On the replicator dynamics of lexical stress: modelling rhythmically driven diachronic change in terms of evolutionary game theory

Abstract

This paper addresses the problem that diverse stress patterns are attested among English words such as *hoˈtel* – *ˈlentil*, *ˈenvoy* – *iˈdea*, *ˈresearch*$_N$ – *reˈsearch*$_N$, or *ˈaccess*$_V$ – *acˈcess*$_V$, although they are equivalent in terms of phonotactic structure and morpho-syntactic category. Specifically it discusses whether such cases are accidental or systematic, and demonstrates that the latter is the case. Contrary to extant accounts of lexical stress, it proposes an evolutionary and utterance based approach, under which words are assumed to adopt those stress patterns that produce, on average, optimal rhythmic configurations in actual utterances. This hypothesis is modelled in evolutionary game theory, which allows to make predictions about the stability of stress pattern distributions in word populations. We show that stress pattern diversity among otherwise equivalent polysyllabic word forms depends on the frequency of monosyllables and has been inevitably stable in English since the Middle English period.

1 Introduction

This paper attempts to identify circumstances under which the lexicon of a language will predictably display stress pattern diversity. Although we try to address the question in the most general terms possible, our discussion is inspired by and based on the situation that obtains in English, and we shall keep referring to that specific language to explain the problem itself, to justify our approach and our methodology, and to demonstrate of our hypotheses.

When we say that a language such as English displays stress pattern diversity we mean that it contains polysyllabic words that can assume different stress patterns even though in terms of phonotactic structure, morphological composition, and morpho-syntactic category they must be considered either as equivalent, as the examples in (1), or even as practically identical, as the examples in (2):

(1) Morphosyntactically and phonotactically equivalent items:
   a. Initial stress: ˈ*concert* (N), ˈ*intern* (N), ˈ*envoy* (N), ˈ*convoy* (N), ˈ*parlay* (N), ˈ*lentil* (N);ˈ*stupent* (Adj),ˈ*constant* (Adj)
   b. Final stress: *desˈsert* (N), *conˈcern* (N) *emˈploy* (N), *disˈmay* (N), *arˈray* (N), *iˈdea* (N), *hoˈtel* (N); *stuˈpend* (Adj), *conˈtent* (Adj)

(2) Identical items:
   a. Initial stress: ˈ*research* (N), ˈ*revert* (N), ˈ*robust* (Adj) ˈ*access* (V)
   b. Final stress: *reˈsearch* (N), *reˈvert* (N), *roˈbust* (Adj) *acˈcess* (V)

We consider this kind of diversity to be interesting because it goes against the fact that, for the most part, English words assume unique stress patterns that can be predicted quite precisely if one takes their phonotactic structure, their morphological composition, and their morphosyntactic category into account. This has been repeatedly demonstrated in a large body of research (see e.g. Anderson 1986; Burzio 1994; Chomsky & Halle 1968; Fudge 1984, 1999; Giegerich 1985; Halle 1998; Hayes 1982; Hyde 2007, 2011; Kager 1989; Kiparsky 1979; Liberman & Prince 1977; Marvin 2003; Poldauf & Lee 1984; Sainz 1988; Schane 2007; van Oostendorp 2010). Thus, and although work on English word stress has reached a high level of descriptive adequacy and coverage, items like the examples in (1) and (2) still represent a residue of obvious irregularities that remain to be accounted for. Our paper attempts to provide such an account.

We introduce and defend two hypotheses. The first, and more fundamental one, is that the constraints which word stress reflects, and which are essentially rhythmic, apply primarily on the phrase level rather than on the word level itself. The fact that rhythmic constraints on the phrase allow one to predict stress patterns by considering the word level alone, implies that words anticipate and adapt to the phrases in which they occur. Secondly, and more specifically, we propose that the emergence and the stable establishment of stress pattern diversity among polysyllables depends on the proportion of monosyllables in speech. We derive the second hypothesis on the basis of the first through a simulation in terms of an evolutionary game. The main part of our paper is dedicated to motivating, describing and interpreting that game in detail. Basically, it constructs polysyllabic words as 'players' and stress patterns as 'strategies'. In each round, two words combine to form a phrasal sequence. Apart from the 'players' themselves, phrasal sequences may also involve monosyllabic items. If the phrase built in an encounter is rhythmically well formed (observing, for example, the constraint that feet should be binary), the words earn a 'reward' and their stress patterns are stabilized. If, on the other hand, the sequence is arrhythmic, the stability of their stress pattern decreases.

First, however, we need to clarify how exactly we conceptualize the problem. What is important to point out is that we understand the conditions under which a language will admit and establish stress pattern diversity in its lexicon to be *general* conditions on the distribution of stress patterns in the lexicon. These general conditions differ from the specific reasons why individual lexical items come to assume their specific patterns, given that diversity is generally licensed. Applied to the case of English, this means that we do not intend to explain, for example, why *lentil* and *intern* take initial stress, while *hotel* and *concern* take final stress. What we want to explain instead is why English allows, or possibly requires, a subset of lexical items to take *either* initial *or* final stress.[1] In other words, the question of why specific words have specific stress patterns must not be confused with the question of why a language allows, or requires, stress pattern diversity among words of the same type, and this paper addresses the latter question.

Our game predicts that a mix of stress patterns will establish itself stably in a population of polysyllables if the proportion of monosyllabic items in speech is high enough. This is compatible with the fact that English is indeed highly monosyllabic and suggests that our hypotheses qualify as a potential explanation of stress pattern diversity, although it obviously does not yet confirm them. As our simulation is an *evolutionary* game, however, it also allows one to deduce testable, albeit global, predictions about the ways in which stress pattern distributions may change diachronically. We show that these predictions fit the historical development of English stress patterns well, which we interpret as a corroboration of our hypotheses.

Although the literature on stress assignment is vast[2], the question we address does not yet seem to have been discussed, but of course, our study relates to, builds on, and complements various strands of extant research. Our own background is in natural phonology (cf. e.g. Dressler 1985; Donegan & Stampe 1979; Donegan & Stampe 2009; Dziubalska-Kołaczyk 2009; Stampe 1979) and in theories of language as a culturally evolving system (cf. e.g. Altmann et al. 2009; Blevins 2004; Brighton et al. 2005;

---

[1] Crucially, it would also be wrong to interpret the general question as a mere summary of the specific questions it appears to cover. Instead, it is qualitatively different and belongs to a more fundamental level of explanation. Thus, even if one answers all the specific questions, one will not have answered the general one – nor vice versa. For instance, one might explain the final stress in hotel with reference to its French origin, and/or possibly the associations of elegance and luxury that a Romance stress pattern may evoke. Even if such an explanation were true (which may or may not be the case), it would not answer the deeper question of why English at all admits diverse stress patterns for words such as lentil, hotel, etc. After all, not all languages do. Polish, for example, has obligatory penultimate stress, also in hotel.) ¬– Likewise, one will not have explained why lentil stresses the initial syllable and hotel the final one, if one knows why English allows both patterns on words of that type.

[2] For English, see, once again the references on page 2.

Christiansen & Chater 2008a, 2008b; Croft 2013, 2000; Lass 1997; McMahon & McMahon 2013; Ritt 2004; Smith & Kirby 2008; Smith & Salmons 2008; Tallerman & Gibson 2012; Wedel 2009; Pierrehumbert 2012; Steels 2011).

Thus, sharing views already expressed by Hermann Paul (1920) and revived in recent decades (see the references above), we regard constituents of linguistic competence – in the context of this paper: the segmental and prosodic *gestalts* of words – as being instantiated in populations of speaker minds, and as owing their existence to successful transmission via communication and language acquisition. To be transmitted, lexical items (as, indeed, all competence constituents) need to be articulated, perceived, and processed in discourse, and their transmission is sensitive to physiological and cognitive constraints on these processes.

Following a large body of linguistic and biological literature (see e.g. Hayes 1984, or Fitch 2013 and the references therein) we take a deeply rooted preference for linguistic utterances to be rhythmically structured to be one of those constraints. As pointed out, our study rests on the idea that lexical word stress patterns represent adaptations to rhythmic well-formedness constraints on the phrases they build when uttered. That idea is not new. The impact of rhythmic, and ultimately utterance based, constraints on the lexical and grammatical constituents of specific languages has been demonstrated, for example, by Donegan & Stampe 1983, Fullwood 2014, Kelly 1988, 1989 or 1992, and Schlüter 2005. Also, our paper relates to studies of rhythmically induced accent shifts as in ˈ*Tenne*ˌ*see* (←ˌ*Tenne*ˈ*see*) ˈ*air, or* ˈ*thir*ˌ*teen* (← *thir*ˈ*teen*) ˈ*men* (cf. e.g. Liberman & Prince 1977; Hayes 1984). Unlike the latter studies, however, we do not restrict our focus to cases where phrase level rhythm actually reverses lexically established prominence relations, but explore the possibility that lexical stress patterns normally reflect such constraints already, which, we think, explains why observable stress shifts represent the exception rather than the norm.

The revival of evolutionary approaches to language has motivated the increased use quantitative methods developed for the study of dynamical systems in other domains, particularly evolutionary biology.[3] Their fruitfulness has encouraged us to approach our own research question also with such a method. The particular one we have chosen is evolutionary game theory, because our question is about

[3] There is no space, here, for reviewing the great diversity of relevant research. For the sake of illustration, the methods that have come to be employed include, for example, computational phylogenetics, e.g. Dunn et al. (2011), agent based modelling (Steels 2011), iterated learning experiments (Scott-Phillips & Kirby 2010), and analytical methods, such as, population dynamics (Baxter & Croft forthc.), or evolutionary game theory (Zuidema & Boer 2009).

general principles governing the distribution of stress patterns in the lexicon that result from *interactions* among the items that make it up. Since evolutionary game theory counts as an established tool for addressing issues of that kind, we feel confident in our choice.[4]

Needless to say, the view that constituents of linguistic competence are shaped by factors whose immediate effect is on discourse[5] relates our work to utterance based theories as represented, for example, in the work of Joan Bybee (e.g. 2001; 2007; Bybee et al. 1998) or Janet Pierrehumbert (e.g. Altmann et al. 2009; Pierrehumbert 2012). Since our concern is not with individual words, however, but with global properties of the lexicon, frequency effects on specific items do not figure centrally in our study, nor does the question whether words are cognitively represented as exemplars or otherwise. In that respect, our study also differs from statistical investigations of stress pattern diversity such as, for example, Domahs et al. (2014), which derive the stress patterns assumed by English compounds probabilistically but quite precisely, from the interaction of a large variety of factors.

A recent study which is closely related to our own concerns – in terms of both approach and method – is by Sonderegger and Niyogi (2010, 2013). They approach the functionally motivated contrast between stress patterns among English nouns and verbs in terms of dynamical-systems modelling. Although we focus on stress pattern diversity that is *not* morpho-syntactically motivated, the results of our model are in many respects compatible with the predictions of Sonderegger and Niyogi.[6]

Our paper is structured as follows. First (in section 2), we discuss extant approaches to word stress and why they do not account for stress pattern diversity as attested in English. We argue that this is because established stress assignment algorithms take isolated lexical items as their input, although the structures they build are essentially rhythmic and rhythm is a property of phrases rather than words. We suggest an alternative conceptualisation of stress assignment, in which rhythmic well-formedness constraints apply on phrases rather than words, while words adopt the patterns that satisfy these constraints best – on average – in all the phrases in which they occur. In the same section, we also discuss some of the implications of a phrase-based theory of word stress and what kind of predictions can be derived on its basis. We show that, while it makes the task of predicting the stress patterns of

---

[4] In the central part of this study, we explain and justify our method in more detail, and compare it to possible alternatives as well.

[5] Rather than on the cognitive representation of linguistic knowledge.

[6] We discuss this in more detail below (see 4.4).

individual words highly complex, it allows one to derive hypotheses about the global distribution of stress patterns by means of mathematical models.

Section 3 prepares our research question for being modelled. It explains how phrase-based stress assignment can be construed as an evolutionary process in which constraints on phrase level rhythm select for a stable distribution of word stress patterns in the lexicon so that uttered phrases are maximally eurythmic. It motivates the choice of evolutionary game theory for modelling the dynamics of that process, deals with the abstractions that this requires, and discusses their implications for interpreting the game as a viable model of actual languages.

Section 4, which is the main section of this study, describes the game itself both from a linguistic and from a mathematical perspective and interprets its results. It derives the hypothesis that lexical phonotactics will license stress pattern diversity if the lexicon contains a large proportion of monosyllables, and demonstrates that this hypothesis qualifies as a potential explanation of the diversity attested in English. Section 5 shows that the diachronic predictions inherent in the model fit the actual evolution of English word stress as well, thereby corroborating the validity of the model. Section 6 finally summarizes our observations and points out possible directions for further research.

2 English word stress and some questions it (still) raises

English polysyllabic words have their stress patterns assigned lexically, so that knowing a word implies knowing which of its syllables to stress: in *ˈfather* it is the first syllable, in *iˈdea* it is the last one, in *aˈgenda* it is the penultimate one, etc. In the majority of cases, English lexical stress is also 'immobile'. The syllable that is lexically stressed normally emerges as prominent in all utterances of a word. Stress shifts in words such as *fifteen* – which is stressed on the last syllable in isolation and in expressions like *She's fifˈteen*, but on the first syllable in phrases like *ˈfifteen years* – affect only a minority of items (cf. also *Chinˈese* vs. *ˈChinese ˈWhispers*, *Berˈlin* vs. *ˈBerlin ˈWall*, *Prinˈcess* vs. *ˈPrincess ˈAnn*, or *ˌTennesˈsee* vs. *ˈTennesˌsee ˈair*, *ˌPennsylˈvania* vs. *ˈPennsylˌvania ˈLegisˌlature* etc.).[7]

[7] Note that apparent 'stress shifts' in pairs such as *ˈformal – forˈmal+ity*, *conˈstruct*$_V$ – *ˈconstruct+Ø*$_N$, or *ˈacid – aˈcid+ic*, do not really imply that stress is movable in words such as *formal*, *construct* or *acid*. Rather, the difference suggests that each of the members of the pairs represents a word in its own right and has its own stress pattern. This is also true for layered conceptualisations of the lexicon (e.g. in Lexical Phonology, cf. e.g. Kiparsky (1979), or stratal OT, which model derivational relations cyclically. Such models employ procedural

Although English stress patterns represent properties of lexical items, they are not unpredictable, but seem to be systematically related to other lexical properties, some of them phonotactic, some morpho-syntactic. Thus, for many English words the position of stress can be predicted if one knows the number and the weight of the syllables they contain, and whether they are nouns, verbs, adjectives, etc..

Extant theories of word stress provide many proposals for formalizing these relations, and they differ from theory to theory (see, again, e.g. Anderson 1986; Burzio 1994; Chomsky & Halle 1968; Fudge 1984, 1999; Giegerich 1985; Halle 1998; Hayes 1982; Hyde 2007, 2011; Kager 1989; Kiparsky 1979; Liberman & Prince 1977; Marvin 2003; Poldauf & Lee 1984; Sainz 1988; Schane 2007; van Oostendorp 2010). What most proposals have in common, however, is that they consider lexical items in isolation. In order to capture that some aspects of a word's structure imply others, they take the former to be 'underlying' and the latter to be derived (by rules or ranked constraints). Always, the stress pattern that is derived for a word results from (a) its assumed phonotactic and morphological structure and its word class, and (b) whatever algorithm a model happens to use for building output structures. Thus, the stress pattern a word receives is fully implied in its own lexically specified phonotactic and morpho-syntactic properties.

Although extant models have reached high levels of sophistication and adequacy, none of them has managed to fit the facts fully. In particular, there is a non-negligible number of English words that are fully equivalent in terms of morphological and syllabic structure, syllable weight, and belong to the same morph-syntactic categories, but are nevertheless stressed in different ways. Some examples are

(3) [ˈ**len**.til]$_N$ vs. [ho.ˈ**tel**]$_N$
[ˈ**al**.ge.bra]$_N$ vs. [ven.ˈ**de**.tta]$_N$
[ˈ**Ca**.na.da]$_N$ vs. [ba.ˈ**na**.na]$_N$
[ˈ**con**.cert]$_N$ vs. [con.ˈ**cern**]$_N$
[ˈ**en**.voy]$_N$ vs. [i.ˈ**dea**]$_N$
[ˈ**stu**.pent]$_{Adj}$ vs. [stu.ˈ**pend**]$_{Adj}$

Furthermore, also individual words are sometimes attested with two different stress patterns, such as the proper stress doublets in

metaphors for representing what are in effect static relationships. Also on their basis, however, word internal stress patterns are not changeable post-lexically anymore once they are lexically derived.

(4) [ˈ**re**.search]$_N$ vs. [re.ˈ**search**]$_N$
[ˈ**per**.fume]$_V$ vs. [per.ˈ**fume**]$_V$
[ˈ**ac**.cess]$_V$ vs. [ac.ˈ**cess**]$_V$
[ˈ**de**.fense]$_N$ vs. [de.ˈ**fense**]$_N$
[ˈ**a**.zure]$_{Adj}$ vs. [a.ˈ**zure**]$_{Adj}$
[ˈ**ro**.bust]$_{Adj}$ vs. [ro.ˈ**bust**]$_{Adj}$

It is obviously impossible to accomodate word pairs like the ones in (3) and (1) in a model that derives stress patterns from the phonotactic structure and the word class of isolated lexical inputs. There is simply no way in which a deterministic algorithm that predicts a specific stress pattern for a specific type of lexical input can at the same time predict another one without losing its consistency.

The question is what this means. One interpretation would be that stress placement distinguishes between 'regular' and 'irregular' words, just as past tense formation does among verbs (regular *–ed* vs. irregular *swim-swam, put-put*, etc.), or plural formation among nouns (regular *–es* vs. irregular *sheep-sheep*, *man-men*, or *cactus-cacti, etc*). From this perspective, one can restrict oneself to identifying productive rules of stress assignment, while acknowledging that not all words reflect them. Irregular patterns do not have to be considered as a problem. This view is, in fact, widely taken (e.g. in Hayes 1982), and finds support in the fact that at least some of the 'irregular' stress patterns occur in loan words like *hotel*, *vendetta* or *banana*. While representing a possible way of dealing with diversity among stress patterns, however, the acknowledgement that not all words are regularly stressed fails to explain why that should be the case.

This paper attempts to do so. Since stress pattern diversity as attested in (3) and (1) cannot be derived by algorithms that apply on isolated lexical items, a different way of accounting for it is required. As indicated, we pursue the hypothesis that word stress does not primarily reflect properties of words themselves, but instead represents an adaptation to constraints that apply on phrases rather than words. There are several reasons why this hypothesis appears promising.

First, consider occasional phrase level stress shifts in cases such as *fifˈteen* vs. *ˈfifteen ˈyears*, *Chinˈese* vs. *ˈChinese ˈwhispers*, *Berˈlin* vs. *ˈBerlin ˈWall*, *Prinˈcess* vs. *ˈPrincess ˈAnn*, *ˌTennesˈsee* vs. *ˈTennesˌsee ˈair*, *ˌPennsylˈvania* vs. *ˈPennsylˌvania ˈLegisˌlature* etc. The type of stress pattern diversity created by them has received considerable attention, and the agreement is that it reflects a conflict between lexically assigned stress and rhythmic well-formedness constraints on phrases (cf. e.g. Hayes 1984; Kager 1989; Kiparsky 1979). While such shifts clearly represent positive evidence that phrase level rhythm can affect word stress, the argument does not work the other way. In particular, it would be wrong to conclude that words whose stress never shifts (i.e. the majority) are immune to rhythmic phrase level constraints. It is equally possible, and in fact more likely, we think, that their lexical stress patterns already reflect these constraints so well that stress shifts are not required.

The plausibility of this interpretation is corroborated by the fact that the structures built by established stress assignment algorithms are in fact rhythmic units. Thus, consider a rule that figures in most accounts of English word stress. Although it comes in various versions and contexts, the following representation (based on Hayes 1982: 238) is as good as any.

(5) English Stress Rule (ESR):

> Starting from the end of the word, move leftwards and stress the first syllable if it is heavy (i.e. if it ends in VV or VC). If it is not, stress the next one, irrespective of its weight.

In combination with so-called extrametricality conventions (by which word final consonants and sometimes word final syllables come to be ignored – more on them below), rule like this derive stress adequately for many English items, such as

(6) *staˈlac~~tite~~, aˈgen~~da~~, ˈalge~~bra~~; inˈhabi~~t~~, conˈtai~~n~~*

Essentially, the rule builds left headed feet that are binary either in terms of syllables (as in ˈ[*happy*]$_{\Sigma}$, ˈ[*alge*]$_{\Sigma}$ ~~*bra*~~*, or in*ˈ[*habi*]$_{\Sigma}$ ~~*t*~~) or in terms of moras (as in [*cool*]$_{\Sigma}$, *sta*ˈ[*lac*]$_{\Sigma}$ ~~*tite*~~*,* or *a*ˈ[*gen*]$_{\Sigma}$ ~~*da*~~). This is even more explicit in functionally corresponding OT constraints, in which the term 'foot' is deliberately used:

(7) Foot Binarity (FTBIN): Feet are binary at some level of analysis (μ, σ).

(Prince & Smolensky 2002: 50)

That word structure should at all be rhythmic, however, is puzzling, because words cannot establish rhythm by themselves (cf. Vennemann 1986). After all, rhythm is "characterized as the repetition of patterned sequences of elements, often varying in prominence." (Fraisse, 1974, 1982, quoted in in Hay & Diehl 2007: 113), and words are normally too short for sequences to be repeated in them. Repetition cannot occur below the phrase level. While Vennemann expressed this already in 1989, it is being increasingly recognized, and Lee & Gibbons, for instance, report

> converging evidence that the principle of rhythmic alternation applies across words, [and] that the domain within which the principle applies is the phonological phrase [...]. (2007: 448)

Thus, if English words appear to build trochees, the reason is probably that they do this in response to constraints on the phrases they form when combined.

A third argument involves the phenomenon of 'extrametricality'. For predicting English word stress adequately, for example, rules or constraints like those in (5) or (1) are not enough. Additionally, they require a set of conventions by which final consonants, as well as the final syllables of English nouns and some adjectives are made 'invisible' to stress assignment algorithms, i.e. 'extrametrical' (cf. e.g. Liberman & Prince 1977, Hayes 1982, or Hyde 2007 and 2011).

Thus, in verbs such as *torment* and *inhabit* – for which the stress rule in (3) by itself would predict *torˈment* (correctly) but *inhaˈbit* (falsely) –, the final /t/s are 'extrametrical', i.e. they don't count. This leaves the structures *tor.men* and *in.ha.bi*. Since the last syllable in *in.ha.bi* is now light, stress rules skip it, and stress the next one, i.e. *ha*, yielding *inˈhabit*, as required. – Additionally, in nouns and a number of (typically derived) adjectives also all final syllables are extrametrical. Thus, in words such as *agenda*, *diplomat*, *adjectival*, or *nominal*, stress rules see only *a.gen*, *di.plo*, *a.djec.ti*, and *no.mi*. On these structures they work normally, stressing the heavy final *gen* in *a.gen*, but not the light final *plo* in *di.plo*, etc., yielding, as expected, *aˈgenda*, *ˈdiplomat*, *adjecˈtival*, and *ˈnominal*.

While the usefulness of extrametricality conventions has come to taken more or less for granted, however, the question of why they should be required at all has received little attention. As soon as one begins to pursue it, however, one is invariably directed to the phrase level once again.[10]

Take word final consonants, for example. Normally, a syllable that ends in a consonant counts as heavy, and there is no obvious reason why a rule like (5), which is sensitive to weight, should ignore it. In phrases, however, a word final consonant may come to be followed a vowel – as in *inhabit every continent,* for example. In such cases, it may come to be re-syllabified, yielding, for example, [ɪn . hæ . <u>bi . tev</u> . rɪ] from [ɪn . hæ . <u>bit . ev</u> . rɪ]. This reduces the weight of the word final syllable, rendering it light in the case of *inhabit*. At the same time, the first syllable of *every* is stressed, and figures as a foothead in phases like ˈ*every* ˈ*continent*. Thus, if **inhaˈbit* were stressed on the final syllable, phrases could arise in which a syllable that is light by re-syllabification (in this case [bi]) form a monosyllabic foot at the same time, as in *[ˈɪn . hæ . <u>ˈbi . ˈtev</u> . rɪ]. Such a sequence is ungrammatical in English and would be highly arrhythmic, and its occurrence is effectively prevented by the strategy of disregarding final consonants in word stress assignment. Thus, the otherwise puzzling convention of final-consonant extrametricality serves a clearly definable purpose, and since it serves it so well, it is plausible to conclude that it is actually there for that purpose (cf. Dennett 1987). This suggests that the function of

[10] Thus, already Heinz Giegerich (1985) suggested that the 'extrametricality' assumptions required to make about the right edge of words might reflect the impact of their potential phrase level contexts.

lexical stress assignment algorithms in general is to adjust word structure for their use in rhythmically structured phrases and that the constraints they reflect apply on the phrase level.

A similar argument can be made for the apparent 'extrametricality' of final syllables in English verbs. It seems to be the case that since Middle English times verbal stems have been followed by syllabic inflectional endings significantly more often than nominal stems (see Ritt 2012a: 163), and for Modern English, Kelly (1988 & 1989) has demonstrated this as well. If verbal roots are systematically followed by one more unstressed syllable than nouns, however, they will also be further away, on average, from the next potential foot head than nominal roots. If phrases are subject to rhythmic well-formedness constraints, this means that feet will be preferred to be – at least roughly – equally long in terms of syllables. In order to achieve this, of course, it makes perfect sense to stress verbal roots further to the right than nominal roots. Since this is what final-syllable extrametricality in verbs achieves, it is once again reasonable to assume that it is there for that purpose: it helps words to anticipate and to satisfy a rhythmical constraint that applies on the phrase level, in this case the preference for feet to be isochronous.

In sum, there is considerable, if indirect, evidence that some of the most important constraints determining words stress pattern in the English lexicon apply on phrases, and that words reflect them only indirectly because they anticipate their effects and adopt the structures best adapted to them.

The next question is, however, what predictions the acknowledgement that word stress patterns are likely to reflect phrase level constraints produces. An obvious one is that the assignment of stress on any specific word must depend on the stress patterns assumed by the words in its phrasal context, and that the same necessarily applies to those words as well. Therefore, a theory of phrase-based word-stress assignment will have to take mutual interdependencies between items into account. This means that for predicting the stress pattern of any single word one needs – at least in principle – to consider all lexical items in a language, as well as the frequency with which each of them co-occurs with the any of the others. This renders the task of predicting the stress patterns of individual items computationally too complex to be practicable. On the other hand, however, computational tools such as agent-based simulation, or evolutionary game theory, are specifically designed for modelling the dynamics of populations in which the historical stability of their members depends on their interaction. While such models say little about the fates of individuals, they allow one to derive predictions about the relative frequencies of subpopulations. Since the distribution of stress patterns in the lexicon can be construed in terms of subpopulations and their relative frequencies, it is clearly addressable by means of mathematical models and on the assumption that words adopt their stress patterns by interacting with one another under rhythmic constraints that apply on the phrases in which they interact.

3 An **evolutionary approach to lexical stress pattern distribution**

The general problem we address is whether there are conditions under which words of a single type – for example, the type *robust*, i.e. monomorphemic adjectives with a superheavy final syllable – will assume more than a single type of stress pattern.[11] For making the problem addressable by evolutionary game theory, we conceptualise it in evolutionary terms. For the purpose, we take a strictly replicator-based view. Following in this respect Croft (2000, 2013), Dawkins (1976), Pierrehumbert (2012), or Ritt (2004), we think of the lexicon as a population of transmittable constituents, which acquire evolutionary stability if they are faithfully transmitted through communication and language acquisition. In discourse, words combine to form phrases and these phrases are subject to rhythmic well-formedness constraints. The better and the more often the expression of a word satisfies these constraints, the fitter, or the more evolutionarily stable its stress pattern will be. Since phrases normally involve more than a single word, the fitness of the stress pattern on any word therefore depends not only on the rhythmic well-formedness constraints on phrases as such, but also on the stress patterns assumed by the words it combines with.

To see what we mean by this, consider a hypothetical language that consists only of disyllabic CVCV items, so that any utterance will be a sequence of $[CVCV]_{Wd}$ $[CVCV]_{Wd}$ $[CVCV]_{Wd}$ $[CVCV]_{Wd}$ and so on. Assume, for the sake of the argument, that there is only a single criterion of rhythmical well-formedness: a strictly alternating rhythm is good, and any deviation from strict alternation is bad, no matter if it involves clashes (σˈσ ˈσσ) or lapses (ˈσσ σˈσ). It is easy to predict that in such a language all words will wind up either as trochees or as iambs, because as soon as a majority of word tokens assumes one of the two patterns, the other one will produce suboptimal rhythmic structures more often than not. Thus, words will eventually be forced into absolute conformity.

What is also easy to see, however, is that predictions will be more difficult to make as soon as things become even slightly more complex. What distribution of stress patterns among lexical items would a preference for alternating rhythm predict, for example, for a language that includes not only disyllables, but also lexical monosyllables (which normally carry stress) and monosyllabic function words (which are

[11] As has been pointed out, this question is different from the question what stress pattern the specific adjective *robust*, or indeed any specific word, will assume. The latter problem is much more adequately addressable by means of statistical analyses of large bodies of usage data in which a large variety of factors can be considered (see e.g. Domahs et al. (2014), or Kelly (2004)).

usually expressed as rhythmic dips)? Question like that have no intuitive answer, and require formal modelling.

One of the tools for modelling interactions of this type is evolutionary game theory.[12] While it is not the only tool available for our problem, it is highly transparent and compares favourably in this respect to agent-based computer simulations. At the same time, it puts considerable constraints on the number of variables that can be taken into account and requires problems to be reduced to their very essence. Otherwise, they become intractable by analytical modelling. While computer simulations are more powerful in this respect, the very general way in which our problem can be formulated allows us to choose the more transparent option.

Since we are interested in a principled answer to a general question, we do not need to model a specific language. Instead, we can derive predictions from our hypothesis by modelling a highly idealized artificial mini-language that is simplified in various respects. It contains only monosyllabic and disyllabic items. Major class items can be monosyllabic or disyllabic and will always carry lexical stress. Function words, which are always monosyllabic, never do. Among disyllabic major class items no further distinctions are made. All of them have the same phonotactic structure and can be used as nouns, verbs, adjectives, and do not differ in informativity or stylistic value. The phrases that our game considers involve disyllabic items that occur either in immediate neighbourhood of one another, or are separated by one monosyllabic item, which may either be a lexically unstressed function word, or a lexically stress major class word.[13]

The only items in our language that have a choice with regard to their stress pattern are major class disyllables. Since they are all of the same phonotactic, morphological and syntactic type, they represent a completely uniform population in that respect. We also assume no lexical solidarities of any kind. Instead, all disyllables are equally likely to occur with any of the others.

---

[12] Some previous applications in linguistics are referred to in section 4.

[13] Although the syntax of our mini language allows them, we disregard phrases in which disyllables are separated by longer sequences of monosyllables, because we consider the disyllables as too far removed to exert a strong rhythmic influence on one another.

As far as the rhythmic quality of utterance sequences is concerned, we employ only a single criterion, requiring strict alternation of stressed and unstressed syllables.[14] Any deviation from that ideal pattern will be considered suboptimal.

In any single round of the game, two disyllabic 'players' meet in one of the three syntactically possible contexts, i.e. either next to one another or separated by a single lexical or functional monosyllable. Each disyllabic player can assume either initial or final stress as its 'strategy'. Then, the rhythmic quality of the resulting stress pattern is evaluated, and a 'payoff' is accordingly distributed among the players. That payoff then determines the 'fitness' (and thereby the 'evolutionary stability') of the chosen stress pattern. Then, evolutionary game theory allows us to calculate which stress patterns, or 'strategies' will become stably established in the population of disyllables. In principle, there can be the following outcomes to the game: (a) all disyllables choose initial stress, (b) all disyllables choose final stress, or (c) a specific mix of initially and finally stressed words turns out to be stable. Clearly, the relevant question is whether and under what conditions outcome (c) is produced.

Note that, for the purposes of our discussion, the simplicity of the artificial language we have constructed represents a clear advantage, because it allows us to be certain that the evolutionary dynamics our model predicts result from no other factors than the ones we have actually modelled. In contrast, real language usage always reflects a large number of diverse factors, some system internal, others not. For example, transitive verbs might prefer final rather than penultimate stress because they are more likely than intransitives to be followed by a noun phrase beginning with an unstressed article (see Fullwood 2014). Likewise, the stress pattern on a specific word might be due to lexical solidarities or quasi-formulaic sequences it is involved in. Thus, initially stressed ˈ*research* might be motivated by the frequency of such phrases as ˈ*research* ˈ*exercise* or ˈ*research* ˈ*outcome*, or the rise of initially stressed ˈ*hotel* might reflect the popularity of the song ˈ*Hotel* ˌ*Cali*ˈ*fornia* by The Eagles, or it might be the case that the unexpected final stress in words such as *ca*ˈ*fé*, or *set*ˈ*tee* is motivated by a desire to mark them, stylistically, as foreign and thus 'special'. Also, in natural languages, lexical items hardly ever fully equivalent to one another even in terms of their phonotactic make up. Instead they are usually distinguished by fine grained differences in terms of syllable weight, segmental structure, etc. The point is that our model prevents such factors from interfering. In a sense, abstracting them is a radical way of controlling for them, and allows us to isolate our central question: Are there circumstances under which

---

[14] In Optimality Theory, this criterion involves two constraints, namely *CLASH: No adjacent strong beats on the grid, and *LAPSE: No two adjacent unstressed syllables. See also Alber (2005); Fullwood (2014); Kager (2005). Prince (1980)

the interaction of words in phrases that are under rhythmical well-formedness constraints makes the emergence of stress pattern diversity among words of the same type inevitable – even if they are identical *in all other respects*?

As we demonstrate in the next section, such circumstances do indeed seem to exist. As our game suggests, stress pattern diversity will be established among polysyllables if the number of monosyllables exceeds a certain threshold. It also predicts, more specifically, that the relative proportions of final and initial stress depends on the relative frequencies of lexical and functional monosyllables.

4 The stress game: structure and analysis

We address our problem in terms of evolutionary game theory (henceforth EGT, Maynard Smith & Price 1973, Smith 1982), modelling, in particular, 'replicator dynamics' (see e.g. Hofbauer & Sigmund 1998). Applying EGT to problems in diachronic phonology or morpho-syntax is not new *per se*. For example, Nowak and colleagues (Nowak & Komarova 2001, Nowak et al. 2002, Mitchener 2003, Mitchener & Nowak 2004, Nowak 2006) have studied the replicator dynamics of generative grammars in detail. In their models entire systems are construed as replicators that compete for existence in a population of speakers. Also Yang (2000) and Niyogi (2006) have proposed inherently generative dynamical-systems models based on various learning algorithms.

On the other hand, item based replicator models have been developed less frequently, although the concept of linguistic replicators or "linguemes" (Croft 2000) has gained currency during the last two decades (see also Ritt 2004, Jäger & Rosenbach 2008, Baxter et al. 2009, McCrohon 2012).[15] A notable exception is Jäger (2008), investigating the dynamics of phonemes in the vowel space in terms of EGT. He construes exemplars, i.e. stored phonetic events (cf. Pierrehumbert 2001, Wedel 2006) as linguistic replicators.[17] A more general, but also item-based model derived from EGT is Nowak's model of word dynamics (Nowak 2000, Nowak et al. 2000, Solé 2011). However, in this model linguistic replicators

---

[15] The concept is not uncontested, however. For a critical discussion see, for example, Smith (2012) and the contributions by Jäger's (2008b).

[17] Interestingly, Jäger's results agree remarkably well with results of agent based models (de Boer 2000, 2002; see also Mühlenbernd & Wahle 2016).

reproduce independently of one another, so that that it cannot simulate the interactions that we are interested in.

Not all dynamical-systems approaches to phonological change are based on EGT, of course. Wang et al. (2004), for example, study Lotka-Volterra type dynamics in the phonological evolution of a lexicon (thereby including the dynamics of stress patterns implicitly). Their model could, in principle, take word co-occurrence on the utterance level into account, but since they are primarily interested in the so-called 'snowball effect', i.e. the accelerating spread of phonological change through the lexicon, they do not address the problem of stable diversity (see however Sherman's 1975 related discussion of the diffusion of noun-verb stress alternation). – A dynamical-systems model that does deal specifically with stress pattern diversity, on the other hand, has been suggested by Sonderegger and Niyogi (2013). It is speaker-based rather than item-based, and models the effects of 'mistransmission', i.e. the probability of an originally finally stressed word to be re-categorized as initially stressed, and *vice versa*. The model indeed predicts a stable mix of stress patterns.[18]

In short, the methods employed in our study are well established in linguistic research, but have so far not been applied to the specific problem we discuss here. In the following, we briefly introduce EGT (4.1), and describe the specific game we have developed (4.2). In 4.3 we analyse the game and interpret it.

4.1 Evolutionary game theory

Evolutionary game theory is an extension of game theory developed for studying the dynamics of strategy distributions among populations in a series of abstract games (Maynard Smith & Price 1973, Hofbauer & Sigmund 1998, Nowak 2006). In a single game, two players meet, select strategies and interact. Depending on the combination of strategies, both players receive a payoff. In traditional game theory, the set of possible strategies is finite, and the payoff is a pre-defined function allocating numerical values to strategy combinations.

Like all mathematical models, games are neutral regarding their empirical interpretation. Thus, in a game involving two strategies, players could, for example, be animals choosing between the strategies 'fight' and 'flee'), fund managers opting to either 'buy' or 'sell' stock, or, as in our case, words that assume either 'initial' or 'final stress'.

[18] More on differences between Sonderegger & Niyogi's model and ours is said below (4.3)

Games can be visualized in payoff matrices. In the case of two strategies (S_1 and S_2) that are both available to both players, the payoff matrix A reads

$$\mathrm{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \tag{F.1}$$

i.e. both players get payoff a_{11} if they choose S_1; if a player using strategy S_1 meets a player using strategy S_2, the former gets a_{12} while the latter gets a_{21}; and if both players choose S_2, both obtain a_{22}.

A special property of EGT is that its focus is on player populations that play a series of games rather than on individuals encountering each other once. Individual players in the population are taken to meet one another randomly, always sticking to their chosen strategies. In EGT payoff is then interpreted as reproductive success or fitness (Nowak 2006: 49). Crucially, players do not actually pick strategies. Instead, they inherit them from their progenitors, and pass them on to their offspring. As a result, players who receive more payoff than their opponents produce more descendants, and the relative frequency of their strategy in the population rises.

Since individuals are fully characterized by their strategies, populations can be divided into disjoint subpopulations, one for each strategy. Assuming a very large population, we denote the proportion of S_1 players by x and the proportion of S_2 players by $y = 1 - x$. In the context of this paper, x denotes the proportion of initially stressed disyllables, and y the proportion of finally stressed ones.

Box 1: Replicator dynamics

Replicator dynamics (Hofbauer & Sigmund 1998: 67f.; Nowak 2006: 45f.) model frequency dependent selection (Heino et al. 1998) as a dynamical system in continuous time, in which the rates of change of the frequencies x and y, i.e. their increase or decrease per time unit, are given by

$$\dot{x} = x(f_1(x,y) - \phi(x,y)) \tag{F.2a}$$

$$\dot{y} = y(f_2(x,y) - \phi(x,y)) \tag{F.2b}$$

where $f_1(x,y) = a_{11}x + a_{21}y$ and $f_2(x,y) = a_{12}x + a_{22}y$ denote the fitness of S_1 players and S_2 players, respectively, and $\phi(x,y) = xf_1(x,y) + yf_2(x,y)$ denotes the average fitness for a given composition of frequencies (x,y). The growth rates of both player types are determined by (a) the frequency of individuals adopting a strategy and (b) the difference between strategy specific fitness and average fitness.

Since $y = 1 - x$, it is sufficient to restrict the model to the frequency of S_1 players. For a 2×2 payoff matrix $A = (a_{ij})$ the two-dimensional system above can be reduced to

$$\dot{x} = x(1-x)\big((a_{11}+a_{22}-a_{12}-a_{21})x + a_{12}-a_{21}\big).$$

Dynamical equilibria, i.e. situations in which the distribution of strategies among players does not change in one or the other direction, are of particular interest. Mathematically, this is the case if $\dot{x} = 0$. Then the population finds itself in an equilibrium, which shall be denoted by $\hat{x}$. For the equation above there are three potential equilibria: two 'pure' ones $\hat{x}_1 = 1$ (all players play strategy S_1, e.g. all words take initial stress) and $\hat{x}_2 = 0$ (all players play strategy S_2, e.g. all words take final stress), as well as an 'internal equilibrium'

$$\hat{x}_{\text{int}} = \frac{a_{22}-a_{12}}{a_{11}+a_{22}-a_{12}-a_{21}} \tag{F.3}$$

as long it is positive. In that configuration, some words take initial stress and others final stress. A system at equilibrium will not change by itself. It may do so, however, if it is perturbed by some system external factor. Whether or not it does depends on its 'stability'. In replicator dynamics, stable equilibria are referred to as 'evolutionarily stable strategies' (ESS).

Assessing the stability of of $\hat{x}_1$, $\hat{x}_2$ and $\hat{x}_{\text{int}}$ is simple: If $a_{11} > a_{21}$ and $a_{12} > a_{22}$, then $\hat{x}_1$ is stable, $\hat{x}_2$ is unstable, and there is no internal equilibrium $\hat{x}_{\text{int}}$. The dynamics result in a population of S_1 players only (i.e. initial stress throughout the lexicon). If $a_{11} < a_{21}$ and $a_{12} < a_{22}$, then $\hat{x}_1$ is unstable while $\hat{x}_2$ is stable. This leads to a population consisting exclusively of S_2 players (i.e. final stress throughout the lexicon). If $a_{11} > a_{21}$ and $a_{12} < a_{22}$, both $\hat{x}_1$ and $\hat{x}_2$ are stable, and there exists an internal equilibrium that is unstable. The dynamics will eventually lead to a population consisting of either of the two player types (i.e. the lexicon will eventually include only words with initial stress, or only words with final stress). Finally, a stable internal equilibrium $\hat{x}_{\text{int}}$ exists, if $a_{11} < a_{21}$ as well as $a_{12} > a_{22}$. In that case it is always better for both players to have opposing strategies. Then, the dynamics will inevitably result in a stable mix of S_1 players and S_2 players (i.e. a lexicon with some words stressed initially, and others stressed finally). The fraction of S_1 players (i.e. the proportion of words with initial stress) in this stable equilibrium depends on the quotient above.[19]

Changes in these proportions over time are modelled in terms of the replicator dynamics (see Box 1). For games with two strategies, the long term evolution of the population can be predicted from the payoff matrix A. In particular, the conditions for a stable mix of S_1 players and S_2 players can be

[19] Note that the long term dynamics of an evolutionary game are fully determined by the entries in the payoff matrix.

determined: if $a_{11} < a_{21}$ as well as $a_{12} > a_{22}$, i.e. if pairs of words benefit from having different stress patterns, the replicator dynamics will converge on a stable internal equilibrium $\hat{x}_{\text{int}}$, with a fraction of $\hat{x}_{\text{int}}$ S_1 players and a fraction of $1 - \hat{x}_{\text{int}}$ S_2 players.

In the context of our discussion, the most important question is whether and under what conditions a stable internal equilibrium is produced in which two different strategies (i.e. two different stress patterns) coexist. – It is worth pointing out, however, that our game will predict more than just that.

Since it predicts the evolutionary trajectory of fraction x deterministically (on the basis of system internal selection processes), our model also predicts, for each initial distribution of strategies, how it will look after a given time. This means furthermore that if, at any point in time, we have reason to assume a change in x brought about by factors that are not captured by the model itself, the model predicts how the fraction x will evolve after that change has occurred. In the case of stress patterns, system external factors that affect their relative frequencies in the lexicon, may include language contact, and the incorporation of foreign vocabulary in particular. We come back to that below.

4.2 Formulation of the stress game

As specified in Section 2.3 the players in our model are major class disyllables of the form [σσ]. They can adopt either initial stress [ˈσσ] or final stress [σˈσ]. Thus, the populations of words can be divided into players adopting strategy [ˈσσ] (*S₁*), and players adopting strategy [σˈσ] (S_2).

(8) Strategy S_1 ‘initial stress’: [ˈσσ]
    Strategy S_2 ‘final stress’: [σˈσ]

Additionally, our model includes lexical monosyllables, which are stressed, i.e. [ˈσ], and functional ones, which are not, i.e. [σ]. Since they have no choice regarding their stress pattern, they do not represent players either. That we do not consider trisyllabic items, or longer ones for that matter, does not strike us as particularly problematic. One the one hand, the point of our exercise is to prove a principle, and on the other hand, the number of words of more than two syllables is also small in English, where roughly 92.76% ($\pm .02$, 95% confidence interval) of all utterance tokens are either monosyllabic or disyllabic (see Figure 1a; calculations based on the CELEX database, Baayen et al. 1995).

In our game, two players meet to form sequences (Figure 1b). Since our hypothetical language consists of lexical disyllables [σσ], lexical monosyllables [ˈσ], and functional monosyllables [σ], we consider three contexts, in which two players, i.e. two lexical disyllables, can potentially meet:

(9) Context C_1: [σσ][σ][σσ] (ˈ*father and* ˈ*mother*, ˈ*Susan and Mi*ˈ*chelle*, *i*ˈ*deas of* ˈ*grandeur*, etc.)

Context C_2: [σσ][ˈσ][σσ] (ˈ*father* ˈ*loves* ˈ*mother*, ˈ*Susan* ˈ*hates Mi*ˈ*chelle*, *i*ˈ*deas* ˈ*kill* ˈ*people*, etc.

Context C_2: [σσ][σσ][20] (ˈ*father's* ˈ*lover*, ˈ*Susan* ˈ*worries*, *for*ˈ*get Mi*ˈ*chelle*, etc.)

Crucially, the contexts are taken to vary independently of the strategies chosen by the players (i.e. there are no rhythm based lexical solidarities). Instead, the relative frequencies of the contexts in which disyllables meet depends exclusively on the relative frequencies of lexical monosyllables and functional monosyllables in the language. That distribution is denoted by $\pi = (\pi_1, \pi_2, \pi_3)$, where $\pi_i \in [0,1]$ for all i and $\pi_1 + \pi_2 + \pi_3 = 1$. π_1 is the proportion of contexts involving a functional (unstressed) monosyllable, π_2 is the proportion of contexts involving a lexical (stressed) monosyllable, and in a proportion π_3 of all contexts, no monosyllabic item is involved (Figure 1c). Since the composition of π can be altered at will, it is possible to observe its effects on the stable distribution of stress patterns.

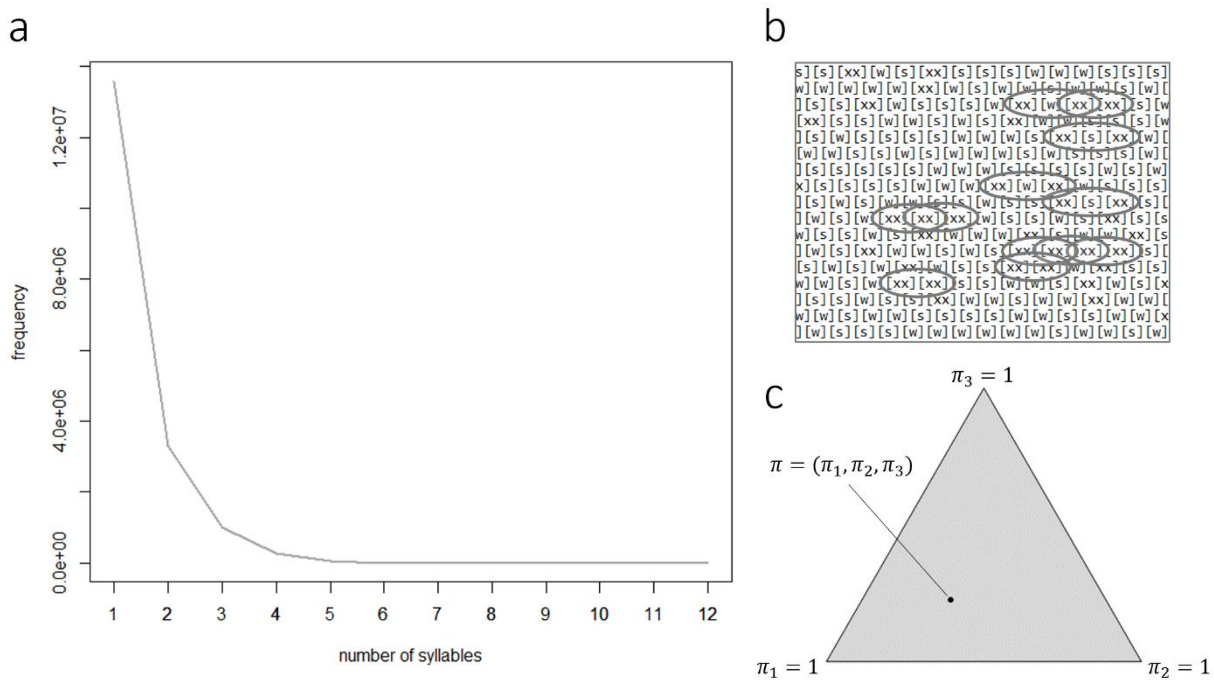

Figure 1. (a) Token frequency of words with different numbers of syllables in English (based on the CELEX database, Baayen et al. 1995). The distribution resembles a power law, where token frequency decreases as the number of syllables rises. Together, monosyllabic and disyllabic items account for about 92.72% of all tokens. (b) An artificial language as a sequence of weak (functional) monosyllables, strong (lexical) monosyllables and disyllables. Subsequences in which two disyllables are close to each other (either neighbouring or separated by at

[20] As indicated, we disregard sequences where disyllables are separated by more than one monosyllable. This keeps complexity of the model manageable, and reflects the plausible assumption that the rhythmic influence two disyllables exert on each other decreases with the distance between them.

most one monosyllable) corresponding to contexts C_1 to C_3 are encircled. (c) The distribution of contexts C_1, C_2 and C_3 as a point $\pi = (\pi_1, \pi_2, \pi_3)$ in the 2-simplex, represented by a triangle.

Lexical stress is taken to be immobile and does not shift to meet rhythmic phrase level constraints, so that both players enter the game with their strategies fixed in advance. Thus, the resulting sequences satisfy phrase level constraints to varying degrees. As indicated, the only constraint our model considers is the preference for strict stress alternation, reflecting *CLASH and *LAPSE, so that two types of violation can occur: sequences of two stressed syllables (clashes), and sequences of two unstressed syllables (lapses). Payoff is then determined by the number of violations in a sequence. For example, the sequence ˈ*robust reˈsearch* contains one violation (a lapse), while *roˈbust ˈstress ˈresearch* contains two (clashes). Thus, reward is modelled as a decreasing function φ of the number violations . We reward optimal rhythmicity (no violation) with $\varphi(0) = \varphi_{\max} > 0$, and attribute $\varphi(2) = 0$ to the rhythmically worst sequences that can occur in our game (i.e. sequences with two violations).

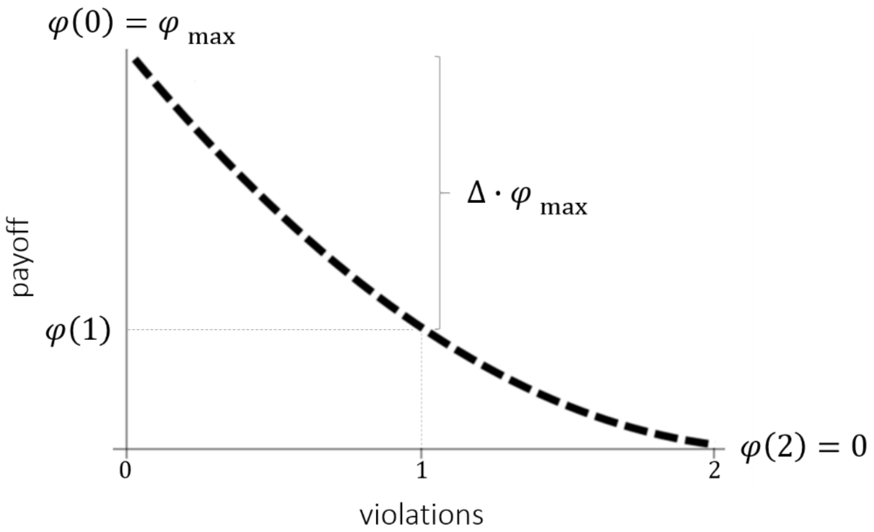

Figure 2. Rhythmicity as a decreasing function of the number of violations produced in a sequence of words. In this example, a single violation already decreases rhythmicity and hence the received payoff to a large extent, and two violations are less than twice as bad as one.

Note that these definitions leave open the question of how the occurrence of a single violation is to be rewarded, as it is not obvious that it should be half as bad as two violations. As will be seen, the issue is crucial for the interpretation of the game. If the first violation diminishes the rhythmic quality of the sequence more strongly than the second one, then $\varphi(1) < \frac{\varphi_{\max}}{2}$, as in Figure 2. This would be adequate if two violations count as less than twice as bad as a single violation. If, on the contrary, the second violation contributes more than the first one to the overall reduction of the rhythmic quality of a sequence, then $\varphi(1) > \frac{\varphi_{\max}}{2}$. In any case, we denote the normalized difference between no violation and one violation by

$$\Delta \coloneqq \frac{\varphi_{\max} - \varphi(1)}{\varphi_{\max}}.$$

Another issue is this: while rhythmicity is assigned to whole sequences, game theory demands payoff to be allocated to single players, and it is not at all self-evident that it should be divided equally, because the contributions of words to the rhythmic quality of a foot they build together may not be perceived to be equal by speakers. In English, for example, which counts as trochaic, feet are perceived to begin with a peak and extend to the right. Thus, in the case of both clashes and lapses, it is the left foot that is felt to be deficient, and when rhythm is repaired by shift, it occurs normally in the first word, not in the second one (i.e. ˈ*Pennsyl*ˌ*vania* ˈ*Legis*ˌ*lature* rather than ?ˌ*Pennsyl*ˈ*vania* ˌ*Legis*ˈ*lature*).[21] Thus, we divide payoff into two shares $\alpha \cdot \varphi(\cdot)$, and $(1-\alpha) \cdot \varphi(\cdot)$, where α lies within the unit interval. For the purpose of modelling stress placement in English, we attribute a larger share of the payoff to the first word, and a smaller share to the second one, so that $\alpha > 0.5$.[22]

For each encounter of two words we evaluate the payoff given to each of the two words for both possible orderings. Thus, when $[\sigma\sigma]_A$ and $[\sigma\sigma]_B$ meet, payoffs are first determined separately for the sequences $[\sigma\sigma]_A...[\sigma\sigma]_B$ and $[\sigma\sigma]_B...[\sigma\sigma]_A$, and then added together.

A single round of the game for fixed π, φ, and α unfolds as follows:

[21] For evidence that this is a language specific property see Kiparsky (1966).

[22] In principle, of course, α-values below 0.5 are possible as well, say, for applications of the model to languages other than English.

(10) **1:** Randomly draw two words $[\sigma\sigma]_A$ and $[\sigma\sigma]_B$.

2: Randomly draw a context C_i according to distribution π.

3: Build a sequence $[\sigma\sigma]_A$...$[\sigma\sigma]_B$ according to C_i.

4: Evaluate the payoff depending on the number of rhythmicity violations according to φ.

5: Attribute share α to $[\sigma\sigma]_A$ and share $1-\alpha$ to $[\sigma\sigma]_B$.

6: Repeat steps 4 to 5 for the reversed sequence $[\sigma\sigma]_B$...$[\sigma\sigma]_A$.

7: Add the payoff shares received by $[\sigma\sigma]_A$ and $[\sigma\sigma]_B$ together for each of them.

4.3 Evolutionary dynamics of the lexicon

Our game allows us to model the evolution of a lexicon whose items interact in a series of independent rounds as in (10). What we want to know is if and under what conditions a mix of initially stressed and finally stressed words will be evolutionarily stable. For that purpose, we calculate the payoff matrix for the stress game and analyse it in terms of EGT as outlined in Box 1.

As the entries of the payoff matrix depend on the distribution π, the rhythmicity function φ, and the weight α, the according payoff matrix can be expressed as

$$A^{\pi,\varphi,\alpha} = \begin{pmatrix} a_{11}^{\pi,\varphi,\alpha} & a_{12}^{\pi,\varphi,\alpha} \\ a_{21}^{\pi,\varphi,\alpha} & a_{22}^{\pi,\varphi,\alpha} \end{pmatrix}.$$

Box 2: Derivation of the payoff matrix

Since the payoffs for all entries in the matrix $A^{\pi,\varphi,\alpha}$ are determined in the same way, we demonstrate the procedure only for one of them. For that purpose we chose $a_{11}^{\pi,\varphi,\alpha}$, which represents the payoffs for an encounter of two initially stressed words $[ˈ\sigma\sigma]_A$ and $[ˈ\sigma\sigma]_B$.

In context C_1 the sequences $[ˈ\sigma\sigma]_A[\sigma][ˈ\sigma\sigma]_B$ and $[ˈ\sigma\sigma]_B[\sigma][ˈ\sigma\sigma]_A$ are formed. Both produce a single violation (a lapse) and incur a rhythmicity score of $\varphi(1) = \varphi_{\max} - \Delta\varphi_{\max}$. In the first sequence $[ˈ\sigma\sigma]_A$ receicves a payoff of $\alpha(\varphi_{\max} - \Delta\varphi_{\max})$ while $[ˈ\sigma\sigma]_B$ gets $(1-\alpha)(\varphi_{\max} - \Delta\varphi_{\max})$. In the second sequence, $[ˈ\sigma\sigma]_A$ gets $(1-\alpha)(\varphi_{\max} - \Delta\varphi_{\max})$ and $[ˈ\sigma\sigma]_B$ gets $\alpha(\varphi_{\max} - \Delta\varphi_{\max})$. In total, $[ˈ\sigma\sigma]_A$ receives $\alpha(\varphi_{\max} - \Delta\varphi_{\max}) + (1-\alpha)(\varphi_{\max} - \Delta\varphi_{\max}) = \varphi_{\max} - \Delta\varphi_{\max}$. The same holds for $[ˈ\sigma\sigma]_B$.

In context C_2 the sequences $[ˈ\sigma\sigma]_A[ˈ\sigma][ˈ\sigma\sigma]_B$ and $[ˈ\sigma\sigma]_B[ˈ\sigma][ˈ\sigma\sigma]_A$ are formed. Again, both produce a single violation (in this case a clash) and incur a rhythmicity score of $\varphi(1) = \varphi_{\max} - \Delta\varphi_{\max}$. As above, both $[ˈ\sigma\sigma]_A$ and $[ˈ\sigma\sigma]_B$ obtain $\varphi_{\max} - \Delta\varphi_{\max}$.

In context C_3 the sequences [ˈσσ]A [ˈσσ]B and [ˈσσ]B [ˈσσ]A are formed. None of the two produces a violation, so that both incur scores of $\varphi(0) = \varphi_{\max}$. Hence, both [ˈσσ]A and [ˈσσ]B obtain $\alpha\varphi_{\max} + (1-\alpha)\varphi_{\max} = \varphi_{\max}$.

Since the proportion of encounters in each of the three contexts reflects distribution π, the entry of the payoff matrix can be expressed as

$$a_{11}^{\pi,\varphi,\alpha} = \pi_1(\varphi_{\max} - \Delta\varphi_{\max}) + \pi_2(\varphi_{\max} - \Delta\varphi_{\max}) + \pi_3\varphi_{\max}.$$

After payoff determination (see Box 2) we find that

$$a_{11}^{\pi,\varphi,\alpha} = a_{22}^{\pi,\varphi,\alpha} = (\pi_1 + \pi_2)(\varphi_{\max} - \Delta\varphi_{\max}) + \pi_3\,\varphi_{\max},$$

$$a_{21}^{\pi,\varphi,\alpha} = \pi_1\alpha\varphi_{\max} + \pi_2(1-\alpha)\varphi_{\max} + \pi_3\alpha(\varphi_{\max} - \Delta\varphi_{\max}) \text{ and}$$

$$a_{12}^{\pi,\varphi,\alpha} = \pi_1(1-\alpha)\varphi_{\max} + \pi_2\alpha\varphi_{\max} + \pi_3(1-\alpha)(\varphi_{\max} - \Delta\varphi_{\max}).$$

By inserting the entries $a_{ij}^{\pi,\varphi,\alpha}$ $(i,j = 1,2)$ into formula F.3 (Box 1), the replicator dynamics can be shown to exhibit an internal equilibrium, which can be simplified as

$$\hat{x}_{\text{int}} = \frac{\alpha + \pi_2 - 2\alpha\pi_2 - (1 - 2\pi_3 + \alpha\pi_3)\Delta}{1 - (2 - 3\pi_3)\Delta}. \quad \text{(F.2)}$$

Notably, the size of the internal equilibrium no longer depends on the maximal reward $\varphi_{\max}$ but only on directly interpretable variables. It is stable, i.e. an internal ESS, if $a_{11}^{\pi,\varphi,\alpha} < a_{21}^{\pi,\varphi,\alpha}$ as well as $a_{12}^{\pi,\varphi,\alpha} > a_{22}^{\pi,\varphi,\alpha}$ (Box 1). By replacing these variables with the corresponding terms from the above derived payoff matrix and solving both equalities for π_3, it can be shown that x_{int} is indeed an internal ESS as long as

$$\pi_3 < \frac{\pi_1\alpha + \pi_2(1-\alpha) - (\pi_1 + \pi_2)(1-\Delta)}{1 - (1-\alpha)(1-\Delta)} \quad \text{(F.5)}$$

as well as

$$\pi_3 < \frac{\pi_1(1-\alpha) + \pi_2\alpha - (\pi_1 + \pi_2)(1-\Delta)}{1 - (1-\alpha)(1-\Delta)} \quad \text{(F.6)}$$

is fulfilled. This means that the proportion of sequences in which two disyllables follow each other immediately defines a threshold for the existence of internal ESSs. In other words, stress pattern diversity among disyllables will only be stable if there is a sufficiently high number of monosyllabic items. If the above conditions are fulfilled, however, it is inevitable.[23]

While the prediction that stress pattern diversity depends on the number of monosyllables is perhaps the most striking result of our simulation, it has other interesting implications as well. Thus, its prediction of inevitable stress pattern diversity seems to depend, also, on two other variables. First, it depends on the distribution of payoff shares among two the two disyllables in a phrase (already discussed above on page 22), and secondly, the relative weighting of the number of violations of rhythmic wellformedness in a phrase, i.e. on the question of how much worse two violations of rhythmic well-formedness conditions should count in comparison to a single one. This dependence can be demonstrated in the following way.

F.5 and F.6 above define a region in the 2-simplex, in which the internal ESS can be located. Its size depends on the syntagmatic weighting parameter α as well as on Δ, illustrated in Figure 3. In this composite plot, the horizontal dimension corresponds to the weight α which determines the payoff distribution between the first and the second disyllable. As can be seen in the composite plot, α determines the slope of the surface of internal ESSs. This is also evident by taking the directional derivative of $\hat{x}_{\text{int}}$ (F.4) with respect to π_2, i.e. the ratio of constellations involving lexical monosyllables, assuming that π_3 is fixed. Then, for some strictly positive constant C,

$$\frac{\mathrm{d}\hat{x}_{\text{int}}}{\mathrm{d}\pi_2} = (1-2\alpha)\cdot C \lesseqgtr 0 \text{ iff } \alpha \lesseqgtr \tfrac{1}{2}. \qquad \text{(F.7)}$$

Thus, if the first word gets a larger share than the second one (i.e. $\alpha > 0.5$), then an increase in lexical (and stressed) monosyllables leads to an increase in $\hat{x}_{\text{int}}$, i.e. the proportion of initially stressed disyllables. This is intuitive: if the foot built by first word in a sequence counts as more important, then the frequent occurrence of stressed monosyllables between disyllables, will select more strongly for trochaic patterns among preceding disyllables than for iambics pattern among subsequent ones. The reverse holds, of course, for $\alpha < 0.5$. If both words obtain an equal share, the only possible internal ESS is exactly $\hat{x}_{\text{int}} = 0.5$. Notably, α-values close to 0.5 produce an internal ESSs, no matter whether monosyllables are lexical and stressed, or functional and unstressed. Conversely, if α is close to 0 or 1, i.e. if (nearly) all of the payoff goes to only one of the two disyllables, the dynamics will produce (almost) pure equilibria corresponding to uniform stress placement. This is also intuitive: if monosyllables affect only one of their disyllabic neighbours, selection will either act for trochees or for

[23] Note that our model (reassuringly) confirms the intuitively obvious fact (see above page 12) that languages that have only disyllabic items will be either purely trochaic, or purely iambic. This results when $\pi_3 = 1$.

iambs, depending on which of their neighbours they affect, and on whether there are more lexical or more functional monosyllables. – As discussed above (page 22), the assumption that we consider most plausible for English is that a larger share of the payoff should be attributed to the first of two disyllables in a phrase (corresponding to the last two columns in Figure 3).

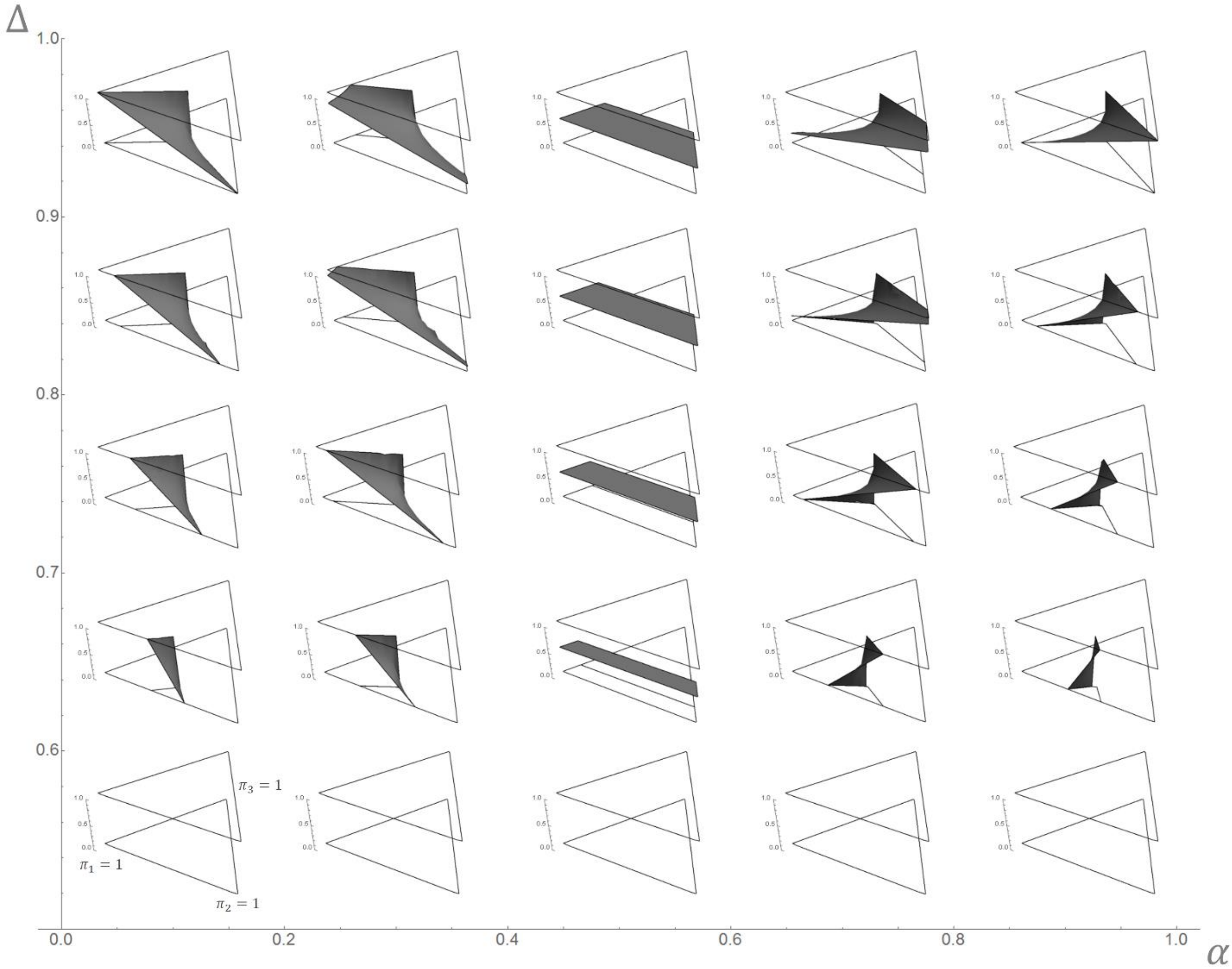

Figure 3. Plots of the ESS $\hat{x}_{\text{int}}$ (gray surface) depending on α, Δ, and π. Each plot shows the size of x_{int} for each $\pi = (\pi_1, \pi_2, \pi_3)$ in the triangular 2-simplex. The lower triangle corresponds to $x = 0$, i.e. only final stress, while the upper triangle represents disyllable populations in which there is only initial stress, i.e. $x = 1$. The horizontal axis measures α, i.e. the relative weight of the first word in an utterance. If $\alpha < 0.5$ the second word gets a larger amount of the resulting payoff, while if $\alpha > 0.5$ the first words obtains a larger share. If both words obtain an equal share $\hat{x}_{int} = 0.5$. The vertical axis measures the difference Δ between the payoffs obtained in the case of one and two violations, respectively. If $\Delta = 0.5$ both violations have an equal impact on the received payoff. If $\Delta = 1$, it does not make a difference at all whether a sequence features one or two stress violations. For $\Delta \leq 0.5$ there is no internal ESS.

The second implication of (F.5) and (F.6) is that, for both inequalities to hold, two violations must be perceived as less than twice as bad as one violation. This is so because π_3 is nonnegative and both denominators are positive. Consequently, both numerators, and hence their sum, must be positive. This can, in turn, be shown to require $\Delta > \frac{1}{2}$, and hence $\varphi(1) < \frac{1}{2} \cdot \varphi_{\max}$ (see Figure 2). It is also illustrated by the vertical dimension in Figure 3. The closer Δ gets to 0.5, smaller the support of x_{int}, i.e. the smaller the region in which initial stress and final stress stably coexist. This behaviour of the model is also intuitive: in sequences with monosyllables two violations will only occur when the disyllables have different stress patterns, as in [σˈσ][ˈσ][ˈσσ] (a double clash), or in [ˈσσ][σ][σˈσ] (a double lapse). When both disyllables have the same stress pattern no more than a single violation can ever be produced ([ˈσσ][ˈσ][ˈσσ], [σˈσ][ˈσ][σˈσ], [ˈσσ][σ][ˈσσ], [σˈσ][σ][σˈσ]. If now two violations are treated as much worse than single violations, the latter set of sequences will lead to much higher rewards. Thus, stress pattern diversity will never be stable, even if the number of monosyllables in a language is high.

This is a crucial issue, which deserves discussion. Recall that there are three possibilities: (a) two violations count as exactly twice as bad, as a single one, (b) they count as more than twice as bad, or (c) they count as less than twice as bad. While the first possibility (a) might strike one as the most intuitive and innocent one (because 1+1=2), it can in fact be ruled out on probabilistic grounds, because the possibility of two violations being *exactly* twice as grave as a single one represents a single conceivable case, and faces an infinite number of conceivable alternatives. This leaves (b) and (c), and we think there are good reasons to assume that it is indeed (c) that holds.

For the sake of the argument, compare the phrases

(11) a.	1 violation:	*CLASH:	*Mi ˈchelle's ˈfather*	[σˈσ] [ˈσσ]
		*LAPSE:	*ˈSusan and ˈMike*	[ˈσσ] [σ] [ˈσ]
b.	2 violations:	*CLASH:	*Mi ˈchelle's ˈold ˈfather*	[σˈσ] [ˈσσ],
		*LAPSE:	*ˈSusan and Mi ˈchelle*	[ˈσσ] [σ] [ˈσ]

The double violations in (11) are produced by three adjacent syllables that are either all stressed or all unstressed. While this still makes them, technically, double violations, there is the possibility of mitigating their effect by changing the rhythmic status of the middle syllable either only perceptually, or even also in actual articulation. Thus, *Mi ˈchelle's ˈold ˈfather* can be parsed, optionally, as *Mi ˈchelle's ˌold ˈfather*, and *ˈSusan and Mi ˈchelle* as *ˈSusan ˌand Mi ˈchelle*. Thus, through the demotion of lexically stressed *old* and the promotion of lexically unstressed *and* rhythm may be partially restored, albeit at the cost of backgrounding a content word, or forgeounding a functional one. Evidence that such adjustments do naturally occur can be found in verse, for example (see also Attridge 1982). Consider for instance the following lines from Shakespeare, in which lexical *fair* is demoted, and grammatical *of* promoted so that viable pentameters are produced.

(12) Aˈrise, ˌfair ˈsun, and ˈkill the ˈenvious ˈmoon (*Romeo and Juliet*, 2/1/5),
ˈNow is the ˈwinter ˌof our ˈdisconˈtent (*Richard the Third*, 1/1/1)

Crucially, the violations in (11) cannot be similarly mitigated without producing new ones. Therefore, we conclude that double violations can plausibly be considered as less than twice as bad than single ones, even though they should still count as worse, since (a) demotion and promotion are only optional and since (b) they come at the cost of (unnaturally) backgrounding a content word, or foregrounding a functional one. Thus, we feel confident in working on the assumption that $\Delta > \frac{1}{2}$, and hence $\varphi(1) < \frac{1}{2} \cdot \varphi_{\max}$.

We are now in a position to summarize our observations and draw first conclusions. As has been shown, our game does indeed produce stable internal equilibria. Thus it predicts that, under specific conditions, stress pattern diversity will inevitably be stably established in a lexicon. Thereby, we have have answered one of our central questions: if lexical stress assignment reflects constraints on the rhythmic quality of the utterances, then there are conditions under which words of the same phonotactic structure and the same morpho-syntactic class will necessarily adopt a variety of different stress patterns rather than just a single one. This will do happen irrespective of any possible further motivations, such as differences among words in terms of syllable weight, collocational preferences, etymological origins, or stylistic values. Thus, the problem of extant theories of word stress assignment, i.e. that they invariably face a subset of words for which stress can be predicted only probabilistically, has received a principled account. In addition, our model has allowed us to derive clear hypotheses about the conditions under which stress pattern diversity will arise:

(13) Mixed stressing patters exist in an evolutionarily stable way,

a. if $\Delta > \frac{1}{2}$, i.e. if two violations (i.e. a double clash or a double lapse) are perceived as less than twice as bad as a single violation, and
b. if π_3 is sufficiently small, i.e. if there are relatively few sequences of two neighbouring disyllables.
c. In case a stable mixed equilibrium $\hat{x}_{\text{int}}$ measuring the fraction of initially stressed items exists, then it increases (decreases) in the fraction of sequences involving monosyllabic lexical items if $\alpha > \frac{1}{2}$ ($\alpha < \frac{1}{2}$), i.e. if in a pair of two subsequent disyllables the first one (second one) obtains a larger share of the received payoff.

We have provided arguments for condition (a), consider it true that condition (b) is also met in English, where the proportion of monosyllabic items exceeds the proportion of all polysyllables taken together (see Figure 1), and we have shown evidence that the sufficient condition in (c) is true for English in the sense that the first word in a sequence receives more payoff than the second one. Therefore, we conclude that our hypothesis that word stress assignment reflects an adaptation to constraints on the rhythm of phrases qualifies as a plausible explanation of stress pattern diversity in the English lexicon. In order to corroborate it further, we hold the predictions it implies for the diachrony of stress pattern distribution against what is known about the evolution of word stress in English.

First, however, we briefly compare the findings of our model with studies by Sonderegger and Niyogi (2010, 2013), in which English stress is also approached on the basis of dynamical systems theory. In contrast to our own game, they model the distribution of stress patterns in a population of speakers[25] rather than in an evolving lexicon, and focus on the potential effects of 'mistransmission'. Reassuringly, Sonderegger and Niyogi's model converges with ours in many respects. This is particularly so in the case of x_{int} being an ESS, where our model predicts the same as theirs if the respective payoff differences between choosing the same strategy as the competing player and choosing the converse strategy are interpreted as Sonderegger and Niyogi's mistransmission rates. In their model, *a* is the rate of misinterpreting finally stressed items as initially stressed items, and *b* is the rate of misinterpreting initially stressed items as finally stressed ones.[27] Sonderegger and Niyogi (2013: 277-278) then show that there is a stable internal equilibrium at $\alpha_* = b/(a+b)$, where α_* measures the mean probability of a speaker using word final stress. If we now set $a = \left(a_{12}^{\pi,\varphi,\alpha} - a_{22}^{\pi,\varphi,\alpha}\right) \cdot C > 0$ and $b = (a_{21}^{\pi,\varphi,\alpha} - a_{11}^{\pi,\varphi,\alpha}) \cdot C > 0$, *C* being some positive constant, then

$$\alpha_* = \frac{a_{21}^{\pi,\varphi,\alpha} - a_{11}^{\pi,\varphi,\alpha}}{a_{21}^{\pi,\varphi,\alpha} - a_{11}^{\pi,\varphi,\alpha} + a_{21}^{\pi,\varphi,\alpha} - a_{11}^{\pi,\varphi,\alpha}} = 1 - \frac{a_{12}^{\pi,\varphi,\alpha} - a_{22}^{\pi,\varphi,\alpha}}{a_{21}^{\pi,\varphi,\alpha} - a_{11}^{\pi,\varphi,\alpha} + a_{21}^{\pi,\varphi,\alpha} - a_{11}^{\pi,\varphi,\alpha}}. \tag{F.8}$$

Assuming that the population of speakers, and thus also the population of utterances, is homogeneously mixed, it follows that $\alpha_* = \hat{y}_{int} = 1 - \hat{x}_{\mathrm{int}}$, so that due to (F.3) both their model and ours result in equivalent equilibria.

Although, the two models are equivalent in these respects, there are nevertheless aspects in which they differ. Thus, Sonderegger and Niyogi (2013) do not say much about the factors that actually

[25] Using a learning algorithm that is driven by probability matching and admits mistransmission.

[27] Note, that in contrast to Sonderegger and Niyogi (2013)'s model, the parameters *a* and *b* are conceptualized as rates rather than as probabilities, since the present replicator dynamics represent a dynamical system in continuous time. Moreover, it is worth pointing out that Sonderegger and Niyogi actually studied the word-stress dynamics of single lexical items (e.g. specific nouns or verbs) rather than that of an entire set of words of the same phonotactic and morpho-syntactic type. This is, however, only a matter of reinterpreting the variables in the dynamical system, which shall not pose much of a problem: instead of single words one can just as well look at classes of morpho-syntactically and phonotactically equivalent items.

condition mistransmission rates. In contrast, our model makes very specific proposals about the conditions under which one stress pattern gets transmitted more successfully than its alternative. In that sense, it not only corroborates Sonderegger and Niyogi's findings but elaborates them as well. Also, the game theoretic approach taken here is more encompassing, and covers the dynamics modelled by the mistransmission-based system as a special case among various possible scenarios. The strength of the replicator equation is that it cannot only account for stable coexistence of stress patterns but also for scenarios in which one of them becomes dominant as well as for bi-stability among stress patterns (see Box 1).[28] For changing the qualitative behaviour of their model so that bi-stability becomes possible, Sonderegger and Niyogi (2013: 279-280) need to resort to a different underlying learning mechanism, which in turn does not admit stable coexistence. Thus, in a number of ways, the game theoretic approach taken here is more powerful and can account for a variety of evolutionary dynamics with the same set of internal mechanisms.[30] Which of them actually unfolds, does not depend on properties of our model as such, but is predicted to follow from differences between languages, or language states, such as the relative frequencies of monosyllabic and disyllabic words. As pointed out, and as we shall see immediately, this makes it possible to derive testable, albeit general, hypotheses.

5 Further corroboration by diachronic evidence

As we have so far seen, our game provides a consistent account of stress pattern diversity as attested in Present Day English. It therefore lends plausibility to the hypothesis that word stress patterns may represent adaptations to constraints on rhythmic well-formedness that apply on phrases. This hypothesis explains why their distribution cannot be fully predicted by theories that derive word stress for isolated lexical item, because it depends to a considerable degree on the way in which words can be combined and come to interact in utterances. This view is compatible with utterance based and evolutionary theories of language. At the same time, however, it allows one to make general predictions about language systems. One of them is that the stability of stress pattern diversity depends

---

[28] These are the scenarios in which the inequalities $a_{11}^{\pi,\varphi,\alpha} > a_{21}^{\pi,\varphi,\alpha}$ or $a_{12}^{\pi,\varphi,\alpha} < a_{22}^{\pi,\varphi,\alpha}$ (or both in the case of bi-stability) are fulfilled. Crucially, there are settings for the defining parameters π, φ and α that entail one or the other inequality. For example, $\pi_1 = 1$ entails final stress to dominate, $\pi_2 = 1$ leads to domination of initial stress, and $\pi_3 = 1$ immediately implies bistability.

[30] Similarly, Yang's (2000) speaker-based diachronic model driven by variational learning only allows for either strategy to be stable and attracting, but cannot account for either bi-stability or stable coexistence.

crucially on the number of monosyllables in the lexicon, and that stress pattern distribution will therefore reflect diachronic changes in the monosyllabicity of a language if such changes occur.

In the following we show that the specific predictions our game allows one to derive fit the long term evolution of English word stress in a way that we consider to be reassuring. The history of English stress has of course been investigated widely, and in the context of this paper no more than a global summary can be reported. For the purposes of our argumentation, however, this will be good enough.

Old English stress was uniformly root initial (Minkova 1997: 137, see also Hutton 1998, Mitchell & Robinson 2007). Also, Old English morphology was still highly inflecting (at least in comparison to Middle and Modern English, so that it contained a large proportion of polysyllabic word forms. However, already during later Old English times, unstressed inflectional syllables underwent reduction and deletion processes, leading to a gradual but marked increase in monosyllables. Thus, what would have been a stable pure equilibrium in which initial stress dominated, would at one point have become evolutionarily unstable, although initial stress may still have been the rule. If a pure equilibrium of initial stress is unstable, however, this means that a lexicon will cease to display pure initial stress, if it comes to adopt words that display final stress. This is exactly what happened in English, however, when in the wake of the Norman Conquest, finally stressed loans from French were adopted in large numbers. (cf. e.g. Lass 1992, Minkova 1997, Dresher & Lahiri 2005, Mitchell & Robinson 2007, Ritt 2012, as well as already Tamson 1898). Although, early Romance mostly adopted the native initial stress pattern before long, they produced considerable surface variation, as can be seen for example in the works of Chaucer, in stress doublets like those in (14).

(14) [ˈci.té]$_N$ vs. [ci.ˈté]$_N$
[ˈdi.vers]$_{Adj}$ vs. [di.ˈvers]$_{Adj}$
[ˈhus.bond]$_N$ vs. [hus.ˈbond]$_N$
[ˈA.pril]$_N$ vs. [A.ˈprile]$_N$
[ˈout.lawe]$_N$ vs. [out.ˈlawe]$_N$
[ˈho.nour]$_N$ vs. [ho.ˈnour]$_N$
[ˈfor.tune]$_N$ vs. [for.ˈtune]$_N$
[ˈser.vise]$_N$ vs. [ser.ˈvise]$_N$

Eventually, however, such surface variability was reduced again, and the English principle of assigning stress to words lexically and keeping its position immobile in their utterance expressions reasserted itself, so that items whose stress patterns had varied on the surface came to adopt one specific pattern in their lexical representation, and would maintain it in most of their realisations as well. Crucially, however, the stress patterns that came to be lexicalised did not all reflect the same set of stress assignment principles, so that French loans like *cité* or *beautée* wound up with initial stress, i.e. as *ˈcity* and *ˈbeauty*, while others, such as *deˈgree*, retained final stress. Ever since, the English lexicon has continued to contain items that are equivalent to one another both in terms of phonotactic structure and in terms of word class, but that nevertheless have different stress patterns. The only changes that have been observable during the Modern period involve frequent, albeit anything but system wide, shifts of stress towards the left, for example in words like *ˈaddress*$_{\text{N}}$ (< *adˈdress*$_{\text{N}}$) *ˈbalcony* (< *balˈcony*), *ˈcompact*$_{\text{Adj}}$ (< *comˈpact*$_{\text{Adj}}$) or recent *ˈhotel* (< *hoˈtel*) (see also Minkova 1997).

Now, as Figure 4 below demonstrates, the specific developments that English stress patterns seem to have undergone, correspond extremely well to evolutionary dynamics predicted by our model for a language in which the number of (lexical) monosyllables increases over time.This figure shows a so-called bifurcation diagram from two different perspectives which plots the equilibria of the evolutionary dynamics, i.e. the stable as well as unstable steady-state distributions of initially and finally stressed items, in correlation with the number of monosyllables in the language. The upper and lower triangular planes represent populations in which all disyllables are either initially stressed (upper plane), or finally stressed (lower plane). The tips of the two triangles that are marked $\pi_3 = 1$, i.e. $\pi = (0,0,1)$, represent equilibria in which there are no monosyllables at all, so that when you move away from them, this indicates an increase in the number of sequences containing monosyllables (either grammatical ones, if you move towards $\pi_1 = 1$, or lexical ones, if you move towards $\pi_2 = 1$). Finally, the twisted surface connecting the upper and lower planes represents states in which mixed populations occur. Dark gray areas represent evolutionarily stable equilibria, light gray areas instable ones.

Now, think of English starting at point (1) on the upper plane in Figure 4a and 4b, somewhere in the dark gray area. This point characterises a language with pure and evolutionarily stable initial stress, which corresponds to the situation that obtained in Old English. Next, assume a move of π towards (0,1,0), i.e. assume that the number of monosyllables increases, as indeed happened in the development of Old English through the phonetic erosion of final syllables and the concomitant simplification of the inflectional system (e.g. functional *ande* > *and*, or lexical *gode* > *god*). This increase does, at first, not produce variable stress, but after crossing a critical threshold (the boundary between the dark gray and the light gray area), it destabilizes the state of pure initial stress, as the language moves into the light gray area (2) on the upper surface. Being evolutionarily unstable, however, the dominance of pure initial stress in the population will come to a catastrophic end as soon as a number of finally stressed items enter the population. Again, this is indeed what happened to English when it adopted French loan words in great numbers from the beginning of the 12th century. As soon as this happens, our model predicts that a mix of initially and finally stressed words should become evolutionarily stable, as the population falls down to land near point (3) on the twisted surface inhabited by mixed populations visible in Figure 4b.

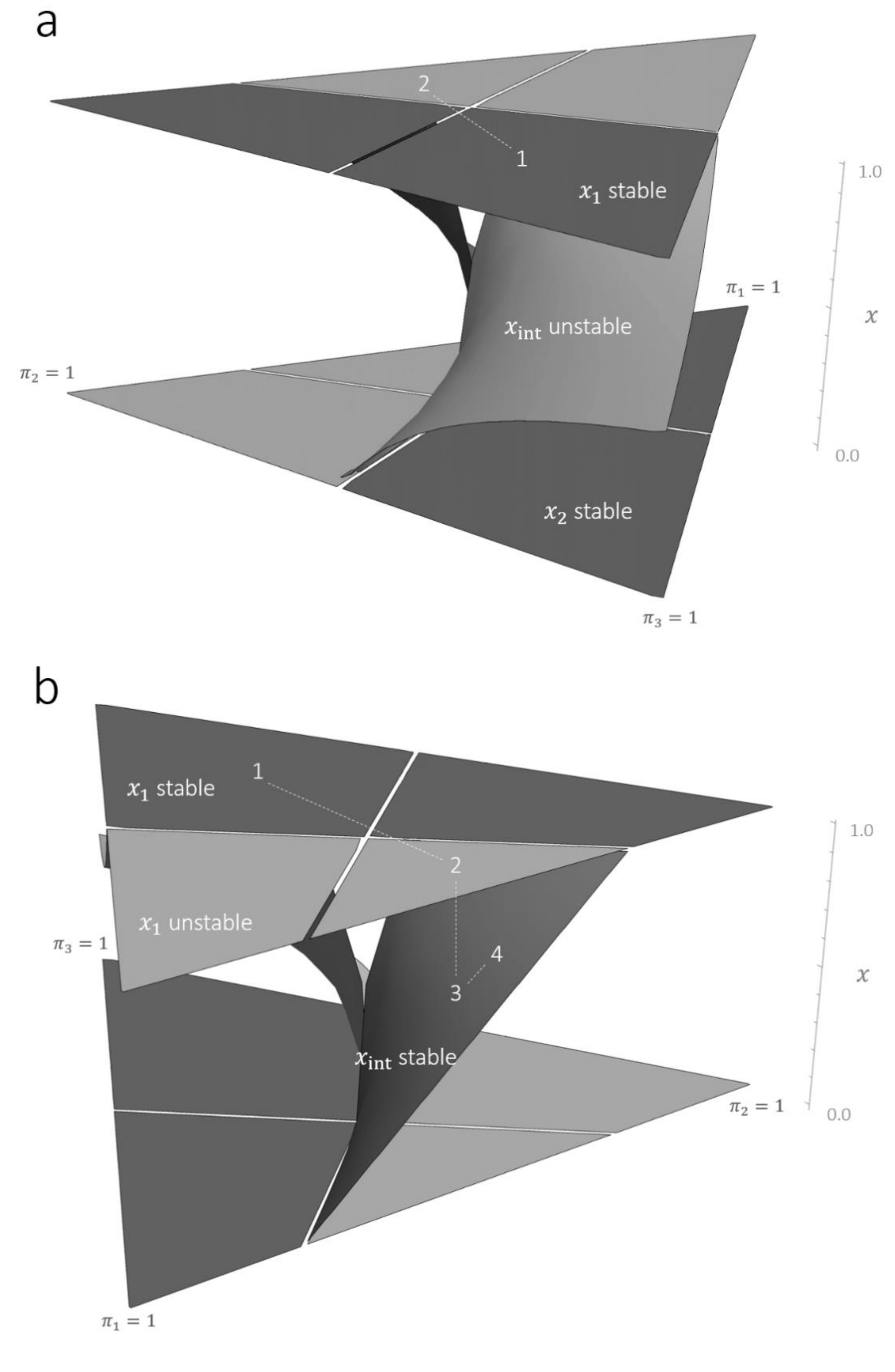

Figure 4. Bifurcation diagram showing the equilibria $\hat{x}$ of the replicator dynamics as a function of the distribution of contexts $\pi = (\pi_1, \pi_2, \pi_3)$, α and Δ remaining fixed, from two different perspectives (a, b). Possible distributions of contexts lie within the 2-simplex, which is represented by the horizontal triangles. The vertical axis measures the fraction of initially stressed words $\hat{x}$. Dark gray denotes stable while light gray denotes unstable equilibria. Numbers (1-4) indicate the approximate positions of English in its diachronic development from ME to PDE.

Finally, our models predicts that, if the number of lexical monosyllables keeps increasing, the population will evolve steadily towards the upper plane again, i.e. towards a situation in which purely initial stress becomes once again the most frequently employed stressing strategy (4). Once again, this seems to be exactly what has been going on in English during the last centuries: as monosyllabicity increased among lexical words (see already Jespersen 1912, 1928), more and more polysyllabic words underwent stress shifts towards the left (for example, *ˈaddress*$_{N}$ (< *adˈdress*$_{N}$) *ˈbalcony* (< *balˈcony*), *ˈcompact*$_{Adj}$ (< *comˈpact*) or recent *ˈhotel* (< *hoˈtel*) (see also Minkova 1997).

Since our model was originally designed to predict, in very general terms, under what conditions stress pattern diversity can be predicted to establish itself in the lexicon of a language such as Present day English, we consider the fact that it predicts the dynamics in the historical evolution of English word stress so surprisingly well too, as an independent and rather strong corroboration of the hypotheses on which our model rests.

6 Conclusion and outlook

This paper has introduced a new way of thinking about stress assignment in languages such as English, in which stress is assigned lexically and usually gets faithfully expressed in all utterance realisations of individual words. What we have tried to account for is the existence of variable stress patterns among words of the same phonotactic and morpho-syntactic types. Although focussing on word stress, we have broken with the established tradition of deriving it from the properties of individual lexical items in isolation. Instead, we have employed evolutionary game theory to propose a model in which the assignment of stress patterns to word types reflects rhythmic constraints on utterance sequences rather than on individual items. One of the main arguments for this approach was that rhythmic constraints can be naturally grounded on the utterance level, but have no place in word or stem level phonology.

Our model is highly abstract, which has allowed us to factor out a number of factors that might in principle be adducible for explaining stress pattern variability such as word specific collocations, or stylistic factors. Therefore, we have been able to demonstrate that variable stress patterns among words of the same phonotactic and morpho-syntactic types can indeed follow, if lexical stress reflects constraints on the rhythmic well-formedness of the utterance sequences that words build with one another. More strongly, we have specified conditions under which this will inevitably be the case, and argued that these conditions do indeed hold in Present Day English. Thereby, we have accounted for a set of empirical facts which so far have been merely acknowledged but left without a principled explanation.

More specifically and most importantly, we have shown that there seems to be a causal connection between the degree to which languages allow and demand the stress patterns of polysyllabic items to be variable and the number of monosyllables in a language. Building on this insight, we have also demonstrated that our model is capable of making predictions about the evolutionary dynamics of stress systems, and that these show a surprising fit with the actual developments that English word stress has undergone during the last millennium.

Although our model has remained highly abstract and general, we have demonstrated its explanatory power, the general viability of our approach, and its advantages over established theories of word stress in languages such as English. Naturally, much work can and needs still to be done with regard to the refinement of the model (such as differential payoffs for lapses and clashes, a wider selection of word structures in terms of syllable counts, or including explicit interaction among various morpho-syntactic classes), and the specificity of the predictions to be derived from it. From what it manages to do even in its present stage, however, we conclude that its architecture will prove sufficiently robust for further research to be based on it.

7 References

Alber, Birgit. 2005. Clash, Lapse and Directionality. *Natural Language & Linguistic Theory* 23(3). 485–542.

Altmann, Eduardo, Janet Pierrehumbert & Adilson Motter. 2009. Beyond word frequency: bursts, lulls, and scaling in the temporal distributions of words. *PloS one* 4(11). e7678.

Anderson, John M. 1986. Suprasegmental dependencies. In Jacques Durand (ed.), *Dependency and non-linear phonology* (Croom Helm linguistics series), 55–133. London, Dover, N.H: Croom Helm.

Attridge, Derek. 1982. *The Rhythms of English Poetry*. London: Taylor & Francis.

Baayen, H. R. Piepenbrock & L. Gulikers. 1995. *CELEX2 LDC96L14*. Web download.

Baxter, G. J. R. A. Blythe, W. Croft & A. J. McKane. 2009. Modeling language change: An evaluation of Trudgill's theory of the emergence of New Zealand English. *Language Variation and Change* 21(2). 257–296.

Baxter, Gareth & William Croft. forthc. Modeling language change across the lifespan: individual trajectories in community change. *Language Variation and Change*.

Blevins, Juliette (ed.) (2004). *Evolutionary phonology*: *The emergence of sound patterns*. Cambridge, New York: Cambridge University Press.

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

Boer, Bart de. 2002. Evolving sound systems. In A. Cangelosi & D. Parisi (eds.), *Simulating the evolution of language*, 79–97.

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

Burzio, Luigi. 1994. *Principles of English stress* (Cambridge studies in linguistics 72). Cambridge [England], New York: Cambridge University Press.

Bybee, Joan. 2001. *Phonology and language use.*. Cambridge: Cambridge University Press.
Bybee, Joan. 2007. *Frequency of use and the organization of language*. Oxford: Oxford University Press.
Bybee, Joan L. Paromita Chakraborti, Dagmar Jung & Joanne Scheibman. 1998. Prosody and Segmental Effect Some Paths of Evolution for Word Stress. *Studies in Language* 22(2). 267–314.
Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper and Row.
Christiansen, Morten H. & Nick Chater. 2008a. Brains, genes, and language evolution: A new synthesis. *Behavioral and Brain Sciences* 31(05).
Christiansen, Morten H. & Nick Chater. 2008b. Language as shaped by the brain. *Behavioral and Brain Sciences* 31(05).
Croft, William. 2000. *Explaining language change*: *An evolutionary approach* (Longman linguistics library). Harlow, England, New York: Longman.
Croft, William. 2013. Language use and the evolution of languages. In Kenny Smith & Philippe Binder (eds.), *The language phenomenon*, 93–120. Berlin [a.e.]: Springer.
Dawkins, Richard. 1976. *The selfish gene*. New York, Oxford: Oxford University Press.
Dennett, Daniel C. 1987. *Intentional Stance*. Cambrdige, Mass.: MIT Press.
Domahs, Ulrike, Ingo Plag & Rebecca Carroll. 2014. Word stress assignment in German, English and Dutch: Quantity-sensitivity and extrametricality revisited. *The Journal of Comparative Germanic Linguistics* 17(1). 59–96.
Donegan, Patricia J. & David Stampe. 1979. The study of Natural Phonology. In D.A Dinnsen (ed.), *Current Approaches to Phonological Theory*, 126–173. Bloomington: Indiana University Press.
Donegan, Patricia J. & David Stampe. 1983. Rhythm and the holistic organization of language structure. In John F. Richardson, Mitchell Marks & Amy Chukerman (eds.), *Papers from the parasession on the interplay of phonology, morphology and syntax*, 1st edn. 337–353. Chicago: Chicago Linguistic Society.
Donegan, Patricia J. & David Stampe. 2009. Hypotheses of Natural Phonology. *PSiCL Poznań Studies in Contemporary Linguistics*.
Dresher, Elan B. & Aditi Lahiri. 2005. Main stress left in Early Middle English. *Historical Linguistics 2003: Selected Papers from the 16th International Conference on Historical Linguistics, Copenhagen, 11-15 August 2003.* 257.
Dressler, Wolfgang U. 1985. *Morphonolgy*: *The dynamics of derivation*. Ann Arbor: Karoma.
Dunn, Michael, Simon J. Greenhill, Stephen C. Levinson & Russell D. Gray. 2011. Evolved structure of language shows lineage-specific trends in word-order universals. *Nature* 473(7345). 79–82.
Dziubalska-Kołaczyk, Katarzyna. 2009. NP Extension: B&B Phonotactics. *Poznań Studies in Contemporary Linguistics* 45(1). 55–71.
Fitch, W. T. 2013. Rhythmic cognition in humans and animals: distinguishing meter and pulse perception. *Frontiers in Systems Neuroscience* 7. 68.
Fudge, E. 1999. Words and feet. *Journal of Linguistics* 35(2). 273–296.
Fudge, Erik C. 1984. *English word-stress*. London, Boston: Allen & Unwin.
Fullwood, Michelle A. 2014. Asymmetric Correlations between English Verb Transitivity and Stress. *Annual Meeting of the Berkeley Linguistics Society* 40. 125–138.
Giegerich, Heinz J. 1985. *Metrical phonology and phonological structure*: *German and English* (Cambridge studies in linguistics 43). Cambridge, New York: Cambridge University Press.
Halle, M. 1998. The stress of English words 1968-1998. *Linguistic Inquiry* 29(4). 539–568.
Hayes, Bruce. 1982. Extrametricality and English Stress. *Linguistic Inquiry* 13(2). 227–276.
Hayes, Bruce. 1984. The Phonology of Rhythm in English. *Linguistic Inquiry* 15. 33–74.
Heino, Mikko, Metz, Johan A. J. & Veijo Kataiala. 1998. The enigma of frequency-dependent selection. *Trends in Ecology and Evolution* 13. 367–370.

Hofbauer, J. & K. Sigmund. 1998. *Evolutionary games and population dynamics*. Cambridge: Cambridge University Press.
Hutton, J. 1998. Stress in Old English, giet ongean. *LINGUISTICS* 36(5). 847–885.
Hyde, Brett. 2007. Non-finality and weight-sensitivity. *Phonology* 24(02). 287–334.
Hyde, Brett. 2011. Extrametricality and nonfinality. In Marc van Oostendorp, Colin J. Ewen, Elizabteh Hume & Keren Rice (eds.), *The Blackwell companion to phonology* (Blackwell companions to linguistics series). Malden, MA: Wiley-Blackwell.
Jäger, Gerhard. 2008. Applications of game theory in linguistics. *Language and Linguistics compass* 2(3). 406–421.
Jäger, Gerhard & Anette Rosenbach. 2008. Priming and unidirectional language change. *Theoretical Linguistics* 34(2). 85–113.
Jespersen, Otto. 1912. *The growth and structure of the English language.* 2nd edn. Leipzig: Teubner.
Jespersen, Otto. 1928. *Monosyllabism in English* (Proceedings of the British Academy 14). London: H. Milford.
Kager, René. 1989. *A Metrical Theory of Stress and Destressing in English and Dutch*. Dordrecht: Foris.
Kager, René. 2005. The factorial typology of rhythmic licensing constraints. *Phonological Studies* 5. 47–155.
Kelly, Michael H. 1988. Rhythmic alternation and lexical stress differences in English. *Cognition* 30(2). 107–137.
Kelly, Michael H. 1989. Rhythm and language change in English. *Journal of Memory and Language* 28(6). 690–710.
Kelly, Michael H. 1992. Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. *Psychological Review* 99(2). 349–364.
Kelly, Michael H. 2004. Word onset patterns and lexical stress in English. *Journal of Memory and Language* 50(3). 231–244.
Kiparsky, Paul. 1966. Über den deutschen Wortakzent. In Alexander V. Isačenko, Paul Kiparsky & Manfred Bierwisch (eds.), *Untersuchungen über Akzent und Intonation im Deutschen*, 71–98. Berlin: Akademie Verlag.
Kiparsky, Paul. 1979. Metrical structure assignment is cyclic. *Linguistic Inquiry* 10. 421–441.
Lass, Roger. 1992. Phonology and morphology. In Norman Blake (ed.), *The Cambridge History of the English Language*: *Vol. 2: 1066-1476* (The Cambridge History of the English Language Ser), 23–155. New York: Cambridge University Press.
Lass, Roger. 1997. *Historical linguistics and language change* (Cambridge studies in linguistics 81). Cambridge, New York: Cambridge University Press.
Liberman, Mark & Alan Prince. 1977. On Stress and Linguistic Rhythm. *Linguistic Inquiry* 8(2). 249–336.
Marvin, Tatjana. 2003. *Topics in the stress and syntax of words* : Massachusetts Institute of Technology.
Maynard Smith, John & George R. Price. 1973. The logic of animal conflict. *Nature* 246. 15–18.
McCrohon, Luke. 2012. The two-stage life cycle of cultural replicators. *Theoria et Historia Scientiarum* 9. 149–170.
McMahon, April M. & Robert McMahon. 2013. *Evolutionary linguistics* (Cambridge textbooks in linguistics). Cambridge [England], New York: Cambridge University Press.
Minkova, Donka. 1997. Constraint ranking in Middle English stress-shifting. *English Language and Linguistics* 1. 135–175.
Mitchell, Bruce & Fred C. Robinson. 2007. *A guide to Old English*, 7th edn. Malden, MA: Blackwell Pub.
Mitchener, W. G. 2003. Bifurcation analysis of the fully symmetric language dynamical equation. *Journal of Mathematical Biology* 46(3). 265–285.

Mitchener, W. G. & Martin A. Nowak. 2004. Chaos and language. *Proceedings of the Royal Society of London-B* 271(1540). 701–704.
Mühlenbernd, R. & J. Wahle. 2016. Self-organization In Sound Systems: A Model Of Sound Strings Processing Agents. In S. G. Roberts, C. Cuskley, L. McCrohon, L. Barceló-Coblijn, Fehér O. & T. Verhoef (eds.), *The Evolution of Language: Proceedings of the 11th International Conference (EVOLANG11)*.
Niyogi, Partha. 2006. *The computational nature of language learning and evolution*. Cambridge, MA.
Nowak, M. A. 2000. The basic reproductive ratio of a word, the maximum size of a lexicon. *Journal of theoretical biology* 204(2). 179–189.
Nowak, M. A. J. Plotkin & V. Jansen. 2000. The evolution of syntactic communication. *Nature* 404(6777). 495–498.
Nowak, Martin. 2006. *Evolutionary dynamics: exploring the equations of life*. Harvard: Belknap.
Nowak, Martin A. & Natalia L. Komarova. 2001. Towards an evolutionary theory of language. *Trends in cognitive sciences* 5. 288–295.
Nowak, Martin A. Natalia L. Komarova & Partha Niyogi. 2002. Computational and evolutionary aspects of language. *Nature* 417. 611–617.
Paul, Hermann. 1920. *Prinzipien der Sprachgeschichte*, 5th edn. Halle: Niemeyer.
Pierrehumbert, J. 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: John Benjamins Pub. Company.
Pierrehumbert, Janet. 2012. The Dynamic Lexicon. In Abigail Cohn, Marie Huffman & Cécile. Fougeron (eds.), *Handbook of Laboratory Phonology*, 173–183. Oxford: Oxford University Press.
Poldauf, Ivan & William R. Lee. 1984. *English word stress*: *A theory of word-stress patterns in English*, 1st edn. Oxford, New York: Pergamon Press.
Prince, Alan. 1980. A metrical theory for Estonian quantity. *Linguistic Inquiry*(11). 511–562.
Ritt, Nikolaus. 2004. *Selfish sounds and linguistic evolution*: *A Darwinian approach to language change*. Cambridge: Cambridge University Press.
Ritt, Nikolaus. 2012. Middle English phonology in the digital age: What written corpora can tell us about sound change. In Terttu Nevalainen & Elizabeth C. Traugott (eds.), *The Oxford handbook of the history of English* (Oxford handbooks in linguistics), 161–167. Oxford, New York: Oxford University Press.
Sainz, Susana. 1988. *A noncyclic analysis of English word stress*.
Schane, Sanford. 2007. Understanding English word accentuation. *Language Sciences* 29(2-3). 372–384.
Schlüter, Julia. 2005. *Rhythmic grammar*: *The influence of rhythm on grammatical variation and change in English* (Topics in English linguistics 46). Berlin, New York: Mouton de Gruyter.
Scott-Phillips, Thomas C. & Simon Kirby. 2010. Language evolution in the laboratory. *Trends in cognitive sciences* 14(9). 411–417.
Sherman, Donald. 1975. Noun-verb stress alternation: an example of the lexical diffusion of sound change in English. *LINGUISTICS* 13(159). 43–72.
Smith, John M. 1982. *Evolution and the Theory of Games* : Cambridge University Press.
Smith, Kenny & Simon Kirby. 2008. Cultural evolution: implications for understanding the human language faculty and its evolution. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 363(1509). 3591–3603.
Smith, Laura C. & Joseph Salmons. 2008. *Historical Phonology and Evolutionary Phonology.* Diachronica. Amsterdam: Benjamins 25(3).
Solé, Ricard. 2011. *Phase transitions*. Princeton: Princeton University Press.

Sonderegger, Morgan & Partha Niyogi. 2010. Combining Data and Mathematical Models of Language Change. In, *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, 1019–1029: Association for Computational Linguistics.

Sonderegger, Morgan & Partha Niyogi. 2013. Variation and change in English noun/verb pair stress: data and dynamical systems models. In Alan Yu (ed.), *Origins of Sound Change*, 262–284.

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

Steels, Luc. 2011. Modeling the cultural evolution of language. *Physics of Life Reviews* 8(4). 339–356.

Tallerman, Maggie & Kathleen R. Gibson (eds.) (2012). *The Oxford handbook of language evolution* (Oxford handbooks in linguistics). Oxford, New York: Oxford University Press.

Tamson, George J. 1898. *Word-stress in English A short treatise on the accentuation of words in Middle-English as compared with the stress in Old and Modern English* (Studien zur englischen Philologie H. 3). Halle an der Saale.

van Oostendorp, Marc. 2010. Stress and metrical structure. http://www.vanoostendorp.nl/pdf/051115.pdf (5 July, 2013).

Vennemann, Theo. 1986. *Neuere Entwicklungen in der Phonologie*. Berlin [a.e.]: Mouton.

Wang, William, Jinyun Ke & James Minett. 2004. Computational studies of language evolution. *Computational linguistics and beyond*. 65–106.

Wedel, Andrew. 2006. Exemplar models, evolution and language change. *The Linguistic Review* 23(3). 247–274.

Wedel, Andrew. 2009. Variation, multi-level selection and conflicts between phonological and morphological regularities. In James P. Blevins & Juliette Blevins (eds.), *Analogy in grammar*: *Form and acquisition* (Oxford linguistics), 83–101. Oxford, New York: Oxford University Press.

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

Zuidema, Willem & Bart de Boer. 2009. The evolution of combinatorial phonology. *Journal of Phonetics* 37(2). 125–144.