Language Dynamics and Change

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

--Manuscript Draft--

| | |
|---|---|
| **Manuscript Number:** | LDC-1147R1 |
| **Full Title:** | Linguistic diversification as a long-term effect of asymmetric priming: an adaptive-dynamics approach |
| **Short Title:** | Diversification as an effect of asymmetric priming |
| **Article Type:** | Full Length Article |
| **Corresponding Author:** | Andreas Baumann<br>Universitat Wien<br>Vienna, AUSTRIA |
| **Corresponding Author's Institution:** | Universitat Wien |
| **First Author:** | Andreas Baumann |
| **Order of Authors:** | 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 |
| **Funding Information:** | Austrian Science Fund (P27592) — Mr. Andreas Baumann<br>Österreichische Forschungsgemeinschaft (06/15355) — Mr. Andreas Baumann |

Dear editors and reviewers,

thanks a lot for the great number of helpful comments on our manuscript. We took all of them into consideration and improved our manuscript accordingly. Most of our work went into improving the structure of our paper and making the results more accessible. We would like to highlight the main steps that we have taken to do so:

1. Most of the mathematical derivations were moved to the appendix and their results are verbally given in the main body of the text.
2. We added two additional boxes which summarize the key assumptions and results for both applications (lexical/sublexical).
3. We provide better (psychological/cognitive) interpretations for the model parameters (c, r, mu, alpha etc.) whenever they are introduced and give short definitions when they are mentioned later again.
4. We linked our model to the literature on weak cognitive biases and universals in language evolution.

As a result of restructuring our manuscript, we could manage to push the word count of the main body of the text (excl. abstract, appendix, captions, reference list) below the 10,000 word threshold (9,914 at the moment).
Please, find more detailed information on how we addressed every single comment that we received below (formatted in **>>bold** fonts).

---

Reviewer #1: The authors propose an original theoretical model of language change based on a mathematical model of population dynamics. Their aim is to support the claim that asymmetric priming can explain aspects of language change, such as linguistic diversification and reductionist tendencies. They designed a Lotka-Volterra model in which asymmetric priming features as a competition coefficient.  This coefficient is in fact a function of formal (phonological) substance, which they designed in a way consistent with the asymmetric priming hypothesis (more substance primes less substance). Using this competition coefficient as well as several other aspects (such as token frequency), they model for example how strong priming can result in formal reduction, while weak priming can result in linguistic diversification and stable co-existence.

The ideas presented in this paper are very interesting and relevant to the current theoretical questions regarding asymmetric priming and language change. Their simulations are consistent with empirical observations, which means that the paper bridges a gap between theory and empirical investigation (which is actually quite prevalent with asymmetric priming). I would like to see this article published, however I think that a major problem is that its current version is quite difficult to fully understand as it involves many advanced mathematical notions that most linguists (including myself) are not familiar with, such as the Lotka-Volterra differential equations. Some of the discussion is focused on the mathematical analysis of these equations, which I think is fairly challenging for an audience of linguists. This is why I would like to offer several suggestions below to make the paper more accessible.

Major remarks:

The paper would benefit from being friendlier to an audience that will be less than fully familiar with the mathematical aspects of the model. I appreciate that you made an appendix with some of these aspects and I would like to encourage you to continue in this direction. Sections 1 to 2 are quite clear and require no major changes (see below for minor remarks).

Section 3.1 is generally quite accessible and shows a high degree of formalism as every element of the model is clearly introduced. But it gets difficult towards the end, specifically at the paragraph on page 7 which begins with "Let us continue with analyzing the system." Is it really necessary to discuss the different equilibria? I think that what is relevant here for a linguistic audience is to know that equilibrium is a possibility (and under which conditions), without necessarily getting the technical details about it. For example, I found Figure 2a/b to be very explicit about stable co-existence. I think that having equations (1), (2), (3a) and (3b) mentioned in the paper is OK because they are the main elements of the model, but the equilibrium bit should be simplified. A middle ground would be to move all the mathematical aspects of the equilibria to the appendix.

**>> We restructured the description of the model in 3.1 by providing more explanations (environmental constraints, p. 6; competition term, p. 7) and moving the more mathematical parts to Appendix A1, as suggested. Particularly the more technical discussion of the ecological equilibria is now in A1.**

Section 3.2 introduces the asymmetric competition term which is one of the key points of the model. First, I am actually not sure to which kind of similarity μ refers to? Are we talking about a Levenshtein distance, semantic similarity, or something else?

**>> We refer to formal similarity. We added a clarifying note as well as a psychological motivation for using that parameter (together with some references). The way in which similarity is measured is now described in footnote 6.**

Second, is it important to discuss that  defines the inflection points of the function? You mention that it will be important later on, but at this stage this is more confusing than anything else. A more pedagogical approach might be required to inform the reader why this is important.

**>> In fact, it is only important for the technical derivation of the condition for branching. We moved this discussion to A1. Thanks for pointing this out!**

Section 3.3 is very well written and easy to follow. You discuss the technical assumptions of adaptive dynamics and how they are met in the case of linguistic evolution, along with some relevant references. I found it very informative.

Section 3.4 discusses stable diversification and the evolutionary branching point. In a similar way to section 3.1 and equilibria, I'm wondering whether the discussion could be more basic without involving the actual calculations, which might be moved to the appendix? The example with Figure 2 is very accessible and should absolutely stay.

**>> Section 3.4 was simplified so that it now only shows inequality (4). We decided to leave the set of inequalities in 3.4 and not to remove it to the appendix because it represents a key result of the paper (the paragraph below describes that the inequality entails that branching only occurs if priming is not too strong, a result we also mention in the conclusion section). However, all details about its derivation are in the appendix.**

Also note that there is a numbering problem as (4) is used both on this page (10) and on page 8.

**>> Fixed.**

Section 4.1 is also rather difficult to understand. Again, the figures were very helpful. The top two paragraphs of page 13 are somewhat confusing and I would suggest to simplify them or to have them in the appendix. What I think matters is that you can have a branching point and a stable situation. The actual calculations are obviously important, but I don't think that they should be front stage.

**>> Mathematical derivation moved to appendix A4 and replaced by a quick verbal description. Also, the discussion of r(s) was simplified, and we dropped some mathematical details (domain of the function etc.; now only in A4).**

I found the discussion of Figure 4 to be more accessible, because we can actually see what the situation looks like when the values of each parameter changes.

**>> We made this part of the discussion more accessible by repeatedly mentioning what the parameters \alpha, \pi, \tau, \mu refer to.**

The last part of Section 4.1 is very good as it compares the simulated evolution in Figure 5 with actual datasets in Figure 6. This is where the paper shows its value for linguistic research, as it can bridge a gap between theory and empirical investigation in this area. I should however mention that there might be a mistake in the legend of Figure 6; if circles are for the lexical variants and crosses are for the morphonotactic variants, then the dashed line corresponds to the lex variants and the solid line corresponds to the mpt variants. The way you phrased it, using "respectively", implies the reverse, which got me very confused because then the empirical dataset actually shows the opposite of the simulated dataset. I hope that this is just a minor mistake and not that I completely misunderstand something.

**>> Yes, this was a typo; thanks for reading the manuscript so carefully!**

The comments regarding Section 4.2 are similar to those of Section 4.1. The beginning of the section is difficult (in particular page 16, lines 10-32), but the reminder is accessible.

**>> The first couple of paragraphs were restructured, as in 4.1. Mathematical formalisms were simplified, formal derivation of the results moved to Appendix A5 and results only reported verbally.**

**In the discussion of grammaticalization, we left the re-definition of the competition term in section 4.1, though, because with think that the reader really needs to know what happens (essentially inversion of the competition coefficient, motivated by Hilpert & Correia Saavedra 2016).**

Again, comparing the simulation to actual datasets is well-appreciated. Since the going to example might not be from a previous study, it may be useful to add a quick note regarding the way auxiliary vs. main verb going to were retrieved.

**As our word count is terribly limited we have decided not to elaborate extensively on our corpus investigation and query language used. The COHA/COCA interface is an interfaced used so often by Corpus linguists that we assume that the query language is well known ;  the readers simply need to trust us that we searched for the constructions professionally. However, we added the respective queries from the corpus interface in Fig.8; the queries are quite informative on how we distinguished the two constructions from each other in terms of their syntactic distribution.**

Finally, the conclusion might be improved by adding some other potential uses of the model (or uses of similar models) for the study of language change. I think the examples in sections 4.1 and 4.2 were convincing and it would be beneficial to give some extra perspectives, if possible.

**>> We now mention two additional fields of historical research (phonemic split; constructional diversification) in the conclusion section of which we think that our model can be of some explanatory value.**

Minor remarks:

Note that the line numbers were slightly off in the margin of the draft that I received, so my numbers might be a bit approximate.

Page 1, lines 57-58: I find the phrasing "In a nutshell, the 'priming triggers language change' argument goes like this" to be a bit colloquial, which seems odd given the tone in the rest of the article.

**>> Changed the phrase.**

Page 2, line 9: A reference to the "ease of effort" argument might be a useful addition. It also appears on page 4 without at least one explicit reference.

**>> Added sources (Zipf 1949; Martinet 1955; Hawkins 2004).**

Page 2, footnote 1: I have trouble understanding what the footnote is about. You mention that your model does not differentiate between priming and inhibition, but that it accounts for this, which I assume is done through the c(si-sj) function. So far so good. What I do not understand is the comparison with Hilpert and Correia Saavedra (2016). They had an experiment in which the response times showed a negative priming effect, but this was an empirical finding, not really a question whether a model can include positive or negative priming?

**>> We deleted that footnote; the relationship between asymmetric inhibition and asymmetric priming (and how c(...) accounts for it) is discussed on 3.1 in more detail.**

Page 3, lines 12-19: The end of the introduction is a bit underwhelming as it points out the limitations of the model. Surely, these are important, but maybe not right there. I would suggest that instead of saying "we will illustrate the empirical applicability of the model in an exemplary manner", you should mention that you made simulations with the model that were fairly consistent with previous empirical observations.

**>> We removed this paragraph (to save some space) and mention the plausibility of our results (analyses/simulations) in the final paragraph of section 1.**

Page 5, lines 23-29: I think a bit of caution is required here regarding the affirmation 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" based on Hilpert and Correia Saavedra (2016). They used lexical and grammatical counterparts which shared the same form. So their results are not exactly the best illustration for the phonological aspect. It does however work for the semantic aspect of the claim.

**>> We assume that in Hilpert and Correia's experiment, partially a self-paced reading task, the two tested forms (grammatical vs. lexical) are also different in phonological form (see, Appendix 1 in H&C's article). For example, the reader/speaker will definitely shorten the vowel in the first has:** ***Her sister <u>has (aux)</u> told me that the dog <u>has (lex)</u> fleas.*** **That is why we believe that it also illustrates the phonological/phonetic aspect (see literature on grammaticalization, e.g. Heine & Kuteva 2002 or Hopper & Traugott 2003). Nevertheless, effects caused by phonetic differences might be overridden by effects caused by semantic differences (but this would have to be tested in additional experiments dedicated to that question).**

Page 5, lines 52-53: Your model focuses on formal substance, which I understand as phonological substance given the examples. However, in the previous sections, you did mention that asymmetric priming also relates to semantics. I was just wondering to what extent c could also be a function of semantic substance? You discuss this in section 4.2 by including degrees of grammaticality (at the bottom of page 16), but my question still holds.

**>> In 4.2, semantics is covered by the degree of grammaticalization g, but in an admittedly simplistic way: g is still a one-dimensional trait which is assumed to encompass semantic, grammatical, phonological aspects etc. These limitations are briefly addressed in footnote 14. Clearly, it would be more precise to model every single dimension associated with grammaticalization. Our simplification builds on the assumption adopted in the grammaticalization literature that lexical items can be ranked according to their "degree of lexicality/grammaticality".**

Typos/Formatting:

Page 1, line 47: There is a single quotation mark in 'more substance primes less substance.
Page 5, line 31: "we will show is two things in this paper", there is an extra "is" in the sentence.
Page 18, lines 13-14: a full stop is missing at the end of the first sentence of the paragraph.
Page 18, lines 17-18: there might be an extra "less" before "items". If it's not a mistake, then the sentence is a bit confusing.
Page 18, lines 22-23: In the sentence: "it is the more lexical words which are inhibited less by their lexical counterparts than the other way around", should it not be their grammatical counterparts instead?
Page 30, line 46: typo "lexcial"

**>> Fixed it.**

Reviewer #2: This paper introduces a model for the fate of linguistic items with different amounts of 'formal substance', e.g., the length of a word or its degree of semantic specificity. The are two basic mechanisms at play in the model. The first is an intrinsic growth rate that depends on the formal substance. For example, a more specific item might be more useful in satisfying a communicative goal, so could have an intrinsic growth rate that is larger than a more general item (although the authors keep the relationship between the growth rate and substance general, which allows a wide variety of different situations to be modelled). The second mechanism is inspired by the phenomenon of asymmetric priming where, for example, a more specific item tends to activate more general items more than the other way around. Putting these two mechanisms together, one can find instances where an item with more substance outcompetes one with less substance, instances where the converse is true, and instances where multiple items coexist. The outcome depends on the assumed relationship between the growth rate and the formal substance, and on the strength of asymmetric priming. Roughly the first half of the paper is devoted to motivating and defining the model, and the second half to establishing conditions under which the different outcomes are seen, as well as a couple of applications to specific linguistic phenomena.

My reaction to this paper was mixed. On the one hand, I found the basic premise interesting. As far as I am aware, there has not been much modelling of the interaction between substance and asymmetric priming. Moreover, the connections to the Lotka-Volterra system and evolutionary game theory are well exploited. At the same time I found the paper quite difficult to read, and shall attempt to expand on the reasons why below in the hope that it can be improved in a revision. My main criticism in this regard is that I did not come away with any intuition as to the general conditions under which I would expect to see a particular item outcompete the others or for multiple items to coexist. The main reason for this is the relevant results are presented mostly in mathematical terms, and I found it hard to tie them back to the basic mechanisms in the model described above. (This was further exacerbated by the format in which the paper was presented, in which the figures and explanatory box were positioned as far away from the text as humanly possible; something which I appreciate may have been imposed on the authors by the journal, who should take note if this is the case).

**>> We understood the formatting guidelines of the journal in this way (probably wrongly so); we are sorry for any inconvenience.**

That said, I don't doubt the analysis and the insights that this yields into the specific applications discussed; it's just that I feel that overall the presentation could be clearer.

More specifically, my lack of a confident grasp of the authors' findings stem from the following:

1. Asymmetric priming is discussed in Section 2 as a mechanism where one form promotes another (e.g. p4:39 "an improvement in performance" or p4:43 "pre-activation". In the model, what we actually seem to have is an _inhibition_ of one form by the other, as evidenced by the minus sign before the summation in Eq (1). There is likely to be some sort of equivalence between A promoting B and B inhibiting A, but I think the authors should spell out what this is and why this model is appropriate to describe asymmetric priming.

**>> Yes, this is clearly an important issue. We added a paragraph to section 3.1 (page 7) which addresses this point. As you are saying correctly, there is an equivalence between A promoting B and B inhibiting A asymmetrically. This is so, because environmental interactions in our model are always supposed to have a negative effect on growth (a common assumption in ecological modeling, also in linguistic applications; think of game theoretical models, where overall population size is constrained and usually set to 1, for instance). That is, we do not model priming as a case of mutualism. This should now be much clearer.**

Related to this, a side effect of this choice is that an item then inhibits itself (through c(0)). This is essential to obtain the logistic-like behaviour, but the authors don't appear to offer a psychological motivation for why items should be self-limiting in this way.

**>> This is related to the point above. We assume that resources are limited. In fact, this does not only cover psychological/cognitive aspects (e.g. limited memory) but also linguistic factors (not every utterance requires the use of a certain word, phoneme or diphone) and external factors (there are not infinitely many speakers, speakers die, speakers only produce a limited number of utterances). In 3.1 we list these factors on page 6 under (i-vi). We now refer to these limiting factors more often and say why they are relevant both to one-dimentional and multi-dimensional dynamics.**

2. The model as presented contains a fair number of parameters (mu, tau, kappa etc). This is not in itself problematic, and Box 1 is appreciated. However, the authors assume throughout the text that the reader is able to rapidly internalise the meaning of each of these parameters, as the symbols are quoted frequently and without any reminder of what they represent in psychological terms. There are two steps the authors could take here. First, they could look to see if they need to quote the symbols as often as they do. Second, when a symbol hasn't been used for a while, it might be worth reminding the reader what it represents (e.g., "the asymmetric priming strength $\tau$"). Otherwise the reader has to constantly flip between the text and Box 1. (I found Section 4.1 particularly hard to penetrate in this regard).

**>> We tried to make our presentation more easily accessible by moving the more technical parts to the appendix. When parameter are still present in the main body of the text, we proceeded as suggested and added quick descriptions of them again.**

3. Related to (2), when the authors present key mathematical results, e.g., the equation labeled (iv) on p8, verbal interpretation is often more limited than it needs to be. For example, the inequality that immediately follows this that guarantees stability should have some more psychological interpretation than is currently given. Ideally this should make explicit reference to both the intrinsic growth and asymmetric priming mechanisms that are core to the model.

The discussion below eq (4) is a step in the right direction, but obscured more than it needs to be by again using symbols without naming them. I think it would help if the authors were to go systematically through the paper, determine what they consider to be the key results, highlight them as such, and provide a discussion of the origin of these results in terms of the basic psychological mechanisms that are present in the model, so that the reader can start to intuit the behaviours that emerge and the underlying
reasons for them.

**>> Section 3.2 and the description of the equation therein (now (3)) were extended to include a better psychological interpretation (and some details were moved to Appendix A2). The interpretation of inequality (4) in 3.4 is now made more accessible by explicitly mentioning the first key result of our analysis ("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.", p. 11) and by dropping some technical details (considerations of slope and curvature of r(s*) which cannot so easily interpreted). Apart from that, we added two more boxes (Box 2 and Box 3) which summarize the key features (assumptions and predictions) of the respective models in 4.1 and 4.2.**

4. In the discussion of adaptive dynamics, there appears to be some model for mutation that is being used implicitly, but I didn't understand what this was. Is there some rate at which new mutations are generated?

**>> Mutation (and how this works in the linguistic case) is now discussed more thoroughly in 3.3 (final paragraph and footnote 8). You can find a formal introduction of the mutation rate k in appendix A3 below the canonical equation of adaptive dynamics.**

Can an item with any amount of substance be generated, or are the mutations drawn from a limited number of types? How many types might be simultaneously present?

**>> Mutations are assumed to be small so that evolution is a continuous process (i.e. the step-wise invasion-substitution process is approximated by a continuous model, namely the canonical equation of adaptive dynamics; see appendix A3). Also we assume that mutations are rare (cf. 3.3 and A3). As a consequence, under normal conditions ("As long as D(s_2) is not close to zero"; p 21 in A3) we only have two**

**populations involved in a competition process: mutant and resident (to be precise, in some special cases diversification can lead to more than two stably coexisting variants, but this would go beyond the scope of this paper; see Kisdi 1999 for a detailed account of multiple coexisting populations). We expanded the motivation of both assumptions in 3.3.**

Most of the discussion centres around two types, but the application of section 4.1 seems to refer to a spectrum of many types, and I couldn't reconcile this.

**>> The four types in 4.1 (/ld, rn, rT, rd/) are not supposed to be variants of each other. For that reason, they propagate independently. We added a clarifying note on page 16.**

I think the discussion in Section 3.3 ought to make clear exactly how we should view the mutation process as operating.

**>> We extended 3.3. See comments above.**

Also there are one or two minor points to consider:

5. p3:11-19, the authors suggest that the qualitative behaviour of the model (such as directionality of stable coexistence) is sufficient to demonstrate the empirical applicability of the model. However this ignores timescales, specifically that the rate of change predicted by the model might be vastly discrepant from reality, even if the end-state or trajectory shape is reasonable. It is not clear if the timescale of change can be pinned down in this model, so perhaps this would best be left as an open question.

**>> This is definitely an important issue. We removed that claim. In addition, we now briefly discuss the time scale of the evolutionary model (and its linguistic interpretation) in appendix A3.**

6. p7:51, the authors propose that Eq (1) provides a "sufficiently accurate way of modelling language change" but don't give any indication of what they mean by "accurate" (or how they would be able to tell that this is the case). Perhaps just drop this claim.

**>> Claim removed.**

7. The phrase 'so-called' on p10:5 has connotations (at least in some varieties of English) of 'is referred to by some people incorrectly and/or deceitfully' whereas I think something like 'the quantity known as' is meant. In other words, I don't think 'so-called' should be used when the term in question is what it is actually called.

**>> Fixed.**

8. On p17:42 "loosers" should be "losers" I think.

**>> Fixed.**

9. The reference "Boer, Bart de (2000)..." should I think be "de Boer, Bart (2000)"

**>> Fixed.**

Reviewer #3: Thanks for the opportunity to read this interesting manuscript. I thought it was excellent. Overall, I judged this paper to be of high quality, both in form and content: it is well crafted, and makes an important contribution with useful implications for our understanding of a potentially broad range of linguistic phenomena.

The authors lay out a mathematical model of population dynamics and illustrate its application to lexical and sub-lexical phenomena. The model is designed to capture evolutionary dynamics among competing linguistic tokens as a function of cognition. The focus is on the long-term effects of a basic psychological mechanism -- priming -- in cases where this mechanism results reinforcement effects that are asymmetric among linguistic tokens. In particular, the manuscript analyses focus on asymmetric priming among tokens with more or less formal substance, a general concept that the authors apply to phonological and grammatical phenomena. The authors demonstrate a number of general conditions for reduction and diversification.

I felt the prior literature, both psychological/linguistic and formal, is well described. The abstractions of the model are well documented and justified. The mathetmics appears sound to me. The results are nuanced and numerous, and the conclusions are appropriately qualified in light of the simplifications necessary to derive this kind of model. My only general hesitation is that the paper is extremely dense. As a reader, I would happily sacrifice a few of the more minor results in favour of readability, if space is an issue.

My overall view is that this paper is essentially already of publishable quality, subject to a few minor revisions which I will keep to a minimum. Here are my suggestions:

- in the introductory paragraph on page 2, betwen lines 28 and 41, perhaps do a little more to justify assumptions that relate formal substance to base rate frequency. After the sentence starting on line 37 and ending on line 41, give an example or two.

**>> We added Zipfian duration-frequency relationship as an example (references also provided in the paragraph above).**

- In section 3, page 6, paragraph begininng line 52, give more informal intuition concerning the competition coefficient. While it is well defined formally, I found this section extremely desnse and difficult to follow intuitively. Spell out informally why the derivatives are important, and what the conditions imply semantically.

**>> We moved this part to Appendix A2. It now also provides a more accessible description of what the derivative of c means. Apart from that we restructured section**

**3.1 by adding some more remarks on interpretation and motivation, and by removing the technical discussion of ecological equilibria (now in appendix A1)**

- w.r.t. an assumption in the top-level dynamics (eq. 1) of the model: if i have understood correctly, there is no structure to the co-occurence of linguistic items, with respect to either substance or item class (fianl term x_i * x_j in eq. 1). The biological analogy here is unform-random mixing in the population. Given the focus of the manuscript on interactions among linguistic items, and the centrality of this assumption to the results, I feel this requires some justification, or at least acknowledgment. I'm not asking the authors to re-derrive results with an interaction frequency matrix: this would be too much work and the not neccesary for this paper (in my view). But acknowledge the assumption and its importance in words, or spell out why my impression that this is important is wrong.

**>> This is a good point; we added a paragraph discussing this issue in 3.1.**

- in 3.3, page 9, line 20, provide a reference or two after "in ecologically complex setups"

**>> Examples and reference added (Cushing 1998).**

- *weak* assymetric priming is an important concept throught the manuscript. More generally, the effect of assymetric priming is a specific instance (a novel one to be sure...) of the more general idea that cognitive biases (e.g. assymetric priming) can shape cultural transmission (e.g. linguistic token frequency). There is a literature on this more general treatment that I feel should be cited, and results concerning the effects of weak biases in particular. I'd recommend citing: Griffiths & Kalish 2007, Conitive Science, Language Evolution by Iterated Learning With Bayesian Agents; Kirby et al 2007, PNAS, Innateness and culture in the evolution of language; Thompson et al 2016, PNAS, Culture shapes the evolution of cognition.

**>> Thanks a lot for this comment! We extended the discussion section with a paragraph that discusses this notion in more detail (by also referring to some of the suggested literature). We also provide examples for (putative) diachronic universals, and relate historical literature to this issue.**

- in section 4.1, page 14, sentence ending line 29: is this empirically attested too?

**>> We added two references (specific to phonotactics) and now refer to the more commonly used terms lenition/fortition.**

- page 16, unnumbered eq. on line 23: What is D(s)? Formalised but no explicitly described in the main text as far as i could see.

**>> The whole part was moved to the appendix (A5). D(s) is the fitness gradient defined in A3.**

- sec 4.2, page 17, line 18: remind the reader where the parameter values come from.

towards the end the paper is very dense and I found it difficut to keep track of parameter setting. hold the reader's hand a little!

**>> We made this part more accessible by repeating what the parameters mean and by moving the more technical parts to the appendix (A4 in this case). We also added two boxes (Box 2 and Box 3) which summarize the key assumptions and findings in 4.1 and 4.2, respectively.**

- while most of the paper is well written, there are plenty of typos that should be addressed.

# Title page

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

Andreas Baumann
University of Vienna
Department of English
Spitalgasse 2-4/Hof 8.3
A-1090 Vienna
Austria
andreas.baumann@univie.ac.at

Lotte Sommerer
University of Vienna
Department of English
Spitalgasse 2-4/Hof 8.3
A-1090 Vienna
Austria
lotte.sommerer@univie.ac.at

Word count: 10225 (including footnotes, excluding references, appendix and figures)

Acknowledgements: This work was supported by the Austrian Science Fund (FWF, grant No. P27592) and the Österreichische Forschungsgemeinschaft (ÖFG, grant No. 06/15355). We would like to thank Nikolaus Ritt, Eva Zehentner, Chistina Prömer, Bodo Winter, Martin Hilpert, T. Mark Ellison and Jan Nuits for many valuable comments and discussion of the paper.

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

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 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).

---

[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).

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 $j$,

---

[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.

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 $c(s_i - s_j) < c(s_j - s_i)$.

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

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

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 $s_j$: $\Delta = s_i - s_j \mapsto c\left(s_i - s_j\right)$, $\mathbb{R} \rightarrow \mathbb{R}^+$, which fulfils that $s_i < s_j$ implies $c\left(s_i - s_j\right) < c\left(s_j - s_i\right)$. 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 $\Delta$ (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 $\mu$ 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, $\mu$ operationalizes the scope of priming. The parameter $\tau$ the extent to which priming is asymmetric (it determines the steepness of the curve). If $\tau$ is large both items have a relatively similar impact on each other. If $\tau$ 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.

Fig. 1 here

---

[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.

Box 1 here

### 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

---

[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)).

to apply the framework of adaptive dynamics to diachronic change in linguistics (see also Doebeli 2011 and AUTHORS 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 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

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

---

[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).

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, $\tau$ in the asymmetric-priming term should not be much smaller than $\mu$ (ideally $\tau > \mu$). If this is the case then the curve defining 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 $\tau$ is small, the asymmetric-priming curve is much steeper than if $\tau$ 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.

Fig 2 here

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.

## 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

---

[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).

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

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

---

[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.

where $C$, $\alpha$ and $\pi$ are strictly positive. In this function, $\alpha$ measures articulatory effort and $\pi$ 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 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.

Fig 3 here

Box 1 about here

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 $\alpha$, perceptual effort $\pi$, the similarity threshold $\mu$ defining the scope of priming and strength of asymmetric priming $\tau$ (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 $\alpha$, $\pi$, $\mu$ and $\tau$ 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 $\mu$ and $\tau$ (horizontal axes). Across the single bifurcation plots, perceptual effort $\pi$ increases from the left-most column to the right-most column, while articulatory effort $\alpha$ 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.

Fig 4 here

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 $\mu$ as can be seen from the decreasing values on the vertical axis. Since $\mu$ 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

---

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

[12] As can be seen, there are no repellors or Garden-of-Eden points for the admitted combinations of $\alpha$, $\pi$, $\mu$ and $\tau$. See appendix.

that it applies to pairs of items which are substantially different from another (large $\mu$). In contrast, if asymmetric priming has a narrow scope (small $\mu$), then formal reduction is hampered.

Second, $s^*$ increases in $\tau$, which determines the impact of asymmetric priming. If $\tau$ is small, then asymmetric priming has a strong impact. In that case, items tend to get shortened. If $\tau$ 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 $\mu$ and strength of asymmetric priming $\tau$ play in evolutionary branching (dark gray region in Fig. 4). As can be seen in Fig. 4, $\mu$ must be relatively small in order to enable stable diversification. If $\mu$ 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, $\tau$ must be greater than $\mu$, 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, $\tau$ must not be too large, and if $\tau$ is large then $\mu$ 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.

Box 2 about here

We can simulate the evolution of a diphone’s duration $s$ given articulatory effort $\alpha$, perceptual effort $\pi$, similarity threshold $\mu$ and strength of asymmetric priming $\tau$. Figure 5a

---

[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.

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.

Fig 5 here

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.

Fig 6 here

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 $\kappa$ 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.

Fig 7 here

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.

Fig 8 here

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

---

[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).

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 about here

## 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 $\mu$ is positive (which is plausible because the effect of priming ultimately decreases with dissimilarity) and since $c(\Delta) > 0$ for all $\Delta$. The parameter $\tau$ determines the steepness of the curve defined by $c$. If $\tau$ is small, then the effect of asymmetric priming is very strong. Conversely, if $\tau$ 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 $\tau$ 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 $r$, $\mu$ and $\tau$ 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

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

## A4 Sublexical evolutionary dynamics

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

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

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

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

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 $r'(s) = -\kappa C s^{-\kappa-1} < 0$ 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.

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.

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 |
| $\mu$ | similarity threshold for asymmetric priming (scope of priming); beyond a difference of $\mu$, priming effects become weaker |
| $\tau$ | measure of the strength of asymmetric priming; if $\tau$ is small/large priming has strong/weak effects on processing |
| $\alpha$ | language specific articulatory effort; small $\alpha$ corresponds to a speaker friendly linguistic system (see 4.1) |
| $\pi$ | language specific perceptual effort; small $\pi$ corresponds to a listener friendly linguistic system (see 4.1) |
| $\kappa$ | language specific strength of the inverse relationship between substance and productivity of words (see 4.2) |

Box 2. Sublexical dynamics: key results

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

Box 3. Lexical dynamics: key results

| | *Assumptions* | |
|---|---|---|
| | Substance only | Substance and meaning (degree of grammaticality $g$) |
| 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 $\tau$ | Unidirectional reduction irrespective of $\tau$ | Diversification possible under weak asymmetric priming |
| Effect of scope of asymmetric priming $\mu$ | Unidirectional reduction irrespective of $\mu$ | Diversification possible if priming has a relatively small scope |

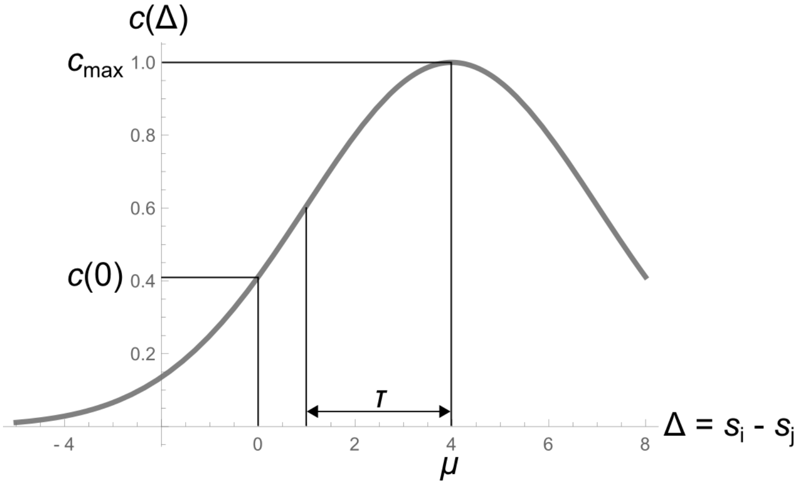


Figure 1. Gaussian function underlying the asymmetric competition term with $c_{\max} = 1, \mu = 4, \tau = 3$.

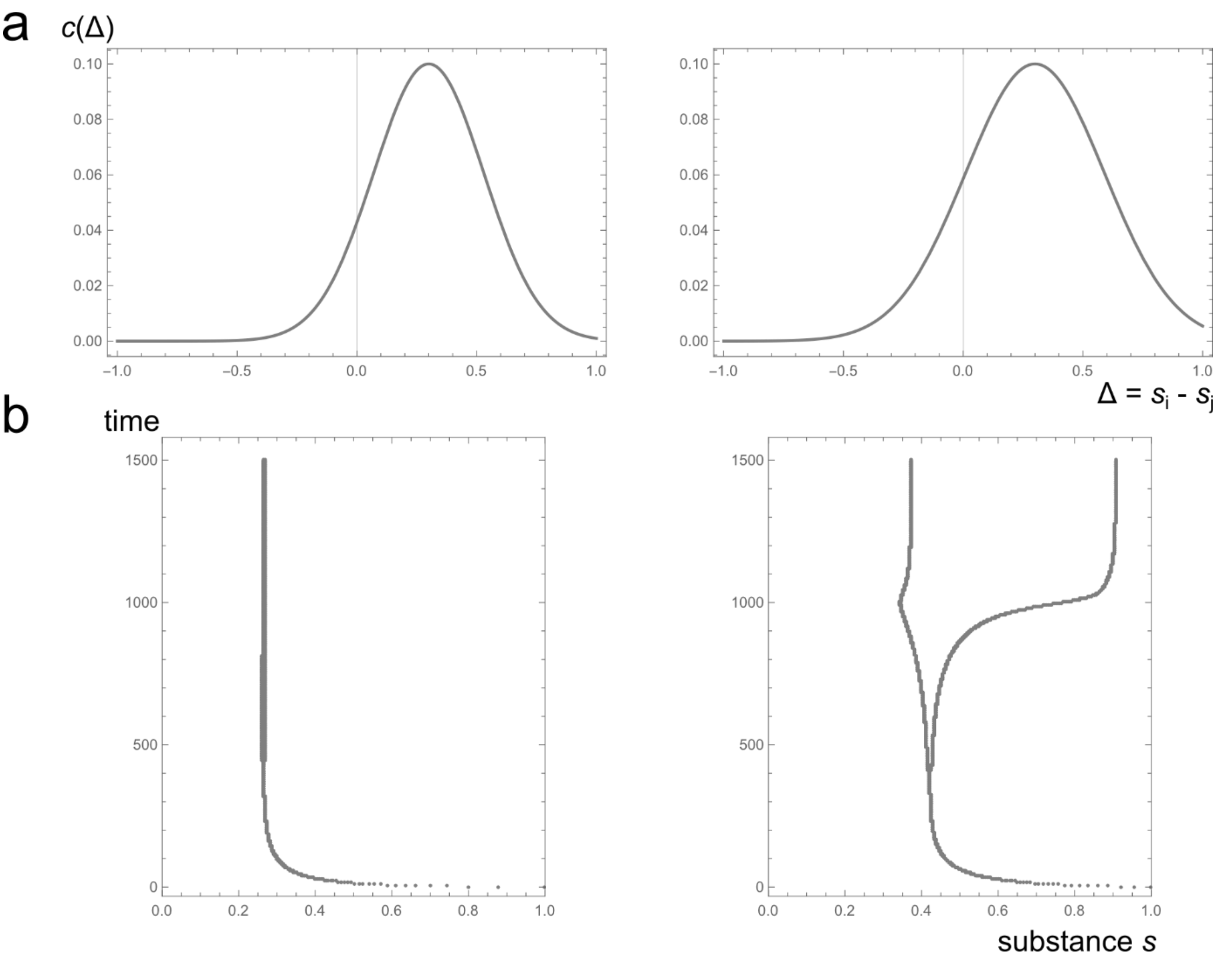


Figure 2. (a) Asymmetric competition terms with $\mu = 0.3$ and $c_{\max} = 0.1$ assuming strong (left; $\tau_{\text{strong}} = 0.23$) and weak (right; $\tau_{\text{weak}} = 0.29$) priming effects, respectively. (b) Evolutionary trajectory of formal substance $s$ based on the canonical equation of adaptive 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).

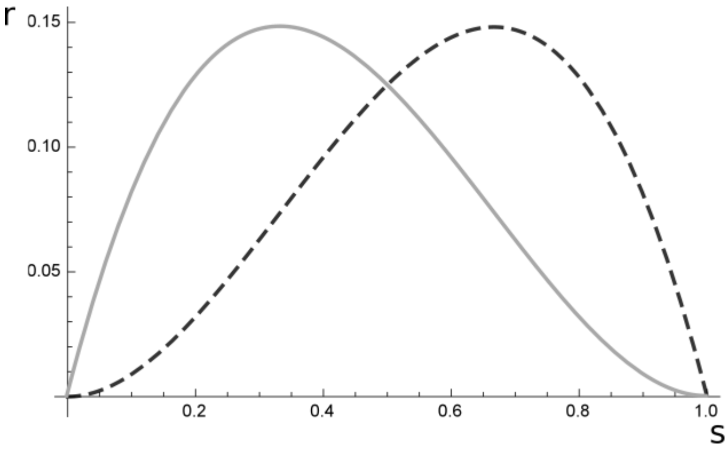


Figure 3. Intrinsic growth rate $r$ as a function of $s$, where $r(s) = Cs^{\alpha}(1-s)^{\pi}$. Solid light gray curve: $\alpha = 1$, $\pi = 2$, i.e. perceptual effort dominates. Dashed dark gray curve: $\alpha = 2$, $\pi = 1$, i.e. articulatory effort dominates. In both cases, $C = 1$.

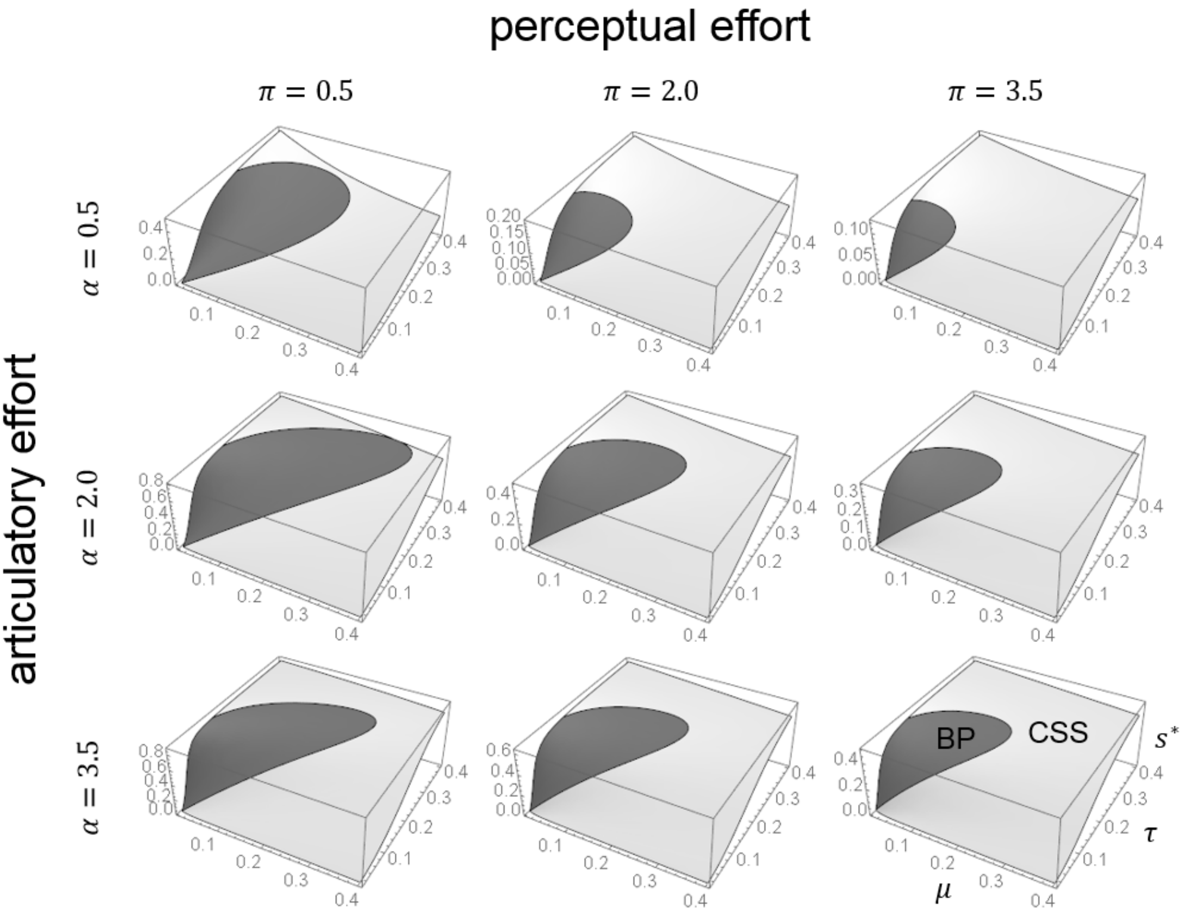


Figure 4. Bifurcation plots of the evolutionary singularity $s^*$ depending on the similarity threshold $\mu$ and priming strength $\tau$. Dark gray areas denote BPs, light gray areas denote CSSs.

Plots are shown for different values of articulatory effort $\alpha$ (rows) and perceptual effort $\pi$ (columns).

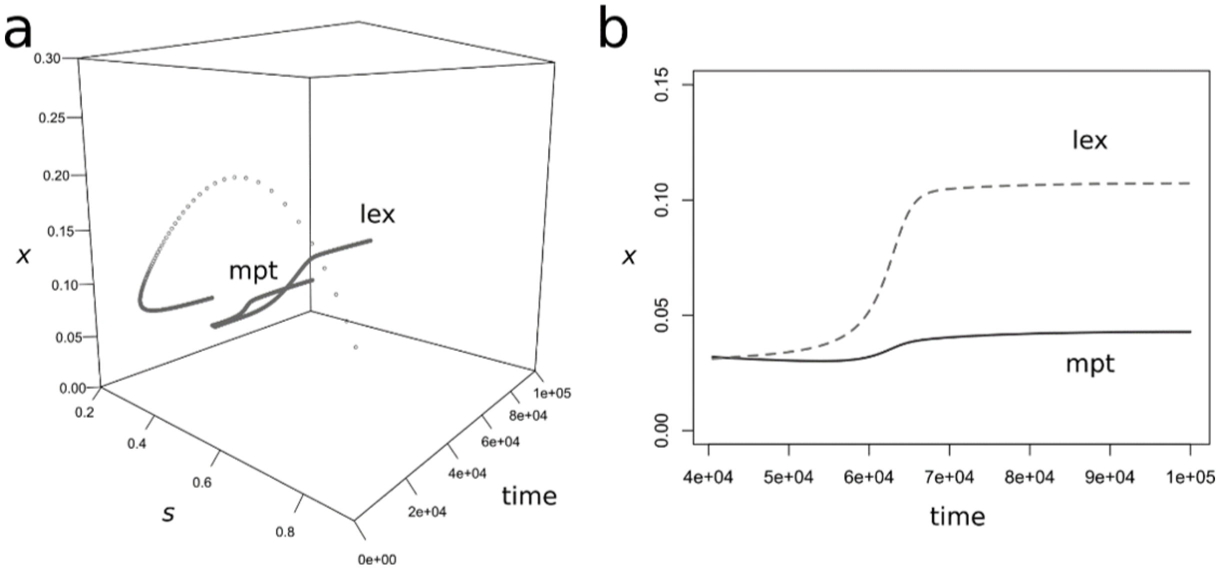


Figure 5. (a) Evolutionary trajectory of $(s, \hat{x}(s))$ before and after branching. Substance s proceeds towards a BP, subsequently followed by branching and coexistence of a shorter (morphonotactic, 'mpt') and a longer (lexical, 'lex') variant (only every 100th point displayed). (b) Frequency trajectories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

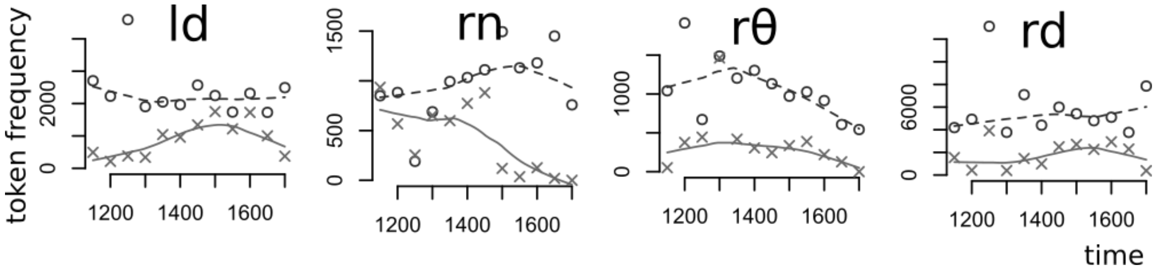


Figure 6. Empirical developments of four word-final consonant-diphone types retrieved from Middle and Early Modern English corpus data. Circles and crosses denote normalized frequencies (p.m.) of morpheme internal (lexcial) and boundary spanning (morphonotactic) diphones, while dashed and solid lines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

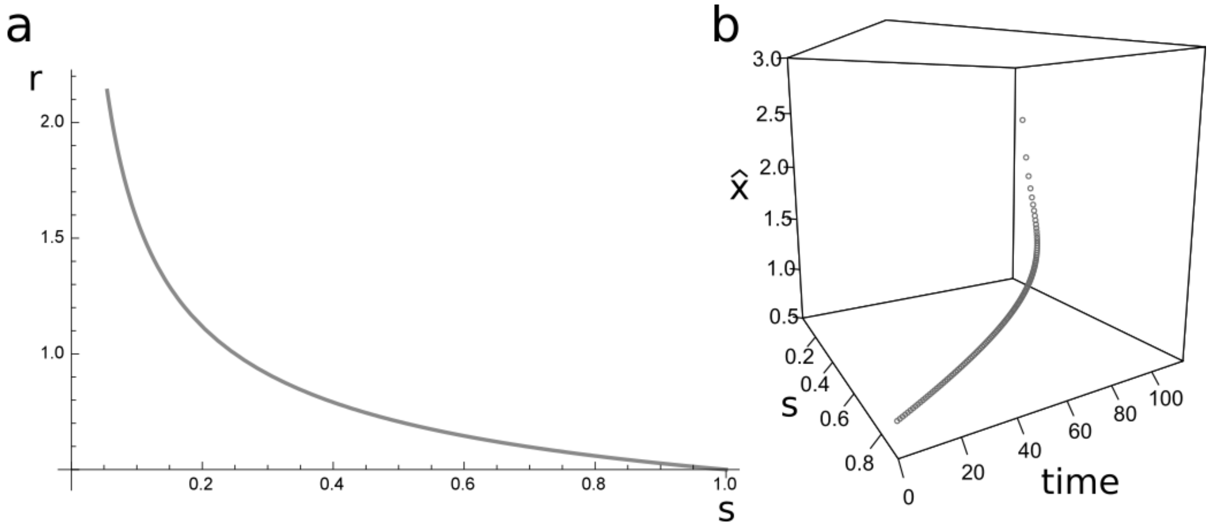


Figure 7. Evolution of formal substance $s$ in grammaticalization under asymmetric formal priming and (a) Zipfian intrinsic growth. (b) Items undergo unidirectional reduction and become increasingly frequent (frequency $\hat{x}$ measured on the vertical axis; $C = 1, \kappa = 0.5, c_{\max} = 1, \mu =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

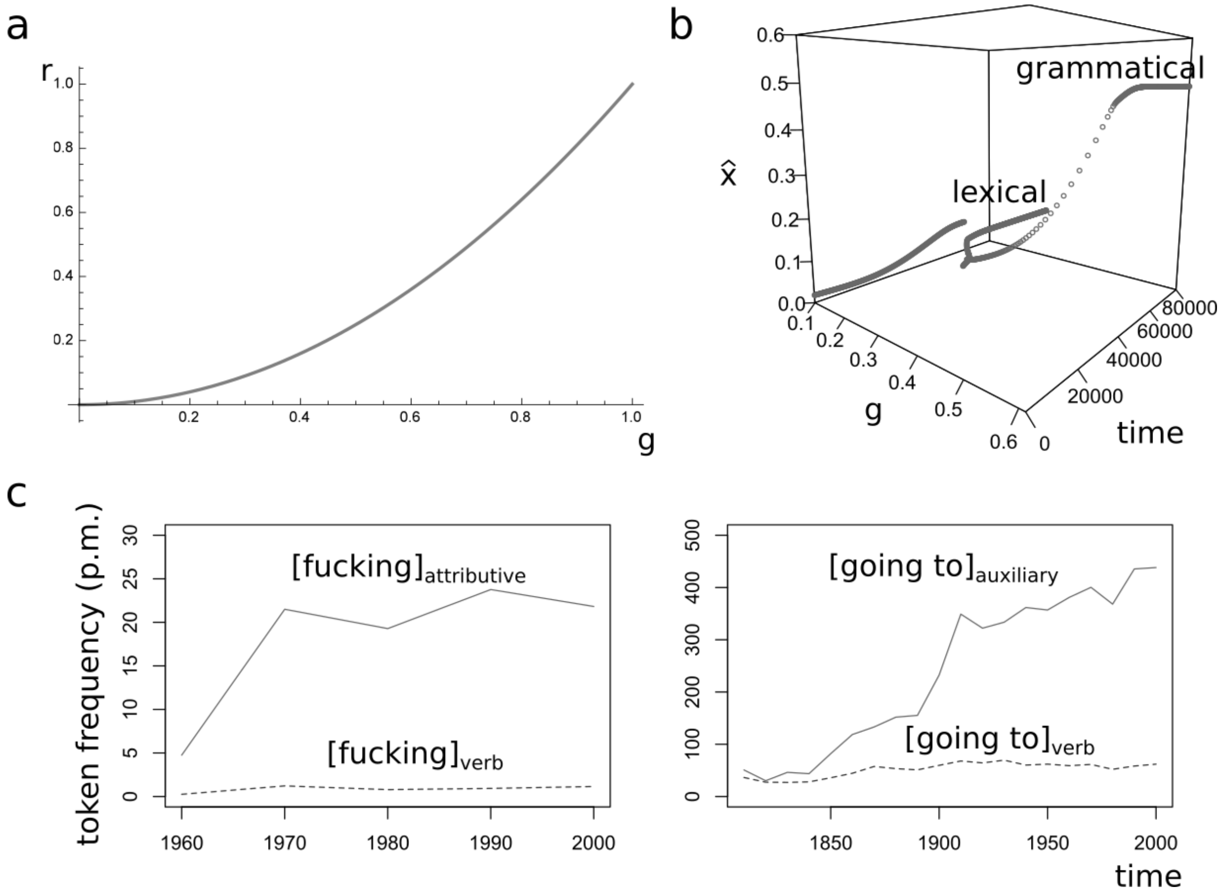
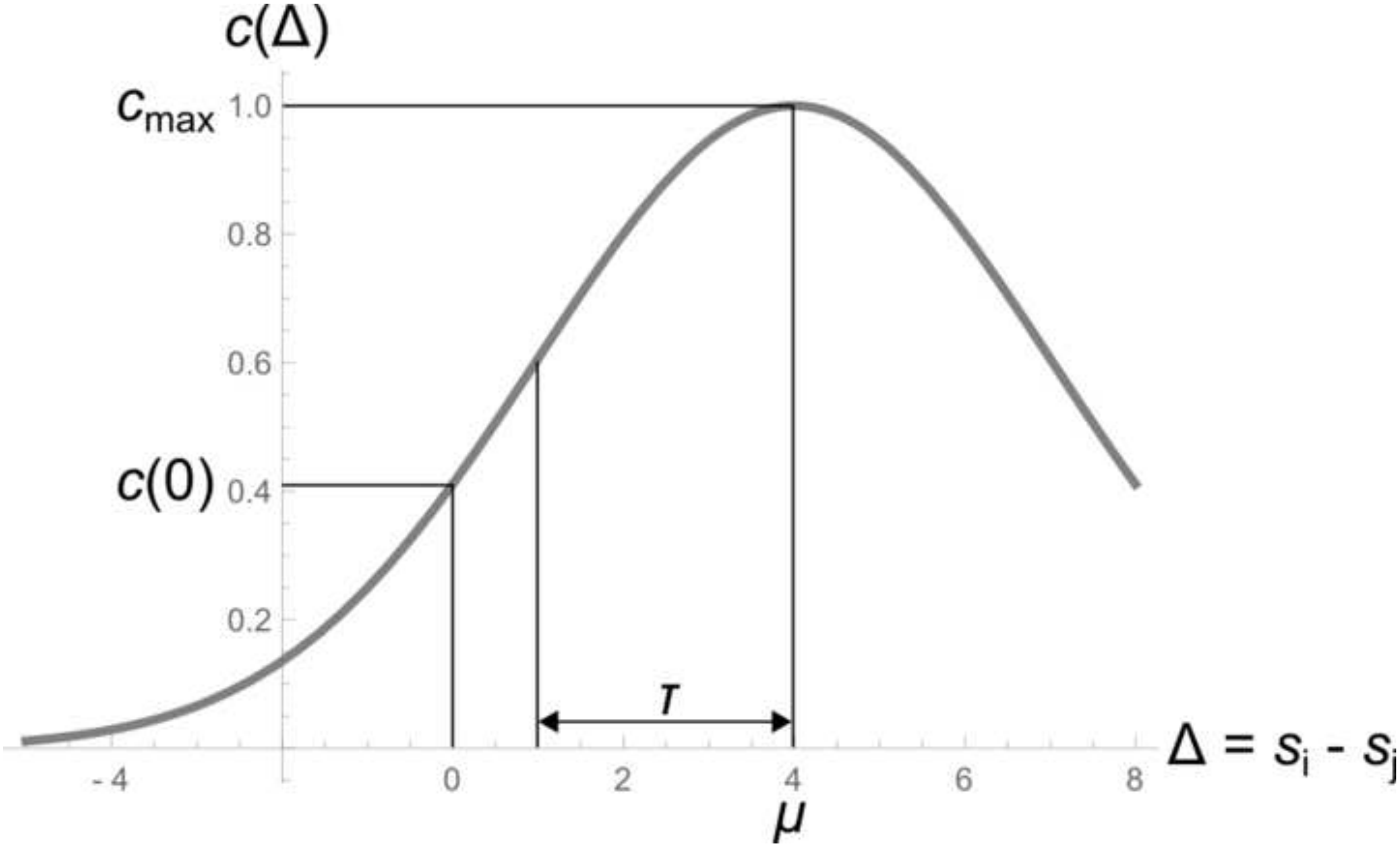
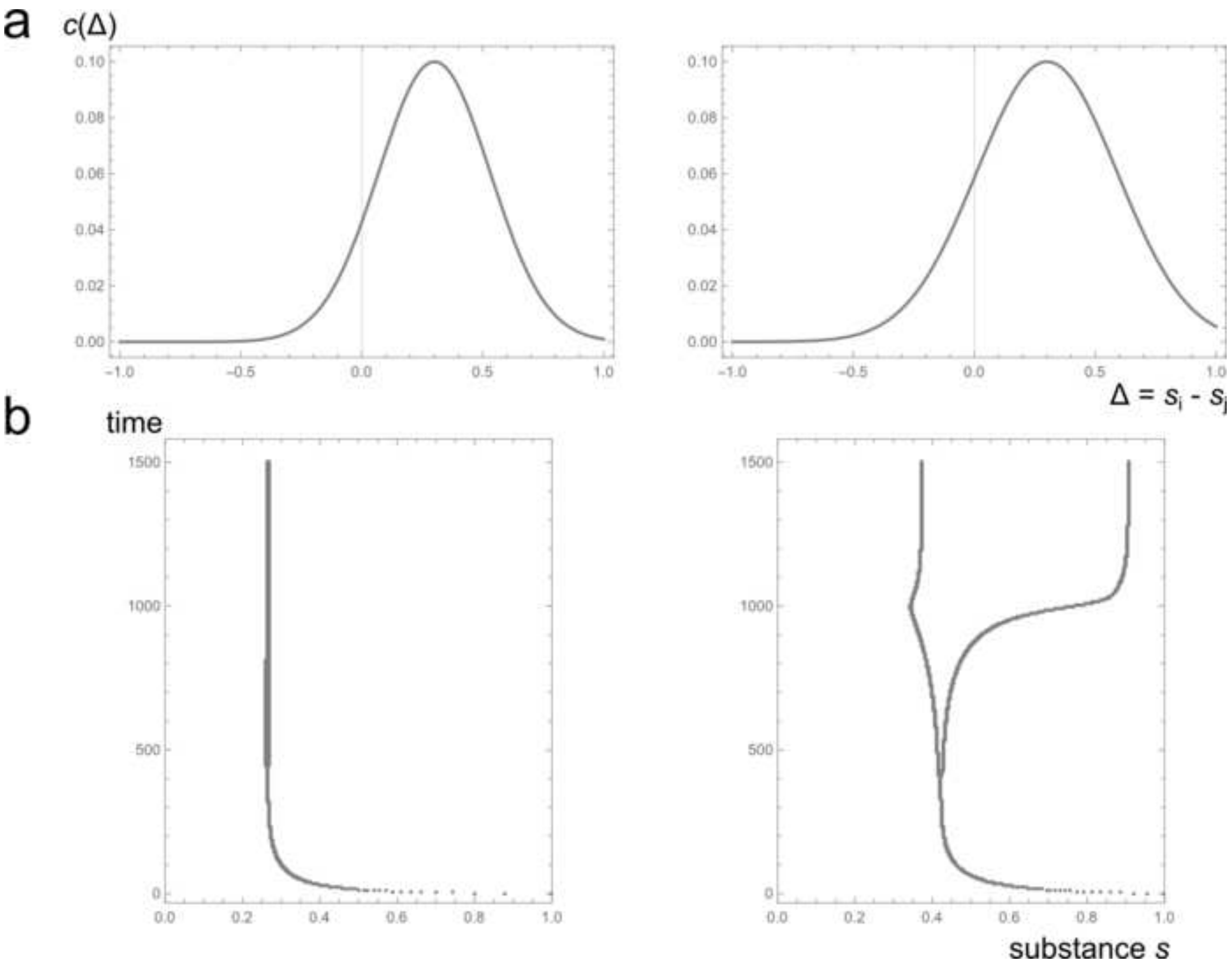


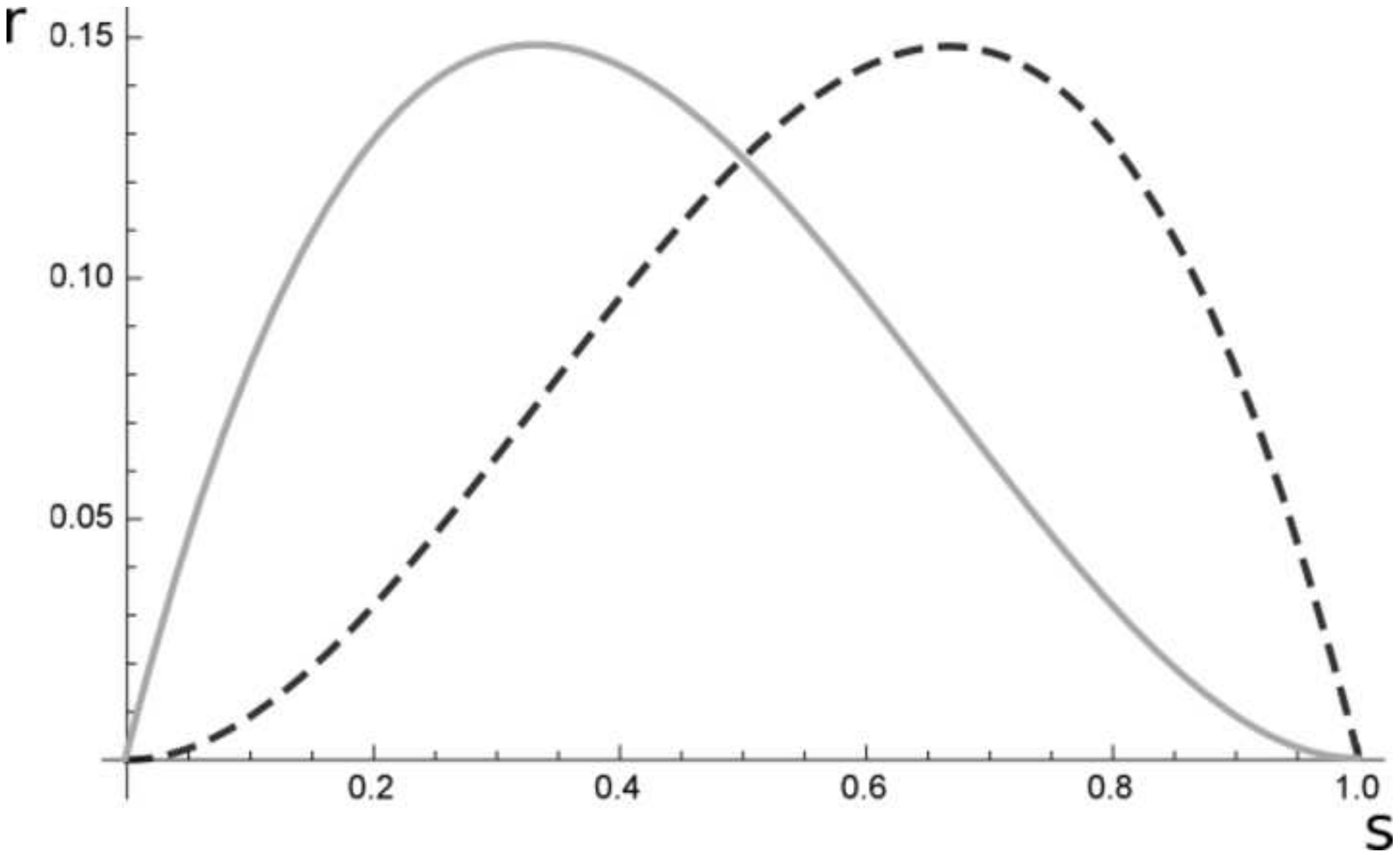
Figure 8. Evolution of the degree of grammaticality $g$ in grammaticalization under asymmetric priming among words $c_{\text{word}}$ and (a) a positive relationship between $g$ and intrinsic growth rate: $r(g) = g^2$. (b) After a period of increasing grammaticality (and decreasing formal substance), 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

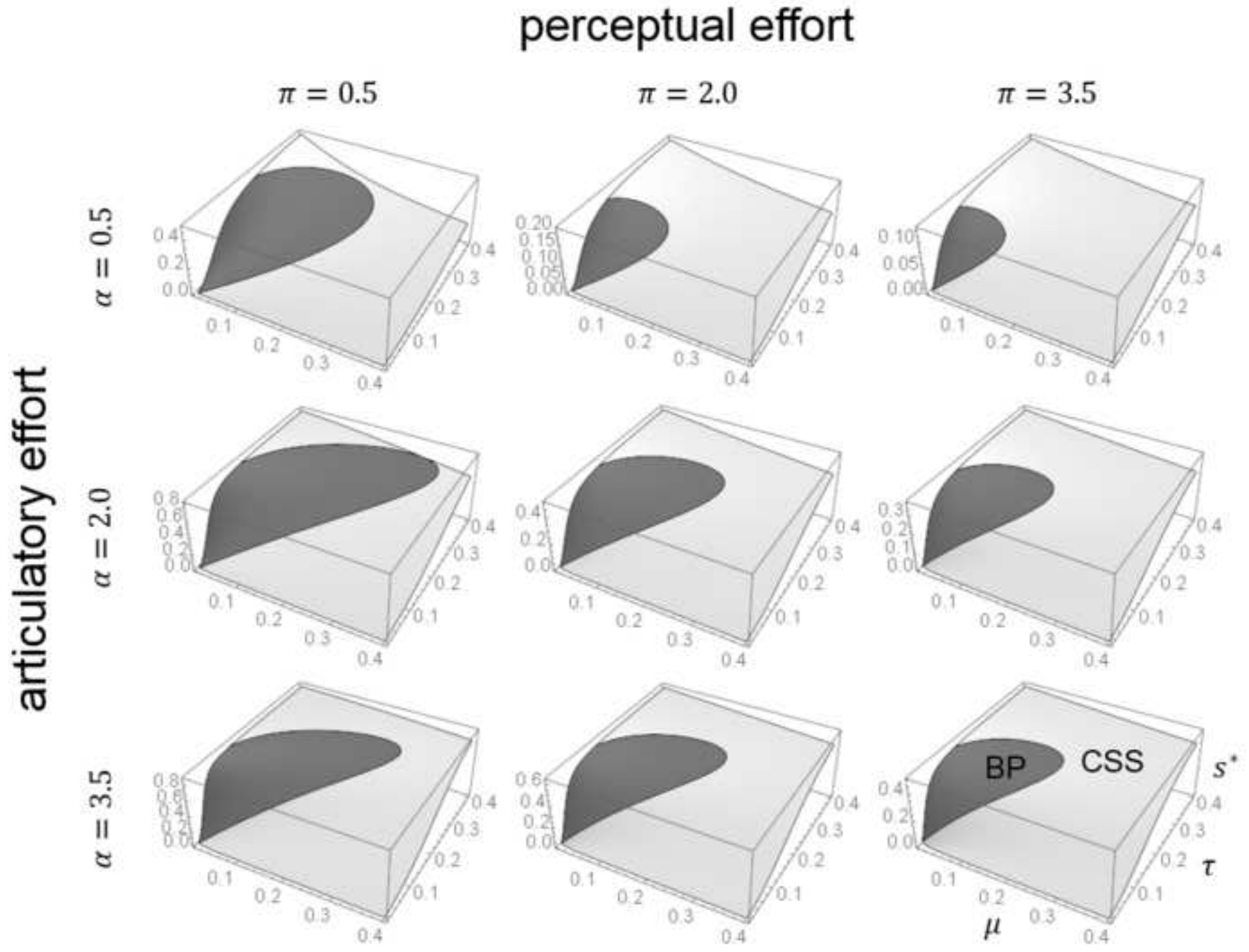
word is more frequent and more reduced than its more lexical cousin. Both trajectories exhibit sigmoid shapes ($c_{\max} = 1, \mu = 0.2, \tau = 0.18$; only every 100th point displayed). (c) Diachronic trajectories of grammaticalized (solid) and lexical (dashed) variants. On the left: attributive (grammaticalized) 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*.

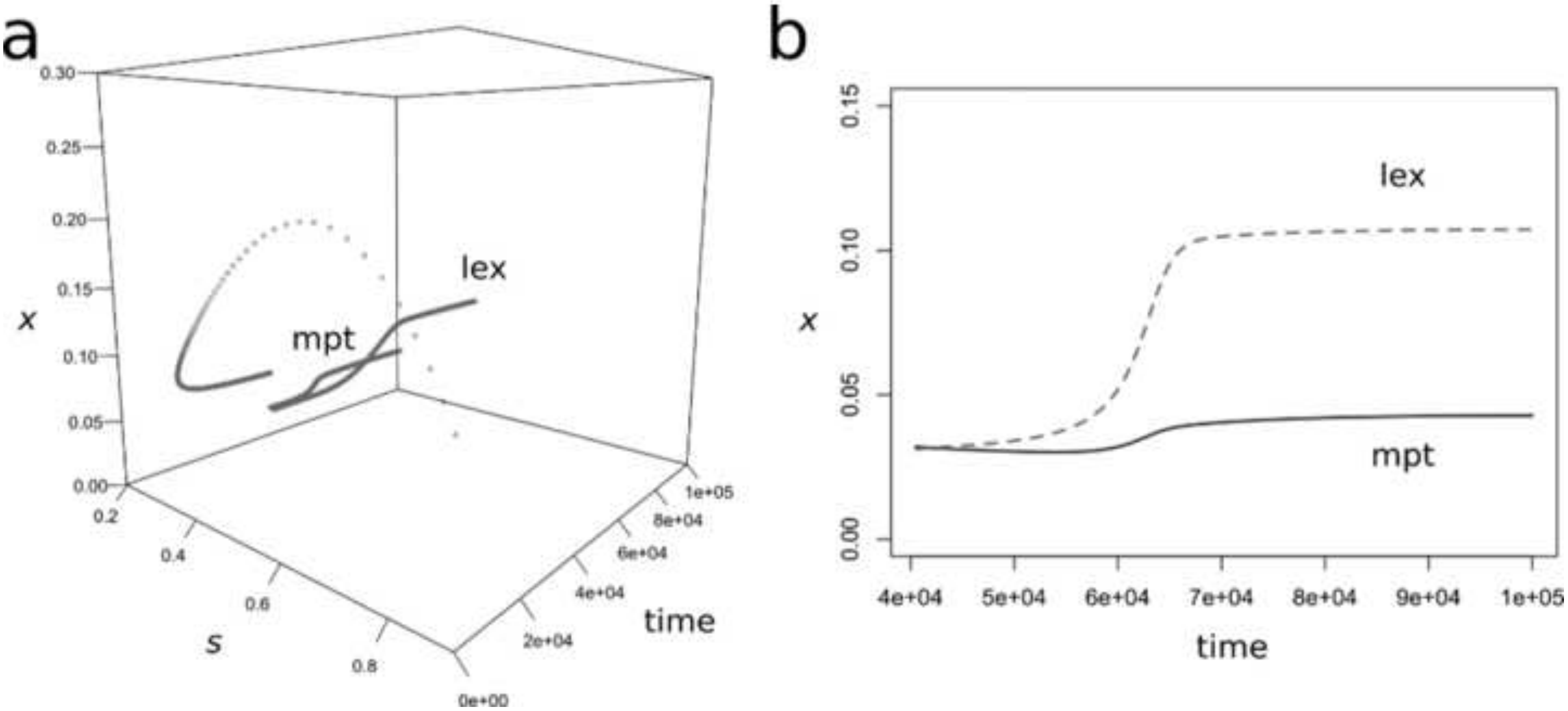


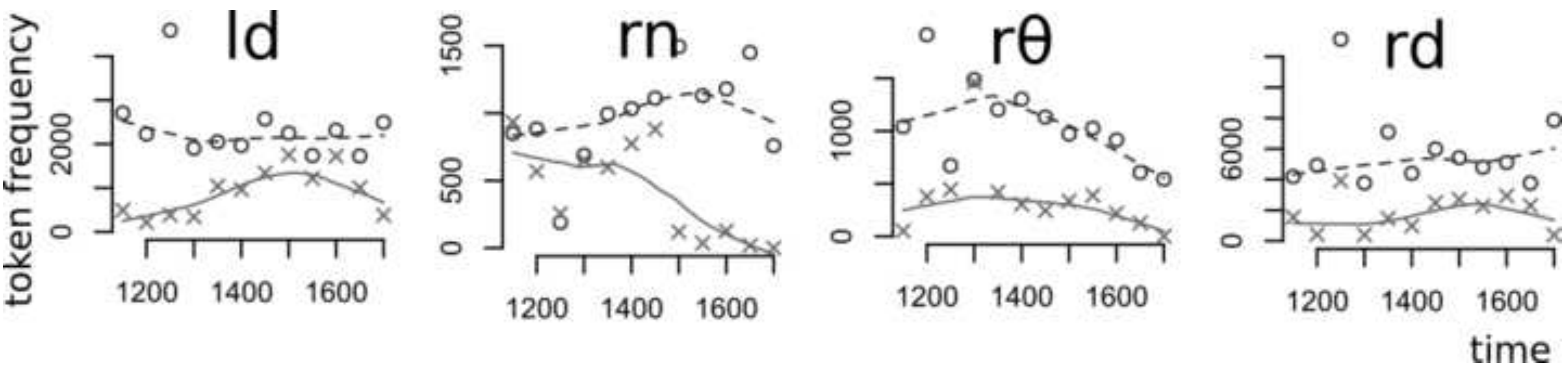
$c(\Delta)$
$c_{max}$
1.0
0.8
0.6
$c(0)$
0.4
0.2
$T$
$\Delta = s_i - s_j$
- 4
0
2
4
6
8
$\mu$

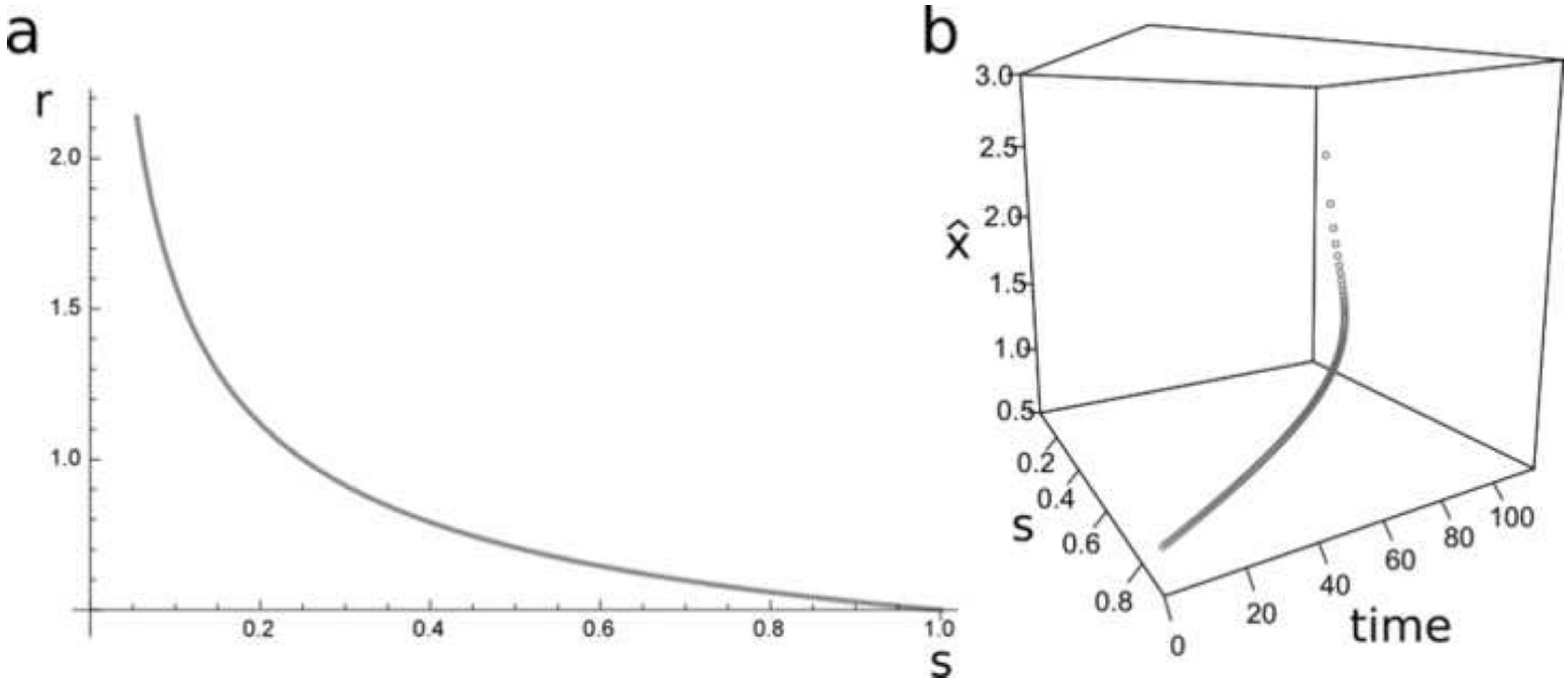

a
c(Δ)
0.10
0.08
0.06
0.04
0.02
0.00
-1.0
-0.5
0.0
0.5
1.0
Δ = s_i - s_j
b
time
1500
1000
500
0
0.0
0.2
0.4
0.6
0.8
1.0
substance s

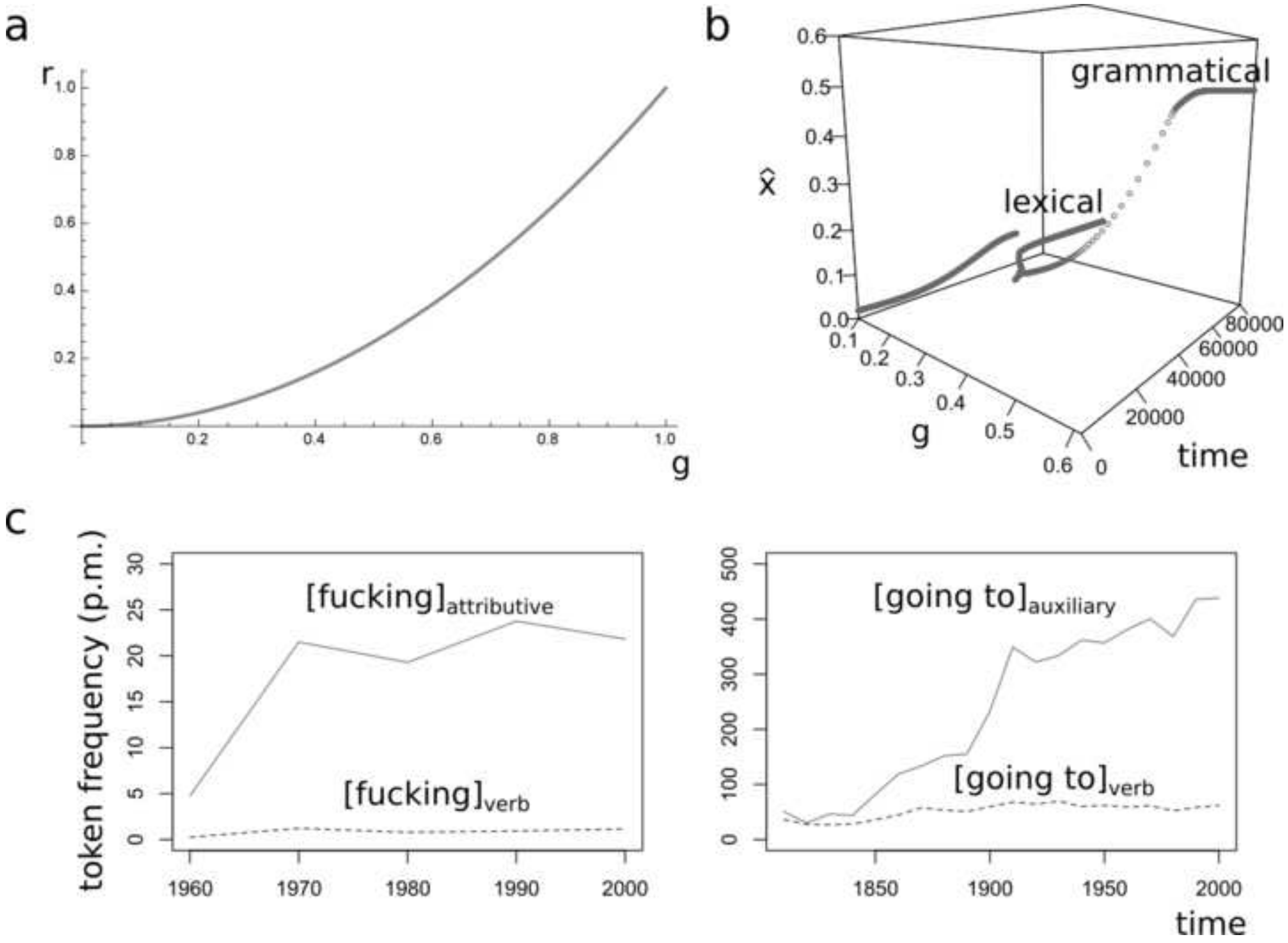

r
0.15
0.10
0.05
0.2
0.4
0.6
0.8
1.0
s


perceptual effort
$\pi = 0.5$
$\pi = 2.0$
$\pi = 3.5$
articulatory effort
$\alpha = 0.5$
$\alpha = 2.0$
$\alpha = 3.5$
0.0
0.2
0.4
0.00
0.05
0.10
0.15
0.20
0.6
0.8
0.1
0.3
0.1
0.2
0.3
0.4
BP
CSS
$s^*$
$\tau$
$\mu$


a
0.30
0.25
0.20
0.15
0.10
0.05
0.00
x
lex
mpt
0.2
0.4
0.6
0.8
s
0e+00
2e+04
4e+04
6e+04
8e+04
1e+05
time
b
0.15
0.10
0.05
0.00
x
lex
mpt
4e+04
5e+04
6e+04
7e+04
8e+04
9e+04
1e+05
time


token frequency
ld
rn
rθ
rd
0
2000
0
500
1500
0
1000
0
6000
1200
1400
1600
time


a
r
2.0
1.5
1.0
0.2
0.4
0.6
0.8
1.0
s
b
3.0
2.5
2.0
$\hat{x}$
1.5
1.0
0.5
0.2
0.4
s
0.6
0.8
0
20
40
60
80
100
time


a
r
1.0
0.8
0.6
0.4
0.2
0.2
0.4
0.6
0.8
1.0
g
b
0.6
0.5
0.4
0.3
0.2
0.1
0.0
x̂
grammatical
lexical
0.1
0.2
0.3
0.4
0.5
0.6
g
0
20000
40000
60000
80000
time
c
token frequency (p.m.)
30
25
20
15
10
5
0
[fucking]attributive
[fucking]verb
1960
1970
1980
1990
2000
500
400
300
200
100
0
[going to]auxiliary
[going to]verb
1850
1900
1950
2000
time

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

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 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).

---

[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).

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 $x_1, x_2, \ldots, x_N$ 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 $i = 1, \ldots, N$. 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 $x_1, x_2, \ldots, x_N$ of *N* formally related linguistic types numbered from 1 to *N*, depending on their respective formal properties $s_1, s_2, \ldots, s_N$ 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 $s_1, s_2, \ldots, s_N$ affecting the development of continuous frequencies $x_1, x_2, \ldots, x_N$. 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 $x_1, x_2, \ldots, x_N$. 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 $j$,

---

[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.

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 $c(s_i - s_j) < c(s_j - s_i)$.

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

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

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 $\Delta$ (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 $\mu$ 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, $\mu$ operationalizes the scope of priming. The parameter $\tau$ the extent to which priming is asymmetric (it determines the steepness of the curve). If $\tau$ is large both items have a relatively similar impact on each other. If $\tau$ 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.

Fig. 1 here

---

[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.

Box 1 here

### 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

---

[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)).

to apply the framework of adaptive dynamics to diachronic change in linguistics (see also Doebeli 2011 and AUTHORS 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 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

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

---

[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).

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, $\tau$ in the asymmetric-priming term should not be much smaller than $\mu$ (ideally $\tau > \mu$). If this is the case then the curve defining 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 $\tau$ is small, the asymmetric-priming curve is much steeper than if $\tau$ 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.

Fig 2 here

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.

## 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

---

[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).

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

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

---

[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.

where $C$, $\alpha$ and $\pi$ are strictly positive. In this function, $\alpha$ measures articulatory effort and $\pi$ 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 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.

Fig 3 here

Box 1 about here

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 $\alpha$, perceptual effort $\pi$, the similarity threshold $\mu$ defining the scope of priming and strength of asymmetric priming $\tau$ (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 $\alpha$, $\pi$, $\mu$ and $\tau$ 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 $\mu$ and $\tau$ (horizontal axes). Across the single bifurcation plots, perceptual effort $\pi$ increases from the left-most column to the right-most column, while articulatory effort $\alpha$ 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.

Fig 4 here

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 $\mu$ as can be seen from the decreasing values on the vertical axis. Since $\mu$ 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

---

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

[12] As can be seen, there are no repellors or Garden-of-Eden points for the admitted combinations of $\alpha$, $\pi$, $\mu$ and $\tau$. See appendix.

that it applies to pairs of items which are substantially different from another (large $\mu$). In contrast, if asymmetric priming has a narrow scope (small $\mu$), then formal reduction is hampered.

Second, $s^*$ increases in $\tau$, which determines the impact of asymmetric priming. If $\tau$ is small, then asymmetric priming has a strong impact. In that case, items tend to get shortened. If $\tau$ 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 $\mu$ and strength of asymmetric priming $\tau$ play in evolutionary branching (dark gray region in Fig. 4). As can be seen in Fig. 4, $\mu$ must be relatively small in order to enable stable diversification. If $\mu$ 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, $\tau$ must be greater than $\mu$, 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, $\tau$ must not be too large, and if $\tau$ is large then $\mu$ 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.

Box 2 about here

We can simulate the evolution of a diphone’s duration $s$ given articulatory effort $\alpha$, perceptual effort $\pi$, similarity threshold $\mu$ and strength of asymmetric priming $\tau$. Figure 5a

---

[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.

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.

Fig 5 here

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.

Fig 6 here

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 $\kappa$ 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.

Fig 7 here

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.

Fig 8 here

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

---

[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).

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 about here

## 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

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

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 $\mu$ is positive (which is plausible because the effect of priming ultimately decreases with dissimilarity) and since $c(\Delta) > 0$ for all $\Delta$. The parameter $\tau$ determines the steepness of the curve defined by $c$. If $\tau$ is small, then the effect of asymmetric priming is very strong. Conversely, if $\tau$ 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 $\tau$ 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’:

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

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 $r$, $\mu$ and $\tau$ 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

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

## 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 $r'(s) = -\kappa C s^{-\kappa-1} < 0$ 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.

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.

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 |
| $\mu$ | similarity threshold for asymmetric priming (scope of priming); beyond a difference of $\mu$, priming effects become weaker |
| $\tau$ | measure of the strength of asymmetric priming; if $\tau$ is small/large priming has strong/weak effects on processing |
| $\alpha$ | language specific articulatory effort; small $\alpha$ corresponds to a speaker friendly linguistic system (see 4.1) |
| $\pi$ | language specific perceptual effort; small $\pi$ corresponds to a listener friendly linguistic system (see 4.1) |
| $\kappa$ | language specific strength of the inverse relationship between substance and productivity of words (see 4.2) |

Box 2. Sublexical dynamics: key results

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

Box 3. Lexical dynamics: key results

| | *Assumptions* | |
|---|---|---|
| | Substance only | Substance and meaning (degree of grammaticality $g$) |
| 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 $\tau$ | Unidirectional reduction irrespective of $\tau$ | Diversification possible under weak asymmetric priming |
| Effect of scope of asymmetric priming $\mu$ | Unidirectional reduction irrespective of $\mu$ | Diversification possible if priming has a relatively small scope |

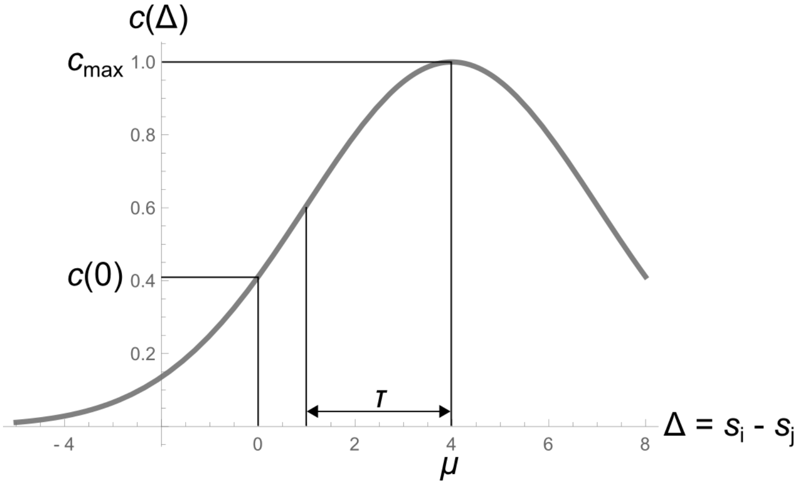


Figure 1. Gaussian function underlying the asymmetric competition term with $c_{\max} = 1, \mu = 4, \tau = 3$.

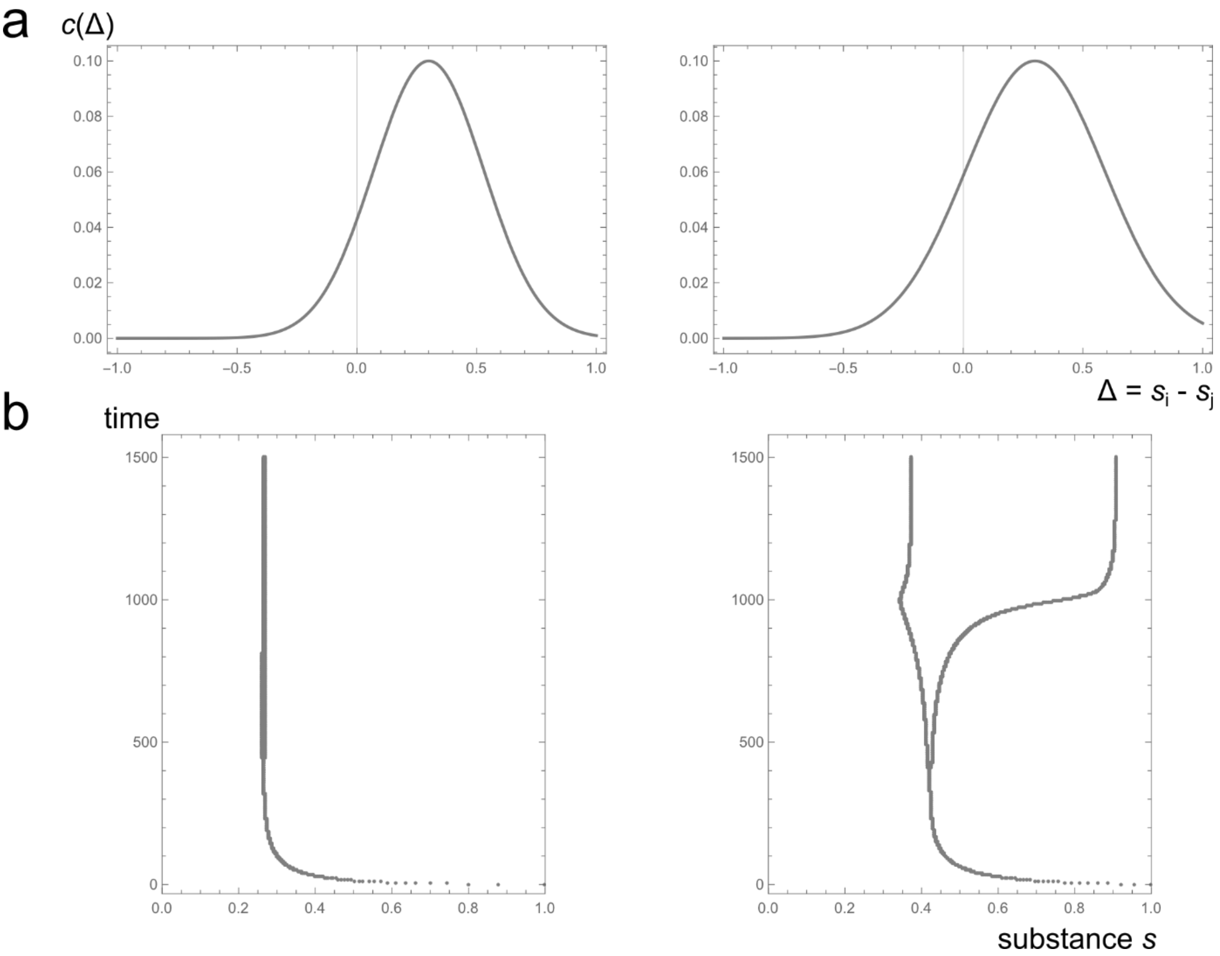


Figure 2. (a) Asymmetric competition terms with $\mu = 0.3$ and $c_{\max} = 0.1$ assuming strong (left; $\tau_{\text{strong}} = 0.23$) and weak (right; $\tau_{\text{weak}} = 0.29$) priming effects, respectively. (b) Evolutionary trajectory of formal substance $s$ based on the canonical equation of adaptive 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).

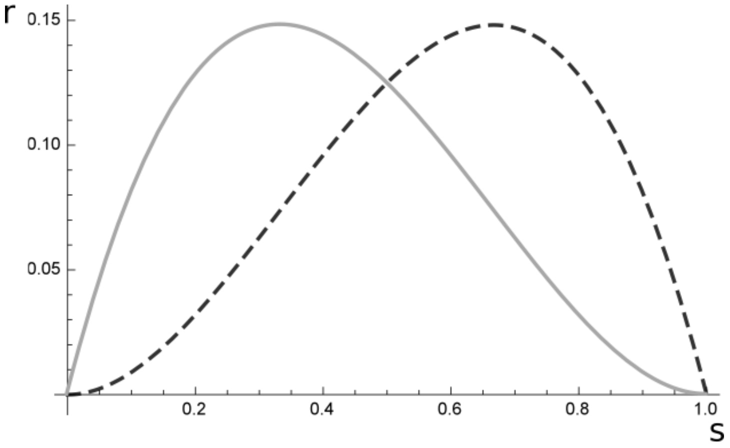


Figure 3. Intrinsic growth rate $r$ as a function of $s$, where $r(s) = Cs^{\alpha}(1-s)^{\pi}$. Solid light gray curve: $\alpha = 1$, $\pi = 2$, i.e. perceptual effort dominates. Dashed dark gray curve: $\alpha = 2$, $\pi = 1$, i.e. articulatory effort dominates. In both cases, $C = 1$.

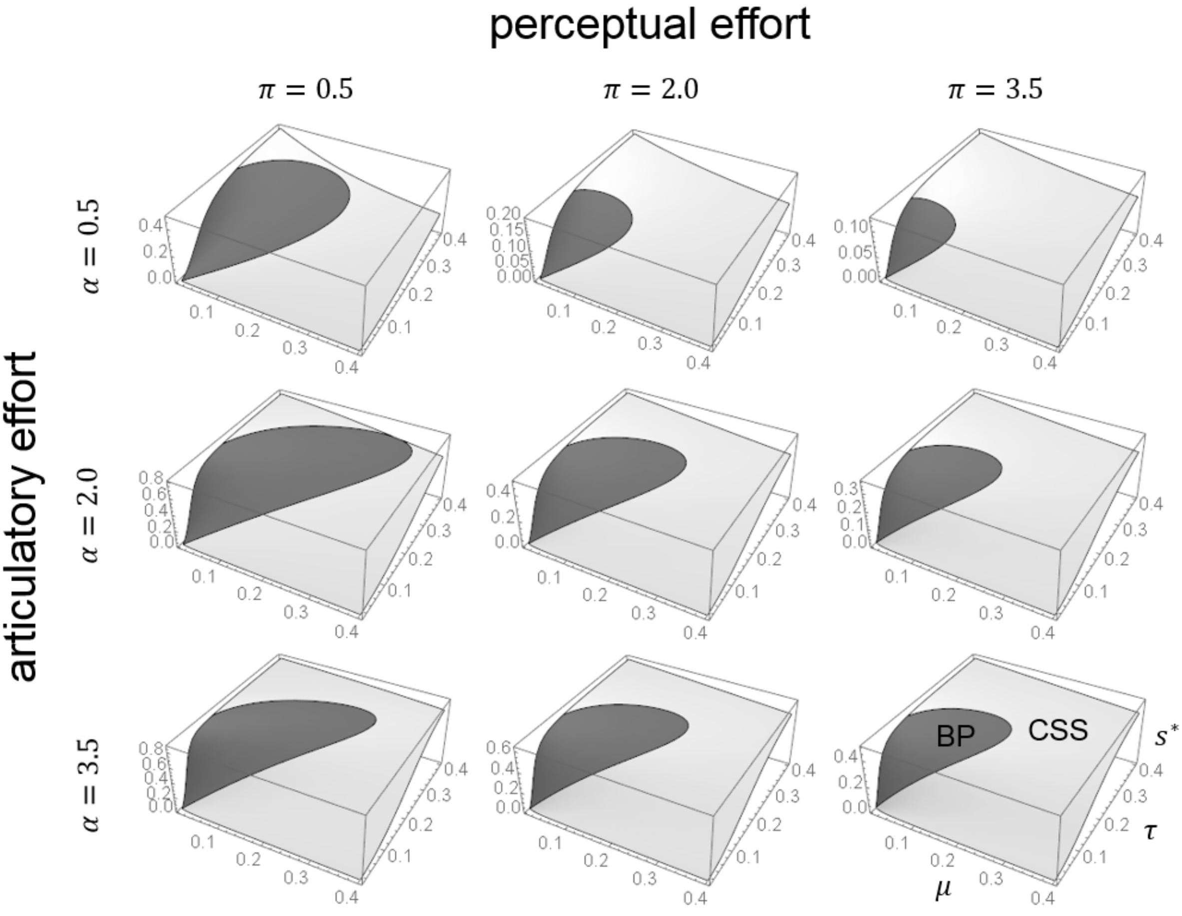


Figure 4. Bifurcation plots of the evolutionary singularity $s^*$ depending on the similarity threshold $\mu$ and priming strength $\tau$. Dark gray areas denote BPs, light gray areas denote CSSs.

Plots are shown for different values of articulatory effort $\alpha$ (rows) and perceptual effort $\pi$ (columns).

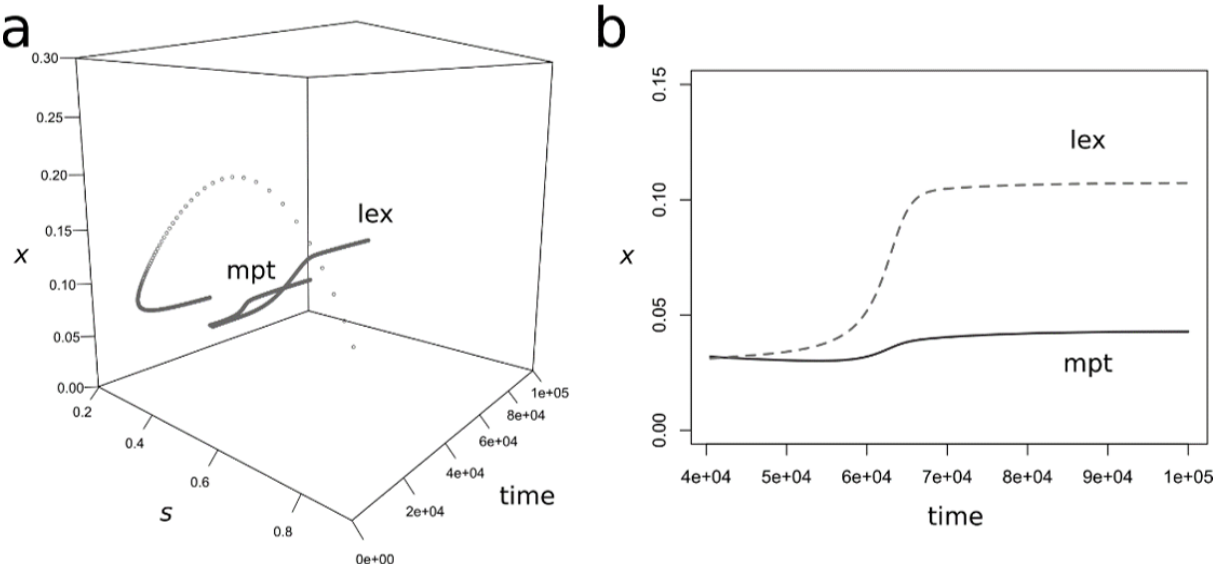


Figure 5. (a) Evolutionary trajectory of $(s, \hat{x}(s))$ before and after branching. Substance s proceeds towards a BP, subsequently followed by branching and coexistence of a shorter (morphonotactic, 'mpt') and a longer (lexical, 'lex') variant (only every 100th point displayed). (b) Frequency trajectories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

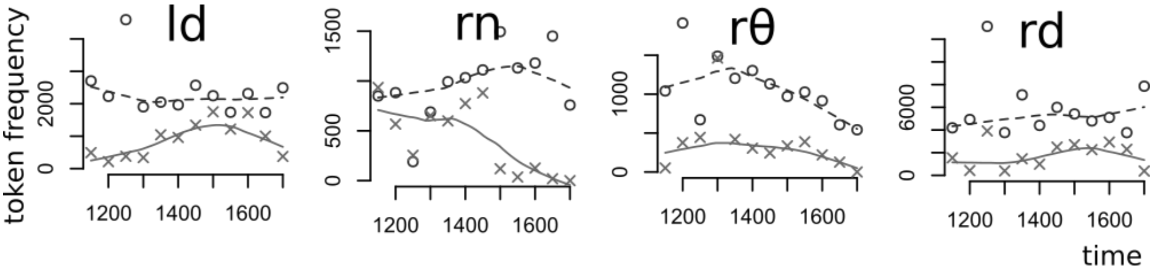


Figure 6. Empirical developments of four word-final consonant-diphone types retrieved from Middle and Early Modern English corpus data. Circles and crosses denote normalized frequencies (p.m.) of morpheme internal (lexcial) and boundary spanning (morphonotactic) diphones, while dashed and solid lines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

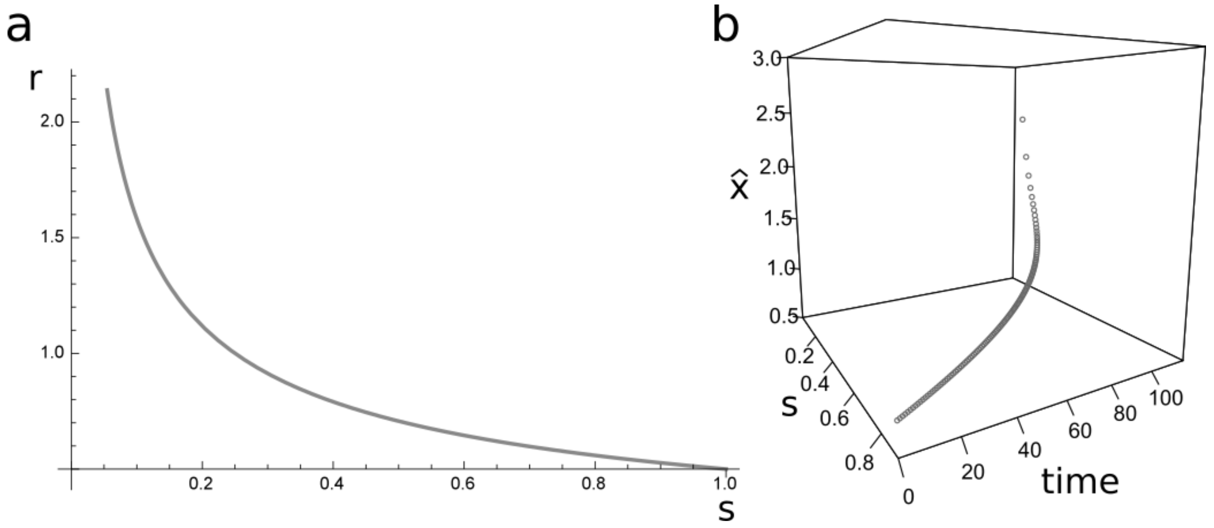


Figure 7. Evolution of formal substance $s$ in grammaticalization under asymmetric formal priming and (a) Zipfian intrinsic growth. (b) Items undergo unidirectional reduction and become increasingly frequent (frequency $\hat{x}$ measured on the vertical axis; $C = 1, \kappa = 0.5, c_{\max} = 1, \mu =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

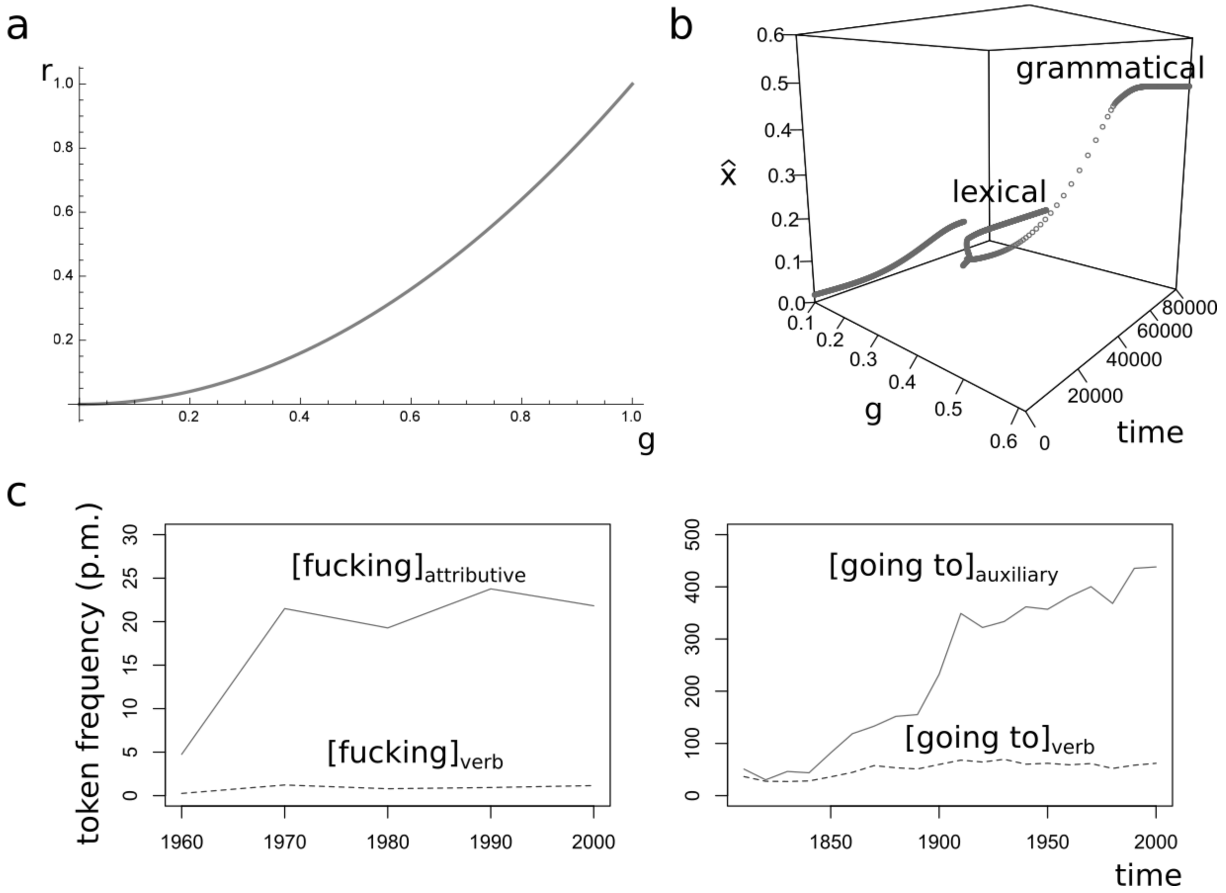


Figure 8. Evolution of the degree of grammaticality $g$ in grammaticalization under asymmetric priming among words $c_{\text{word}}$ and (a) a positive relationship between $g$ and intrinsic growth rate: $r(g) = g^2$. (b) After a period of increasing grammaticality (and decreasing formal substance), 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*.