessed 4 October 2017.

MacNeilage, P.F. and B.L. Davis. 2000. On the origin of internal structure of word forms. *Science* 288, 527–531.

Marecka, M. and K. Dziubalska-Kołaczyk. 2014. Evaluating models of phonotactic constraints on the basis of sC cluster acquisition data. *Language Sciences* 46, 37–47.

Mesgarani, N., C. Cheung, K. Johnson and E.F. Chang. 2014. Phonetic feature encoding in human superior temporal gyrus. *Science* 343, 1006–1010.

Monaghan, P. 2014. Age of acquisition predicts rate of lexical evolution. *Cognition* 133, 530–534.

Nakawaga, S. and I. Cuthill. 2007. Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biological Reviews* 82 (4), 591–605.

Pagel, M., Q.D. Atkinson and A. Meade. 2007. Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature* 449 (7163), 717–720.

Ponelis, F.A. 1989. Ontwikkeling van klusters op sluitklanke in Afrikaans. *South African Journal of Linguistics* 7 (1), 1–5.

Prince, A. and P. Smolensky. 2002. Optimality Theory: Constraint Interaction in Generative Grammar. Rutgers Optimality Archive. http://roa.rutgers.edu/files/537-0802/537-0802-PRINCE-0-0.PDF. Accessed 12 June 2013.

R Development Core Team. 2013. *R: A Language and Environment for Statistical Computing*. Vienna, Austria.

Ritt, N. 2004. *Selfish sounds and linguistic evolution: A Darwinian approach to language change*. Cambridge.

Roberge, P. 1993. The formation of Afrikaans. *Stellenbosch Papers in Linguistics* 27, 1–111.

Roberge, P. 2002. Afrikaans: considering origins. Mesthrie, R. (ed.). *Language in South Africa*. Cambridge [etc.], 79–103.

Roberts, I. 2007. *Diachronic syntax*. Oxford.

Rosenbach, A. 2008. Language change as cultural evolution: Evolutionary approaches to language change. Eckardt, R., G. Jäger, T. Veenestra (eds.). *Variation, Selection, Development: Probing the Evolutionary Model of Language Change*. Berlin, 23–74.

Schouten, M. 1982. T-deletie in de stad Utrecht: schoolkinderen en grootouders. *Forum der Lettern* 23, 282–291.

Seliger, H. and R. Vago (eds.). 1991. *First Language Attrition*. Cambridge.

Szemerényi, O. 1996. *Introduction to Indo-European linguistics*. Oxford.

Välimaa-Blum, R. 2005. *Cognitive phonology in construction grammar: Analytic tools for students of English*. Berlin, New York (Planet communication).

van de Velde, F. 2009. *De nominale constituent: structuuren geschiedenis*. Leuven.

van Veen, T. 1964. *Utrecht tussen oost en west: Studies over het dialect van de provincie Utrecht*. Assen.

Watermeyer, S. 1996. Afrikaans English. Klerk, V.A. de (ed.). *Focus on South Africa*, v. 15. Amsterdam/Philadelphia, 99–148.

Wedel, A. 2006. Exemplar models, evolution and language change. *The Linguistic Review* 23 (3), 247–274.

West, B., K. Welch and A. Gałecki. 2015. *Linear mixed models*. Boca Raton, Florida.

Wikimedia Commons. 2005. File:Pseudoceros dimidiatus.jpg. Author: Richard Ling. https://commons.wikimedia.org/wiki/File:Pseudoceros_dimidiatus.jpg. Accessed 10 March 2014.

Wikimedia Commons. 2012. File:Rufous Piculet (male) - Sasia abnormis.jpg. Author: Michael Gillam. https://commons.wikimedia.org/wiki/File:Rufous_Piculet_(male)_-_Sasia_abnormis.jpg. Accessed 10 March 2017.

Wood, S. 2006a. *Generalized Additive Models: an introduction with R*. Boca Raton, Florida.

Wood, S. 2006b. Low-Rank Scale-Invariant Tensor Product Smooths for Generalized Additive Mixed Models. *Biometrics* 62, 1025–1036.

Yang, C. 2000. Internal and external forces in language change. *Language Variation and Change* 12, 231–250.

Appendix

Table A1. Ordinal scores for articulatory features of first and second consonant in the cluster (analogously as in Dziubalska-Kołaczyk 2014; Dziubalska-Kołaczyk, Pietrala and Aperliński 2014).

cluster	MoA1	MoA2	PoA1	PoA2	Voice1	Voice2
fs	3	3	2	4	0	0
ft	3	1	2	4	0	0
ks	1	3	7	4	0	0
kt	1	1	7	4	0	0
lf	6	3	4	2	1	0
lk	6	1	4	7	1	0
lm	6	5	4	1	1	1
lp	6	1	4	1	1	0
ls	6	3	4	4	1	0
lt	6	1	4	4	1	0
lx	6	3	4	7	1	0
mf	5	3	1	2	1	0
mp	5	1	1	1	1	0
ms	5	3	1	4	1	0
mt	5	1	1	4	1	0
ŋk	5	1	7	7	1	0
ns	5	3	4	4	1	0
ŋs	5	3	7	4	1	0
nt	5	1	4	4	1	0
ps	1	3	1	4	0	0
pt	1	1	1	4	0	0
rk	6	1	4	7	1	0
rm	6	5	4	1	1	1
rn	6	5	4	4	1	1
rp	6	1	4	1	1	0
rs	6	3	4	4	1	0
rt	6	1	4	4	1	0
rx	6	3	4	7	1	0
st	3	1	4	4	0	0
ts	1	3	4	4	0	0
ws	7	3	1	4	1	0
wt	7	1	1	4	1	0
xt	3	1	7	4	0	0

Table A2. Derived scores for the variables introduced in 2.2 (Study 1). Note that all numbers are rounded for two digits and represent raw (i.e. not Box-Cox transformed) data.

cluster	ΔMoA	ΔPoA	ΔVoice	frequency	length	AoA
fs	0	0.67	0	15464	6	8.02
ft	0.33	0.67	0	401	7	6.03
ks	0.33	1	0	2098	7	5.91
kt	0	1	0	588	7	7.51
lf	0.5	0.67	1	743.5	3.5	5.29
lk	0.83	1	1	1383	4	5.55
lm	0.17	1	0	551.5	4	6.01
lp	0.83	1	1	2490	4	4.87
ls	0.5	0	1	1253	7	7.13
lt	0.83	0	1	991	7	6.28
lx	0.5	1	1	5192	6.5	8.18
mf	0.33	0.33	1	512	7	9.78
mp	0.67	0	1	1178	4	5.49
ms	0.33	1	1	19260	4	6.78
mt	0.67	1	1	284	6.5	6.9
ŋk	0.67	0	1	618.5	5	5.72
ns	0.33	0	1	1175	7	6.61
ŋs	0.33	1	1	269	5	7.06
nt	0.67	0	1	664.5	7	5.69
ps	0.33	1	0	159	5	10.56
pt	0	1	0	215.5	5	10.65
rk	0.83	1	1	593	4	5.58
rm	0.17	1	0	3545.5	4	4.3
rn	0.17	0	0	1393	4	8.36
rp	0.83	1	1	2360	7	6.41
rs	0.5	0	1	208.5	5.5	6.91
rt	0.83	0	1	626	5	5.49
rx	0.5	1	1	380	4	5.76
st	0.33	0	0	774	5	5.94
ts	0.33	0	0	651.5	5	6.19
ws	0.67	1	1	1237	4	6.59
wt	1	1	1	946	7	7.9
xt	0.33	1	0	646.5	6	5.67

Table A3. Derived scores for the variables introduced in 3.2 (Study 2). Note that all numbers are rounded for two digits and represent raw (i.e. not Box-Cox transformed) data.

cluster	ΔMoA	ΔPoA	ΔVoice	DutProd	AfrProd
fs	0	0.67	0	2	1
ft	0.33	0.67	0	26	0
ks	0.33	1	0	13	18
kt	0	1	0	34	0
lf	0.5	0.67	1	4	12
lk	0.83	1	1	11	5
lm	0.17	1	0	4	0
lp	0.83	1	1	2	7
ls	0.5	0	1	9	6
lt	0.83	0	1	38	35
lx	0.5	1	1	2	5
mf	0.33	0.33	1	1	0
mp	0.67	0	1	4	3
ms	0.33	1	1	1	5
mt	0.67	1	1	8	0
ŋk	0.67	0	1	16	16
ns	0.33	0	1	38	49
ŋs	0.33	1	1	7	0
nt	0.67	0	1	182	131
ps	0.33	1	0	3	1
pt	0	1	0	2	0
rk	0.83	1	1	13	46
rm	0.17	1	0	8	0
rn	0.17	0	0	4	2
rp	0.83	1	1	7	8
rs	0.5	0	1	18	16
rt	0.83	0	1	73	39
rx	0.5	1	1	9	9
st	0.33	0	0	110	2
ts	0.33	0	0	46	28
ws	0.67	1	1	1	0
wt	1	1	1	3	0
xt	0.33	1	0	114	1

LINGUISTIC STABILITY INCREASES WITH POPULATION SIZE, BUT ONLY IN STABLE LEARNING ENVIRONMENTS

ANDREAS BAUMANN[*1]

[*]Corresponding Author: andreas.baumann@univie.ac.at
[1]Department of English and American Studies, University of Vienna, Vienna, Austria

The effect of population size on linguistic stability and evolution has been investigated in different linguistic domains. The relationship among these factors, however, is not always clear. In this paper, we study a basic population-dynamical model of linguistic spread, derive measures of linguistic stability and fitness, and investigate the effect of population size on these measures. By allowing for stochasticity in the learning process of linguistic constituents, it is shown that a constituent's stability and fitness increases with population size, but that high variability in the learning environment may cause constituent loss, also in large populations. The respective roles of learning and usability are also discussed.

1. Population size and linguistic evolution

Population size has been proposed to affect linguistic structure (Atkinson, Kirby, & Smith, 2015; Hay & Bauer, 2007; Lupyan & Dale, 2010; Nettle, 2012; Wichmann, Rama, & Holman, 2011) as well as rate of linguistic change (Atkinson, 2011; Wichmann & Holman, 2009) and degree of adaptation with respect to cognitive and communicative pressures (Fay & Ellison, 2013). More recently, Bromham et al. (2015) have shown in their empirical study that lexical items are more stable in large populations and that rates of word loss are higher in small populations. Indeed, if linguistic constituents share mechanistic similarities with biological replicators (Croft, 2000; Ritt, 2004) the latter observation is exactly what one would expect as per evolutionary theory (Bromham et al. 2015: 2100).

Purely computational approaches to this problem have been, to our knowledge, primarily limited to simulations (Nettle, 1999; Wichmann, Stauffer, Schulze, & Holman, 2008). More recent advances in mathematical ecology (in particular, stochastic epidemiological dynamics; Gray, Greenhalgh, Hu, Mao, & Pan, 2011; Greenhalgh, Liang, & Mao, 2015) allow for a more analytical assessment. This paper adds to the discussion about the relationship between population size, linguistic stability and evolution by modifying and analyzing an established population-dynamical model of linguistic spread (Cavalli-Sforza & Feldman, 1981; Nowak, 2000; Nowak, Plotkin, & Jansen, 2000; Solé, Corominas-Murtra, & Fortuny, 2010; Wang & Minett, 2005). We focus on the dynamics of single 'linguistic items' or 'constituents' (like phonemes, n-phones, words or constructions) in finite speaker populations. After discussing the deterministic dynamics, we also analyze a stochastic version of the model, which accounts for variability in the process of constituent learning (e.g. varying density of the speaker network due to eco-linguistic factors, or varying usage of the constituent in learner-user interactions). It is shown that the general assumption that linguistic stability increases with population size only holds if variability in the learning process is kept low, and argue that the latter factor provides an interesting mechanism in language evolution.

2. Modeling linguistic spread in finite populations

2.1. Deterministic model

We study a modified version of Nowak's (2000) basic model of linguistic spread. In our version of the model, population size N is restricted to be finite. The model describes the dynamics of a structured population composed of users of a particular linguistic item i (e.g. phoneme, n-phone, word or construction) and learners that do not use it. Let U_i and L_i denote the respective sizes of the (disjoint) subpopulations and let $U_i + L_i = N$. Whenever learners and users meet, the former learn i at a rate λ so that they switch from class L_i to U_i. We assume λ to denote the *learning* rate, where learning of a new form is not necessarily restricted to the first years of language acquisition. Rather, we mean any interaction of individuals one of which does not yet know and use a given item. In Nowak's (2000) model, this rate λ is a function of (a) *network density*, linked to the number of communicative encounters a learner is exposed to, (b) *production rate*, i.e. the extent to which the item is produced, and (c) *learnability*, i.e. the probability that the item is successfully acquired when a learner is exposed to it. Learners and users die at a normalized mortality rate of 1 (so that each time unit equals one speaker generation), and dead learners and users are immediately replaced by new individuals that are added to the learner class so that population size is kept constant. In addition, users can switch back to class L_i at a rate γ when they stop using i ('unlearning'), for instance because they forget the item or because they abandon it in favor of a competing linguistic variant. We suggest that γ is inversely related with the *usability* of i in everyday speech events in which no user-learner interactions are involved. Table 1 summarizes the model parameters.

Table 1. Variables in the model and how they can be interpreted

Variable	*Linguistic and cognitive interpretation*
N	Total size of the population of linguistic agents composed of *U users* and *L learners*
λ	Item-specific *learning rate* in interactions; depends on *network connectivity* (linked to number of communicative encounters), *production rate* (linked to utterance frequency and ease of production), and *learnability* (linked to ease of perception)
γ	Rate at which individuals stop using an item (in addition to speaker death; rate of 'unlearning'); inversely related to factors enhancing *usability* (e.g. ease of memorization or ease of production); assumed to be independent from learner-user interactions
R_0	Expected number of learners that successfully learn an innovation from a single user

The dynamics are determined by a deterministic two-dimensional dynamical system in continuous time which models the respective growth rates of L_i to U_i. In what follows we will omit the index i, for the sake of simplicity, since we only focus on the dynamics of a single item (although the parallel evolution of several items clearly can be studied as well). The model equations read:

$$\begin{aligned} \mathrm{d}L/\mathrm{d}t &= \overbrace{-\lambda LU}^{\text{learning}} + \overbrace{\gamma U}^{\text{unlearning}} - \overbrace{L}^{\text{death}} + \overbrace{N}^{\text{birth}} \\ \mathrm{d}U/\mathrm{d}t &= \underbrace{\lambda LU}_{\text{learning}} - \underbrace{(1+\gamma)U}_{\text{death and unlearning}} \end{aligned} \tag{1}$$

If $\gamma = 0$ and $N = 1$ the dynamical system reduces to the model of linguistic spread in Nowak (2000) and Solé (2011), which is equivalent with a one-dimensional model of logistic growth (although the dynamics can be modeled by a single equation, e.g. only the second one in (1), we stick to the more explicit definition for the sake of clarity).

The qualitative behavior of the model can be predicted by the basic reproductive ratio R_0 which is defined as the expected number of learners that learn an item which has been innovatively introduced into the population by a single user (cf. Nowak 2000, Heffernan, Smith, & Wahl, 2005). If $R_0 > 1$ the dynamics approach a non-trivial equilibrium so that $\hat{U} = N(1 - 1/R_0)$ users know and use the item. That is, the item is stably established in the linguistic community. If, however, $R_0 < 1$ then the dynamics approach an equilibrium in which $\hat{U}_0 = 0$ users know the item. In that case, the item drops out of usage. Thus, the basic reproductive ratio functions as a measure of the stability of a linguistic item. For that reason we treat R_0 as a measure of diachronic stability (technically, R_0 measures the stability of the equilibrium $\widehat{U}_0 = 0$; if $R_0 > 1$ then $\widehat{U}_0$ is unstable so that the population of users persists with probability 1 if any users are added to the population; if $R_0 > 1$ then $\widehat{U}_0$ is stable so that the population of users goes extinct with probability 1).

For the present model, the basic reproductive ratio can be shown to read $R_0 = N\lambda/(1+\gamma)$. The formula can be intuitively understood in the following way. The expected time an individual knowing the item remains in the user class is $1/(1+\gamma)$; based on our assumption that the item is an innovation there are (approximately) N individuals that do not yet know the item; and each learner acquires the item at a rate of λ. Note, crucially, that since the amount of individuals which can acquire an item from a user depends on the number of learners available in the population R_0 depends on population size. Here this dependency is linear, which is an immediate reflex of the assumption that the population is homogeneously mixed so that any user can inform any learner in the population (see Section 3 for some discussion).

We are interested in the role that population size plays for the stability of a linguistic item. The basic reproductive ratio R_0 increases with N since $\lambda/(1+\gamma) > 0$. The larger the population, the less likely is it that R_0 falls below one so that the item would inevitably drop out of usage.

In evolutionary terms, $R_0(\lambda,\gamma)$ can be interpreted as a measure the fitness of a linguistic item (Metz, Mylius, & Diekmann, 1996). Evidently, R_0 increases with λ (because $\partial R_0/\partial\lambda = N/(1+\gamma) > 0$) and decreases with γ (because $\partial R_0/\partial\lambda = -N\lambda/(1+\gamma)^2 < 0$). Thus, items with high learning rates and high usability should be selected for. That is, items are expected to evolve in such a way that they maximize ease of acquisition, production and use (probably governed by some trade-off among these factors). Moreover, the effect of optimizing λ and γ gets stronger the larger the population size N, so that items are expected to be less optimized in small populations.

2.2. Stochastic model

Things get slightly more complicated when variability in the model dynamics is considered. For instance, demographic variability could be accounted for, i.e. fluctuations due to random speaker deaths and births in addition to the deterministic model dynamics. For the class of models (1) belongs to, it has been shown that the effects of demographic variability can be neglected if population size is substantially large (Greenhalgh et al. 2015).[1] Another source of variability might be more relevant to linguistic dynamics, namely that of parametric (or environmental) variability. Here, model parameters fluctuate randomly, thus affecting the behavior of all individuals in the population at the same time. In a linguistic setting, for instance, network density of the entire speech community could vary due to eco-linguistic factors (e.g. migration or areal expansion; cf. Mufwene, 2001; Lupyan & Dale, 2010). Likewise, frequency of use of an item established in a speech community might fluctuate due to socio-linguistic or language-internal factors (e.g. morpho-syntactic or phonological restructuring, or emergence of competing variants for instance in language contact). All of these factors can be argued to have an impact on the linguistic learning process. Thus, we include a stochastic component into the model by extending the rate of transition from class L to class U, denoted by $\tilde{\lambda}$, so that $\tilde{\lambda}\mathrm{d}t = \lambda\mathrm{d}t + \sigma\mathrm{d}W(t)$. Here, $W(t)$ is a Wiener process (random noise) which accounts for fluctuation around λ, and $\sigma \geq 0$ is the variance in the 'learning environment' due to the above-mentioned factors. Thus, σ measures the magnitude of these fluctuations. We consider learning environments with low σ as more stable that those with large σ.[2] By replacing $\lambda\mathrm{d}t$ by $\tilde{\lambda}\mathrm{d}t$ in (1), the model becomes a system of stochastic differential equations (SDE; Allen, 2010):

$$\begin{aligned} \mathrm{d}L &= (-\lambda LU + \gamma U - L + N)\mathrm{d}t - \sigma LU\mathrm{d}W(t) \\ \mathrm{d}U &= (\lambda LU - (1+\gamma)U)\mathrm{d}t + \sigma LU\mathrm{d}W(t) \end{aligned} \quad (2)$$

Clearly, if there is no fluctuation ($\sigma = 0$), (2) reduces to the deterministic model (1). System (2) belongs to the class of Itô SDEs analyzed by Gray et al. (2011). Hence, we can employ the conditions for extinction and persistence derived there. By applying Theorem 4.1 in Gray et al. (2011), the basic reproductive ratio for system (2) can be shown to read

$$R_0 = \underbrace{\frac{\lambda N}{1+\gamma}}_{\text{(i)}} - \underbrace{\frac{\frac{1}{2}\sigma^2 N^2}{1+\gamma}}_{\text{(ii)}} \quad (3)$$

where part (i) equals the basic reproductive ratio of the deterministic system (1) and part (ii) comes from the diffusion term in the SDE (2). Theorem 5.1 in Gray et al. (2011) entails that the system leads to persistence of an item (i.e. stable and positive U), if $R_0 > 1$. If, on the contrary, $R_0 < 1$ and $\sigma \leq \sqrt{\lambda/N}$ (Thm 4.1), or

[1] Based on Greenhalgh et al. (2015, Theorem 4.1), demographic variability only has an additional effect if population size falls below critical size $N_{\text{crit}} = 1/4 + (1+\gamma)/\lambda$.

[2] Note that this notion of stability differs from the one measured by the basic reproductive ratio. While σ measures how constantly transmission of an item takes place, R_0 measures whether or not an item persists in the speaker population.

$\sigma > \sqrt{\lambda/N}$ (Thm 4.3), then the number of users U approaches zero with probability 1, so that the item goes extinct.

Several observations can be made. To begin with, it is not difficult to see that $\mathrm{d}R_0/\mathrm{d}N > 0$ if $\sqrt{\lambda/N} > \sigma$. This means that the stability of an item increases with population size N as long as variability is not too high. In particular, inequality $\sigma > \sqrt{\lambda/N}$ is favored to hold (a) if learning variability σ is large or (b) if population size is high (or both). Thus, severe fluctuations promote the loss of items and impede the establishment of new items in the speaker population (Figure 1). Moreover, for fixed σ, larger population sizes can also have negative effects on the stability of linguistic items. In large populations, even mild fluctuations can yield severe reflexes, as long as they affect the entire linguistic population.

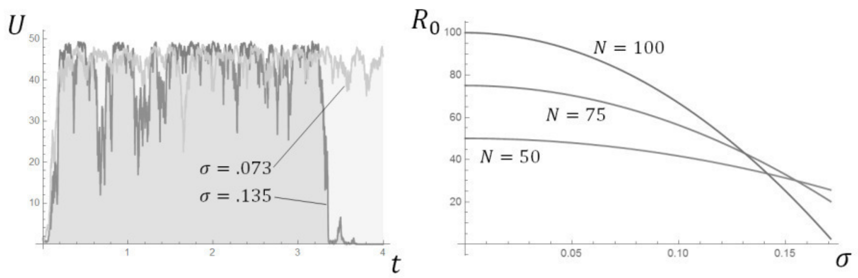


Figure 1. On the left: Itô-process simulations of diachronic developments ($N = 50, \lambda = 0.5, \gamma = 1.5, U(0) = 1$) in two different environments; lower variability ($\sigma = .073$, $R_0 = 7.33$, light gray), and higher variability ($\sigma = .135$, $R_0 = 0.89$, dark gray). After about 3 generations, the item exposed to higher variability in the learning environment goes extinct, as expected. On the right: R_0 as a decreasing function of σ for three different population sizes $N = 50; 75; 100$ ($\lambda = 0.5, \gamma = 1.5$ fixed). For high σ, larger populations yield lower R_0. Computations were done in Mathematica (Wolfram Research, 2016).

What is more interesting is this: a sensitivity analysis reveals information about the relative importance of λ and γ in the optimization of R_0 in the stochastic model. For the respective directional derivatives of $R_0(\lambda, \gamma)$, we have that $\partial R_0(\lambda, \gamma)/\partial(1,0) = N/(1+\gamma) > 0$, and that $\partial R_0(\lambda, \gamma)/\partial(0, -1) = 1/2 \cdot N(2\lambda - \sigma^2 N^2)\log(1+\gamma) > 0$, because $\sqrt{\lambda/N} > \sigma$ if the item already exists stably. Items benefit from increasing λ and decreasing γ (i.e. increasing usability), but in contrast to the former parameter, the effect of decreasing γ suffers from variability in the learning environment. For an item, to put it casually, it pays off to put more effort into improving learning rather than usability if variability is high enough. Improving factors that determine learning does always contribute to an item's success, while effects of increased usability may be vanishingly small in the presence of noise. As in the deterministic case, the effects of optimizing R_0 (i.e. the directional derivatives shown above) get stronger the larger the speaker population (Figure 2).

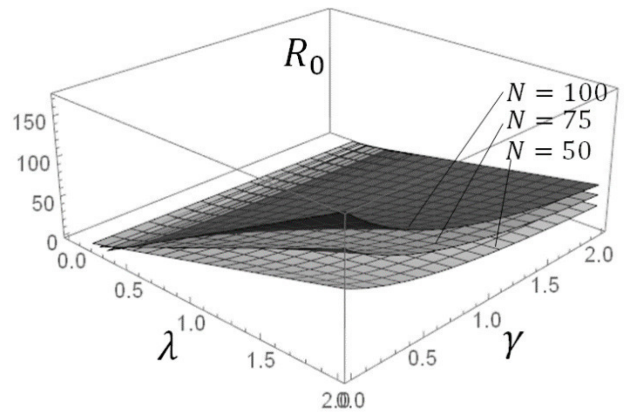


Figure 2. Fitness landscape defined by R_0 as a function of γ and λ for three different population sizes ($N = 50; 75; 100$) in the presence of learning variability ($\sigma = .073$); fitness increases linearly with λ and decreases convexly with γ. Directional slopes (effects of changing parameters) get steeper as population size increases.

3. Discussion and conclusion

By studying systems of ODEs and SDEs, we have shown that population size in general increases the stability of linguistic constituents (cf. Table 2) and thus (a) facilitates their establishment in the speaker population and (b) prevents their loss. This goes in line with Bromham et al. (2015: 2100) who show that "Polynesian languages with larger speaker-population sizes [have] higher rates of gain of new words than their smaller sister languages" and that "languages with a smaller number of speakers [have] higher rates of loss of lexemes". Our results also converge with studies that found a positive correlation between population size and the size of a language's phoneme inventory (see Nettle, 2012 for a review), and by implication phonotactic richness (Maddieson, 2013).

However, the presence of variability in the learning environment decreases stability, and the negative effects of variability get stronger, the larger population size. In the extreme case, this variability can lead to the loss of a constituent (Figure 1, left, dark gray trajectory). As a corollary of this, we can conclude that the establishment of an inventory of constituents (e.g. lexicon of words or phoneme inventory) requires a relatively stable learning environment (cf. McMahon & McMahon, 2013, p. 248). For instance, it can be argued that the small phoneme inventories found in a number of remote Polynesian languages (Trudgill, 2004) might be a reflex of migration and concomitant variability in network density. We argue that complementary to demographic variability (linked to linguistic founder effects as suggested by Atkinson 2011), environmental variability provides another interesting mechanism for explaining linguistic evolution, because it applies even if population size remains constant.[3]

Table 2. Results and model comparison

Feature	*Deterministic model*	*Stochastic model*
Learning environment	Constant ($\sigma = 0$)	Variable ($\sigma > 0$)
Effect of *N* on stability of constituent	Stability increases with population size *N*	Stability increases with *N* if variability σ is small
Effect of *N* on evolution of learning rate λ	Adaptive effects of improving learning increase with *N*	Adaptive effects of improving learning increase with *N*
Effect of *N* on evolution of usability $\sim \gamma^{-1}$	Adaptive effects of improving usability increase with *N*	Adaptive effects of improving usability increase with *N*
Effect of variability on evolution of learning rate λ	Improving learning rate always increases fitness	Improving learning rate always increases fitness
Effect of variability on evolution of usability $\sim \gamma^{-1}$	Improving usability always increases fitness	Effects of improving usability are mitigated by variability σ

One might wonder, what the prediction of the model, that constituent inventories are more likely to shrink in small populations actually means. Clearly, it is not plausible that small populations simply drop constituents like phonemes or lexemes, since some items obviously fulfil specific functions in the linguistic system and cannot be arbitrarily left away. Models like the ones studied in this paper cannot easily account for such details. However, one way of looking at this prediction is this: if constituents vanish (e.g. due to bad adaptation) the language must compensate for this loss, e.g. by adding more complex morpho-syntactic rules. Indeed, this is supported by Lupyan and Dale (2010) who show that small

[3] Indeed, Bybee (2011) has contested demographic variability as the main explanatory link between linguistic evolution and population size.

populations sustain morphologically more complex languages.[4] This argument contrasts the causal directionality proposed by Nettle (2012) who argues that it is the smaller number of contacts in small populations that promotes the acquisition of complex morphology (which, in turn, would allow for a reduced lexicon).

The findings also agree with Fay and Ellison (2013: 7) in the sense that increased population size enhances the optimization of properties associated with linguistic transmission. That is, evolution proceeds faster in large populations. At first sight, this may seem paradox: population size is predicted to increase the stability of an item, but at the same time population size drives linguistic optimization, where an item is effectively replaced by a more successful version of itself. Note, crucially, that the more optimized variant is less likely to get lost.

The analysis of the stochastic model has revealed that constituents always benefit from optimizing factors related to learning while advantages gained from optimizing factors related to usability can be lost due to random fluctuations in the learning environment. Based on this, it can be expected that items are relatively more optimized for being learned easily rather than for ease of use outside of the learning context. This accords with studies that propose a strong connection between (diachronic) stability and ease of acquisition (e.g. Monaghan 2014). It is less compatible with studies stressing the importance of usability and ease of production (i.e. speaker-over-listener dominance) in linguistic transmission (Bybee, 2010; Fay & Ellison, 2013).[5]

Finally, a more technical caveat is in order. The model builds on the assumption that the learning process depends on a mass-action law (i.e., interactions are proportional with the product of the number of learners and users). It has been pointed out (de Jong, Diekmann, & Heesterbeek, 1996), that this assumption does not hold in large populations in realistic ecological scenarios. Consequently, the effect of population size on the basic reproductive ratio is probably overestimated as populations become larger. Accounting for these issues

[4] Note that this observation does not directly follow from the present analysis but rather represents a tentative hypothesis which is compatible with our results. It would be interesting, however, to study a model which includes the possibility of combining items (perhaps similar to the approach adopted by Nowak et al. 2000) to account for complexity. We would like to thank an anonymous reviewer for pointing this out.

[5] This observation, however, might be grounded in the abstract and simplified way in which learning and using constituents is built into the model. Arguably, the rough distinction between factors relevant to learning interactions and those not associated with interactions is very simplistic and must be refined in order to capture learnability and usability more accurately.

eventually requires the implementation of a more complicated network structure.[6] The observations made in this contribution, nevertheless, do not contradict with results from network epidemiology. In large networks, the invasion threshold vanishes under the assumption of a more realistic network structure (small world; scale free). As a consequence of the presence of super spreaders, items can spread easily through large populations (Barabási, 2016). The effects of fluctuations during the learning process in more realistic networks, though, is yet to be looked at more closely.

Acknowledgements

I would like to thank T. Mark Ellison for raising my interest into the discussed phenomenon, and Klaus Hofmann, Niki Ritt as well as three anonymous reviewers for many helpful comments on this contribution. This research was funded by FWF (Austrian Science Fund, grant No. P27592-G18).

References

Allen, L. J. S. (2010). *An introduction to stochastic processes with applications to biology*. Boca Raton, Florida: CRC Press.

Atkinson, M., Kirby, S., & Smith, K. (2015). Speaker Input Variability Does Not Explain Why Larger Populations Have Simpler Languages. *PloS one*, *10*(6), e0129463. https://doi.org/10.1371/journal.pone.0129463

Atkinson, Q. D. (2011). Phonemic diversity supports a serial founder effect model of language expansion from Africa. *Science*, *322*(6027), 346–349.

Barabási, A.-L. (2016). *Network science*. Cambridge: Cambridge University Press.

Bromham, L., Hua, X., Fitzpatrick, T., & Greenhill, S. J. (2015). Rate of language evolution is affected by population size. *Proceedings of the National Academy of Sciences*, *112*(7), 2097–2102.

Bybee, J. (2010). *Language, Usage and Cognition*. Cambridge: Cambridge University Press.

Bybee, J. (2011). How plausible is the hypothesis that population size and dispersal are related to phoneme inventory size? Introducing and commenting on a debate. *Linguistic Typology*, *15*, 147–153. https://doi.org/10.1515/LITY.2011.009

[6] As one of the reviewers of this contribution has rightfully pointed out, reproductive success is – at least partially – a function of network structure, which in turn determines the number of interactions. Here, the functional relationship between population size and network structure is crucial.

Cavalli-Sforza, L. L., & Feldman, M. W. (1981). *Cultural transmission and evolution: aquantitative approach*. Princeton: Princeton University Press.

Croft, W. (2000). *Explaining language change: An evolutionary approach*. *Longman linguistics library*. Harlow, England, New York: Longman.

Fay, N., & Ellison, T. M. (2013). The cultural evolution of human communication systems in different sized populations: usability trumps learnability. *PloS one*, *8*. https://doi.org/10.1371/journal.pone.0071781

Gray, A., Greenhalgh, D., Hu, L., Mao, X., & Pan, J. (2011). A stochastic differential equation SIS epidemic model. *SIAM Journal on Applied Mathematics*, *71*(3), 876–902.

Greenhalgh, D., Liang, Y., & Mao, X. (2015). Demographic stochasticity in the SDE SIS epidemic model. *Discrete and Continuous Dynamical Systems-Series B*, *20*(9), 2859–2884.

Hay, J., & Bauer, L. (2007). Phoneme inventory size and population size. *Language*, *83*(2), 388–400.

Heffernan, J., Smith, R., & Wahl, L. (2005). Perspectives on the basic reproductive ratio. *Journal of The Royal Society Interface*, *2*(4), 281–293. https://doi.org/10.1098/rsif.2005.0042

Jong, M. de, Diekmann, O., & Heesterbeek, J. A. P. (1996). How does transmission of infection depend on population size? In *Epidemic models: their structure and relation to data* (pp. 84–94). Cambridge, UK: Cambridge University Press.

Lupyan, G., & Dale, R. (2010). Language structure is partly determined by social structure. *PloS one*, *5*. https://doi.org/10.1371/journal.pone.0008559

Maddieson, I. (2013). Syllable Structure. In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online.* Max Planck Digital Library. Retrieved from http://wals.info/chapter/12

McMahon, A. M. S., & McMahon, R. (2013). *Evolutionary linguistics*. *Cambridge textbooks in linguistics*. Cambridge [England], New York: Cambridge University Press.

Metz, J.A.J., Mylius, S. D., & Diekmann, O. (1996). When Does Evolution Optimize? On the Relation Between Types of Density Dependence and Evolutionarily Stable Life History Parameters. *IIASA Working Papers*, *96*(004).

Mufwene, S. S. (2001). *The ecology of language evolution*. Cambridge: Cambridge University Press.

Nettle, D. (1999). Is the rate of linguistic change constant? *Lingua*, *108*(2), 119–136.

Nettle, D. (2012). Social scale and structural complexity in human languages. *Phil. Trans. R. Soc. B*, *367*, 1829–1836.

Nowak, M. A. (2000). The basic reproductive ratio of a word, the maximum size of a lexicon. *Journal of theoretical biology*, *204*(2), 179–189. https://doi.org/10.1006/jtbi.2000.1085

Nowak, M. A., Plotkin, J., & Jansen, V. (2000). The evolution of syntactic communication. *Nature*, *404*(6777), 495–498.

Ritt, N. (2004). *Selfish sounds and linguistic evolution: A Darwinian approach to language change*. Cambridge: Cambridge University Press.

Solé, R. V., Corominas-Murtra, B., & Fortuny, J. (2010). Diversity, competition, extinction: the ecophysics of language change. *Journal of The Royal Society Interface*, *7*(53), 1647–1664. Retrieved from http://groups.lis.illinois.edu/amag/langev/paper/sole2010diversitygsc.html

Trudgill, P. (2004). Linguistic and social typology: The Austronesian migrations and phoneme inventories. *Linguistic Typology*, *8*(3), 305–320.

Wang, W., & Minett, J. (2005). The invasion of language: Emergence, change and death. *Trends in Ecology and Evolution*, *20*, 263–269.

Wichmann, S., & Holman, E. (2009). Population size and rates of language change. *Human Biology*, *81*(2), 259–274.

Wichmann, S., Rama, T., & Holman, E. (2011). Phonological diversity, word length, and population sizes across languages: The ASJP evidence. *Linguistic Typology*, *15*(2), 177–197.

Wichmann, S., Stauffer, D., Schulze, C., & Holman, E. (2008). Do language change rates depend on population size? *Advances in Complex Systems*, *11*(3), 357–369.

Wolfram Research, I. (2016). Mathematica. Champaign, Illinois.

Correlates in the evolution of phonotactic diversity: linguistic structure, demographics, and network characteristics

Andreas Baumann & Theresa Matzinger

Abstract. There is an ongoing debate as to whether linguistic structure is influenced by demographic factors. Relationships between these two domains have been investigated on the phonological, morphological and lexical level, mainly drawing on synchronic data and comparative methodology. In this paper, by contrast, we focus on the lesser explored level of phonotactics, and adopt a methodologically orthogonal approach. We investigate the diachronic development of a single lineage, namely English, and compare it with concomitant developments of the demography of the English-speaking population. In addition to linguistic and demographic covariates of phonotactic diversity, we also derive characteristics of the underlying speaker network (network diameter; clustering coefficient). By employing time-series clustering, it is shown that the trajectory of phonotactic diversity most closely matches that of covariates related with density and heterogeneity of the speaker population. Linguistic covariates are less important. We conclude that heterogeneity of the linguistic input a learner is exposed to is the key driving factor in the evolution of phonotactic diversity.

1 Introduction

The structure of a language has been hypothesized to depend on the size of its speaker population (Nettle 2012). This hypothesis has primarily been tested on the phonological level, with a particular focus on the relationship between phonemic richness and population size (Atkinson 2011; Hay and Bauer 2007; Trudgill 2004; Wichmann and Holman 2009; Wichmann et al. 2011). It seems that languages with many speakers tend to have larger phoneme inventories (but see Wichmann et al. 2008 and Bybee 2011 for a more critical assessment). Beyond the phonological level, however, differential relationships seem to hold. For example, it has been shown that on the morphological level populations tend to establish and maintain less complex morphological systems (Atkinson et al. 2015; Lupyan and Dale 2010; see also Bentz and Winter 2013). A more recent study on the lexical level by Bromham et al. (2015) has revealed that words are more likely to get lost in small communities.

Thus, clearly, substantial research into these issues on various linguistic levels has been carried out. However, the phonotactic level, i.e. the domain of sound sequences (Dziubalska-Kołaczyk 2005; Hay and Baayen 2003), which is located between phonology on the one hand and the lexicon and morphology on the other hand has gained much less attention (but see Maddieson 2013; Rama 2013). The present paper fills this gap by assessing the relationship between the diversity of consonant phonotactics and linguistic as well as demographic factors. The domain of phonotactics is interesting because the transmission of sound sequences seems to be rather challenging: sequences of consonants are prone to processes of change as their production and perception is more difficult than that of alternating sequences of consonants and vowels (Berent et al. 2007; Dziubalska-Kołaczyk

2005; Redford 2008). Thus, phonotactic evolution should be particularly sensitive to demographic change.

Apart from the level of investigation, our study furthermore differs from most extant work in terms of methodology and empirical scope. Most of the studies referred to above adopt a comparative approach. That is, characteristics of a number of contemporarily spoken languages are compared to demographic features of their respective speaker populations. For instance, based on a large set of languages a correlation between population size and phoneme inventory size can be detected (e.g. Hay and Bauer 2007). If it is then concluded that the number of phonemes is a function of the number of speakers in the population, then the implicit underlying assumption is that this relationship must have held at any point in time in any single lineage. That is, the respective diachronic developments of the number of phonemes in a language and the size of its speaker population should reflect the very same relationship (provided that the speaker population experienced continuous, i.e. non-catastrophic or 'natural', growth). Testing if the latter holds true requires detailed diachronic data for every language under investigation. These data are unfortunately often not available. Nevertheless, it can be done for single lineages, which is exactly what we provide in our study.

More precisely, we focus on the diachronic development of phonotactic diversity in the American English lineage by investigating the period from 1150 (i.e. the beginning of the Middle English period) up to contemporary English. The shape of the trajectory of phonotactic diversity is then compared against that of potentially related linguistic and demographic factors, such as e.g. morphological structure, population size, population density, and populated area. Moreover, we derive characteristics that are related to the structure of the speaker network, namely a measure of the extent to which individuals in a population form small and (relatively) isolated groups (clustering coefficient) and the diameter of the network (i.e. the maximal number of acquaintances it takes to go from one end of the social network to the other one). While it is clearly impossible to empirically measure these characteristics in large populations (and even more so in the past), they can still be estimated from population size.

Why is it reasonable to take measures such as network diameter, clustering coefficient, populated area or population density, which are directly related to population size, into account? This requires some elaboration. One of the aims of the research on the relationship between demographic and linguistic structure clearly is that this relationship requires plausible underlying mechanisms (Bybee 2011). Several candidates have been proposed for this: for example, Lupyan and Dale (2010) and Bentz and Winter (2013) relate morphological complexity to the number of adult (second language) learners, which is in turn correlated to populated area and population size. Atkinson (2011) and Bromham et al. (2015), on the other hand, argue that founder effects (e.g. demographic bottlenecks) are responsible for differential phoneme-inventory sizes, similar to evolutionary patterns observed in biology. Fay and Ellison (2013) suggest that linguistic structure and population size interact due to increased competition among linguistic items in larger populations (because large populations yield higher variation). Finally, and in a similar vein, Nettle (2012) highlights the role of heterogeneity in the linguistic input a learner is exposed to (although this explanation is ruled out by Atkinson et al. 2015 on the morphological level). Crucially, the latter factor is closely linked to the density and network structure of the speaker population. Dense and mixed populations lead to heterogeneous input while sparse and clustered populations tend to yield homogeneous input. To study aspects of linguistic diversity (as in our case), taking multiple covariates into account thus facilitates detecting mechanistic routes from linguistic structure to population size. As we will show, population density and clustering as well as populated area and network area indeed seem to have a crucial impact on the development of diversity of English consonant phonotactics. Interestingly, linguistic covariates appear to be less relevant.

The paper is structured as follows: in the subsequent section (2) we first discuss processes of change in English phonotactics and introduce two ways of measuring phonotactic diversity as well as the diachronic phonotactic data we made use of (2.1). After that, we discuss linguistic (2.2), demographic (2.3) and network related factors (2.4) and their respective operationalization. All factors enter a time-series clustering analysis which is described in Section 3. The results of this analysis are presented in Section 4. Since we take a relatively large number of potential covariates of phonotactic diversity into account (11 variables in total), it goes beyond the scope of this contribution to discuss all relationships among these covariates in detail. In fact, technically speaking our analysis is based on more than a hundred pairwise comparisons between the respective trajectories of the considered characteristics. We restrict our discussion to those relationships which are relevant to phonotactic dynamics and for which reasonable mechanisms have been proposed. Thus, we finally formulate and discuss seven tentative propositions that can be derived from our analysis in Section 5. Given the methodological architecture of our analysis, it is difficult to make absolute assertions about whether or not two factors, say phonotactic diversity and populations size, hang together. However, what we can do is ranking (groups of) covariates with respect to how closely they relate with the dynamics of phonotactic diversity. For instance, we can say that the trajectories of covariates related with the heterogeneity of the speaker population reflect dynamics in phonotactic diversity better that morphological dynamics do. We think that relative arguments like this nonetheless provide valuable insights into how linguistic structure evolves.

2 Data

In what follows we describe the properties we looked at and the ways in which they have been derived from the data. We use Greek letters (and orange color in plots) for linguistic properties, uppercase Roman letters (and turquoise) for demographic properties, and lowercase Roman letters (and purple) for network related properties. All data are given in Table A2 in the appendix.

2.1 Phonotactic diversity

As stated above, this paper investigates the development of phonotactic diversity in the history of English, from the beginning of the Middle English period up to Contemporary American English. More specifically, we focus on word-final consonant phonotactics, which provide an interesting testing case as the system underwent multiple changes through the history of English (Minkova 1991). These changes include, for example, cluster simplification such as /mb/ > /m/ in *womb* or /xt/ > /t/ in *might*, schwa loss processes as /nəd/ > /nd/ in *stoned*, or sporadic final devoicing /md/ > /mt/ in forms like *dreamt* (Dziubalska-Kołaczyk 2005). Clearly, phonotactic restructuring did not exclusively occur at the end of a word (cf. e.g. /kn/ > /n/ in *knight*). However, being located in a prosodically weak position, word-final diphones are generally more difficult to be acquired, produced, and perceived, which potentially promotes deletion and backgrounding processes and thus change in frequency to a larger extent than word-initially or word-medially (Kirk 2008; Kirk and Demuth 2005). In addition to internal restructuring, language contact and borrowing had a great effect on English word-final phonotactics. This is illustrated e.g. by the increase of /nt/ due to the high productivity of the Latinate suffix *-ment* or the introduction of new cluster types such as morpheme-internal /ts/ in the German loan *blitz*.[1] All of these processes arguably change the frequency distribution and/or the set of phonotactic patters and thereby influence phonotactic diversity.

Our analysis is based on several diachronic corpora (PPCME2, Kroch and Taylor 2000; PPCEME, Kroch et al. 2004; PPCMBE, Kroch et al. 2016; COHA, Davies 2010) covering a period of about 850 years, and taken to reflect the American English lineage. Note that like in any diachronic long-term study, using multiple corpora is unavoidable and entails certain issues such as an increase of authors and genres in the more recent periods covered. This increasing diversity of authors and genres by itself might lead to higher phonotactic richness and higher phonotactic diversity in the more recent data, which can possibly have an undesired effect on the analysis. Why this is however not an issue in our study will be addressed in proposition (3) in our Discussion section.

Our observation period was divided into 17 subperiods of 50 years each (we will use the shorthand notation '1200' for the period from 1200 to 1250, etc.). We used phonological transcriptions and weighted frequencies from ECCE (Ritt et al. 2017) for PPCME2 and PPCEME data (1150-1750), and transcriptions from CMU (Carnegie Mellon Speech Group 2014) for the PPCMBE (1700-1900) and COHA (1800-2000). Note that these phonological transcription data do not include foreign words. As a consequence, these items (which may sporadically occur in the corpora) together with their potentially non-native phonotactic patterns are effectively excluded from the analysis. For each word-final consonant diphone, we extracted per million normalized token frequencies for each subperiod (in the case of overlapping corpora, weighted averages were computed based on the respective corpus sizes). The precise frequencies can be found in Table A1 (see appendix). On the basis of this, we generated discrete frequency distributions $p = (p_1, p_2,...,p_n)$ for each subperiod, where p_i is the (normalized) fraction of diphone i, and n is the number of word-final consonant diphones in that subperiod.

This frequency distribution was then used for deriving two different measures of diversity. First, we computed the diversity number of order 1 (Hill 1973; Tuomisto 2010), which can be defined as $\delta_1 = \exp(H)$, where $H = -\sum_{1\leq i\leq n} p_i \log(p_i)$ is Shannon entropy. This diversity measure is the reciprocal of the weighted geometric mean of the relative

[1] Although not all imported loans resist assimilation; cf. /rx/ > /rk/ in originally Greek *oligarch*.

frequencies p_i, i.e. $\delta_1 = 1/\prod p_i^{p_i}$.[2] As can be seen, δ_1 is largest if all types are equally frequent. Thus, diversity is high if a system has a roughly equal distribution and low if some items are substantially more frequent than others. Second, we measured diversity like Rama (2013) by retrieving the total number of diphone types with $p_i > 0$, δ_0 (equivalent to the diversity number of order 0; Hill 1973).

It can be shown that δ_1 is located between the Simpson index (which is the reciprocal of the arithmetic mean) and the total number of types δ_0 (Hill 1973), i.e. $\delta_1 < \delta_0$. The latter measure, i.e. the diversity number of order 0, is inaccurate in that it possibly overestimates diversity. This is so because rare types (e.g. a diphone type with a single token) contribute as much to δ_0 as frequent types. Given that linguistic items typically follow Zipfian (and crucially non-uniform) frequency distributions (Figure 1, bottom row), this can be problematic. Hence, it pays off to cover the frequency sensitive measure δ_1 as well.

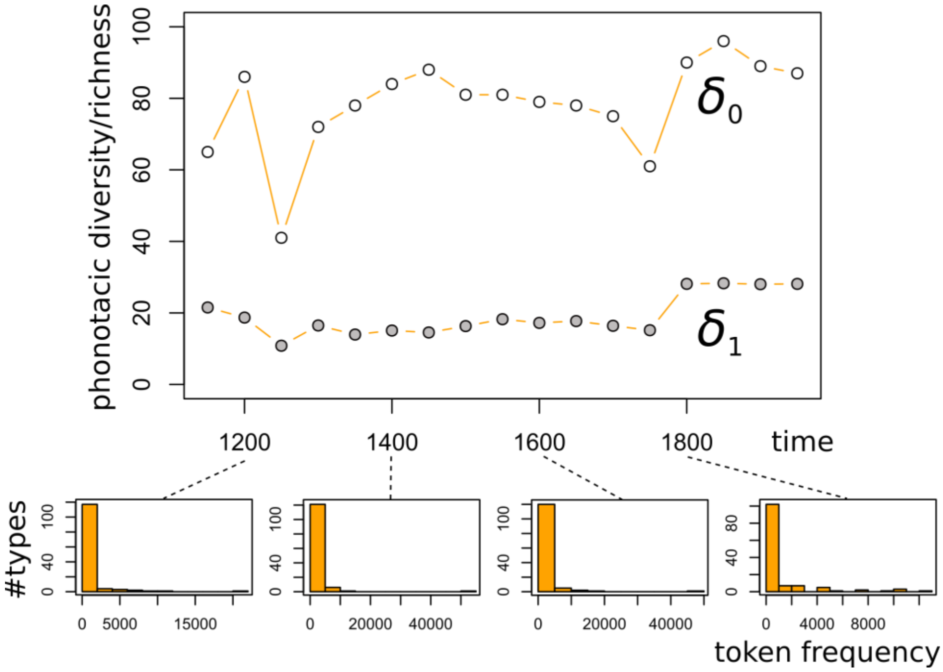


Figure 1. Trajectories of phonotactic richness δ_0 (white dots) and phonotactic diversity δ_1 (gray dots), respectively, from Middle English to Contemporary American English. Bottom row: histograms showing the frequency distributions of cluster types in four periods (1200; 1400; 1600; 1800). All distributions are clearly non-uniform. It can be seen that the right-most distribution (1800) lacks extremely frequent items, which implies higher δ_1.

We derived δ_1 and δ_0 for every subperiod of 50 years in order to obtain the trajectory which describes the diachronic development of phonotactic diversity, shown in Figure 1 (all computations were done with the `entropart` package in R; Marcon and Herault 2013; see Script A1). In what follows we will use the term 'phonotactic diversity' for δ_1 and 'phonotactic richness' for δ_0.

[2] Note that taking Shannon entropy H instead of δ_1 does not qualitatively change our results in any way.

2.2 Linguistic factors

Four main linguistic covariates of phonotactic diversity are covered in this paper: phoneme-inventory size, analyticity, syntheticity, and utterance frequency. Phoneme-inventory size is one of the most prominent features studied in connection with social structure (Atkinson 2011; Hay and Bauer 2007; Trudgill 2004). Furthermore, and highly relevant for the present study, it has been demonstrated that, at least on a comparative scale, there is a weak but statistically robust correlation between consonant-inventory size and phonotactic richness (Maddieson 2013). Since we focus on consonant phonotactics, we ignore vowels and restrict our analysis to the development of the consonant inventory throughout the history of English. To this end, the number of consonants κ (including allophones) was retrieved from the literature for each subperiod (Hogg 1992; Horobin and Smith 2002).

Second, when investigating phonotactic evolution, developments in morphology must be taken into account. This holds especially when considering English diphones in word-final position, since these are often produced by suffixation (Hay and Baayen 2003). To address this, we used syntheticity and analyticity scores from Szmrecsanyi (2012), derived from historical corpora (PPCME2, PPCEME, PPCMBE) to account for morphological co-developments in the diachrony of English. In that study, the syntheticity of a text is defined as the fraction of word forms featuring an affix while analyticity is defined as the fraction of free grammatical morphemes in the text (normalized to 1,000 tokens; Szmrecsanyi 2012: 657). In our case, for each subperiod, syntheticity σ and analyticity α were computed as the mean syntheticity and analyticity of all texts in this subperiod. In the lack of more recent data, syntheticity and analyticity were assumed to have remained constant in the latest two subperiods (1900; 1950).

Finally, the number of cluster tokens ν (normalized per million tokens) in each subperiod was retrieved. That is, ν measures the overall size of the cluster population, functioning as a proxy for utterance frequency, which has been repeatedly shown to be relevant to linguistic evolution (Bybee 2007; Diessel 2007). Note that ν provides information which is not contained in the diversity number of order 1, i.e. δ_1, since the latter is derived from a probability distribution rather than a distribution of actual token frequencies.

2.3 Demographic factors

In order to estimate the size of the English-speaking population N we used data of the population sizes of Great Britain and the U.S. as proxies. Population size estimates for England start in 1086 and for Great Britain around 1700 (Broadberry et al. 2015; Wrigley and Schofield 1981). Census data for Great Britain exist from 1801 up to 2011 (Great Britain Historical GIS Project 2017; Office for National Statistics 2012). Estimated population data for American colonies, roughly corresponding to present day U.S. states, are available from the year 1620 onwards (U.S. Bureau of the Census 1975). More accurate data for the individual U.S. states exist from the year 1790, when the first official census took place, up to 2010 (1790-1990: U.S. Census Bureau 2002; 2000-2010: U.S. Census Bureau 2011).

Besides the population size N, we also used population density $D = N/A$ as a possible demographic explanatory factor for phonotactic diversity, with A specifying the area inhabited by the English-speaking population in the respective periods of interest. The data for the area A was calculated by summing up the total areas of England (Office for National Statistics 2013) and of those U.S. states which were inhabited and for which population data were available in the respective years (U.S. Census Bureau 2012).

This way of collecting demographic data of English speakers comes with certain issues. First, it is not known how many of the people recorded in the censuses are/were actually speakers of English. Second, the areas used for the calculations can only be seen as rough estimates because we do not take into account which parts of the respective states were inhabited at which times, how fast populations spread within the states, or that the borders of the individual colonies/states might have undergone slight changes over the years. Furthermore, we only focus on England and the U.S. and deliberately excluded other English-speaking populations such as Australian, Irish, or Scottish speakers of English, since our specific aim is to model the evolution of the lineage from Middle English to Contemporary American English. However, this issue is in general negligible, as all other English-speaking populations are relatively small compared to the American population. Hence, we would not expect our results to change much if additional variants were considered.

An additional factor that might have influenced phonotactic diversity is the establishment of English as a lingua franca and English as a second language of many speakers all over the world (Crystal 2012). Furthermore, the fraction of non-native speakers of English in England and the U.S. (particularly in early American demographic data) is difficult to quantify. For these reasons, the population sizes and densities used in our calculations might not reflect the actual numbers with absolute accuracy and might disregard certain tendencies. Still, our data provide information about the dimensions of the demographic factors, which will suffice for our purpose.

2.4 Network characteristics

Explanations for the relationship between linguistic structure and population structure are sometimes based on network properties of the population rather than on population size itself. For instance, Nettle (2012) argues that learners in small communities tend to be exposed to more homogeneous (and thus less diverse) linguistic input either because (a) learners in small communities interact with a smaller number of informants (cf. also Hay and Bauer 2007) or because (b) small communities tend to be more clustered. In the language of network science, (a) means that small networks tend to have a smaller average degree, i.e. a smaller average number of links per individual, while (b) means that small networks exhibit a higher clustering coefficient, i.e. a higher tendency to form small groups.

Real-world social networks have been demonstrated to have two important properties. First, they have a relatively small diameter, so that it does not take many steps in the network to get from one individual to another arbitrarily chosen individual (‘small-world property’). Second, the distribution of the number of links per individual is based on a power law (i.e. there are few individuals with many links and many individuals with just a few links; ‘scale free property’). One network model which has both these properties is the Barabási-Albert

(BA) model (Albert and Barabási 2002; Barabási and Albert 1999).[3] It has been shown that the clustering coefficient in BA networks decreases as the size of the network increases so that observation (b) above is sound from a formal perspective.

Looking at network properties more closely makes sense for a second reason. In studies on the relationship between linguistic structure and population size, the latter variable is often log-transformed (Bromham et al. 2015; Hay and Bauer 2007; Lupyan and Dale 2010). While this clearly has methodological reasons in the first place (due to the requirements of (generalized) linear regression modeling), we would like to point out that the logarithm of the size of a network with the small-world property is (approximately) proportional with its diameter. Consequently, apart from being methodologically necessary, log-transformed population size has an immediate interpretation in terms of network structure, namely the maximum number of steps it takes to go from one end of the network to the other one.[4]

We derive two network characteristics from population size N: network diameter d and the network's clustering coefficient c. We assume that the network underlying the English speaker population can be described by the BA-model and consequently base our derivations for the diameter and clustering coefficient on it. We set $d = \log N \,/\, \log\log N$ and $c = (\log N)^2/N$ (Barabási 2016) for each subperiod. Note that $d \propto \log N$ for low N.

3 Calculation

We used model-based autocorrelation-driven time-series clustering to investigate similarities between the derived trajectories (Galeano and Pena 2000). In a nutshell, this involves three steps: (i) fitting a model to each time series; (ii) computing pairwise dissimilarities based on the respective autocorrelation functions; and (iii) clustering the time-series based on the derived dissimilarity matrix. These steps and analytic choices are described in more detail below. All analyses were done in R (R Development Core Team 2013). The corresponding code is provided in Script A2 in the appendix.

3.1 Generalized additive modeling of time series

In general, time-series clustering does not require a model based approach. Similarities between time series can be based on actual scores as well (as opposed to scores predicted from a model). In our case, we find that some of the derived trajectories show fluctuations (in particular syntheticity and analyticity). Since we are primarily interested in general trends rather than noise structure we opted for a model based approach.[5] We used generalized additive models (GAM, Wood 2006) as they allow for an easy implementation of smooth terms which can capture the non-linear (and more generally, non-monotonous) nature of the trajectories. For each of the factors described in the previous section a separate GAM was

[3] Other prominent candidates are the Erdös-Rényi (ER, random) model and the Watt-Strogatz (WS) model. Both fulfil the small-world property, however, they are not scale free. See Watts and Strogatz (1998); Zaki and Meira (2014).

[4] The same holds, *mutatis mutandis*, for average path length.

[5] We would like to point out, however, that the overall results do not change much if actual scores are used.

computed in which time figures as smooth predictor (`mgcv` package in R; Wood 2006).[6] Based on the respective model, we then computed predicted values for each subperiod and each trajectory.

3.2 Autocorrelation based time-series dissimilarity

We employed autocorrelation dissimilarity to assess the extent to which two time-series match. Under the assumption that all data points in the time series are equally important, autocorrelation dissimilarity between two time series is defined as the Euclidian distance between the respective autocorrelation functions (Galeano and Pena 2000; Montero and Vilar 2014). Given our data, this has multiple advantages. First, in contrast to Minkowski distance or Pearson-correlation distance between two time-series vectors, autocorrelation dissimilarity preserves the temporal structure, since the former are invariant with respect to permutations.

Second, in contrast to dissimilarity measures which are based on computing pairwise distances between actual data points (e.g. Minkowski, Fréchet, Dynamic time warping), autocorrelation dissimilarity does not depend on the orientation of the measured variables. If, say, one trajectory goes upwards, while another trajectory more or less mirrors its development by going downwards, the autocorrelation-based approach detects a similarity. That is, the procedure is invariant with respect to flipping the scale of the observed feature. This is not the case with Dissimilarity measures built on pairwise distances between data points.

Third, autocorrelation dissimilarity implicitly normalizes scores (since it is the correlation coefficients of the lagged time series given by the autocorrelation function which are compared to each other). This is important given the different scales of the features observed in this study (e.g. word count vs. people per area). In our analysis, autocorrelation similarity was based on the subperiod-wise values predicted from the GAMs rather than actual data points. The `TSclust` package (Montero and Vilar 2014) was used to compute a dissimilarity matrix in this way.

3.3 Hierarchical time-series clustering

In the third and final step, hierarchical clustering with complete linkage was then applied to the derived dissimilarity matrix. Other agglomerative clustering methods (weighted/unweighted average, centroid, minimum variance; (Murtagh and Contreras 2012; Zaki and Meira 2014) led to qualitatively identical results (only showing quantitatively different branch lengths). Single linkage resulted in a long chain, which we consider as less informative, but the main results of this study still apply. Median linkage moved phonotactic richness to a separate exterior branch. We thus consider the choice of the clustering method as relatively robust. For terminological clarity, we will refer to clusters in the resulting dendrogram as patches.

---

[6] In order to prevent over- as well as underspecification, we used thin-plate regression splines which may be shrunk to zero and selected the initial number of basis functions *k* with the help of gam.check under the condition that 2*k* is less than the number of data points (i.e. 17).

4 Results

A dendrogram together with the trajectories of all variables is shown in Figure 2a. Judging from the scree plot (Figure 2b), three major patches of variables can be identified: (i) number of cluster tokens, analyticity and syntheticity; (ii) population size and number of consonants; (iii) phonotactic richness (potentially forming a separate patch), clustering coefficient, phonotactic diversity, population density, populated area and network diameter.

A couple of remarks are in order. Patch (i) consists exclusively of linguistic variables. Here, number of tokens and analyticity exhibit similar trajectories, while the development of analyticity is inversely related. However, linguistic variables can be found in all clusters. Demographic variables are in patch (ii) and (iii), while network characteristics are located only in patch (iii). Population size and number of consonants in patch (ii) closely correspond to each other, albeit in an inverse way. Within patch (iii), phonotactic diversity, network diameter, population density and populated area are patched closely together. Subsequently, clustering coefficient and finally phonotactic richness is added to patch (iii). Phonotactic richness is more loosely linked to patch (iii), which is reflected by the relatively high curvature at position 4 in the scree plot (see Figure 2b; note that as reported before, the location of phonotactic richness in the dendrogram was sensitive to the choice of the clustering algorithm). As can be seen, the development of the clustering coefficient is inversely related to that of the remaining variables in cluster (iii). We will discuss and interpret these observations below.

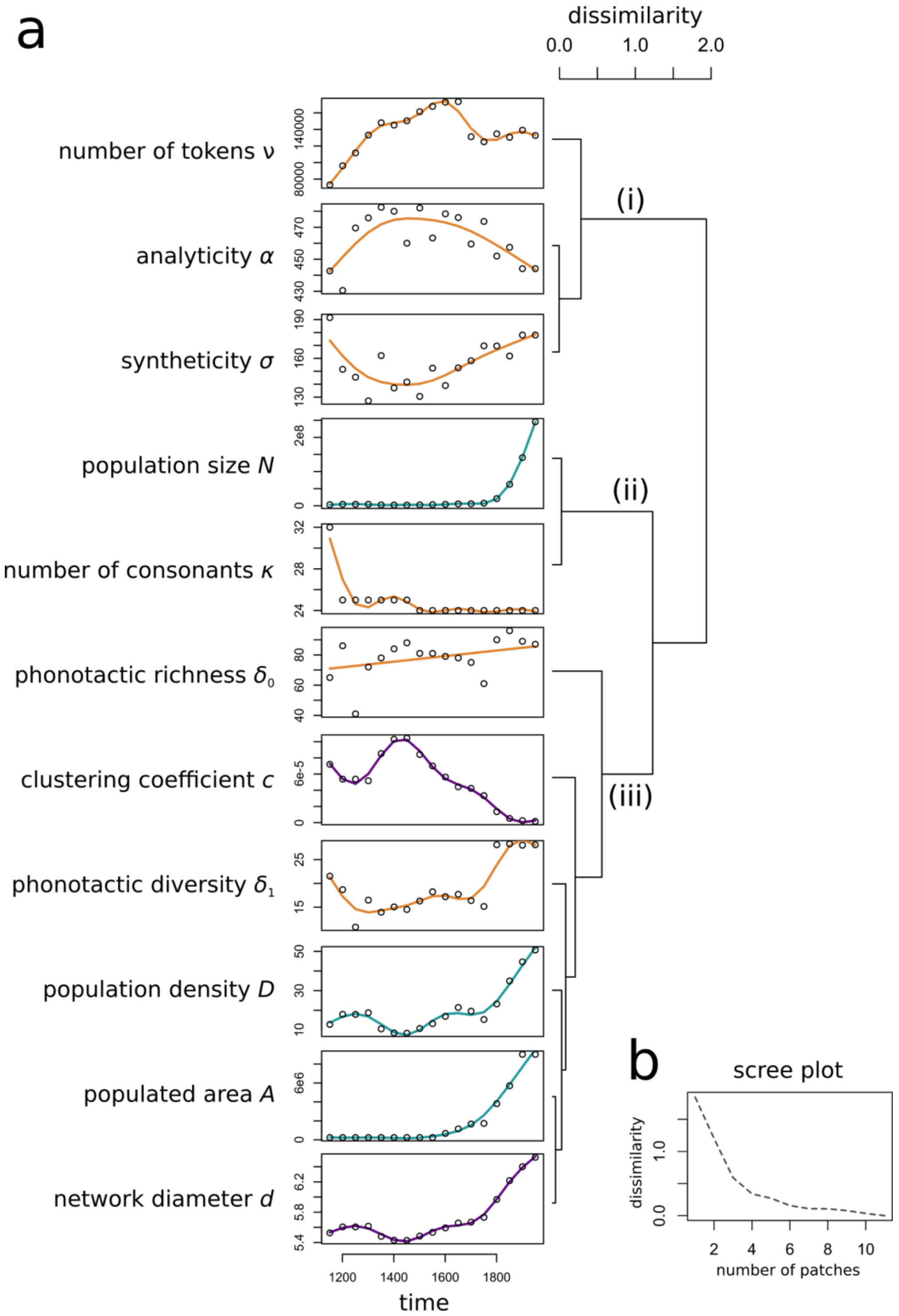


Figure 2. (a) Dendrogram resulting from time-series clustering. Patches of similar trajectories are numbered from (i) to (iii). (b) Scree plot for the dendrogram justifying the choice of patches.

5 Discussion

We summarize our results in seven separate tentative propositions, focusing on phonotactic diversity and phonotactic richness, as well as on potential mechanisms that might be responsible for the observed relationships. From (1) to (7), we move from relatively tight to loose relationships, reflecting the patches in Figure 2.

(1) *Phonotactic diversity increases as populations expand.* Populated area and network diameter are similar in that they reflect the expansion of the speaker population. However, while populated area reflects geographic distance, network diameter measures the maximum social distance in a population. We suggest that two different mechanisms are responsible for this correlation. First, it might be the case that widely spread populations tend to establish heterogeneous phonotactic systems because the likelihood of communicative events - and therefore the necessity to share the same code - decreases as distance increases. In this way, innovations can be stably established on one end of the speaker network, and subsequently spread through the whole population through social hubs (Labov 2007; Pei et al. 2014; Rogers 2003; Stein 2011). Second, the increase in phonotactic diversity might be simply attributed to increased language contact (which correlates with populated area). Obviously, English as a globally spoken language is heavily exposed to linguistic contact (Crystal 2012). As mentioned above, import via loans is a relevant factor in the evolution of the phonotactic inventory.

(2) *Phonotactic diversity is high in mixed populations*. The clustering coefficient reflects to what extent members of a network form small and relatively isolated groups. Low clustering means that the population is more mixed, and therefore leads to heterogeneous linguistic input (cf. Nettle 2012). Likewise, high population density entails more heterogeneous linguistic input because if many people live in a small area, chances of being linguistically isolated are small. Arguably, heterogeneous linguistic input then leads to higher phonotactic diversity, because speakers are more likely to be exposed to innovations (an equivalent mechanistic relationship can be found in epidemiology: diseases spread faster in dense and mixed populations; Anderson and May 1991).

(3) *Phonotactic diversity is loosely related with phonotactic richness.* At first sight, this result is surprising given that both measures are operationalizations of diversity. However, as explained in 2.1, phonotactic richness is not sensitive to token frequencies and consequently overestimates diversity. While the development of phonotactic diversity is roughly U-shaped, we only see a slight increase in phonotactic richness. Thus, phonotactic diversity provides a more nuanced picture. On average, the size of the phonotactic inventory remained constant throughout time but individual clusters varied in terms of utterance frequency. These results also let us conclude that the potential pitfalls of using multiple corpora (as outlined in 2.1) are not an issue in our case. If the higher number of authors and genres in the more recent periods of investigation were to influence phonotactic richness and phonotactic diversity and if we assume novel cluster types to occur with low token frequencies, we would expect to see a higher increase in phonotactic richness than in phonotactic diversity in these periods. This is however not the case in our data.

(4) *Phonotactic diversity and richness slightly depend on the size of the consonant inventory*. At first sight, this is surprising, since in our study phonotactic

diversity/richness and number of consonants are inversely related, whereas Maddieson (2013) found a positive significant relationship between phonotactic richness and number of consonants. However, this relationship was shown to be only weak (cf. also Fenk-Oczlon and Fenk 2016, who failed to show a statistically robust relationship between number of phonemes and syllable complexity). The causal relationship between phonotactic diversity/richness and phonemic richness is not entirely clear. On the one hand, it seems plausible that more consonants allow for a larger variety of combinations of consonants. On the other hand, however, it can be argued that the pressure of maintaining consonants in the phonemic system decreases if a language allows for a larger variety of consonant sequences because both consonants as well as consonant sequences can be used to disambiguate on the lexical level. Conversely, languages featuring many consonants are not urged to form consonant sequences, which are articulatorily and perceptually cumbersome. Thus, a trade-off between phonotactic richness and phonemic richness is equally plausible. This is in line with a suggested trade-off among paradigmatic and syntagmatic complexity (Good 2015; Nettle 2012; Sinnemäki 2016). Consequently, we find the slightly negative relationship discovered in our data not particularly astonishing.

(5) *Phonotactic diversity slightly depends on population size*. Here, two different yet related observations can be made. First, the weak relationship between phonotactic diversity (or richness) and population size might be due to the fact that it should be at best indirect in any case. Founder effects (Atkinson 2011; Bromham et al. 2015) only apply to small populations so that there must be other mechanisms which link diversity in a linguistic (sub)system and population size. As outlined above (as well as in 2.4), such mechanisms can be found in network characteristics that are related with population size. The particular network properties we have taken into account are the tendency to form clusters (which decreases in network size) and network diameter (which increases in network size). Both characteristics have been argued to affect the linguistic input of a learner (see (1-2) above). This brings us to the second observation. In general, the diameter of a small-world network of size N is proportional to $\log(N)$ (and in the BA-model at least for $N < 10^4$, Barabási 2016). In fact, most extant empirical studies relating population size with linguistic structure examine log-transformed population size for methodological reasons. Thus, the detected relationships in these studies can be effectively interpreted as holding between linguistic structure and network diameter. Interestingly, $\log(N)$ has another interpretation: in urban areas, population density is a linear function of the logarithm of N (Craig and Haskey 1978), thus providing another mechanistic route for the relationship between linguistic structure and population size; see proposition (2) about the positive correlation between density and diversity above.

(6) *Phonotactic diversity only remotely depends on overall utterance frequency.* The underlying rationale for why such a relationship could be expected is this: if speakers of a language get used to complex phonotactic sequences (e.g. frequent clusters like /nd/ or /st/), i.e. if the process of concatenating multiple consonants is already entrenched, then the likelihood of adopting new (and particularly similar) types of complex sequences increases (cf. studies on the relationship of cluster repair and

previous exposure to consonant sequences in English and Russian by Berent et al. 2007 and Redford 2008). This in turn increases phonotactic diversity and richness. However, it seems that this was not the case in the development of English coda phonotactics. Why? Assume two hypothetical cluster inventories with the same relative distribution of clusters $p = (p_1, \ldots, p_n)$ (and consequently identical diversity) which only differ in utterance frequency (e.g. all cluster types in inventory 1 are twice as frequent as their counterparts in inventory 2). We suggest two potential explanations for why the less frequent inventory is more likely to increase in diversity. First, it can be argued that a new cluster faces stronger competition in the high-frequency inventory than in the low-frequency inventory. For instance, if a new cluster is articulatorily similar to an established cluster, then the former is more likely to analogically adapt to the established competitor if the latter is highly frequent (Schryver et al. 2008). As a consequence, innovative clusters are expected to extend the inventory (and thus to increase diversity) if overall utterance frequency is low. Second, the eroding effects of frequency might cause loss of (articulatorily difficult) consonant clusters, in particular in adult speakers (Bybee 2007, 2010). Thus, phonotactic richness could as well decrease with utterance frequency.[7] Overall, we conclude that utterance frequency can have opposing effects on phonotactic diversity, which may be reflected in the weak relationship between these variables that we found in our study.

(7) *Phonotactic diversity only remotely depends on morphological factors*. Suffixation is clearly involved in the creation of word-final consonant clusters (e.g. by the past tense suffix *-ed*) and it has been argued that morphologically produced clusters enforce the establishment of phonemically identical stem-internal clusters (Hogg and McCully 1987). Consequently, the weak correlation between syntheticity and phonotactic diversity is surprising. We would like to point out, however, that before the population explosion in the 19th and 20th century both trajectories are relatively similar. Lupyan and Dale (2010) demonstrated that morphological complexity (and in particular use of inflection, cf. Figure 2 therein) negatively correlates with population size. The latter might be reflected in the relatively weak correlation between syntheticity and population size in our data. Lupyan and Dale (2010) argue that languages spoken by many (adult) speakers distributed over a large area face stronger learning pressures, which in turn drives morphological simplification. Since a similar argument arguably could as well apply to the sublexical level, we have to assume that phonotactic acquisition poses less problems to adult learners than morphological acquisition does (see also point (1)). Indeed, Diessel (2012) among others has argued that many sublexical innovations are introduced by adolescent or adult speakers.

---

[7] There is another more technical argument. Assume that for some $C > 1$, $\left(u_1^{(1)}, \ldots, u_n^{(1)}\right) = C \cdot (u_1^{(2)}, \ldots, u_n^{(2)})$, where $u_i^{(1)}$ and $u_i^{(2)}$ denotes the utterance frequency of cluster i in inventory 1 and 2, respectively. For the overall frequencies of cluster tokens holds $\nu^{(1)} < \nu^{(2)}$. Let us say that a new cluster type labeled $n + 1$ successfully enters the language e.g. in some imported loan. It can be expected that this loan is at least initially used at the same utterance frequency u_{n+1}, regardless of the present phonotactic inventory. But then u_{n+1} increases phonotactic diversity δ_1 more if overall utterance frequency is low, because for the respective proportions of cluster $n + 1$ in the two inventories holds $u_{n+1}/\nu^{(1)} > u_{n+1}/\nu^{(2)}$.

It can be seen that phonotactic richness is more strongly related with demographic factors than with linguistic ones. This should become clear from the respective strengths of the relationships discussed in (1-7). It is evidently not population size itself which affects phonotactic diversity but rather intermediate factors related to population size – namely density, area, clustering and social distance – which contribute to phonotactic diversification. Two causal mechanism that are directly linked to these intermediate factors are (a) differential heterogeneity and variability in the linguistic input a learner of a language receives (Fay and Ellison 2013; Nettle 2012) and (b) language contact (Mufwene 2001; Trudgill 2001, 2004).[8] Our results support both of them.

What we would like to stress at this point are the consequences for using phonotactic richness as a measure of linguistic time depth (i.e. the 'age' of a language family), as proposed by Rama (2013). Rama (2013) suggests that differences in phonotactic richness work well for measuring time depth because it "represents the net result of phonological erosion, morphological expansion and fusion". However, if demographic factors such as population density have a larger impact on phonotactic richness than changes in the phonological or morphological domain do, then time-depth estimates derived from differences in phonotactic richness should be corrected for demographic differences in order to prevent them from merely reflecting differential increase in the density of speaker populations. Although we have here only demonstrated this for the English lineage, we expect that similar issues are present across languages.

6 Conclusion

In this study we investigated linguistic, demographic and network related correlates of English phonotactic diversity in its evolution from the Middle English period to Contemporary English, with a particular focus on word-final consonant phonotactics. Consonant clusters are articulatorily and perceptually challenging linguistic items so that any pressures on their transmission are expected to be more clearly visible compared to less complex sound sequences. Deriving two measures of diversity, we found that the diachronic trajectory of phonotactic diversity in English matches best with developments in demography as well as network characteristics. The relationship between phonotactic diversity and linguistic covariates, on the other hand, is considerably weaker.

Our research thus contributes to the general discussion about the link between linguistic and social structure (Atkinson et al. 2015; Atkinson 2011; Bromham et al. 2015; Bybee 2011; Hay and Bauer 2007; Wichmann and Holman 2009). It suggests that demographic properties such as population density or populated area as well as network characteristics are closely related with certain linguistic subsystems (in our case phonotactics). In this regard, we have argued (very much in line with Nettle 2012) that it is factors that determine the heterogeneity of phonotactic input a learner is exposed to which are most relevant to phonotactic diversity. This stands in contrast with the morphological level,

[8] Import of non-native phonotactic structures via language contact obviously affects the linguistic input for the subsequent generation of learners. Hence, both (a) and (b) can be argued to feed the same mechanism.

where it has been shown that homogeneity in the input has no significant impact on complexity (Atkinson et al. 2015). Explicitly cognitive and linguistic covariates such as increased pressure on learnability due to increased contact, utterance frequency (and entrenchment), or the size of the set of segments available to compositional processes (here: consonants) seem to be generally less relevant. This provides further support for the claim that the development of a particular linguistic system should not be studied in isolation (Mufwene 2001; Trudgill 2001, 2004).

On a methodological level, we would like to make a case for employing more nuanced measures of diversity in linguistic systems. We have demonstrated that diversity measures that take token frequency into account are more informative (and likely fit better to what is intuitively understood as diversity). Evidently, using these measures comes at the cost of requiring more data (corpus data; word-frequency lists; phonological transcriptions).

Our approach is clearly limited in that we only looked at one single lineage, while most extant research is comparative, investigating up to thousands of languages. Although doing so would be desirable for diachronic investigations as well, this is impeded by the fact that diachronic data as well as information on the demographic developments of their respective speaker populations is simply not available due to multiple reasons. In this regard, our approach should be understood as complementary in that it neglects the comparative dimension in favor of the diachronic dimension. Ideally, both dimensions should be treated in tandem, but given the relatively sparse diachronic data currently at hand, this is a matter of future research.

References

Albert, R. and Barabási, A.-L. (2002) 'Statistical mechanics of complex networks', *Reviews of Modern Physics*, 74/1: 47–97.

Anderson, R. and May, R. (1991) *Infectious diseases of humans*: *Dynamics and control*. New York: Oxford University Press.

Atkinson, M., Kirby, S. and Smith, K. (2015) 'Speaker Input Variability Does Not Explain Why Larger Populations Have Simpler Languages', *PloS one*, 10/6: e0129463.

Atkinson, Q. D. (2011) 'Phonemic diversity supports a serial founder effect model of language expansion from Africa', *Science*, 322/6027: 346–9.

Barabási, A.-L. (2016) *Network science*. Cambridge: Cambridge University Press.

Barabási, A.-L. and Albert, R. (1999) 'Emergence of Scaling in Random Networks', *Science*, 286/5439: 509–12.

Bentz, C. and Winter, B. (2013) 'Languages with More Second Language Learners Tend to Lose Nominal Case', *Language Dynamics and Change*, 3/1: 1–27.

Berent, I., Steriade, D., Lennertz, T. and Vaknin, V. (2007) 'What we know about what we have never heard: Evidence from perceptual illusions', *Cognition*, 104/3: 591–630.

Broadberry, S., Campbell Bruce M. S., Klein, A. and Overton, M. *et al.* (2015) *British Economic Growth 1270-1878*. Cambridge: Cambridge University Press.

Bromham, L., Hua, X., Fitzpatrick, T. and Greenhill, S. J. (2015) 'Rate of language evolution is affected by population size', *Proceedings of the National Academy of Sciences*, 112/7: 2097–102.

Bybee, J. (2007) *Frequency of use and the organization of language*. Oxford: Oxford University Press.
—— (2010) *Language, Usage and Cognition*. Cambridge: Cambridge University Press.
—— (2011) 'How plausible is the hypothesis that population size and dispersal are related to phoneme inventory size? Introducing and commenting on a debate', *Linguistic Typology*, 15: 147–53.
Carnegie Mellon Speech Group (2014) 'Carnegie Mellon Pronouncing Dictionary', <http://www.speech.cs.cmu.edu/cgi-bin/cmudict>.
Craig, J. and Haskey, J. (1978) 'The Relationships Between the Population, Area, and Density of Urban Areas', *Urban Studies*, 15/1: 101–7.
Crystal, D. (2012) *English as a Global Language*. Cambridge u.a.: Cambridge Univ. Press.
Davies, M. (2010) *The corpus of historical American English*: *COHA*. Brigham: BYE, Brigham Young University.
Diessel, H. (2007) 'Frequency effects in language acquisition, language use, and diachronic change', *New ideas in psychology*, 25: 108–27.
—— (2012). 'New Perspectives, Theories and Methods: Diachronic change and language acquisition'. In: A. Bergs and L. J. Brinton (eds) *Historical linguistics of English*:, pp. 1599–613. Berlin [a.e.]: De Gruyter Mouton.
Dziubalska-Kołaczyk, K. (2005). 'Phonotactics of consonant clusters in the history of English'. In: A. Bertacca (ed.) *Historical Linguistic Studies of Spoken English*:, pp. 3–21. Pisa: University of Pisa: PLUS.
Fay, N. and Ellison, T. M. (2013) 'The cultural evolution of human communication systems in different sized populations: usability trumps learnability', *PloS one*, 8.
Fenk-Oczlon, G. and Fenk, A. (2016). 'Complexity trade-offs between the subsystems of language'. In: E. Ifejirika (ed.) *The advent of Christianity in Umuchu*: *Achievements and challenges (1916-2016)*, pp. 43–65. Anambra State, Nigeria: MABCOM Systems.
Galeano, P. and Pena, D. (2000) 'Multivariate analysis in vector time series', *Resenhas IME-USP*, 4/4: 383–403.
Good, J. (2015) 'Paradigmatic complexity in pidgins and creoles', *Word Structure*, 8/2: 184–227.
Great Britain Historical GIS Project (2017) *Great Britain Historical GIS*. Portsmouth: University of Portsmouth.
Hay, J. and Baayen, H. (2003) 'Phonotactics, parsing and productivity', *Rivista di Linguistica*, 15: 99–130.
Hay, J. and Bauer, L. (2007) 'Phoneme inventory size and population size', *Language*, 83/2: 388–400.
Hill, M. (1973) 'Diversity and evenness: a unifying notation and its consequences', *Ecology*, 54/2: 427–32.
Hogg, R. M. (1992) *A Grammar of Old English, Volume 1: Phonology*. Oxford: Wiley-Blackwell.
Hogg, R. M. and McCully, C. B. (1987) *Metrical Phonology*: *A Course Book*. Cambridge: Cambridge University Press.
Horobin, S. and Smith, J. J. (2002) *An Introduction to Middle English*. Oxford: Oxford University Press.

Kirk, C. (2008) ‘Substutution Errors in the production of word-initial and word-final consonant clusters‘, *Journal of Speech, Language and Hearing Reserach*, 51/31-48.
Kirk, C. and Demuth, K. (2005) ‘Asymmetries in the acquisition of word-initial and word-final consonant clusters‘, *Journal of child language*, 32: 709–34.
Kroch, A., Santorini, B. and Delfs, L. (2004) ‘Penn-Helsinki Parsed Corpus of Early Modern English‘, <http://www.ling.upenn.edu/hist-corpora/>.
Kroch, A., Santorini, B. and Diertani, A. (2016) ‘The Penn Parsed Corpus of Modern British English (PPCMBE2)‘, <http://www.ling.upenn.edu/ppche/ppche-release-2016/PPCMBE2-RELEASE-1>.
Kroch, A. and Taylor, A. (2000) ‘Penn-Helsinki Parsed Corpus of Middle English‘, <http://www.ling.upenn.edu/hist-corpora/>.
Labov, W. (2007) ‘Transmission and Diffusion‘, *Language*, 83/2: 344–87.
Lupyan, G. and Dale, R. (2010) ‘Language structure is partly determined by social structure‘, *PloS one*, 5.
Maddieson, I. (2013). ‘Syllable Structure’. In: M. S. Dryer and M. Haspelmath (eds) *The World Atlas of Language Structures Online*: : Max Planck Digital Library.
Marcon, E. and Herault, B. (2013) *entropart, an R package to partition diversity,*
Minkova, D. (1991) *The history of final vowels in English*: *The sound of muting*. Berlin, New York: M. de Gruyter.
Montero, P. and Vilar, J. (2014) ‘TSclust: An R Package for Time Series Clustering‘, *Journal of Statistical Software*, 62/1.
Mufwene, S. S. (2001) *The ecology of language evolution*. Cambridge: Cambridge University Press.
Murtagh, F. and Contreras, P. (2012) ‘Algorithms for hierarchical clustering: An overview‘, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2/1: 86–97.
Nettle, D. (2012) ‘Social scale and structural complexity in human languages‘, *Phil. Trans. R. Soc. B*, 367: 1829–36.
Office for National Statistics (2012) ‘2011 Census: Population Estimates for the United Kingdom‘, <http://webarchive.nationalarchives.gov.uk/20160108132257/http://www.ons.gov.uk/ons/dcp171778_292378.pdf> accessed 1 Sep 2017.
—— (2013) ‘Region and Country Profiles, Key Statistics and Profiles‘, <http://webarchive.nationalarchives.gov.uk/20160106190020/http://www.ons.gov.uk/ons/rel/regional-trends/region-and-country-profiles/region-and-country-profiles---key-statistics-and-profiles--october-2013/index.html> accessed 1 Sep 2017.
Pei, S., Muchnik, L., Andrade, J. S. and Zheng, Z. *et al.* (2014) ‘Searching for superspreaders of information in real-world social media‘, *Scientific reports*, 4: 5547.
R Development Core Team (2013) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
Rama, T. (2013) ‘Phonotactic diversity predicts the time depth of the world’s language families‘, *PloS one*.
Redford, M. A. (2008) ‘Production constraints on learning novel onset phonotactics‘, *Cognition*, 107/3: 785–816.

Ritt, N., Prömer, C. and Baumann, A. (2017) *Evolution of Consonant Clusters in English (ECCE): a diachronic phonotactic database*. Vienna: Department of English and American Studies,University of Vienna.

Rogers, E. (2003) *Diffusion of innovations*. New York: Free Press.

Schryver, J. de, Neijt, A., Ghesquière, P. and Ernestus, M. (2008) 'Analogy, Frequency, and Sound Change. The Case of Dutch Devoicing', *Journal of Germanic Linguistics*, 20/02: 1.

Sinnemäki, K. (2016). 'Complexity trade-offs in core argument marking'. In: E. Ifejirika (ed.) *The advent of Christianity in Umuchu*: *Achievements and challenges (1916-2016)*, pp. 67–88. Anambra State, Nigeria: MABCOM Systems.

Stein, R. A. (2011) 'Super-spreaders in infectious diseases', *International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases*, 15/8: e510-3.

Szmrecsanyi, B. (2012). 'Analyticity and syntheticity in the history of English'. In: T. Nevalainen and E. C. Traugott (eds) *The Oxford handbook of the history of English*:, pp. 654–65. Oxford: Oxford University Press.

Trudgill, P. (2001) *Sociolinguistic variation and change*. Edinburgh: Edinburgh University Press.

—— (2004) 'Linguistic and social typology: The Austronesian migrations and phoneme inventories', *Linguistic Typology*, 8/3: 305–20.

Tuomisto, H. (2010) 'A consistent terminology for quantifying species diversity? Yes, it does exist', *Oecologia*, 164: 53–860.

U.S. Bureau of the Census (1975) *Historical Statistics of the United States, Colonial Times to 1970, Bicentennial Edition, Part 2*. Washington D.C.: Government Printing Office.

U.S. Census Bureau (2002) *Historical Census Statistics on Population Totals by Race, 1790 to 1990, and by Hispanic Origin, 1970 to 1990, for the United States, Regions, Divisions, and States. U.S.* Washington D.C.: U.S. Census Bureau.

—— (2011) *2010 Census Briefs, Population Distribution and Change: 2000 to 2010*. Washington D.C.: U.S. Census Bureau.

—— (2012) *2010 Census of Population and Housing, Population and Housing Unit Counts, CPH-2-1, United States Summary*. Washington D.C.: U.S. Government Printing Office.

Watts, D. J. and Strogatz, S. H. (1998) 'Collective dynamics of 'small-world' networks', *Nature*, 393/6684: 440–2.

Wichmann, S. and Holman, E. (2009) 'Population size and rates of language change', *Human Biology*, 81/2: 259–74.

Wichmann, S., Rama, T. and Holman, E. (2011) 'Phonological diversity, word length, and population sizes across languages: The ASJP evidence', *Linguistic Typology*, 15/2: 177–97.

Wichmann, S., Stauffer, D., Schulze, C. and Holman, E. (2008) 'Do language change rates depend on population size?', *Advances in Complex Systems*, 11/3: 357–69.

Wood, S. (2006) *Generalized Additive Models: an introduction with R*. Boca Raton, Florida: Chapman & Hall/CRC.

Wrigley, E. A. and Schofield, R. (1981) *The Population History of England, 1541–1871. A reconstruction*. Harvard: Harvard University Press.

Zaki, M. J. and Meira, W. (2014) *Data mining and analysis*: *Fundamental concepts and algorithms*. New York: Cambridge University Press.

Appendix

A1 Data

Table A1. Retrieved token frequencies of English inventory of word-final consonant-cluster types for all 50-year periods as described in Section 2.1. Frequencies are normalized per million words. Clusters shown in CMU phonological transcription (see supplementary materials).

Table A2. Data for each period as described in Section 2 (see references therein). Linguistic variables: phonotactic diversity δ_1; phonotactic richness δ_0; analyticity α; syntheticity σ; number of cluster tokens ν; number of consonant phonemes κ; population size *N*; population density *D*; populated areas *A*; network diameter *d*; clustering coefficient *c*. All numbers rounded.

period	δ_1	δ_0	α	σ	ν	κ	N	D	A	d	c
1150	21.52	65	442.68	191.52	73100	32	3100000	12.78	242495	5.53	7.21E-05
1200	18.69	86	430.51	151.53	96136	25	4345000	17.92	242495	5.61	5.38E-05
1250	10.83	41	469.55	145.42	111673	25	4345000	17.92	242495	5.61	5.38E-05
1300	16.49	72	475.87	127.13	133179	25	4540000	18.72	242495	5.62	5.18E-05
1350	13.95	78	482.52	162.14	148069	25	2550000	10.52	242495	5.48	8.53E-05
1400	15.07	84	480.00	137.13	145218	25	2050000	8.45	242495	5.43	1.03E-04
1450	14.54	88	460.02	141.71	150368	25	2020000	8.33	242495	5.43	1.04E-04
1500	16.31	81	482.07	130.55	161364	24	2590000	10.68	242495	5.48	8.42E-05
1550	18.24	81	463.30	152.41	167807	24	3200000	13.20	242495	5.53	7.01E-05
1600	17.21	79	478.39	138.96	172892	24	4118483	16.91	677643	5.59	5.63E-05
1650	17.71	78	476.06	152.67	173524	24	5429848	21.42	1153805	5.66	4.43E-05
1700	16.40	75	459.55	158.15	131175	24	5716758	19.54	1672452	5.67	4.23E-05
1750	15.16	61	473.64	169.66	125101	24	7524409	15.30	1732896	5.73	3.33E-05
1800	28.13	90	451.98	169.59	134656	24	20941646	23.23	3815044	5.97	1.36E-05
1850	28.29	96	457.47	161.81	130509	24	62593266	34.93	5706271	6.22	5.10E-06
1900	28.03	89	444.11	178.05	139045	24	140827280	44.67	9018472	6.40	2.50E-06
1950	28.11	87	444.11	178.05	132686	24	245316179	50.72	9018472	6.52	1.50E-06

A2 Code

Script A1. R code used for computation of phonotactic diversity δ_1, phonotactic richness δ_0 and number of tokens ν as described in 2.1 and 2.2. It requires R package `entropart` (Marcon and Herault 2013). The script takes Table A1 as input.

```
#### 2.1 Phonotactic diversity ####
library(entropart)              #loads entropart package
frequencies.df= read.table("<path>/Table_A1.txt",sep=",",header=TRUE)
```

```
                                 #import Table A1 (specify <path>)

ntime=dim(frequencies.df)[2]-1                            #number of periods
diversity0=0; diversity1=0; nclustertokens=0

for(i in 1:ntime){
  diversity1[i]=exp(bcShannon(Ns=frequencies.df[,i+1])) #\delta_1
  diversity0[i]=sum(frequencies.df[,i+1]>0)             #\delta_0
  nclustertokens[i]=sum(frequencies.df[,i+1])           #frequency \nu
}
```

Script A2. R code used for the analysis described in Section 3. It requires R packages `mgcv` (Wood 2006) and `TSclust` (Montero and Vilar 2014). The analysis takes Table A2 as input.

```
#### 3.0 Preparations ####
library(mgcv)                             #mgcv package for computing gams
library(TSclust)                          #package for timeseries clustering
data.df=read.table("<path>/Table_A2.txt",sep=",",header=TRUE)
                                          #import Table A2 (specify <path>)

#### 3.1 Generalized additive modeling of time series ####
nvar=dim(data.df)[2]                       #number of variables
ntime=dim(data.df)[1]                      #number of periods
time=1:ntime                               #time vector
for(i in 1:nvar){
  var=data.df[,i]
  mdl=gam(var~s(time,k=8,bs="ts"))         #gam with thin-plate splines
                                           #(allowing for shinkage);
                                           #at most k=8 basis functions
  gam.check(mdl)                           #model diagnostics:
                                           #k too low if p<0.05
  acf(resid(mdl),main=paste("ACF",colnames(data.df)[i],sep="."))
                                           #model diagnostics: prints ACF
  predvar=predict(mdl)                     #compute predicted values
  data.df$predvar=predvar                  #add predicted values
  colnames(data.df)[nvar+i]=paste("predict",colnames(data.df)[i],sep=".")
                                           #rename column
}
data.df=as.data.frame(as.matrix(data.df)) #dropping any labels

#### 3.2 Autocorrelation based time-series dissimilarity ####
predict.vars=(nvar+1):(2*nvar)            #set of predicted variables
predict.df=data.df[,predict.vars]         #extracts only predicted variables
tsdist=diss(predict.df, "ACF")            #ACF time-series clustering

#### 3.3 Hierarchical time-series clustering ####
tsclust=hclust(tsdist,method="complete")  #hierarchical clustering with
                                           #complete linkage
plot(tsclust,hang=-1)                      #plotting dendrogram
dendrogram.height=0                        #scree plot
for(i in 2:nvar) dendrogram.height[i]=tsclust$height[i-1]
plot(nvar:1,dendrogram.height,type="l")
```

Appendix

List of contributions

This section provides a concise overview of all publications and manuscripts collected in this thesis. In total, this dissertation consists of eight articles, two of which are already published and an additional one being accepted for publication. Below, I list all manuscripts together with details on their status and remarks on my own contribution to them. Contributions to the articles by co-authors are also indicated.

1. Baumann, A., Ritt, N., submitted. The basic reproductive ratio as a link between acquisition and change in phonotactics.
    - Submitted to Cognition 04/2017; revision requested 08/2017; revised version submitted 11/2017
    - Contribution: conceptual and mathematical modeling, empirical data analysis, writing
2. Baumann, A., Sommerer, L., submitted. Linguistic diversification as a long-term effect of asymmetric priming: an adaptive-dynamics approach.
    - Submitted to Language Dynamics and Change 08/2017; revision requested 12/2017; revised version submitted 01/2018
    - Contribution: conceptual and mathematical modeling, empirical data analysis, writing; overview on structural priming and grammaticalization and diachronic corpus study on grammaticalization by L. Sommerer
3. Baumann, A., Kaźmierski, K., 2016. A dynamical-systems approach to the evolution of morphonotactic and lexical consonant clusters in English and Polish. Yearbook of the Poznan Linguistic Meeting 2, 115–139.
    - Submitted to PLM Yearbook 06/2016; accepted 09/2016; published 09/2016
    - Contribution: conceptual and mathematical modeling, writing; collection of Polish data by K. Kaźmierski
4. Baumann, A., Ritt, N., Prömer, C., 2016. Diachronic dynamics of Middle English phonotactics provide evidence for analogy effects among lexical and morphonotactic consonant clusters. Papers in Historical Phonology 1, 50–75.
    - Submitted to PiHPh 07/2016; accepted 09/2016; published 12/2016
    - Contribution: data analysis, writing; data collection by N. Ritt, C. Prömer and myself
5. Baumann, A., Kaźmierski, K., submitted. Assessing the effect of ambiguity in compositionality signaling on the processing of diphones.
    - Submitted to Language Sciences 03/2017; revision requested 06/2017; revised version submitted 01/2018

- Contribution: experimental design, data analysis, writing; experiment conducted by K. Kaźmierski

6. Baumann, A., Wissing, D., submitted. Stabilizing determinants in the transmission of phonotactic systems: diachrony and acquisition of coda clusters in Dutch and Afrikaans.
   - Submitted to Stellenbosch Papers in Linguistics 10/2017
   - Contribution: data analysis, writing; overview of history of Afrikaans and Afrikaans data provided by D. Wissing
7. Baumann, A., accepted. Linguistic stability increases with population size, but only in stable learning environments, in: The Evolution of Language: Proceedings of the 12th International Conference (EVOLANG12).
   - Submitted to Evolang 08/2017; accepted 12/2017
   - Contribution: single authored
8. Baumann, A., Matzinger, T., submitted. Correlates in the evolution of phonotactic diversity: linguistic structure, demographics, and network characteristics.
   - Submitted to Journal of Language Evolution 11/2017
   - Contribution: modeling, data analysis, writing; collection of demographic data by T. Matzinger

References

Abrams, D.M., Strogatz, S.H., 2003. Linguistics: Modelling the dynamics of language death. Nature 424 (6951), 900. 10.1038/424900a.

Aldrich, H.E., Hodgson, G.M., Hull, D.L., Knudsen, T., Mokyr, J., Vanberg, V.J., 2008. In defence of generalized Darwinism. J Evol Econ 18 (5), 577–596. 10.1007/s00191-008-0110-z.

Alter, A.L., Oppenheimer, D.M., 2006. Predicting short-term stock fluctuations by using processing fluency. Proceedings of the National Academy of Sciences 103 (24), 9369–9372. 10.1073/pnas.0601071103.

Anderson, R., May, R., 1991. Infectious diseases of humans: Dynamics and control. Oxford University Press, New York.

Atkinson, Q.D., 2011. Phonemic diversity supports a serial founder effect model of language expansion from Africa. Science 322 (6027), 346–349.

Barabási, A.-L., 2016. Network science. Cambridge University Press, Cambridge.

Baumann, A., forthc. Linguistic stability increases with population size, but only in stable learning environments, in: The Evolution of Language: Proceedings of the 12th International Conference (EVOLANG12).

Baumann, A., Kaźmierski, K., submitted. Assessing the effect of ambiguity in compositionality signaling on the processability of diphones.

Baumann, A., Kaźmierski, K., 2016. A dynamical-systems approach to the evolution of morphonotactic and lexical consonant clusters in English and Polish. Yearbook of the Poznan Linguistic Meeting 2, 115–139.

Baumann, A., Kaźmierski, K., 2017. Lazy speakers and distant sounds: on the role of articulatory difference in phonotactic production, acquisition, and change. Phonetics and Phonology in Europe 2017.

Baumann, A., Matzinger, T., submitted. Correlates in the evolution of phonotactic diversity: linguistic structure, demographics, and network characteristics.

Baumann, A., Prömer, C., Kaźmierski, K., Ritt, N., 2016a. A Lotka-Volterra model of the evolutionary dynamics of compositionality markers, in: The Evolution of Language: Proceedings of the 11th International Conference (EVOLANG11).

Baumann, A., Ritt, N., submitted. The basic reproductive ratio as a link between acquisition and change in phonotactics. Cognition.

Baumann, A., Ritt, N., 2017. On the replicator dynamics of lexical stress: accounting for stress pattern diversity in terms of evolutionary game theory. Phonology.

Baumann, A., Ritt, N., Prömer, C., 2016b. Diachronic dynamics of Middle English phonotactics provide evidence for analogy effects among lexical and morphonotactic consonant clusters. Papers in Historical Phonology 1, 50–75.

Baumann, A., Sommerer, L., submitted. Linguistic diversification as a long-term effect of asymmetric priming: an adaptive-dynamics approach.

Baumann, A., Wissing, D., submitted. Stabilizing determinants in the transmission of phonotactic systems: diachrony and acquisition of coda clusters in Dutch and Afrikaans. SPiL (Stellenbosch Papers in Linguistics).

Beekhuizen, B., Bod, R., Zuidema, W., 2013. Three design principles of language: The search for parsimony in redundancy. Language and Speech 56 (Pt 3), 265–290. 10.1177/0023830913484897.

Bentz, C., Winter, B., 2013. Languages with More Second Language Learners Tend to Lose Nominal Case. Language Dynamics and Change 3 (1), 1–27.
Berent, I., Steriade, D., Lennertz, T., Vaknin, V., 2007. What we know about what we have never heard: Evidence from perceptual illusions. Cognition 104 (3), 591–630. 10.1016/j.cognition.2006.05.015.
Blevins, J., 2012. Duality of patterning: Absolute universal or statistical tendency? Language and cognition 4 (04), 275–296. 10.1515/langcog-2012-0016.
Blevins, J., Féry, C., van de Vijver, Ruben, 2003. The Independent Nature of Phonotactic Constraints: An Alternative to Syllable-Based Approaches. In: Féry, C., van de Vijver, Ruben (Eds.) The Syllable in Optimality Theory. Cambridge University Press, Cambridge, pp. 375–404.
Bod, R., 2009. From exemplar to grammar: A probabilistic analogy-based model of language learning. Cognitive Science 33 (5), 752–793. 10.1111/j.1551-6709.2009.01031.x.
Boer, B. de, Sandler, W., Kirby, S., 2012. New perspectives on duality of patterning: Introduction to the special issue. Language and cognition 4 (4), 251–259. 10.1515/langcog-2012-0014.
Bowern, C., 2010. Historical linguistics in Australia: Trees, networks and their implications. Philos. Trans. R. Soc. Lond., B, Biol. Sci. 365 (1559), 3845–3854. 10.1098/rstb.2010.0013.
Brighton, H., Smith, K., Kirby, S., 2005. Language as an evolutionary system. Physics of Life Reviews 2 (3), 177–226.
Bromham, L., Hua, X., Fitzpatrick, T., Greenhill, S.J., 2015. Rate of language evolution is affected by population size. Proceedings of the National Academy of Sciences 112 (7), 2097–2102.
Brown, R., 1973. A first language: The early stages. Harvard University Press, Harvard.
Bryson, J.J., 2008. Embodiment versus memetics. Mind & Society 7 (1), 77–94. 10.1007/s11299-007-0044-4.
Bryson, J.J., 2009. Representations underlying social learning and cultural evolution. IS 10 (1), 77–100. 10.1075/is.10.1.06bry.
Bybee, J., 2007. Frequency of use and the organization of language. Oxford University Press, Oxford.
Bybee, J., 2010. Language, Usage and Cognition. Cambridge University Press, Cambridge.
Bybee, J., 2011. How plausible is the hypothesis that population size and dispersal are related to phoneme inventory size? Introducing and commenting on a debate. Linguistic Typology 15, 147–153. 10.1515/LITY.2011.009.
Bybee, J.L., 1994. A view of phonology from a cognitive and functional perspective. Cognitive Linguistics 5, 285–306.
Calderone, B., Celata, C., Korecky-Kröll, K., Dressler, W.U., 2014. A computational approach to morphonotactics: evidence from German. Language Sciences 46, 59–70.
Cavalli-Sforza, L.L., Feldman, M.W., 1981. Cultural transmission and evolution: aquantitative approach. Princeton University Press, Princeton.
Celata, C., Korecky-Kröll, K., Ricci, I., Dressler, W.U., 2015. Phonotactic processing and morpheme boundaries: wordfinal /Cst/ clusters in German. Italian Journal of Linguistics 27 (1), 85–110.
Christiansen, M.H., Onnis, L., Hockema, S.A., 2009. The secret is in the sound: From unsegmented speech to lexical categories. Developmental science 12 (3), 388–395. 10.1111/j.1467-7687.2009.00824.x.

Clark, A., 2013. Embodied, embedded, and extended cognition. In: Frankish, K., Ramsey, W.M. (Eds.) The Cambridge handbook of cognitive science. Cambridge University Press, Cambridge, pp. 275–291.
Clements, G.N., 1990. The role of the sonority cycle in core syllabification., in: Papers in Laboratory Phonology I: Between the grammar and the physics of speech. Cambridge Univ. Press, Cambridge, pp. 282–333.
Coetzee, A., 2014. Grammatical change through lexical accumulation: voicing cooccurrence restrictions in Afrikaans. Language 90 (3), 693–721.
Cowley, S.J., 2014. Linguistic embodiment and verbal constraints: Human cognition and the scales of time. Frontiers in psychology 5, 1085. 10.3389/fpsyg.2014.01085.
Croft, W., 2000. Explaining language change: An evolutionary approach. Longman, Harlow, England, New York, xv, 287.
Croft, W., 2002. The Darwinization of Linguistics. Selection 3, 75–91.
Croft, W., 2013. Language use and the evolution of languages. In: Smith, K., Binder, P. (Eds.) The language phenomenon. Human communication from milliseconds to millennia. Springer, Heidelberg [a.e.], pp. 93–120.
Daland, R., Pierrehumbert, J., 2011. Learning diphone-based segmentation. Cognitive Science 35, 119–155.
Dawkins, R., 1976. The selfish gene. Oxford University Press, New York, Oxford.
Dercole, F., Rinaldi, S., 2008. Analysis of evolutionary processes. Princeton University Press, Princeton, New York.
Deumert, A., 2003. Bringing speakers back in?: Epistemological reflections on speaker-oriented explanations of language change. Language Sciences 25 (1), 15–76. 10.1016/S0388-0001(01)00014-6.
Dieckmann, U., Law, R., 1996. The dynamical theory of coevolution: a derivation from stochastic ecological processes. Journal of Mathematical Biology 34, 579–612.
Diessel, H., 2007. Frequency effects in language acquisition, language use, and diachronic change. New ideas in psychology (25), 108–127.
Diessel, H., Hilpert, M., 2016. Frequency Effects in Grammar. Linguistics: Oxford Research Encyclopedias. Oxford University Press, Oxford.
Dietz, K., 1993. The estimation of the basic reproduction number for infectious diseases. Statistical methods in medical research 2, 23–41.
Doebeli, M., 2011. Adaptive diversification. Princeton University Press, Princeton.
Donegan, P.J., Nathan, G.S., 2014. Natural Phonology and Sound Change. In: Honeybone, P., Salmons, J. (Eds.) Oxford Handbook of Historical Phonology. Oxford University Press, Oxford.
Dressler, W.U., 1989. Markedness and naturalness in phonology: The case of natural phonology. In: Tomić, O.M. (Ed.) Markedness in synchrony and diachrony. Mouton de Gruyter, Berlin, New York, pp. 111–120.
Dressler, W.U., 2003. Naturalness and morphological change. In: Joseph, B.D., Janda, R.D. (Eds.) The handbook of historical linguistics. Blackwell, Malden, MA, pp. 461–471.
Dressler, W.U., Dziubalska-Kołaczyk, K., 2006. Proposing Morphonotactics. Wiener Linguistische Gazette 73, 69–87.
Dressler, W.U., Dziubalska-Kołaczyk, K., Pestal, L., 2010. Change and variation in morphonotactics. Folia Linguistica Historica 31, 51–68. 10.1515/flih.2010.003.
Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., Mehler, J., 1999. Epenthetic vowels in Japanese: A perceptual illusion? Journal of Experimental Psychology: Human Perception and Performance 25 (6), 1568–1581.

Eckert, P., Labov, W., 2017. Phonetics, phonology and social meaning. J Sociolinguistics 21 (4), 467–496. 10.1111/josl.12244.
Ernestus, M., 2014. Acoustic reduction and the roles of abstractions and exemplars in speech processing. Lingua 142, 27–41.
Everett, C., Blasí, D.E., Roberts, S.G., 2016. Language evolution and climate: The case of desiccation and tone. Journal of Language Evolution 1 (1), 33–46. 10.1093/jole/lzv004.
Ferreira, V.S., Bock, K., 2006. The functions of structural priming. Language and Cognitive Processes 21 (7-8), 1011–1029.
Frisch, S., Pierrehumbert, J., Broe, M., 2004. Similarity Avoidance and the OCP. Natural Language & Linguistic Theory 22 (1), 179–228.
Geritz, S.A.H., Kisdi, E., Meszéna, G., Metz, J.A.J., 1998. Evolutionarily Singular Strategies and the Adaptive Growth and Branching of the Evolutionary Tree. Evolutionary Ecology 12, 35–57.
Gray, A., Greenhalgh, D., Hu, L., Mao, X., Pan, J., 2011. A stochastic differential equation SIS epidemic model. SIAM Journal on Applied Mathematics 71 (3), 876–902.
Gray, R.D., Atkinson, Q.D., 2003. Language-tree divergence times support the Anatolian theory of Indo-European origin. Nature 426 (6965), 435–439. 10.1038/nature02029.
Greenhalgh, D., Liang, Y., Mao, X., 2015. Demographic stochasticity in the SDE SIS epidemic model. Discrete and Continuous Dynamical Systems-Series B 20 (9), 2859–2884.
Grice, M., Savino, M., Caffò, A., Roettger, T.B., 2015. The tune drives the text–Schwa in consonant-final loan words in Italian. The Scottish Consortium for ICPhS.
Grodzinsky, Y., Nelken, I., 2014. Neuroscience. The neural code that makes us human. Science (New York, N.Y.) 343 (6174), 978–979. 10.1126/science.1251495.
Hay, J., Baayen, H., 2005. Probabilistic phonotactics and morphological productivity. http://www.ling.canterbury.ac.nz/jen/documents/haybaayenh_esse.htm.
Hay, J., Pierrehumbert, J., Beckman, M., 2004. Speech perception, well-formedness, and the statistics of the lexicon. Papers in laboratory phonology VI, 58–74.
Heffernan, J., Smith, R., Wahl, L., 2005. Perspectives on the basic reproductive ratio. Journal of The Royal Society Interface 2 (4), 281–293. 10.1098/rsif.2005.0042.
Heine, B., Kuteva, T., 2007. The Genesis of Grammar. Oxford University Press, Oxford.
Henrich, J., Boyd, R., 2002. On Modeling Cognition and Culture: Why cultural evolution does not require replication of representations. Journal of Cognition and Culture 2 (2), 87–112. 10.1163/156853702320281836.
Hethcote, H., 1989. Three basic epidemiological models. In: Levin, S., Hallam, T., Gross, L. (Eds.) Applied Mathematical Ecology.
Hofbauer, J., Sigmund, K., 1998. Evolutionary games and population dynamics. Cambridge Univ. Press, Cambridge [u.a.], XXVII, 323 S.
Hogg, R.M., McCully, C.B., 1987. Metrical Phonology: A Course Book. Cambridge University Press, Cambridge.
Hopper, P.J., Traugott, E.C., 2003. Grammaticalization. Cambridge University Press, Cambridge.
Hull, D.L., 1988. Science as a process: An evolutionary account of the social and conceptual development of science. The University of Chicago Press, Chicago [etc.], XIII, 586.
Jäger, G., 2008a. Applications of game theory in linguistics. Language and Linguistics compass 2 (3), 406–421.

Jäger, G., 2008b. Language evolution and George Price's ‘‘General theory of selection. Language in flux: Dialogue coordination, language variation, change and evolution. London: College Publications, 53–82.
Jäger, G., Rosenbach, A., 2008. Priming and unidirectional language change. Theoretical Linguistics 34 (2), 85–113.
Jusczyk, P.W., Houston, D.M., Newsome, M., 1999. The beginnings of word segmentation in English-learning infants. Cognitive Psychology 39 (3-4), 159–207. 10.1006/cogp.1999.0716.
Kandler, A., Unger, R., Steele, J., 2010. Language shift, bilingualism and the future of Britain's Celtic languages. Philos. Trans. R. Soc. Lond., B, Biol. Sci. 365 (1559), 3855–3864. 10.1098/rstb.2010.0051.
Kemps, R., Wurm, L., Ernestus, M., Schreuder, r., Baayen, H., 2005. Prosodic cues for morphological complexity in Dutch and English. Language and Cognitive Processes 20, 43–73.
Kershenbaum, A., Blumstein, D.T., Roch, M.A., Akçay, Ç., Backus, G., Bee, M.A., Bohn, K., Cao, Y., Carter, G., Cäsar, C., Coen, M., DeRuiter, S.L., Doyle, L., Edelman, S., Ferrer-i-Cancho, R., Freeberg, T.M., Garland, E.C., Gustison, M., Harley, H.E., Huetz, C., Hughes, M., Hyland Bruno, J., Ilany, A., Jin, D.Z., Johnson, M., Ju, C., Karnowski, J., Lohr, B., Manser, M.B., McCowan, B., Mercado, E., Narins, P.M., Piel, A., Rice, M., Salmi, R., Sasahara, K., Sayigh, L., Shiu, Y., Taylor, C., Vallejo, E.E., Waller, S., Zamora-Gutierrez, V., 2016. Acoustic sequences in non-human animals: A tutorial review and prospectus. Biological reviews of the Cambridge Philosophical Society 91 (1), 13–52. 10.1111/brv.12160.
Kirby, S., 2012. Language is an Adaptive System: the role of cultural evolution in the origins of structure. In: Tallerman, M., Gibson, K.R. (Eds.) The Oxford handbook of language evolution. Oxford University Press, Oxford, New York, pp. 589–605.
Kirby, S., Dowman, M., Griffiths, T.L., 2007. Innateness and culture in the evolution of language. Proceedings of the National Academy of Sciences 104 (12), 5241–5245. 10.1073/pnas.0608222104.
Korecky-Kröll, K., Dressler, W.U., Freiberger, E.M., Reinisch, E., Mörth, K., Libben, G., 2014. Morphonotactic and phonotactic processing in German-speaking adults. Language Sciences 46, 48–58.
Kristiansen, G., 2006. Towards a usage-based cognitive phonology. International Journal of English Studies 6 (2), 107–140.
Kuperman, V., Ernestus, M., Baayen, H., 2008. Frequency distributions of uniphones, diphones, and triphones in spontaneous speech. The Journal of the Acoustical Society of America 124, 3897–3908.
Kuperman, V., Stadthagen-Gonzalez, H., Brysbaert, M., 2012. Age-of-acquisition ratings for 30,000 English words. Behavior Research Methods 44, 978–990.
Ladd, D.R., 2014. Simultaneous Structure in Phonology. Cambridge University Press, Cambridge.
Laham, S.M., Koval, P., Alter, A.L., 2012. The name-pronunciation effect: Why people like Mr. Smith more than Mr. Colquhoun. Journal of Experimental Social Psychology 48 (3), 752–756. 10.1016/j.jesp.2011.12.002.
Lakoff, G., 1990. The Invariance Hypothesis: Is abstract reason based on image-schemas? Cognitive Linguistics 1 (1), 39–74. 10.1515/cogl.1990.1.1.39.
Lakoff, G., 1993. Cognitive phonology. In: Goldsmith, J. (Ed.) The last phonological rule. Chicago University Press, Chicago, pp. 117–145.

Langacker, R.W., 2008. Cognitive grammar: A basic introduction. Oxford University Press, Oxford, New York, 1 online resource (x, 562.
Loebell, H., Bock, J.K., 2003. Structural priming across languages. LINGUISTICS 41, 791–824.
Lupyan, G., Dale, R., 2010. Language structure is partly determined by social structure. PloS one 5. 10.1371/journal.pone.0008559.
MacNeilage, P.F., Davis, B.L., 2000. On the origin of internal structure of word forms. Science 288, 527–531.
Maddieson, I., 2013. Syllable Structure. In: Dryer, M.S., Haspelmath, M. (Eds.) The World Atlas of Language Structures Online. Max Planck Digital Library.
McCarthy, J.J. (Ed.), 2004. Optimality Theory in phonology. Blackwell, Malden, MA.
McCrohon, L., 2012. The two-stage life cycle of cultural replicators. Theoria et Historia Scientiarum 9, 149–170.
McMahon, A.M.S., McMahon, R., 2013. Evolutionary linguistics. Cambridge University Press, Cambridge [England], New York, xix, 309.
McQueen, J., 1998. Segmentation of Continuous Speech Using Phonotactics. Journal of Memory and Language 39, 21–46.
Mesgarani, N., Cheung, C., Johnson, K., Chang, E.F., 2014. Phonetic feature encoding in human superior temporal gyrus. Science 343, 1006–1010.
Mitchener, W.G., Nowak, M.A., 2004. Chaos and language. Proceedings of the Royal Society of London-B 271 (1540), 701–704.
Mompeán, J., 2006. Cognitive phonology in cognitive linguistics. International Journal of English Studies 6 (vii-xii).
Mompeán-González, J., 2004. Category overlap and neutralization: The importance of speakers' classifications in phonology. Cognitive Linguistics 15, 429–469.
Monaghan, P., 2014. Age of acquisition predicts rate of lexical evolution. Cognition 133, 530–534.
Moreton, E., 2002. Structural constraints in the perception of English stop-sonorant clusters. Cognition 84 (1), 55–71. 10.1016/S0010-0277(02)00014-8.
Nathan, G.S., 2006. Is the Phoneme Usage-Based? – Some Issues. International Journal of English Studies 6 (2), 173–194.
Nathan, G.S., 2007. Phonology. In: Geeraerts, D., Cuyckens, H. (Eds.) The Oxford handbook of cognitive linguistics. Oxford University Press, Oxford, New York, pp. 611–631.
Nettle, D., 2012. Social scale and structural complexity in human languages. Phil. Trans. R. Soc. B 367, 1829–1836.
Newberry, M.G., Ahern, C.A., Clark, R., Plotkin, J.B., 2017. Detecting evolutionary forces in language change. Nature 551 (7679), 223–226. 10.1038/nature24455.
Niyogi, P., 2006. The computational nature of language learning and evolution, Cambridge, MA.
Nowak, M.A., 2000. The basic reproductive ratio of a word, the maximum size of a lexicon. J. Theor. Biol. 204 (2), 179–189. 10.1006/jtbi.2000.1085.
Nowak, M.A., 2006. Evolutionary dynamics: exploring the equations of life. Belknap, Harvard.
Nowak, M.A., Komarova, N.L., 2001. Towards an evolutionary theory of language. Trends in cognitive sciences 5, 288–295.
Otto, S., Day, T., 2007. A biologist's guide to mathematical modeling in ecology and evolution. Princeton University Press, Princeton.

Page, K., Nowak, M.A., 2002. Unifying evolutionary dynamics. Journal of Theoretical Biology (219), 93–98.
Pagel, M., 2009. Human language as a culturally transmitted replicator. Nature reviews. Genetics 10 (6), 405–415. 10.1038/nrg2560.
Pagel, M., Atkinson, Q.D., Meade, A., 2007. Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. Nature 449 (7163), 717–720.
Pickering, M.J., Ferreira, V.S., 2008. Structural priming: a critical review. Psychological Bulletin 134 (3), 427.
Pierrehumbert, J., 2001. Exemplar dynamics: Word frequency, lenition, and contrast. In: Bybee, J., Hopper, P.J. (Eds.) Frequency effects and the emergence of lexical structure. Benjamins, Amsterdam [a.e.], pp. 137–157.
Pierrehumbert, J., 2016. Phonological Representation: Beyond Abstract Versus Episodic. Annual Review of Linguistics 2, 33–52.
Plag, I., Homann, J., Kunter, G., 2015. Homophony and morphology: The acoustics of word-final S in English. Journal of Linguistics, 1–36.
Rama, T., 2013. Phonotactic diversity predicts the time depth of the world's language families. PloS one. 10.1371/journal.pone.0063238.
Ramachandran, V.S., Hubbard, E.M., 2001. Synaesthesia - A window into perception, thought and language. Journal of Consciousness Studies 8, 3–34.
Repetti, L., 2012. Consonant-final loanwords and epenthetic vowels in Italian. Catalan journal of linguistics 11, 167–188.
Ritt, N., 2004. Selfish sounds and linguistic evolution: A Darwinian approach to language change. Cambridge University Press, Cambridge.
Ritt, N., Kaźmierski, K., 2015. How rarities like gold came to exist: on co-evolutionary interactions between morphology and lexical phonotactics. English Language and Linguistics (19), 1–29. 10.1017/S1360674315000040.
Ritt, N., Prömer, C., Baumann, A., 2017. Evolution of Consonant Clusters in English (ECCE): a diachronic phonotactic database. Department of English and American Studies,University of Vienna, Vienna.
Rosenbach, A., 2008. Language change as cultural evolution: Evolutionary approaches to language change. In: Eckardt, R., Jäger, G., Veenestra, T. (Eds.) Variation, Selection, Development: Probing the Evolutionary Model of Language Change. De Gruyter Mouton, Berlin, pp. 23–74.
Saffran, J.R., Aslin, R.N., Newport, E.L., 1996. Statistical learning by 8-month-old infants. Science (New York, N.Y.) 274 (5294), 1926–1928.
Schmid, H.-J., 2016. Linguistic entrenchment and its psychological foundations. In: Schmid, H.-J. (Ed.) Entrenchment and the psychology of language learning. Mouton de Gruyter, Berlin, pp. 435–452.
Shatzman, K., Kager, R., 2007. A role for phonotactic constraints in speech perception. Proceedings of the 16th International Congress of Phonetic Sciences, 1409–1412.
Solé, R.V., 2011. Phase transitions. Princeton University Press, Princeton.
Solé, R.V., Corominas-Murtra, B., Fortuny, J., 2010. Diversity, competition, extinction: the ecophysics of language change. Journal of The Royal Society Interface 7 (53), 1647–1664.
Sonderegger, M., Niyogi, P., 2013. Variation and change in English noun/verb pair stress: data and dynamical systems models. In: Yu, A. (Ed.) Origins of Sound Change, pp. 262–284.

Song, H., Schwarz, N., 2009. If it's difficult to pronounce, it must be risky. Psychological Science 20 (2), 135–138. 10.1111/j.1467-9280.2009.02267.x.
Sperber, D., 1985. Anthropology and psychology: Towards an epidemiology of representations. Man, 73–89.
Stampe, D., 1979. A dissertation on natural phonology. Garland Pub, New York, xxv, 84.
Steels, L., 2011. Modeling the cultural evolution of language. Physics of Life Reviews 8 (4), 339–356. 10.1016/j.plrev.2011.10.014.
Steffensen, S.V., Fill, A., 2014. Ecolinguistics: The state of the art and future horizons. Language Sciences 41, 6–25. 10.1016/j.langsci.2013.08.003.
Taylor, J., 2002. Cognitive grammar. Oxford University Press, Oxford.
Taylor, J.R., 1995. Linguistic categorization. Prototypes in linguistic theory. Clarendon Press, Oxford.
Thelen, E., Schöner, G., Scheier, C., Smith, L.B., 2001. The dynamics of embodiment: A field theory of infant perseverative reaching. Behav. Brain Sci. 24 (1), 1–34. 10.1017/S0140525X01003910.
Thompson, B., Kirby, S., Smith, K., 2016. Culture shapes the evolution of cognition. Proceedings of the National Academy of Sciences of the United States of America 113 (16), 4530–4535. 10.1073/pnas.1523631113.
Topolinski, S., Zürn, M., Schneider, I.K., 2015. What's in and what's out in branding? A novel articulation effect for brand names. Frontiers in psychology 6, 585. 10.3389/fpsyg.2015.00585.
Tremblay, A., Baayen, H., Derwing, B., Libben, G., Tucker, B.V., Westbury, C., 2012. Empirical evidence for an inflationist lexicon. Yearbook Phraseol 3, 109–126.
Trubetzkoy, N.S., 1969. Grundzüge der Phonology. Vandenhoeck & Ruprecht, Göttingen.
Trudgill, P., 2001. Sociolinguistic variation and change. Edinburgh University Press, Edinburgh.
Trudgill, P., 2004. Linguistic and social typology: The Austronesian migrations and phoneme inventories. Linguistic Typology 8 (3), 305–320.
Tsur, R., 2006. Size–sound symbolism revisited. Journal of Pragmatics 38 (6), 905–924. 10.1016/j.pragma.2005.12.002.
Ulbrich, C., Alday, P.M., Knaus, J., Orzechowska, P., Wiese, R., 2015. The role of phonotactic principles in language processing. Language, Cognition and Neuroscience 31 (5), 662–682. 10.1080/23273798.2015.1136427.
Välimaa-Blum, R., 2009a. Phonotactic constraints in cognitive phonology. Corela 7 (1). 10.4000/corela.137.
Välimaa-Blum, R., 2009b. The phoneme in cognitive phonology: episodic memories of both meaningful and meaningless units? CogniTextes 2.
van Bergem, D.R., 1993. Acoustic vowel reduction as a function of sentence accent, word stress, and word class. Speech Communication 12 (1), 1–23. 10.1016/0167-6393(93)90015-D.
Vennemann, T., 1988. Preference laws for syllable structure and the explanation of sound change. De Gruyter, Berlin.
Vitevitch, M.S., Luce, P.A., Charles-Luce, J., Kemmerer, D., 1997. Phonotactics and Syllable Stress: Implications for the Processing of Spoken Nonsense Words. Language and Speech 40 (1), 47–62.
Wang, W., Minett, J., 2005. The invasion of language: Emergence, change and death. Trends in Ecology and Evolution 20, 263–269.

Wedel, A., 2006. Exemplar models, evolution and language change. The Linguistic Review 23 (3), 247–274.

Winter, B., Sókuthy, M., Perlman, M., 2017. 'r' is for rough: Iconicity in English and Hungarian Surface Descriptors. CLS-MPI Iconicity Focus Group Workshop. http://iconicity.ruhosting.nl/wp-content/uploads/2017/06/iconicity-workshop-abstracts-booklet.pdf.

Yang, C., 2000. Internal and external forces in language change. Language Variation and Change 12, 231–250.

Zehentner, E., 2017. Ditransitives in Middle English: On semantic specialisation and the rise of the dative alternation. English Language 11, 1–27. 10.1017/S1360674316000447.

Supplementary material

Supplementary data for Baumann & Ritt (submitted):

Baumann, A., 2017. Phonotactics and R0 data. phaidra.univie.ac.at/o:572645.

Supplementary data for Baumann & Kazmierski (submitted):

Baumann, A., 2018. Ambiguous diphones: Experiment 1. phaidra.univie.ac.at/o:654000.

Baumann, A., 2018. Ambiguous diphones: Experiment 2. phaidra.univie.ac.at/o:654001.

Supplementary data for Baumann & Matzinger (submitted):

Baumann, A., 2018. Phonotactic diversity: Table A1. phaidra.univie.ac.at/o:654002.

Curriculum vitæ

Personal information

Name	Mag. Mag. Andreas Baumann
Date of birth	30 July 1985 (Vienna)
Nationality	Austrian
Marital status	married

Education and working experience

12/2017	Erasmus+ Lectureship at the Adam Mickiewicz University in Poznan, Poland (Erasmus+ International Teaching Mobility Programme)
Since 02/2015	Project assistant at the Department of English and American Studies, University of Vienna, Evolution of consonant clusters in English project (FWF grant No. P27592-G18, ÖAD grant No. PL8/2014)
10/2014–01/2015	Visiting researcher at the University of Stellenbosch (ZA) Department of General Linguistics South African Center of Excellence in Epidemiological Modelling and Analysis (SACEMA)
Since 09/2014	University lecturer at the Department of English and American Studies, University of Vienna
04/2013	Mag. rer. nat. in Mathematics (final examination passed with distinction), diploma thesis supervised by Claus Rüffler, Mathematics and BioSciences Research Group
05/2011–09/2014	University assistant at the Department of English and American Studies, University of Vienna, NatSIDE Research Group
01/2010	Mag. phil. in Theoretical and Applied Linguistics (final examination passed with distinction), diploma thesis supervised by Martin Prinzhorn, University of Vienna
10/2005–04/2013	Studies of Mathematics, University of Vienna
10/2005–01/2010	Studies of Theoretical and Applied Linguistics, University of Vienna
10/2004–09/2005	Community service, St John Ambulance (JUH Wien)
06/2004	School-leaving examination (passed with distinction)
09/1996–06/2004	Grammar school specializing in humanities, Bundesgymnasium XVIII, Klostergasse 25, 1180 Vienna
09/1992–06/1996	Primary school, Volksschule, Scheibenbergstraße 63, 1180 Vienna

Editorial and organizational experience

09/2017	Co-organizer, Diachronic phonotactics workshop, Vienna, 7-8 September 2017
Since 04/2014	Member of the Evolang permanent organizing committee
04/2014	Local organizer, 10th International Conference on the Evolution of Language (Evolang X), Vienna, 14–17 April 2014
01/2013–09/2014	Co-editor in chief, Vienna English Working Papers in Linguistics

12/2012	Co-organizer, 1st Memetics Workshop, Vienna, 07 December 2012
05/2011–09/2014	Editorial team member, Folia Linguistica Historica and Vienna English Working Papers in Linguistics
11/2009	Co-organizer, 2nd Austrian Student Conference in Linguistics (ÖSKL), Vienna, 21–22 November 2009

Prizes and grants

08/2017	*Travel grant: international communication*, Österreichische Forschungsgemeinschaft (ÖFG, grant No. 06_15355)
09/2016	*Best poster award* (second prize), 49th SLE, Naples.
08/2016	*VivA travel grant*, Virtuele Instituut vir Afrikaans.
04/2016	*Evolang poster prize* (first prize), Evolang XI, New Orleans.

Publications

In review	Baumann, A.; Matzinger, T. Correlates in the evolution of phonotactic diversity: linguistic structure, demographics, and network characteristics. *Journal of Language Evolution.*
In review	Baumann, A.; Wissing, D. Stabilizing determinants in the transmission of phonotactic systems: on the emergence of the Afrikaans consonant-cluster inventory. *Stellenbosch Papers in Linguistics*.
In review	Baumann, A.; Sommerer, L. Diversification as a long-term effect of asymmetric priming: an adaptive-dynamics approach. *Language Dynamics and Change.*
In review	Baumann, A.; Ritt, N. The basic reproductive ratio as a link between acquisition and change in phonotactics. *Cognition*.
In review	Baumann, A.; Kazmierski, K. Assessing the effect of ambiguity in compositionality signaling on the processing of diphones. *Language Sciences*.
Forthcoming	Baumann, A. Linguistic stability increases with population size, but only in stable learning environments. *The Evolution of Language (Proc. Evolang 12).*
2017	Baumann, A.; Ritt, N. On the replicator dynamics of lexical stress: accounting for stress pattern diversity in terms of evolutionary game theory. *Phonology* 34: 439-471.
2016	Baumann, A.; Prömer, C.; Ritt, N. 2016. Diachronic dynamics of Middle English phonotactics provide evidence for analogy effects among lexical and morphonotactic consonant clusters. *Papers in Historical Phonology* 1: 50-75.
2016	Kazmierski, K.; Wojtkowiak, E.; Baumann, A. Coalescent assimilation across word-boundaries in American English and in Polish English. *Research in Language* 3: 235–262.
2016	Baumann, A.; Kazmierski, K. A dynamical-systems approach to the evolution of morphonotactic and lexical consonant clusters in English and Polish. *Yearbook of the Poznan Linguistic Meeting* 2: 115–139.
2016	Baumann, A.; Prömer, C.; Kazmierski, K.; Ritt, N. A Lotka-Volterra model of the evolutionary dynamics of compositionality markers. In

	S.G. Roberts, C. Cuskley, L. McCrohon, L. Barceló-Coblijn, O. Feher & T. Verhoef (eds.) *The Evolution of Language: Proceedings of the 11th International Conference (EVOLANG11)*.
2015	Baumann, A.; Prömer, C.; Ritt, N. Identifying therapeutic changes by simulating virtual language stages: a method and its application to Middle English coda phonotactics after schwa deletion. *VIEWS* 24: 1-31.
2013	Baumann, A. *Evolutionary dynamics in structured populations: a life-cycle approach in continuous time*. Diploma thesis. Vienna University.
2012	Ritt, N.; Baumann, A. Transferring mathematics to English studies. In: M. Manfred and H. Schendl (eds.). *Transfer in English studies*: 219-237. Vienna: Braumüller.
2009	Baumann, A. *The triggering learning algorithm and the problem of local maxima*. Diploma thesis. Vienna University.

Conference papers and guest talks

09/2017	With N. Ritt, C. Prömer, K. Kazmierski. *Phonotactic word form shapes are selected to be morphotactically indicative*. Diachronic phonotactics workshop, Vienna.
08/2017	With L. Sommerer. *Layering an effect of asymmetric priming*. 23rd ICHL, San Antonio, TX.
08/2017	With C. Prömer. *Interpolating diachronic phonotactic data: On the logistic spread of Middle English schwa loss*. 23rd ICHL, San Antonio, TX.
07/2017	With K. Kazmierski. *Perceptual effects of ambiguity in the long-term development of boundary-signaling consonant clusters: combining experiments and dynamical systems in (mor)phonotactic research*. Dynamic Modeling Workshop, Cologne.
07/2017	With L. Sommerer. *Layering as a long-term effect of asymmetric priming.* 14th ICLC, Tartu.
07/2017	*Early acquired sound sequences spread, late ones don't.* 14th ICLC, Tartu.
06/2017	With K. Kazmierski. *Lazy speakers and distant sounds: on the role of articulatory difference in phonotactic production, acquisition and change*. PaPE, Cologne.
05/2017	*Manner of articulation is the primary articulatory pressure in the formation of phonotactic systems: evidence from the acquisition and diachrony of Dutch and Afrikaans*. Adam Mickiewicz University in Poznań, Poland.
09/2016	With K. Kazmierski. *Assessing the effect of compositionality-signaling variability on the segmentation function and processability of diphones*. Experimental Approaches to Perception and Production of Language Variation, Vienna.
08/2016	With K. Hofmann. *The effect of differential stress patterns on age-of-acquisition ratings in English*. SLE, Naples.

08/2016	With C. Prömer, N. Ritt. *Identifying therapeutic changes by constructing virtual language stages*. SLE, Naples.
08/2016	*Stabilizing determinants in the transmission of phonotactic systems: on the emergence of the Afrikaans consonant-cluster inventory*. Afrikaans Grammar Workshop, Johannesburg.
04/2016	With E. Zehentner, N. Ritt, C. Prömer. *A Game Theoretic Account Of Semantic Subjectification In The Cultural Evolution Of Languages.* Evolang XI, New Orleans.
04/2016	With C. Prömer, K. Kazmierski, N. Ritt. *A Lotka-Volterra model of the evolutionary dynamics of compositionality markers*. Evolang XI, New Orleans.
11/2015	With C. Prömer, N. Ritt. *Diachronic reflexes of frequency effects among word final (mor)phonotactic consonant clusters in Middle and Early Modern English*. 3rd International Workshop on Phonotactics and Phonotactic Modeling, Vienna.
09/2015	With K. Kazmierski: *A structured-population dynamical-systems approach to the evolution of morphonotactic and lexical consonant clusters.* PLM (45th Poznań Linguistic Meeting): Poznań, Poland.
12/2014	*Mathematical modelling of language change: what is the point of all this?* University of Stellenbosch, South Africa.
11/2014	*Languages as eco-evolutionary systems: why it sometimes still makes sense to compare linguistic change with disease.* SACEMA, Stellenbosch, South Africa.
04/2014	*Rhythm-driven evolutionary dynamics of lexical stress in natural languages.* EvoMus Workshop, Evolang MA/PhD,X: Vienna, Austria.
08/2013	*When uniform stress placement collapsed: the history of English word stress told by Evolutionary Game Theory*. ICHL 21: Oslo, Norway.
05/2013	*Eco-evolutionary dynamics of linguistic replicators*. Ways to Protolanguage 3: Wrocław, Poland.
12/2012	*Grammaticalizing to death: evolutionary suicide of OE/ME case markers.* Memetics Workshop: Vienna, Austria.
10/2012	*Grammatikalisierung, Optimierung und Unidirektionalität.* OELT 39 (Österreichische Linguistiktagung): Innsbruck, Austria.
09/2012	*Struggle for rhythm: an evolutionary approach to stress assignment.* PLM (43rd Poznań Linguistic Meeting): Poznań, Poland.
07/2012	*Grammaticalization and evolutionary optimization.* NRG 5 (New Reflections on Grammaticalization): Edinburgh, Scotland.
03/2012	*Grammatikalisierung als evolutionärer Optimierungsprozess.* 2nd Workshop 'Sprachwissenschaftliche Dissertationsprojekte der Wiener Germanistik': Vienna, Austria.
09/2011	With Nikolaus Ritt: *Lexical stress, game theory and utterance rhythm.* SLE 44 (Societa Linguistica Europaea): Logroño, Spain.
07/2010	*Der Triggering Learning Algorithm, lokale Maxima und negative Evidenz.* Eberhard Karls University, Tübingen, Germany.

11/2009	*An approach to negative evidence in formal language acquisition models.* 2nd ÖSKL (Österreichische Studierendenkonferenz in Linguistik): Vienna, Austria.

Teaching

12/2017	*Statistical modeling and multimodel inference* (workshop at Adam Mickiewicz University in Poznan, Poland)
12/2017	*Introducing statistics module* (BA/MA, Adam Mickiewicz University in Poznan, Poland)
2017 (SS)	*Introduction to statistical modeling for linguists* (MA, 122252 AR)
2017 (WS) 2016 (WS)	*Approaching statistical problems in English language studies with R* (Staff seminar, 120015/120010 SE)
2015 (WS) 2014 (WS)	*Introduction to statistics for linguists* (MA, 122252 AR)
Since 2013 (WS)	Statistics module in *Research methods in linguistics* (MA, 129005 AR)
2015 (SS)	*Introducing statistics* (guest lecture in *Introduction to information and research literacy*, BEd, 120422 VO)
2014 (SS) 2013 (SS) 2012 (WS)	*Proseminar Linguistics 1* (BA, 122040 PS)
2013 (WS)	*Quantitative and statistical methods* (MA, 129008 AR)
2012 (SS)	*Introduction to Evolutionary Linguistics* (MA, 122255 AR, co-teaching)

Abstract

When we speak, we do not use sound sequences arbitrarily. For instance, no English word ends in the sequence /mb/, while many words end in /ks/ (like in *box*), yet we do not find any words that begin with /ks/. The study of the rules and tendencies that determine which sound sequences are permitted, ruled out, or preferred in a language is called ‘phonotactics’.

Some of the constraints on the phonotactic setup of natural languages reflect weak biases in processing, whose effects accumulate when languages are transmitted in vast numbers of parallel and iterated acquisition and interaction processes, and as a consequence become visible in language history. Thus, history provides evidence not only of articulatory and auditory constraints on language production and perception, but also on cognitive constraints on the processing of language.

In this dissertation project, I study how sound sequences like /mb/ and /ks/ replicate and spread through languages and populations of speakers. I focus on the question of why some sequences do so more successfully than others. The project tries to account for the evolution of sound sequences in terms of a number of factors on different levels of organization. These include (a) the properties of phonotactic expressions in speech, (b) articulatory and perceptual constraints, (c) cognitively grounded constraints, (d) the impact of the systemic-linguistic environment such as elements in their phonological context or morphological structure, and (e) constraints imposed by the population structure speakers are embedded in.

Empirically, the project is based on evidence derived from language acquisition studies, from digital diachronic and synchronic text corpora in different languages, as well as from experimental research. Theoretically, it is grounded in the framework of evolutionary linguistics and conceptualizes languages as systems of culturally transmitted, replicating constituents of linguistic knowledge. The methods that I apply are derived mostly from evolutionary ecology and epidemiology and consist to a large extent in the mathematical modeling of dynamical systems.

By analyzing these models and testing them against empirical language data, I demonstrate the relevance of factors like ease of acquisition, perceptual contrast, heterogeneity and stability of linguistic input, and morphological parsing on the long-term evolution of phonotactic knowledge.

Zusammenfassung

In gesprochener Sprache werden Lautfolgen nicht beliebig verwendet; so endet beispielsweise kein englisches Wort auf /mb/ und es beginnt auch kein Wort mit /ks/, wohingegen viele Wörter auf /ks/ (wie in *box*) enden. Die ‚Phonotaktik' befasst sich mit den Regularitäten und Tendenzen davon, welche Lautfolgen in natürlichen Sprachen gebraucht werden.

Phonotaktische Einschränkungen natürlicher Sprachen resultieren oft aus relativ schwachen Tendenzen in der Verarbeitung phonotaktischer Strukturen. Wenn sprachliches Wissen in einer Vielzahl parallel und wiederholt stattfindender Interaktionen weitergegeben wird, werden diese schwachen Tendenzen sichtbar, sodass die Geschichte einer Sprache nicht nur artikulatorische und perzeptuelle Faktoren in der Sprachproduktion und -perzeption sondern auch kognitive Faktoren in der Sprachverarbeitung aufzeigt.

Das vorliegende Dissertationsprojekt befasst sich damit, wie sich Lautfolgen wie etwa /mb/ oder /ks/ in Sprachsystemen und Sprecherpopulationen ausbreiten, wobei die Frage danach im Mittelpunkt steht, warum manche Lautfolgen diachron erfolgreicher sind als andere. In diesem Projekt wird die Evolution von Lautfolgen aus unterschiedlichen Gesichtspunkten und auf verschiedenen Organisationsebenen betrachtet. Dies umfasst (a) Eigenschaften sprachlicher Äußerungen, (b) artikulatorische und perzeptuelle Faktoren, (c) kognitive Faktoren, (d) Faktoren, die die linguistisch-systemische Umgebung sprachlicher Elemente betreffen (wie etwa der phonologische oder morphologische Kontext) und (d) der Einfluss der Populationsstruktur, welcher Sprecher ausgesetzt sind.

Aus empirischer Sicht greift das Projekt auf Sprachdaten aus dem Erstspracherwerb und diachrone sowie synchrone Textkorpora unterschiedlicher Sprachen zurück, sowie auf experimentell gewonnene Daten. Aus theoretischer Sicht lässt sich das Projekt in die evolutionäre Linguistik einbetten, wobei Sprache als System kulturell weitergegebener replizierender Konstituenten linguistischen Wissens aufgefasst wird. Die dabei verwendeten Methoden entstammen Großteils der evolutionären Ökologie und Epidemiologie; die mathematische Modellierung mittels dynamischer Systeme spielt dabei eine wichtige Rolle.

Durch die Analyse der Modelle und den Abgleich ihrer Vorhersagen mit empirischen Sprachdaten wird die Relevanz verschiedenster Faktoren wie etwa von Lernbarkeit, perzeptuellem Kontrast, Heterogenität sprachlichen Inputs oder morphologischer Dekomposition in der Langzeitentwicklung phonotaktischen Wissens aufgezeigt.