ISSN 2074-9864 (online)

VIENNA ENGLISH WORKING PAPERS

HTTP://ANGLISTIK.UNIVIE.AC.AT/VIEWS/

VOLUME 24 **2015**

ANDREAS BAUMANN, CHRISTINA PRÖMER & NIKOLAUS RITT

Identifying therapeutic changes by simulating virtual language stages: a method and its application in the study of Middle English coda phonotactics after schwa deletion

Published online: 29 December 2015

Identifying therapeutic changes by simulating virtual language stages: a method and its application in the study of Middle English coda phonotactics after schwa deletion

Andreas Baumann, Christina Prömer & Nikolaus Ritt, Vienna†

This article presents a new method for testing hypotheses in diachronic linguistics. It consists in the construction of hypothetical language stages that reflect the immediate effects of changes which are supposed to have triggered therapeutic responses in the language because (some of) their outputs were sub-optimal with regard to universal or language specific constraints. We demonstrate the method in terms of the example of Middle English schwa loss and its effects in the domain of coda-cluster phonotactics. We show that schwa loss created codas that were phonologically dispreferred and created ambiguities in the phonotactic representation of morphological word structure. We also show that the problems brought about through schwa loss were (partly) resolved in the later development of the language.

1. Introduction

Our paper introduces a method of testing hypotheses about the historical development of languages. Our proposal reflects a way of thinking that is well established and implicit in most reasoning about language change: anybody who gives a reason why language (X) has come to assume some attested state (A), necessarily implies that the same reason also explains why it has not come to assume any of the virtual alternative states (¬A) instead. What we propose is simply to make this line of reasoning more accessible to systematic testing by simulating virtual language states and by comparing them to the ones that actually have come about. We show that in some cases it is not only possible and illuminating to proceed by this method, but in fact necessary for identifying effects that would otherwise be difficult to detect. The way in

†The authors' e-mails for correspondence: andreas.baumann@univie.ac.at, christina.proemer@univie.ac.at, nikolaus.ritt@univie.ac.at. The work reported in this paper was supported by the Austrian Science Fund (FWF): grant No. P27592-G18.

which we propose to proceed involves, essentially, the following steps: first, we select a specific change that is known to have occurred in a language. Then, we take corpus data from a period before the change and simulate its effects on them. This yields a set of data which we take to reflect a virtual post-change stage, in which no other change occurred except the one whose effects we artificially created. Next, we search the theoretical literature for predictions about properties of that virtual stage that were brought about by the change whose effects we simulated. Finally, we test these predictions by comparing the virtual post-change data to actual post-change data. If observable differences are indeed as predicted, we take the theories to be corroborated.

In order to make our argumentation easy to follow, we illustrate the method in terms of a concrete example, namely the hypothesis that the loss of schwas in unstressed final syllables in Middle English not only increased the number of word final consonant clusters (for instance by changing earlier /ˈkli:mbə/ 'climb' into /ˈkli:mb/, or earlier /failəd/ 'failed' into /faild/), but also triggered a number of subsequent developments that reflect the predictable effects (a) of quasi-universal[1] constraints on the production, perception, and transmission of those clusters, and (b) of a semiotically grounded preference for phonotactic configurations that signal the morphological structure of word forms and thereby facilitate their decomposition in processing.

In section 2 of this paper we describe schwa loss and theories concerning the clusters it produced. Section 3 introduces the method and discusses the rationale on which it is based, sections 4 and 5 demonstrate how it applies in the case of schwa loss and the clusters it produced, and section 6 reports and discusses the results it delivers.

First, however, a few remarks are made in order to explain why the study of language change should at all be in need of new methods for hypothesis testing, and, more generally, in order to contextualize our paper.

The historical study of languages is often based on the general premise that "everything is the way it is because it got that way"[2]. As is well known, of course, this approach raises the vexing question of whether the events that amount to the histories of languages are at all causally related to one another, so that they can be explained or understood, or whether they represent mere sequences of "one damn thing after the other" (Toynbee 1954 [reprint 1961]: 151), in which case they can only be acknowledged and described.[3] The question is crucial and philosophically challenging. But even if one puts meta-theoretical considerations aside

[1] Because physiologically grounded.
[2] An often-quoted saying attributed to the evolutionary biologist D'Arcy Wentworth Thompson (1860-1948).
[3] Arguments against historical explanation in linguistics were first made popular by Ferdinand de Saussure (1916), but see also Lass (1980), Adamska-Sałaciak (1989).

and simply starts on the assumption that the historical development of languages is lawful, the practical problem remains that hypotheses about any laws that might govern linguistic evolution are very hard to test. On the one hand, this is because the study of history is by definition retrospective, but on the other hand it is also because languages are complex and variable systems, and embedded in contexts that are equally complex and variable. Thus, (the history of) any specific language differs, predictably, from (those of) all others, and the number of factors that are potentially responsible for such differences is vast. This means that one cannot falsify a hypothesis about a development in one specific language by simply checking if its predictions are borne out in the history of another one, even if the two languages appear to be comparable in all relevant respects. The reason why this rarely works is that one can never be sure that a specific development in a specific language might not reflect factors that one did not consider when developing the hypothesis to be tested. Thus, if a hypothesis derived from an observed development in one language predicts similar developments in other languages, and if these do not undergo them although they satisfy the relevant conditions, this does not necessarily mean that the hypothesis is wrong, and if one believes that hypotheses should be falsifiable, this is bad.[4]

When laws are immune to falsification by individual exceptions, the next best thing is to treat them as probabilistic, and to test them statistically, i.e. against many different data sets. For instance, the probabilistic law that smoking increases one's chance of contracting cancer counts as corroborated by the fact that more smokers than non-smokers get cancer, even though there may be smokers who never do. In the historical study of languages, however, also statistical methods are hard to apply. This is because the number of languages whose histories are well documented is actually quite small (at least compared to the number of smokers, for example), and the number of languages or language stages in which the specific conditions are met that allow the testing of specific hypotheses will necessarily be – even considerably – smaller. Therefore, methods by which available data can be used for testing

[4] For example, it strikes one as intuitively plausible that English has lost the rich system of nominal and verbal inflections still attested in Old English because, as a Germanic language, it had word stress fixed on the first syllable of word roots, so that inflectional endings received little energy in pronunciation, were difficult to perceive and to learn, and eventually ceased to be transmitted altogether. Plausible as the story sounds, however, the hypothesis that it seems to imply, namely that fixed root initial stress will lead to the loss of inflectional suffixes, is very difficult to falsify. One cannot invalidate it by pointing to German, for example, whose inflectional morphology is still quite rich even though, like English, it is a Germanic language and its ancestor had stressed fixed on the first root syllable. This is because the history of English differs in various, and plausibly relevant respects from that of German. English went through long periods of intensive contact with other languages, for example, and was hardly written for one and a half centuries after the Norman Conquest of 1066. These contextual, system-external factors may plausibly have increased the effect of word stress on the inflectional system, and their absence may have reduced it in the case of German. Thus, German does not falsify the hypothesis that fixed initial stress causes inflectional suffixes to be lost.

hypotheses about language change are (or: should be) in great demand, and this motivates our paper, which introduces one.

Since our method represents a specific way of preparing empirical data for (primarily quantitative) analysis, its usefulness does not depend on a specific theory of language or language change. Nevertheless, it might be adequate to outline – if only briefly – our own way of conceptualizing the phenomena that historical linguists study and that our method targets.

We think of languages as complex evolutionary systems (Brighton et al. 2005, Chater & Christiansen 2010, Cowan et al. 1994, Croft 2000, Deacon 1997, Gell-Mann 1994, Hurford 2007, 2011, Kirby 2012, Ritt 2004). This means that their constituents represent populations that are distributed among the minds of the speakers who share them, and expressed in speakers' utterances. In order to remain stably established in a population, the constituents of languages need to be transmitted among speaker generations through communication and language acquisition. Thus, generally speaking, what explains their existence in any language at any specific time is that they have emerged at one point and have been successfully transmitted since. What one therefore needs to understand are the factors that determine the chance of a constituent to be transmitted. We think those factors can be divided into three general classes: (i) socio-cultural factors, (ii) physiological constraints and (iii) context-dependent factors. Of the three, the impact of at least the latter two strikes us as accessible to explanation.

Let us first consider the first class of factors, however, and discuss why their effects are likely to resist explanation. The fact that languages are sensitive to their socio-cultural contexts reflects that they are conventional and arbitrary systems (Saussure 1959) and that they serve such social purposes as the establishment of group identity and group coherence. This means that the likelihood of any constituent to be successfully transmitted and established in a group will depend on its happening to be 'cherry-picked' as a group marker. To the extent that it is vital for speakers to be recognizable and accepted as group members, they will conform to any linguistic convention that is socially established – at least as long as it is at all learnable and not absurdly dysfunctional. At the same time, the purpose of establishing group identity necessarily requires linguistic codes to be idiosyncratic and, at least to some extent, unpredictable, because otherwise they would be too easy to figure out by outsiders (Dunbar 1996). To the extent that unpredictability represents a design feature of linguistic codes, it is inevitable that they should be unpredictable by linguists as well. In that sense, some aspects of language must remain forever inaccessible to scientific explanation.

The same is not true with regard to the other two classes of factors that affect the transmittability of linguistic constituents. It is plausible, for example, that their transmission should be sensitive to (ultimately) physiological constraints, because although they may be implemented in human minds, linguistic constituents need to be expressed in speech or writing, their acoustic or visual expressions need to be perceived, and they need to be reconstructed or inferred from their expressions. No conventional code is likely to establish itself whose elements are inexpressible, imperceptible, or unlearnable. This means that the transmission of linguistic constituents must be constrained by the make-up of the sensori-motor system and the cognitive-ideational system, and these constraints can clearly be considered as universal – at least for all practical purposes and to the extent that all humans are physiologically similar (Donegan & Stampe 1979, Donegan & Stampe 2009, Dressler 1985, Prince & Smolensky 2002).

However, although physiological constraints on constituent transmission are likely to be universal, their effects on any specific constituent must at the same time be context dependent as well, and the elements that a constituent will most frequently find in its context are other constituents of the language (Ritt 2004; Albright 2008). In other words, all elements in the linguistic system of which a constituent happens to be a part, make up a separate, third class of factors that may potentially affect its transmittability.

For instance, any phoneme must be perceptible to be transmitted, and the perceptibility of consonants is greatest when they are uttered before vowels. Similarly, the perceptibility of any segment is greater if it is expressed in a syllable on which the prosodic system of a language confers stress than if it is expressed in a systematically unstressed syllable. Therefore, it is not surprising that CV syllables are more frequent than others in the languages of the world (Maddieson 2011), and that the variety of segments that occur in unstressed syllables is comparably reduced. To give another example, the communicative reward for expressing inflectional morphemes will be smaller in languages where thematic roles are additionally signaled through word order than in languages where they are not. Thus, a plausible preference for minimizing effort in articulation and processing will impede the successful transmission of inflectional case morphemes in languages that have, or are developing, syntactically fixed word order.

In short, any natural language will consist of constituent populations whose elements are not only adapted to more or less universal, physiological constraints on their transmission, but, crucially, also to one another. This implies that changes in one set of constituents in a system are likely to affect the transmission of other constituents as well. If the successful transmission of constituent (A) depends on the presence of constituent (B), and if constituent (B) is replaced by (B’), then constituent (A) may cease to be transmitted as successfully as it was, and may disappear or be (at least partially) replaced by (A’), if (A’) transmits better in the context of (B’) than (A).

Of course, the degree to which two constituents support or impede each other's transmission will obviously depend itself on the universal, physiological constraints we discussed above. Thus, theories of such universal constraints or preferences should be able to predict how changes in one constituent will affect the fate of those whose transmittability depends on it.

The method we describe in this paper is designed to test predictions derived on the basis of this very general argumentation. It thus focuses specifically on changes that are potentially triggered by the loss (or the replacement) of constituents on which the transmission of others depends. To the extent such changes re-establish system states in which constituents are well-adapted to each other (i.e. mutually supportive of one another's transmission), they can be (and have been) called 'therapeutic' (cf. Aitchison 2001).

For the sake of illustration, and as already pointed out, we describe our method in terms of a specific episode in the history of English phonology, namely the disappearance of schwas in unstressed final syllables, its immediate effects on the inventory of consonant clusters, and some of the changes that occurred in its wake. Our goal is to examine if they represent the 'therapeutic' responses that theories of universal, physiologically and cognitively grounded preferences would predict.

2. Schwa loss and why it represents a suitable test case

Schwa loss is an Early Middle English sound change by which schwas were systematically deleted from final unstressed syllables. It first affected open syllables where it produced forms such as /ma:k/ 'make' from earlier /ˈma:kə/, or /ˈkli:mb/ 'climb' from earlier /ˈkli:mbə/. In a second phase, it affected checked syllables (i.e. syllables containing codas, e.g. CVC), producing forms like /fa:md/ 'famed' from /fa:məd/, or /faild/ 'failed' from /failəd/. In open syllables, schwa loss started in the late tenth century, and was completed by the middle of the fifteenth, while checked syllables started to display schwa loss from the thirteenth century onwards, but it was not before the end of the sixteenth century that the change was completed (Minkova 1991, 2009).

What makes schwa loss particularly suitable for illustrating our method is that some of its effects on word-form phonotactics were blatantly suboptimal in terms of well-established phonological and semiotic preferences, and seem to have called for 'therapeutic' responses in the sense outlined above. In particular, schwa loss considerably increased the frequency of complex codas, in terms of both types and tokens. Thus, before schwa loss, the clusters in forms of *climben* 'climb' had been realized medially rather than finally in most forms of the verb, and the clusters emerging in forms like *fayled* 'failed' had not been clusters at all. As indicated, what makes this drastic increase in the frequency of coda clusters interesting is that extant linguistic theories contain clear predictions about their historical stability.[5] Thus, consonant clusters in general, and coda clusters in particular, count as highly marked or dispreferred (Cairns & Feinstein 1982; Clements & Keyser 1983; Vennemann 1988; Restle & Vennemann 2001; Dressler & Dziubalska-Kołaczyk 2006; Dziubalska-Kołaczyk 2009; Flack 2009). This is primarily so because the lack of contrast between their elements makes them hard to perceive, particularly so in the prosodically weak coda position. This makes them also difficult to acquire, and suggests that they should be historically unstable.[6]

At the same time, there is evidence that coda clusters are produced more faithfully, and learnt (and therefore also transmitted) more easily when there is a morphological boundary between their elements, that is to say when they are produced by morphological operations (Dressler & Dziubalska-Kołaczyk 2006; Dressler et al. 2014, Ritt 2010; Ritt & Kazmierski 2015) . This is the case in the past tense form *failed*, for example, but not in lexical *field*. Two reasons have been proposed for this. On the one hand, the individual elements in morphologically produced (or morphonotactic) clusters also occur independently of each other, and in contexts where they are easier to perceive. For example, the /l/ of the /ld/ cluster in *failed* occurs also in forms such as *failing* or *failure*, and the /d/ also occurs in forms such as *played*. On the other hand, morphonotactic clusters may signal the morphological complexity of the forms they occur in. The /md/ in *seemed* does so unambiguously, for example, because it never occurs in simple word forms. This may facilitate the decomposition of such complex forms by listeners, and may motivate speakers to pronounce the final clusters faithfully: exploiting their signaling function will help them to communicate more effectively. Thus, it can be hypothesized that morphonotactic coda clusters should be more stable than purely ‘lexical’ clusters, i.e. clusters at the end of simple word forms.

Finally, interesting predictions have been made about the interaction between lexical and morphonotactic clusters, particularly about constellations involving homophony. For example, the signaling function of coda clusters may turn into a disadvantage when they send the wrong signal. This can be the case when a cluster is both produced through a morphological operation and occurs in simple lexical forms as well, such as the cluster /ld/, which occurs both in past tense forms like *failed* and in simple words like *field*. For such cases, the prediction has been made that the number of such structural homophones will tend to be reduced (Dressler & Dziubalska-Kołaczyk 2006, Dressler et al. 2010). This could theoretically be ef-

[5] Note that we do not discuss the causes of schwa loss itself, as this would go beyond the scope of this paper. Instead, we take schwa loss for granted and focus exclusively on its effects.
[6] The smaller the auditory distance between their elements is, the less stable they should be.

fected in various ways, such as by deleting the final /d/ in lexical items, while retaining it (or keeping to re-produce it by rule) in complex word forms, or simply by avoiding the use of the simple forms involved in the homophony.

While there are reasons for assuming that homophony between lexical and morphonotactic clusters will tend to be reduced, however, there are also reasons for predicting the opposite. Thus, it is well known that children start to acquire language by memorizing word forms – simple or complex, it does not matter – as unanalyzed lexical chunks, and they acquire more frequent forms before rarer ones. If a productive rule such as regular past tense formation produces word forms ending in /nd/ frequently enough, it is easily conceivable that children should acquire them before becoming aware of their complexity, and this may induce them to infer that it is perfectly fine for simple words to end in /nd/. Thereby, the acquisition of truly lexical /nd/ words may be facilitated as well (Jusczyk et al. 2002).

Thus, there are two further (and complimentary) hypotheses to be tested: First, the hypothesis that has come to be known as the Strong Morphonotactic Hypothesis (SMH), which claims that a cluster which occurs both lexically and morphonotactically will experience a selective repair process affecting exclusively lexical or complex sequences. Second, a hypothesis one might call the Hypothesis of Mutual Support (HMS), which claims that lexical clusters will remain stable as long as there are a sufficient number of complex homophones to support them.[7]

In sum, the situation created by schwa loss lends itself well for testing hypotheses about the historical stability of coda clusters: it produced a large number of them, it produced both morphonotactic and lexical ones, and it produced homophonous clusters in some cases, but not in others. Thereby it produced a constellation of phenomena that conflicted, in many respects, with phonological and semiotic preferences and for which therapeutic changes affecting the involved constituents can be predicted.

The method introduced in this paper is specifically designed to test whether ‘therapeutic’ changes, such as the ones predicted to occur after schwa loss, are indeed identifiable. Crucially, and as already indicated above, it goes beyond a mere analysis of the history of coda clusters as it unfolded in the wake of schwa loss. Instead, we first construct a corpus of virtual (or hypothetical) post-schwa-loss data, which are exactly like pre-schwa-loss data except that the effects of schwa loss have been simulated on them. Only then do we compare them to actual post-schwa-loss data and see if the differences are as our theories predict. In the next section, we explain why it is necessary to take this additional step.

[7] And potentially *vice versa*.

3. The method and what motivates it: general considerations

3.1 Linguistic changes on different time-scales

As indicated, we think of languages as evolving complex systems. The notion of complex systems refers to the fact that languages consist of a large number of (mutually or unidirectionally) interdependent and interacting constituents with very diverse properties. Evolution implies that certain properties of these entities, such as – most crucially – their abundances in the speaker population, change. The question that we pursue in this paper is whether a specific language change can be shown to have triggered (a set of) subsequent changes. Specifically, we ask whether predictions about the coda clusters produced by schwa loss – such as that they should undergo cluster simplifications, or that homophony between morphonotactic and lexical clusters should be reduced – were indeed borne out. One problem that one faces when addressing questions such as these is that in the real world changes do not neatly succeed one another. Instead the periods of their implementation frequently overlap, and some changes may even start and be completed while other changes, which started earlier, still continue to unfold. Thus, when one attempts to distinguish triggering changes from the ones that they trigger, one needs to take into account that changes can proceed at different speeds and must therefore also distinguish between changes that unfold on different time scales. For the purposes of this paper, we call changes that take a long time to come to completion 'large-scale changes' and changes that are completed within relatively short time spans 'small-scale changes'.

In the history of English, word-final schwa loss can be regarded as 'triggering large-scale change'. It qualifies as a 'large-scale change' because it gradually spread through the language during a period of more than 500 years (Minkova 1991; Dobson 1957). In addition, at least for the purposes of our paper, it can be regarded as a 'triggering change'[8], because we are interested in the changes that it may have triggered.[9] 'Triggered small-scale changes', on the other hand, are defined as immediate responses to the differences in their context – or 'environment' – such as those brought about by a triggering large-scale change. Additionally, they are completed within much shorter periods.

A large-scale change comes to be a triggering change if it exposes linguistic constituents such as strings of sounds, morphemes or words to cognitively or physiologically (e.g. articulatorily) grounded pressures of universal constraints or preferences in such a way that

[8] Even though it may itself have been triggered, we do not concern ourselves with its causes here. Instead, we simply regard it as having occurred for whatever reason.

[9] Note that the crucial feature of 'large-scale changes' is the length of the period from their onset to their offset, which is supposed to be comparably large as opposed to that of other language change phenomena. Whether or not a large-scale change is also triggering other changes is a different matter.

their transmittability and thereby their stability in the system is compromised. The linguistic entities that come to be exposed to such pressures may then undergo changes themselves – qualitatively or in terms of frequency –, so that the relevant constraints cease to be violated and a stable equilibrium is restored, or at least re-approached. Such triggered small-scale changes can be considered as ‘therapeutic’ in the sense that they happen to configurations of linguistic constituents which do not optimally support one another’s transmission under cognitively or physiologically grounded constraints, and increase their adaptedness to one another.

Consider now that it is possible for triggered small scale changes to be completed before the large scale changes that have triggered them. Take, as a hypothetical example of a small-scale change potentially triggered by schwa loss, the process of word-final obstruent devoicing: as soon as schwa loss causes previously medial consonants to surface word-finally, they are exposed to the well-established constraint against voicing in codas. Plausibly, they may respond by undergoing the small-scale change of word-final devoicing. If final devoicing spreads at a faster rate than schwa loss, all final obstruents might wind up fully devoiced throughout the lexicon, even while schwa loss itself is not fully implemented. As we shall see, this possibility represents a problem for the identification of therapeutic small scale changes triggered by large scale changes, and plays a central role in motivating the particular method we propose.

3.2 How large-scale changes and their potential effects are usually explored and why this does not always work

Assume, for the sake of the argument, that a specific small scale change has occured. Let us call it change X, and let us say that it affected the frequency of variable *x*. Say, furthermore that we think that change X was triggered by the large scale change Y. In order to check if that hypothesis may at all be true, we need to collect data at least from the period before Y started (pre-Y data) and from the period after Y was completed (post-Y data). We would then measure the frequency of *x* in both pre-and post-Y-data. If Y was the trigger of X, then the effects of X should be detectable as a significant difference between the frequencies of *x* in the two data sets (Figure 1a). Of course, this would still not show that our hypothesis actually *is* true, but only that it *might* be, because the difference between the frequencies of *x* may have any other causes than the large scale change Y. Clearly, we would have to employ other quantitative methods for assessing how likely it is that X was actually triggered by Y. Nevertheless, the comparison of pre- and post-Y data would show whether our hypothesis is at all worthy of further pursuit.

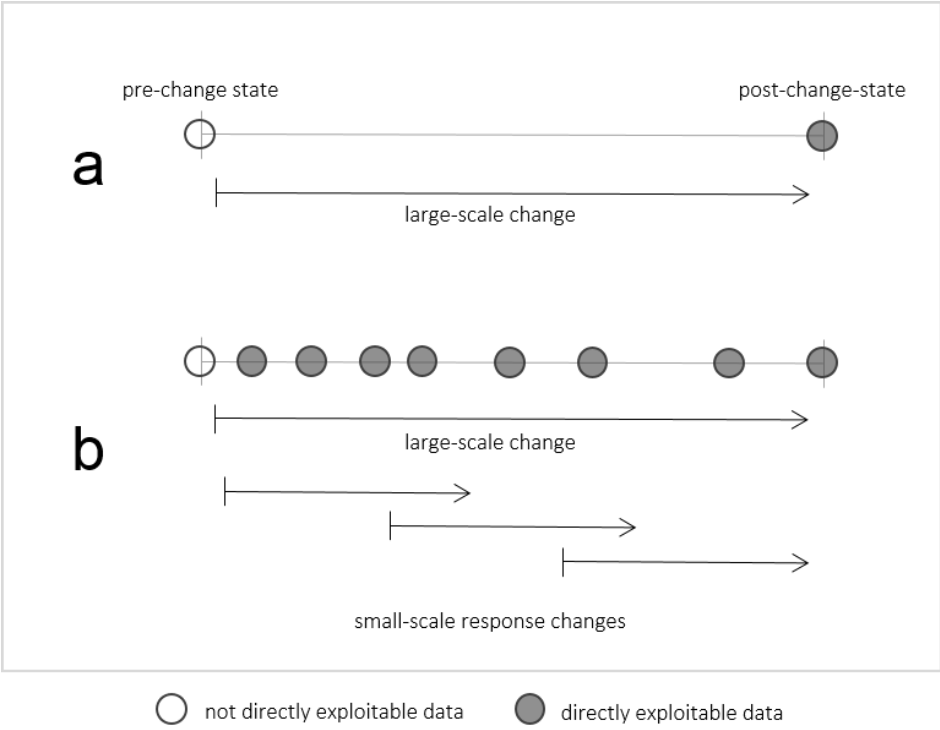

Figure 1. Schematic representation of the data which are necessary for exploring (a) large- and (b) small-scale language changes.

Of course the situation is better when we also have data from stages within the period during which Y was implemented. In that case, we can measure the frequencies of *x* for each of those intermediate stages. If we observe that changes in the frequency of *x* correlate with changes in the degree to which Y was implemented, then this would make the hypothesis of a causal relation more plausible (Figure 1b) – even though, of course, other potential causes would still have to be considered in order to assess the actual impact of the large-scale change Y on the small-scale change X.

So much for the normal procedure. Depending on the availability and the nature of the data at our disposal, however, two practical problems may arise and complicate matters. First, we may not have sufficiently exact knowledge about intermediate stages of the period during which a gradual change got implemented, but only about the pre-change stage and the post-change stage. Second, the variable presumably changing in response to a large scale trigger may only be detectable in data from the post-change stage, but not in data from the pre-change stage. This can be the case, for example, if the variable is only first produced by the large-scale change itself.

In cases where both problems pose themselves, the standard method of comparing pre-change, post-change, and intermediate stages can obviously not be applied. This is because we don't have anything to compare the post-change state with, nor do we have any information about intermediate stages, so that any correlations in diachronic development cannot be observed either.

Unfortunately, this situation obtains in the case of schwa loss and many of the changes it is assumed to have triggered. To illustrate the problem, consider the question if the number of words involved in homophones such as the one between the /ld/ in *field* and the /ld/ in *failed* was reduced after schwa loss first produced it. Clearly, we cannot compare pre-schwa loss data to post-schwa loss data, because schwa loss only produced the homophones in question. Nor can we look for correlations in data from different phases during the implementation of schwa loss, since its implementation is extremely difficult to reconstruct. Particularly in checked syllables schwa loss was implemented gradually, remained optional between the late thirteenth and the sixteenth centuries, and is practically impossible to infer from spelling evidence.

As we shall show in the next sections, our method is specifically designed to make cases such as this accessible to systematic quantitative study. Since it creates a hypothetical language stage that displays all effects of an assumed trigger, but none of the effects of the responses it may have triggered, it neatly separates the effects of an assumed large-scale trigger from any of the potentially therapeutic small-scale changes that may have been triggered either after or during its implementation.

3.3 Simulating hypothetical language stages

Think of a language as a set of constituents, a subset of which is affected by a large-scale change – in our case schwa loss. This implies that there will be two successive language stages which necessarily differ with regard to the set of constituents that undergoes the change. While the two stages may differ with regard to any number of other constituents, only some of these differences may reflect the impact of therapeutic response changes, and they are the ones we need to identify. Let us call the two stages the **actual pre-change stage** and the **actual post-change stage**, where 'change' refers to the large-scale trigger in whose effects we are interested (Figure 2a). Generally speaking, each of the two actual language stages is represented by a set of data, and contains subsets representing different constituent types. Now we construct a third set of language data, which is meant to represent a ***virtual*** **post-change stage**. This data set is identical to the actual pre-change data, except for the subset representing a single constituent type, i.e. the type affected by the large-scale change. With regard to that constituent set we consider the large-scale change to be fully implemented.

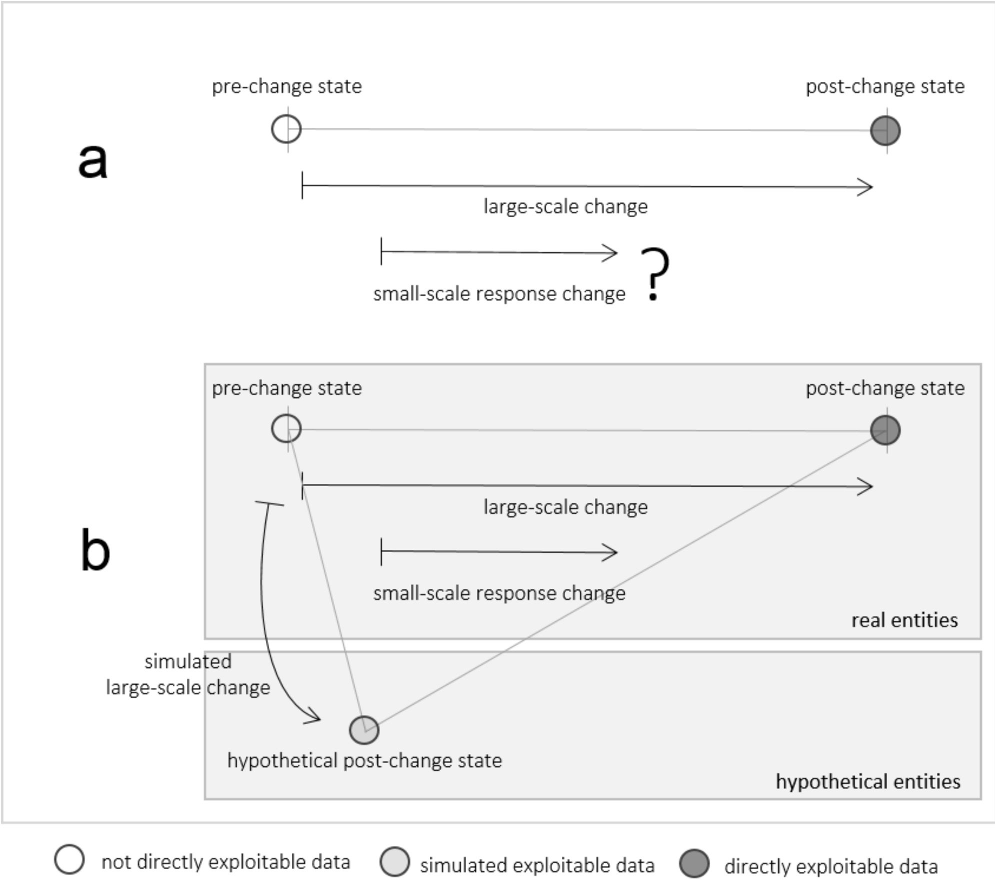

Figure 2. (a) Without temporally fine grained data small-scale response changes cannot be detected. (b) Schematic representation of the simulation of hypothetical language stages.

The virtual language data constructed in this manner can be conceptualized as representing a hypothetical stage in which the large-scale trigger was implemented fully and instantaneously. In terms of its properties, it therefore represents a post-trigger stage, while in terms of relative chronology it represents a stage *before* the onset of any therapeutic small-scale changes that may have been triggered (Figure 2b). This allows us to address two questions by straightforward quantitative methods. First, we can check whether an assumed large scale trigger did at all increase the degree to which a language conflicted with universal or language specific constraints or preferences, and might therefore have necessitated repair through therapeutic responses. This should become visible as a measurable difference between the frequencies of problematic items in actual pre-change and virtual (or hypothetical) post-change data [10]. Second, we can check whether the expected therapies were at all administered: to the extent that the outputs of a large scale trigger conflicted with universal or language specific constraints and preferences, and on the assumption that such conflicts were indeed repaired through therapeutic changes, the number of problematic items should be

[10] In the case of schwa loss it is not really necessary to test this. It is after all obvious that it increased the number of complex word-form codas and that complex codas are highly marked, or dispreferred.

higher (or at least not lower) in the virtual post-trigger data than in any of the data sets representing actual stages after the onset of the trigger. This prediction can be tested quite straightforwardly by comparing the number instances in virtual and actual post-trigger data – as explained in more detail in the next section.

3.4 Falsifying and corroborating hypotheses about therapeutic small-scale changes

In order to test hypotheses about therapeutic changes, and as pointed out, we need to check if the degree to which virtual post-trigger data violate universal or language specific constraints is indeed lower than the degree to which the same constraints are violated in actual post-trigger data. Since – like any change – also therapeutic changes may be implemented gradually, the measurable difference should rise in correlation with the temporal distance between the virtual post-trigger data set and the actual ones to which one compares it. Thus, if therapeutic changes did indeed occur in response to the large-scale change, the most substantial differences should be measurable between the frequencies of dispreferred items in the virtual post-trigger data and actual post-trigger data[11], respectively. If there is no significant difference between the two, the occurrence of a therapeutic response can be ruled out and the hypothesis that predicts it discarded.

Note at this point that the only purpose for which our method can be responsibly used is the falsification of previously established hypotheses about therapeutic changes. It can neither verify them[12], nor does it allow their inference by induction. This is because differences between the frequencies of constituents in the two language stages may in principle reflect other causes than the need to repair suboptimal effects of a large-scale trigger. Thus, the method proposed in this paper can indeed only discard hypotheses about potential therapeutic response-changes, not identify any.[13] Nevertheless, and in accordance with falsificationist theories of normal science, hypotheses about therapeutic small-scale changes may of course count as corroborated unless falsified.

[11] I.e. data from a period after the triggering change was completed rather than data from any of the periods during which it was still spreading.

[12] This will be obvious to anybody subscribing to Popperian standards of normal science, but it deserves to be pointed out nevertheless.

[13] This has implications for the set of statistical methods that is applicable for testing whether there is no such difference. Standard frequentist techniques, most prominently null-hypothesis significance testing (NHST), cannot be applied straight-forwardly for showing that there is *no* difference between two estimates in order to discard a candidate response change, due to their logical design (Cohen 1994, but see Cumming 2014). Rather, methods from Bayesian statistics (such as credibility-interval analyses) should be resorted to (Lesaffre & Lawson 2012). A discussion of these methods and their application to the present problems goes beyond the scope of this paper.

4. The method at work: Constructing hypothetical post-schwa-loss English

In the previous sections we have outlined our method and explained the rationale behind it in very general terms. This section illustrates how it works and what results it produces when applied to the case of Middle English schwa loss and its potential effects in the domain of coda cluster phonotactics.

As indicated in section 2 above, schwa loss set in at the beginning of the Middle English (ME) period and ended in the Early Modern English (EModE) period (Minkova 1991, Dobson 1957). Its implementation took more than 500 years, and it thus qualifies as a large-scale change as defined in section 3.

As also already pointed out, one of the effects of schwa loss was an increase in the number of word final consonant clusters, and various phonological theories suggest that these violate, or are badly adapted to, plausible constraints on their production, perception and transmission. Additionally, some of these theories also posit differences between morphotactically produced coda clusters on the one hand, and morpheme internal ones, on the other. From the various proposals, fairly detailed predictions can be derived about the historical stability of members of different subtypes in the coda cluster inventory, or, in other words, about therapeutic changes expected to occur to them.

In order to qualify as therapeutic changes, we expect them to occur within a time window that is bounded by the period in which schwa loss had begun to spread as a *terminus-post-quem*, and a time not too long after its completion, i.e. roughly the second half of the 17th century, as a *terminus-ante-quem*. Only if they occur within that period, we shall consider them as plausibly having been triggered by schwa loss[14].

Thus, in order to check whether the effects of therapeutic responses to schwa loss can be identified in post-schwa-loss data, we compare them to a set of virtual data that display the effects of schwa loss but nothing else. The data we use are corpus data derived from the Penn-Helsinki Parsed Corpora of Middle English and Early Modern English texts (Kroch & Taylor 2000; Kroch et al. 2004). For actual post schwa loss data, we look at the latest Early Modern English section (E3, 1640-1710), and for the derivation of virtual post-schwa-loss data we use the earliest Middle English sections (MX1 & M1, 1150-1250).

For constructing hypothetical post-schwa-loss data, we analyze tokens as if schwa loss had affected them fully both in open and in checked final syllables. Since we are interested

[14] Because even if *post hoc* does not imply *propter hoc*, *propter hoc* does require *post hoc*.

only in tokens with final consonant clusters, we search, basically[15], for (a) spellings that end in <CC>, such as *folc* (b) spellings that end in CCV such as *binde* 'bind', *climbe* 'climb' (c) spellings that end in CVC(C) such as *lerned 'learnt', luved 'loved', munec* 'monk', *filledd* 'filled', *muneces* 'monks', and (d) spellings that end in CVC(C)V, such as *luuode* 'loved' *andswerede* 'answered', or *dwellidde* 'dwelt'[16]. Next, we discard all returned tokens on which schwa loss either did not apply, or in which it did not produce consonant clusters, such as *twenti* 'twenty', which contained no deletable schwa, or *mete* 'meat, food', in which schwa loss produced no final cluster. This yields a set of items which either ended in clusters irrespective of the occurrence of schwa loss (such as *folc* 'folk'), or which would end in clusters after the application of schwa loss. We then analyze their endings as if schwa loss had actually applied on them, as illustrated in table 1 below.

Table 1. Construction of the hypothetical post-schwa-loss language stage from actual pre-schwa-loss data.

Spelling type	Item as spelt in corpus	Analyzed as ending in
<*x*CC>	*folc*	[lk]
<*x*CCV>	*binde*	[nd]
	climbe	[mb]
<*x*CVC(C)>	*lerned*	[rnd]
	luved	[vd]
	munec	[nk]
	filledd	[ld]
	muneces	[nks]
<*x*CVC(C)V>	*luuode*	[vd]
	andswerede	[rd]
	dwellidde	[ld]

Thus, after having constructed a virtual post-schwa-loss inventory of word forms ending in consonants clusters, we are ready to compare it to the inventory of such word forms attested in actual post-schwa-loss data.

In the next sections, we first recapitulate what type of predictions about the effects of schwa loss can be derived from current phonological and morphological theories, and then we demonstrate, by way of example, what kind of results are produced when we test them in the way just described.

[15] Actually, we cast our net more widely, so that we should find also cases in which schwas were represented by more than a single vowel graph as in potential **burneyde* 'burnt'. We did not find any such cases in our data, however.

[16] Naturally, we have to take into account the well-known problem that graphs such as <i>, <y>, <u>, <v>, <ȝ>, etc. can represent both vowels and consonants in Middle English.

5. Testable predictions that can be derived from phonological and morphological theories

5.1 Coda clusters should disappear

Among the theories that specify hypotheses relevant to our research question are Natural Phonology and Morphonology in the tradition of Stampe (1979) and Dressler (1985). Putting it very simply, they focus on the economics of language production, perception and processing, "present […] language […] as a natural reflection of the needs, capacities, and world of its users" (Donegan & Stampe 1979: 126f.). From such an essentially economic perspective consonant clusters are suboptimal because the articulatory effort required by their production results, acoustically, in a relatively bad contrast so that their constituents are comparably difficult to perceive. In that sense, clusters are indeed uneconomical. To provide a simple example (because there is no space for elaborating the phonetic argumentation in sufficient detail), in non-rhotic varieties of English the words *tacker* /ˈtækə/, *attack* /əˈtæk/ and *actor* /ˈæktə/ involve the same phonemes, but their arrangement in *tacker* provides for the best perception of each of them, while the one in *actor* makes the /t/ and the /k/ difficult to distinguish and recognize. In codas (as in *act* /ækt/, for example), the perception of consonant clusters is held to be even worse than in onsets or between syllables. Apart from their being generally dispreferred from an economic point of view, the viability of consonants and consonant clusters in syllable codas also depends on the quality of the specific consonants that occur in them. Particularly, their consonantal strength (which is the opposite of their sonority) seems to matter. Thus, the more sonorous a consonant is, the more easily can it be pronounced after a vowel while still producing a passable contrast, while the contrast it forms with a preceding vowel will be the greater, the stronger (or less sonorous) it is. In complex codas the best arrangement seems to involve a highly sonorous first member and a minimally sonorous final member, so that the contrast between the preceding vowel and each of the two consonants is maximal while the articulatory transition between them is still smooth. Thus, /nt/ is better than /nd/, and /nd/ is better than /rl/. This is expressed, for example, in a universal preference law formulated by Theo Vennemann (1988: 21), which states that

> a syllable coda is the more preferred: (a) the smaller the number of speech sounds in the coda [i.e. ideally zero, AB/CP/NR], (b) the less the Consonantal Strength of its offset, and (c) the more sharply the Consonantal Strength drops from the offset toward the Consonantal Strength of the preceding syllable nucleus.

Something equivalent is expressed within Optimality Theory by the assumedly universal *CODA and *COMPLEXCODA (Prince & Smolensky 2002; McCarthy 2004), stating that there should be no codas, and there should be no complex codas respectively.

In sum, most current theories of phonology imply that final consonant clusters, being uneconomic, should also be historically unstable, so that the ones that were produced in the wake of schwa loss should tend to disappear, or to decrease in frequency. Furthermore, the implication is that they should be the more likely to do so, the more strongly they violate preferences like the ones expressed in Vennemann's Coda Law, or similar proposals made in other theoretical frameworks. Note at this point that none of these predictions imply anything about the specific ways in which they should come to be borne out. Crucially, these may not only include regular or sporadic sound changes, but also other developments such as the loss of words or morphemes containing the problematic clusters. In whatever ways a decrease in the frequency of final coda clusters is brought about our method will detect it, because detecting frequency differences is what it is designed for.

5.2 Morphonotactic clusters should be more stable and tend to differ from lexical ones

Even though consonant clusters are generally dispreferred on phonetic grounds they obviously do exist and some of them have remained stably established in English until the present day. Thus, unless the phonological theories discussed above have gotten something terribly wrong, and if "clusters are [indeed] generally dispreferred, [it stands to reason to assume that] they […] survive, [because …] they are sustained by some force counteracting the overwhelming tendency to reduce towards CVs" (Dressler et al. 2010: 52). A hypothesis about such a force has been formulated by Dressler & Dziubalska-Kołaczyk (2006), who argue that consonant clusters can remain stable in spite of their phonological dispreferredness when they serve a morphological signaling function, as /md/ does in the past tense *seem-ed* or as /ks/ does in the plural *kick-s.* The very fact that they are phonologically dispreferred makes them noticeable, and implies that they could be significant, which in word forms like *seemed* and *kicks* they indeed are, marking the word forms as 'special' in being morphologically complex. However, they can obviously fulfil such a function best when they do not also frequently occur in simple lexical items. This applies to final /md/, which never occurs in simple lexical items, but not to final /ks/, which can, although only rarely, also be found in words like *box* or *suffix* and has thus a lower signaling function than /md/ (1).

(1)

	morphonotactic	lexical
a.	*seem-ed* PT	
	dream-ed PT	
	drumm-ed PT	
b.	*kick-s* pl	*suffix*
	cock-s PL	*box*
	lack-s 3SG	*wax*

Dressler and Dziubalska-Kołaczyk hypothesize that if homophones between complex and lexical clusters (as in [1b]) are frequent, then the complex clusters will lose their signaling function and will be as likely to disappear as lexical ones with no signaling function at all (2006: 72). This may also be because such homophones not only deprive morphotactically produced clusters from their signaling function, but impede the successful processing of lexical items as well. Thus, it has been shown that

> any incoming string that shows the critical diagnostic properties of an inflected form – a final coronal consonant (/t/, /d/, /s/, /z/) that agrees in voicing with the preceding segment as in *filled* […] or *milled* – will automatically trigger an attempt at segmentation, […] any stimulus that can be interpreted as ending in a regular inflection […] is responded to more slowly than an unambiguously monomorphemic stimulus (Post et al. 2008: 1).

For the clusters produced through schwa loss, this implies the following predictions. First, morphonotactic clusters without lexical homophones should be comparably stable. Secondly, when clusters are involved in homophones between morphonotactic and lexical types, then either the effects of such homophony should be reduced by making clusters at least predominantly (if not exclusively) morphonotactic or predominantly (if not exclusively) lexical, or the clusters should disappear altogether. In the former case, and unless the initial distribution is skewed in the other direction, the tendency should be towards a relative increase of morphonotactic instances and a relative decrease of lexical instances, rather than vice versa.

5.3 Frequently produced morphonotactic clusters may spawn lexical counterparts

Although the predictions derivable from Dressler and Dziubalska-Kołaczyk's reasoning appear plausible, a case can also be made for a different scenario. Consider that morphonotactic clusters may not exclusively depend on their signaling function to remain stable. If the rules by which they are produced are highly productive they may come to be produced again and again, in spite of their phonological dispreferredness. As pointed out above, this is all the more likely if their constituents also appear in phonologically more preferred environments, such as the past tense /d/, which occurs also after vowels as in *played*. Being frequently produced by a regular morphological process, clusters may not depend exclusively on their signaling function for being historically stable – in spite of the phonological problems they cause. Instead, the frequency with which they keep being (re-)produced by rule will expose children to a large number of exemplars at a very early age. This, then, may not only stabilize the morphotactically produced items themselves, but facilitate the acquisition of purely lexical counterparts, since children will acquire also morphotactically produced clusters as unanalysed chunks first, and thereby expect to find and manage to deal with them in simple lexical items as well. Hogg & McCully (1987: 89) hypothesize that this may explain the stability of VVCd rhymes as in English *child* or *find* (see also Ritt & Kazmierski 2015).

As has been shown in this section, current phonological and morphological theories imply a large number of quite specific predictions about the fate of the coda clusters that Middle English schwa loss produced, all of which are in principle testable by the method we have outlined above. Although promising to be worth it, carrying out all of the tests in practice obviously requires time and effort, because a large set of corpus data needs to be prepared before it can be subjected to statistical analysis. Therefore, this paper only provides a preliminary view of the type of results and the kind of insights to be gained by comparing virtual to actual post-schwa-loss data for identifying therapeutic responses to the change.

6. The fate of final /nd/ clusters: a demonstration by way of example

The cluster /nd/ is one of the most frequent coda clusters in English, and has been stably attested since Old English times in words such as *hand*, *wind* (N), or *and*. Thus, it seems at first sight to defy the prediction that coda clusters should not exist. Its frequency was highly increased through schwa loss, which removed many of the inflectional endings that had rendered stem final /nd/s medial in actual word forms (such as *honde* DAT.SG or *bindan* INF), and which created new final /nd/s in the past tense and past participle forms of verbs whose stems ended in /n/. Some more examples of word forms that came to end in /nd/ through schwa loss are:

(2) *blind+e* > *blind*; *erande* > *erand*; *hælend+e* > *hælend*; *turn+ed+e* > *turnd*

Thus, schwa loss created a situation in which word forms ending in /nd/ could either be past tense or participle forms ending in morphotactically produced /n+d/, lexemes ending in lexical /nd/, or present participles ending in the morpheme {/ənd/}, which occurred in diverse spellings such as <ande>, <onde>, <ende>, <end>, <ind>. We subsume both of the latter types in the category of word forms ending in lexical clusters.

By Dressler and Dziubalska-Kołaczyk's reasoning, homophony between lexical and morphonotactic clusters is dispreferred because it prevents the word forms ending in such clusters from indicating their morphological make-up in their phonotactic shape. The prediction therefore is that the relative number either of word forms ending in lexical clusters, or of word forms ending in morphonotactic clusters (i.e. past tense and past participle forms) should be smaller in actual post-schwa-loss data than in virtual post-schwa-loss data. This would indicate the predicted therapeutic response by which the morphological ambiguity of such word forms, and hence the illformedness of the language in the sense made clear in sections 3.4 and 3.5, would be reduced.

In the first step we took to check this prediction, the respective frequencies of word forms ending in morphonotactic and lexical clusters were measured for hypothetical and actual post-schwa-loss data (Figure 3a). For each stage, an equal distribution of 50% past forms and 50% present forms is considered as the maximally ambiguous and dispreferred configuration. Such a configuration does not help speakers at all to decide whether an encountered cluster signals a past tense form or not. Hence, 'ambiguity' is operationalized as the Pearson correlation distance $1 - r_{\phi}$ (Deza & Deza 2009: 307) where $0 \leq r_{\phi} \leq 1$ denotes the phi-coefficient of the difference between the measured distribution and the (maximally ambiguous) equal distribution (Cohen 1983; Rodgers & Nicewander 1988; Cumming, Geoff 2014). Since the phi-coefficient is a measure of dissimilarity, it is small, i.e. close to 1, whenever the measured distribution is close to an equal distribution of 50% past forms and 50% present forms. Consequently, if the measured distribution is close to an equal distribution, ambiguity – as defined above – will be large, i.e. close to 1.

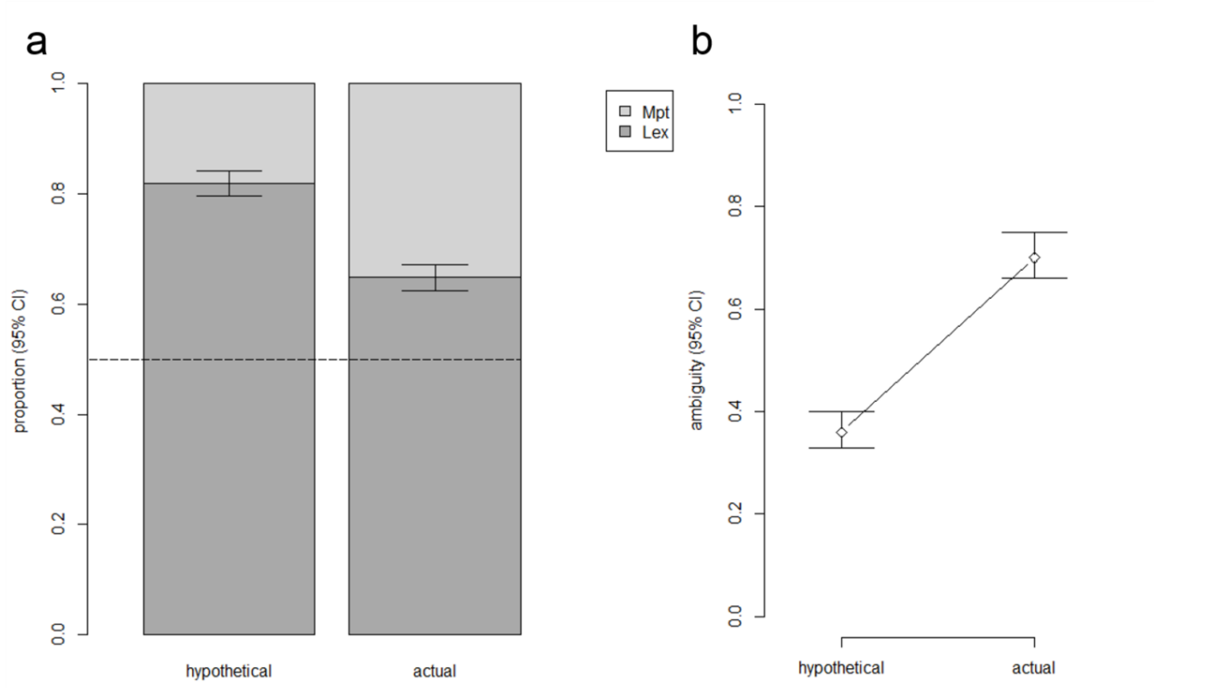

Figure 3. (a) Hypothetical and actual distribution of word final morphonotactic (light gray) and lexical (dark gray) *nd* clusters, respectively ($N_{hyp} = 1162$, $N_{act} = 1676$). The dashed horizontal line denotes a maximally ambiguous distribution. (b) Actual EModE displays a significantly higher ambiguity than hypothetical EModE does.

As figure (3b) reveals, the ambiguity of word forms ending in /nd/ seems to have increased in the wake of schwa loss rather than having been reduced as predicted.

Although this seems, at first sight, to falsify Dressler and Dziubalska-Kołaczyk's reasoning we decided to perform further tests to see whether this conclusion was really warranted. In particular, we reasoned that the morphological structure of a word form might not necessarily be signalled by its coda alone, but by the coda in combination with some of its other phonotactic properties. Thus, in a next step we repeated the procedure exclusively for disyllabic word forms whose morphonotactic instances were past tense or past participle forms of verbs such as *christen*, *beckon*, or *betaken* 'signify', and whose lexical instances were present participles such as *cumind* 'coming'.

The results are displayed in figure (4) below and show that in actual post-schwa-loss data, disyllabic forms ending in /nd/ were past tense forms or past participles in an overwhelming majority of cases, while in virtual post-schwa loss data they are mostly lexical items, or rather present participles. Thus, among disyllables the signaling function of final /nd/ clusters seems to have been both reversed and enhanced: actual post-schwa-loss items are less ambiguously identifiable as past tense or participle forms than virtual post-schwa loss forms would be identifiable as present participles.

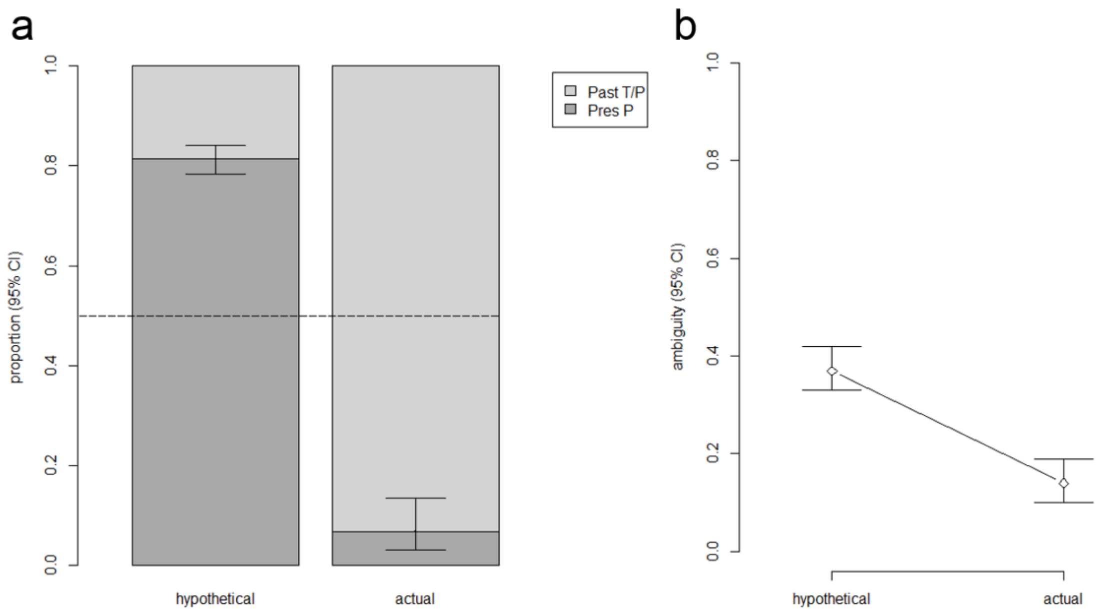

Figure 4. (a) Hypothetical and actual distribution of disyllables ending in *-nd* in past tense or past participle forms (dark gray), and present participles (light gray), respectively. (b) Ambiguity scores for both language stages.

The fact that the results returned in the analysis of disyllabic items differed so strongly from the results gained in a crude and undifferentiated look at the whole, motivated us to repeat the exercise for monosyllables. In their case, however, we distinguished additionally between (a) monosyllables in which the /nd/ clusters were preceded by short vowels, as in *hand* or *tanned*, and (b) monosyllables in which they were preceded by either long vowels, as in *find* or *signed*, or vowel-liquid sequences as in *burned*. The rationale for making this distinction was that we suspected that /nd/ clusters might be more strongly indicative of morphological complexity, the longer the word forms in which they occurred were. After all, the probability of a word form to be complex must generally be higher the longer the word form is. The results of the comparisons of the two word form types are given in figures (5) for word forms ending in V/nd/ and in (6) for clusters ending in V(V|R)/nd/.

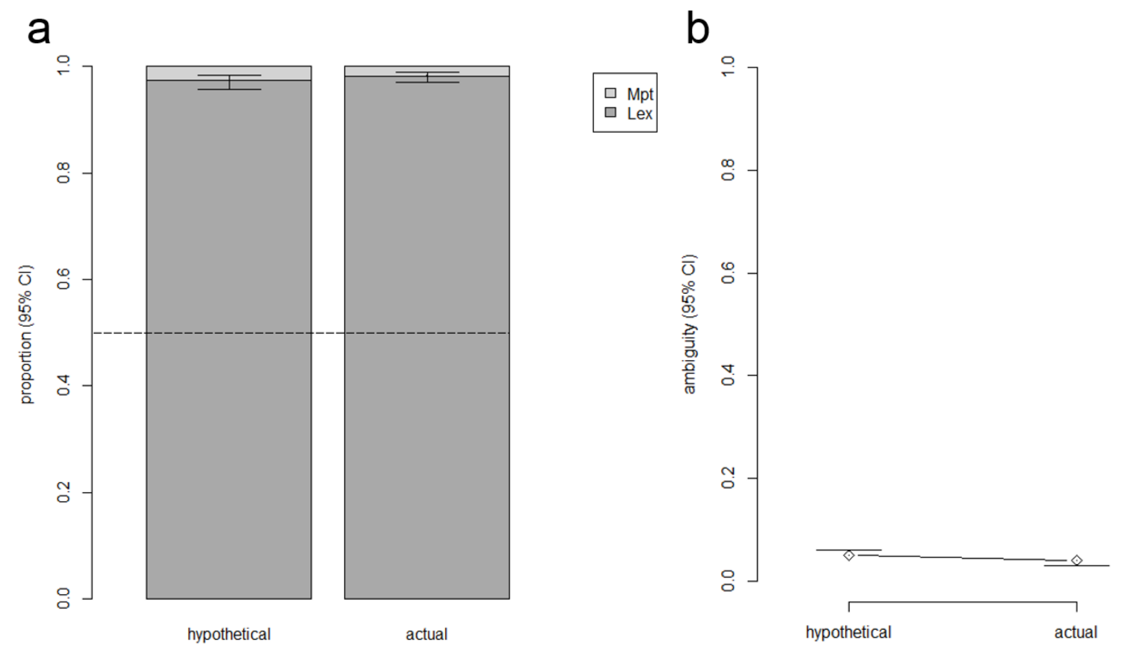

Figure 5. (a) Hypothetical and actual distribution of word final morphonotactic (light gray) and lexical (dark gray) V*nd* clusters, respectively ($N_{hyp} = 642$, $N_{act} = 824$). (b) No significant differences between hypothetical and actual EModE can be found with respect to semiotic ambiguity of V*nd*.

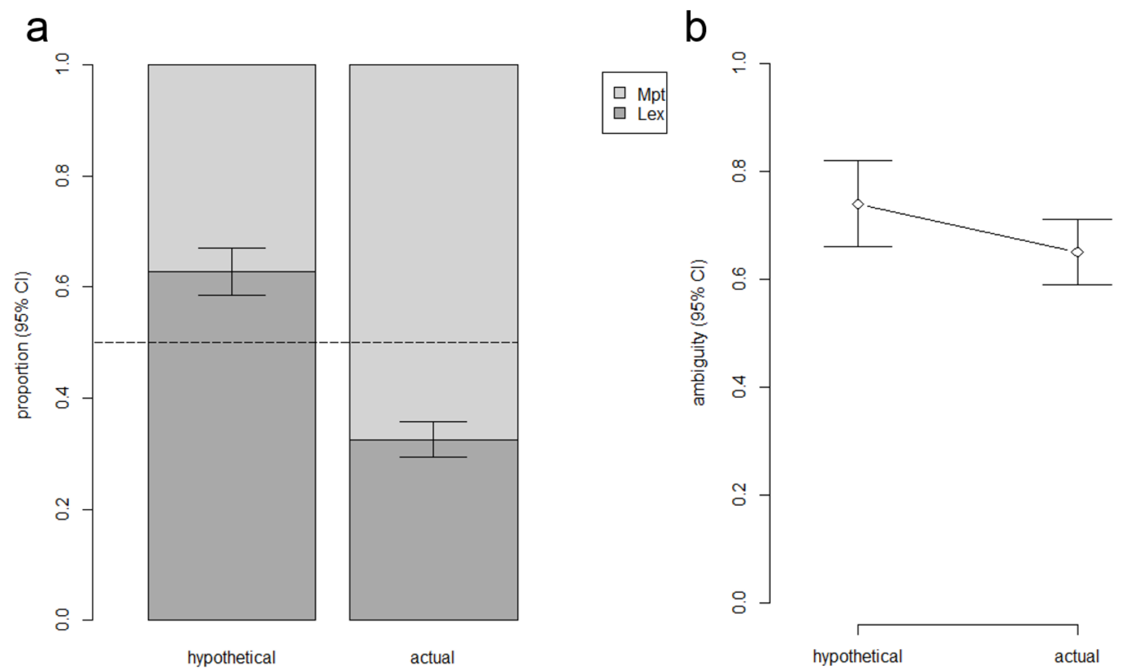

Figure 6. (a) Hypothetical and actual distribution of word final morphonotactic (light gray) and lexical (dark gray) VV*nd* and VR*nd* clusters, respectively (where R stands for a liquid, $N_{hyp} = 520$, $N_{act} = 852$). (b) Again, no significant differences between hypothetical and actual EModE can be found with respect to semiotic ambiguity of V(V/R)*nd*.

For words ending in V/nd/ our results show two things. First, they were overwhelmingly simple, both in virtual and actual post-schwa-loss data. Although, this seems to be in line with the assumption that morphotactically ambiguous word forms are dispreferred, it does not corroborate the hypothesis that their number will be reduced through therapeutic changes. Although actual post-schwa-loss data seem to display a slightly smaller degree of ambiguity than the virtual ones, the measured difference is not significant and may be due to chance.

Very much the same seems to be true for word forms ending in V(V|R)/nd/, as seen in figure (6). At first sight actual post-schwa-loss data appear to differ markedly from virtual ones. V(V|R)/nd/ forms are indicative of morphological complexity rather than of simplicity, but this reversal is not accompanied by a significant decrease in ambiguity. Thus, the figures cannot be interpreted as corroborating the hypothesis that therapeutic changes decreased the morphotactic ambiguity of word forms ending in /nd/.

In short, the first results of our experiment seem to be rather discouraging (at least if one is discouraged when plausible hypotheses fail to be corroborated), but we were nevertheless struck by the impression that the proportion of morphonotactic /nd/ codas seems to have risen strongly both among disyllabic forms and among ‘long’ monosyllabic forms ending in V(V|R)/nd/, while it seems to have stayed very low among ‘short’ monosyllabic forms ending in V/nd/. We therefore decided to compare virtual and actual post-schwa-loss data not only in terms of how ambiguous they would appear to completely unbiased observers (or language users), but how well they would confirm the simple rule that V/nd/ forms are lexically simple, while all longer forms in /nd/ are complex (or, potentially, the converse). The assumption behind this is that speakers may be sensitive to such correlations, infer probabilistic rules from them and employ them in the processing of new data. Thus, the more adequately the distribution of /nd/ forms in a data set reflected such a simple rule concerning its significance, the more preferred we considered it to be, and *vice versa*. This measure of (dis-)preferredness was operationalized in a similar way as ambiguity was defined above, with the only exception that the correlation coefficient r_{ϕ} was now determined for the contingency table comparing morphological complexity and phonological content. For that purpose, phonological content and morphological complexity were defined as two dichotomous variables, where the former allows for the values ‘short’ (phonologically minimal forms ending in /nd/, i.e. monosyllabic C(C)V/nd/ forms) and ‘long’ (all other monosyllabic or disyllabic forms ending in /nd/) while the latter has the values ‘simple’ (e.g. *wind*) and ‘complex’ (e.g. *sinn+ed*).

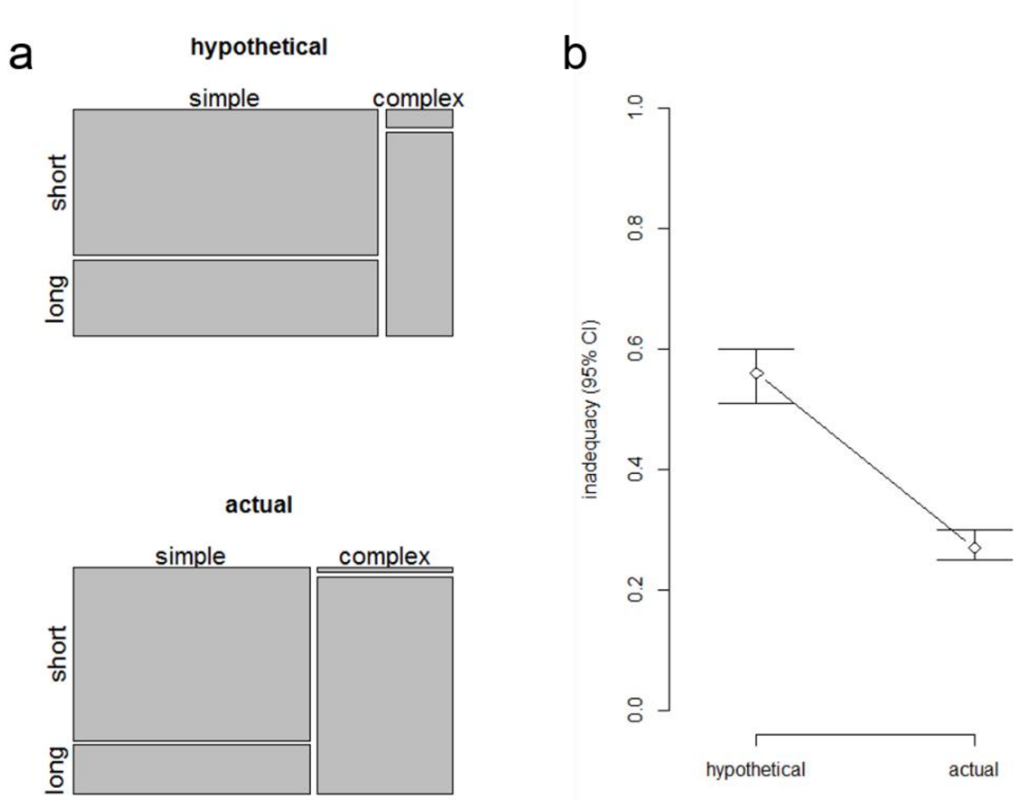

Figure 7 (a) Mosaic plots of the distributions of hypothetical and actual forms ending in *nd* with respect to phonological content and morphological complexity ($N_{hyp} = 1162$, $N_{act} = 1626$). (b) The rule that phonological content (positively) correlates with morphological complexity is less inadequate in actual post-schwa-loss English than in hypothetical post-schwa-loss English.

The comparison of virtual and actual language data reveals a significant difference between the two stages: whereas in virtual post-schwa-loss English the distributional pattern of phonologically long and short forms ending in /nd/ could not have helped speakers much to distinguish morphologically simple from complex forms, this is markedly different in actual Early Modern English: there, phonologically long forms ending in /nd/ indicate morphological complexity reliably (Figure 7). Thus, on the assumption that positive correlations between phonological content and morphological complexity are preferred, it can be argued that among outputs of schwa loss that did not display such a correlation, it was indeed (re-)established through therapeutic responses.

7. Conclusion

What we hope to have shown in this paper is that it is possible to test hypotheses about linguistic change even though the data available to a diachronic linguist are by definition historical and potentially insufficient for conducting quantitative analyses by standard methods. We have suggested a method by which old data can be looked at in a way that brings into sharp profile some of their properties that are not evident from the retrospective point of view normally taken on them. As we hope to have shown, the results to be gained by looking at historical data in the way we have suggested are as unpredictable as genuinely new data. Therefore, we think that our method provides a legitimate way of attempting to falsify hypotheses about language history. In that respect we hope to have contributed to the goal of putting attempts to account for language evolution and change in terms of general (albeit necessarily probabilistic) laws on ever more solid grounds.

Of course, our paper has demonstrated the method in terms of a very specific example, and has shown how its application to Middle English schwa loss allows one to test hypotheses about phonological preferences and the interaction between phonology and morphology in the domain of morphonotactics. We are optimistic, however, that our approach and, particularly, the method of comparing actual history to potential, but unrealized alternatives will prove to be productive when applied in further case studies about the historical stability of consonant clusters in the history of English. Moreover, we see no reason why it should not be applicable to other types of linguistic change as well. Instead, we suppose that the method will be of use in any attempt to account for the way in which linguistic changes hang together, in particular if it is based on the view that the properties, or constituents, of attested natural languages reflect not only constraints grounded in the physiological and cognitive systems in which they are implemented and through which they are expressed and passed on[17], but also constraints emerging from the ways in which they support or impede one another's transmission.

[17] I.e. their speakers.

References

Adamska-Sałaciak, Arleta. 1989. On explaining language change teleologically. *Studia Anglica Posnaniensa* 22. 53–74.

Aitchison, Jean. 2001. *Language change: progress or decay?* Cambridge: Cambridge University Press.

Albright, Adam. 2008. Explaining universal tendencies and language particulars in analogical change. In Jeff Good (ed.), *Language universals and language change,* 144–181. Oxford: Oxford University Press.

Brighton, Henry, Kenny Smith & Simon Kirby. 2005. Language as an evolutionary system. *Physics of Life Reviews* 2(3). 177–226.

Cairns, Charles & Mark Feinstein. 1982. Markedness and the Theory of Syllable Structure. *Linguistic Inquiry* 13. 193–226.

Chater, N. & Morten H. Christiansen. 2010. Language acquisition meets language evolution. *Cognitive Science* 34. 1131–1157.

Clements, George N. & Samuel J. Keyser. 1983. *CV phonology*. Cambridge, MA: Mit Press.

Cohen, Jacob. 1983. The cost of dichotomization. *Applied Psychological Measurement* 7(3). 249–253.

Cohen, Jacob. 1994. The earth is round (p<.05). *American Psychologist* 49(12). 997–1003.

Cowan, G. A. David Pines & David Meltzer (eds.) (1994). *Complexity*: *Metaphors, models, and reality* (Santa Fe Institute studies in the sciences of complexity. Proceedings volume 19). Reading, Mass: Addison-Wesley.

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

Cumming, Geoff. 2014. The New Statistics: Why and How. *Psychological Science* 25(1). 7–29.

Deacon, Terrence W. 1997. *The symbolic species*: *The co-evolution of language and the human brain*. New York, London: Norton.

Deza, Michel M. & Elena Deza. 2009. *Encyclopedia of Distance*. Heidelberg: Springer.

Dobson, Eric. J. 1957. *English pronunciation 1500-1700*. Oxford, London: Clarendon Press; Oxford University Press.

Donegan, Patricia & David Stampe. 2009. Hypotheses of Natural Phonology. *Poznań Studies in Contemporary Linguistics* 45(1). 1–31.

Donegan, Patricia J. & David Stampe. 1979. The study of Natural Phonology. In D.A Dinnsen (ed.), *Current Approaches to Phonological Theory*, 126–173. Bloomington: Indiana University Press.

Dressler, Wolfgang U. 1985. *Morphonology*: *the dynamics of derivation* (Linguistica extranea, studia 12). Ann Arbor, Mich.: Karoma Publishers.

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. Katharina Korecky-Kröll & Gary Libben. 2014. Conflicting vs. convergent vs. interdependent motivations in morphology. In Brian MacWhinney, Malchukov & Andrej and Edith Moravcsik (eds.), *Competing motivations in grammar and usage*. Oxford: Oxford University Press.

Dunbar, Robin I. M. 1996. *Grooming, gossip and the evolution of language*. London: Faber and Faber.

Dziubalska-Kołaczyk, Katarzyna. 2009. NP Extension: B&B Phonotactics. *Poznań Studies in Contemporary Linguistics* 45(1). 55–71.

Flack, Kathryn. 2009. Constraints on onsets and codas of words and phrases. *Phonology* 26(2). 269–302.

Gell-Mann, Murray. 1994. Complex adaptive systems. In G. A. Cowan, David Pines & David Meltzer (eds.), *Complexity*: *Metaphors, models, and reality* (Santa Fe Institute studies in the sciences of complexity. Proceedings volume 19), 17–45. Reading, Mass: Addison-Wesley.

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

Hurford, James R. 2007. *The origins of meaning*: *Language in the light of evolution I* (Studies in the evolution of language 8). Oxford, New York: Oxford University Press.

Hurford, James R. 2011. *The origins of grammar*: *Language in the light of evolution II*. Oxford: Oxford University Press.

Jones, Mari C. & Edith Esch. 2002. *Language change*: *The interplay of internal, external, and extra-linguistic factors* (Contributions to the sociology of language 86). Berlin, New York: Mouton de Gruyter.

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

Kirby, Simon. 2012. Language is an Adaptive System: The role of cultural evolution in the origins of structure. In Maggie Tallerman & Kathleen R. Gibson (eds.), *The Oxford handbook of language evolution* (Oxford handbooks in linguistics), 589–605. Oxford, New York: Oxford University Press.

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/.

Lass, Roger. 1980. *On explaining language change* (Cambridge studies in linguistics 27). Cambridge: Cambridge University Press.

Lesaffre, Emmanuel & Andrew B. Lawson. 2012. *Bayesian Biostatistics*. Chichester: Wiley.

Maddieson, Ian. 2011. Syllable Structure. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online* : Max Planck Digital Library.

McCarthy, John J. (ed.) (2004). *Optimality Theory in Phonology*. Oxford: Blackwell.

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.

Minkova, Donka. 2009. *Phonological weakness in English*: *From Old to present-day English* (Palgrave studies in language history and language change). Basingstoke, England, New York: Palgrave Macmillan.

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.

Prince, Alan & Paul 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 (12 June, 2013).

Restle, D. & T. Vennemann. 2001. Silbenstruktur. In M. Haspelmath, E. König, W. Oesterreicher & W. Raible (eds.), *Sprachtypologie und sprachliche Universalien: Ein internationales Handbuch.* 1310–1336. Berlin: Walter de Gruyter.

Ritt, Nikolaus. 2004. *Selfish sounds and linguistic evolution*: *A Darwinian approach to language change*. Cambridge: Cambridge University Press.

Ritt, Nikolaus. 2010. MiddleEnglish (mor-)phonotactics: system, usage, change. *SELIM: Journal of the Spanish Society of Medieval English Language and Literature* 17. 42–68.

Ritt, Nikolaus & Kamil Kazmierski. 2015. How rarities like gold came to exist: on co-evolutionary interactions between morphology and lexical phonotactics. *English Language and Linguistics*(19). 1–29.

Rodgers, Joseph & Alan Nicewander. 1988. Thirteen ways to look at the correlation coefficient. *The American Statistician* 42(1). 59–66.

Saussure, Ferdinand d. 1959. *Course in general linguistics*. New York: Philosophical Library.

Stampe, David. 1979. *A dissertation on natural phonology* (Outstanding dissertations in linguistics 22). New York: Garland Pub.

Toynbee, Arnold. 1954 (reprint 1961). *A study of history*. London: Oxford University Press.

Vennemann, Theo. 1988. *Preference laws for syllable structure and the explanation of sound change*. Berlin: De Gruyter.