Grammar or lexicon. Or: Grammar and lexicon? Rule-based and usage-based approaches to phonological variation

Hinskens, F.L.M.P.; Hermans, B.J.H.; van Oostendorp, M.

published in
Lingua
2014

DOI (link to publisher)
10.1016/j.lingua.2014.01.005

document version
Peer reviewed version

Available online at www.sciencedirect.com

ScienceDirect

Lingua xxx (2014) xxx–xxx

Lingua

www.elsevier.com/locate/lingua

Editorial

Grammar or lexicon. Or: Grammar and lexicon? Rule-based and usage-based approaches to phonological variation

Abstract

After a brief discussion of the concept of language variation, some of the main characteristics of 'rule-based' and usage-based paradigms are sketched, confined to the domain of phonological variation and its relation with syntax, morphology and the lexicon. Given the large number of different perspectives from which both approaches and their relationships have recently been addressed in the literature, these outlines cannot be exhaustive. Both paradigms will be compared on the basis of research of the variable deletion of word-final /t/ or /d/ in modern varieties of English, Dutch and German. Several strengths and weaknesses of both approaches are succinctly discussed. The contours of some recent hybrid models, including those advanced in the five contributions to the present thematic issue, will be briefly described. Subsequently, two quantitative studies of variable reduction processes in specific varieties of modern Dutch will be summarized. Following a reflection on the potential implications of one of the outcomes of the latter of the two studies, a few thoughts regarding the potential and desiderata for further research conclude this contribution.

Keywords: Language variation; Phonology; Storage; Computation; Hybrid models; Lexically diffuse sound change

Mit dem Wissen wächst der Zweifel – Goethe

1. Introduction

Although usage-based and 'rule-based' approaches to natural language share a number of constituent properties as a result of the fact that both look at language as a cognitive object, they differ in many respects. In essence, usage-based approaches contrast with 'rule-based', formal theory in that they do not assume language users to have abstract grammatical knowledge at their disposal. Instead, they postulate a close, organic connection between linguistic structure and language usage.

With respect to the phonetic/phonological part of language, usage-based models assume that language users store detailed phonetic information about the words of their language each time that they are exposed to them. These models stipulate redundant mental storage of bundles of maximally concrete articulatory, acoustic, grammatical, semantic and pragmatic information concerning single occurrences ('tokens' or 'exemplars') of lexical items, along with characteristics of both the speaker and the situation, organized in 'clouds'. The items are interconnected in multi-tiered networks, in which aspects of form, meaning and usage are the organizing dimensions. Formal theory, on the other hand, conceives linguistic competence as a computational capacity based on internalized representations, rules, processes, constraints, principles and related abstract devices, which are usually categorical and generalize across many cases. The latter traits characterize several generations of models of generative phonology, including Optimality Theory, even though (classic) Optimality Theory is non-derivational, hence not literally rule-based, but rather declarative in nature. The designation 'formal' to refer to these various approaches goes back to De Saussure's definition of *langue* as ''une forme, non une substance'' (1916, Cours, Ch. III); where De Saussure's notion of 'forme' referred to the structure of the relations holding between linguistic elements.

0024-3841/$ – see front matter
http://dx.doi.org/10.1016/j.lingua.2014.01.005

Undoubtedly, the best-known usage-based model of phonology (and morphology) is Bybee (2001, 2002, 2006). Bybee's views are closely connected to proposals that have been made within the framework of Exemplar Theory in phonetics (Johnson, 1997; Pierrehumbert, 2002). For syntax, Construction Grammar and Cognitive Grammar (Langacker, 2008; Goldberg, 2006) are relatively closely related. Cognitive grammar, Exemplar Theory and other usage-based approaches are all inspired by connectionism and one of the central themes in this family of theories is that language structure emerges from language use.

This thematic issue of *Lingua* will focus on the area of language variation in the sound components, with brief excursions on morphology and the lexicon. The present contribution will therefore add a variationist perspective, which was originally connected to, although by no means exclusive to, rule-based models.

After a concise discussion of the notion of language variation (Section 2), we will give rough outlines of rule-based and usage-based accounts (in Sections 3 and 4, respectively). Given the multitude of disciplinary and theoretical angles from which both paradigms and the relationship between them have been tackled in the literature of the past one or two decades, these sketches can only be incomplete. Both approaches will be compared on the basis of studies of the variable deletion of word-final /t/ or /d/ in specific varieties of several Germanic languages. Next, some strengths and weaknesses of both approaches will be briefly discussed (Section 5). In recent years, several hybrid models have been proposed; in Section 6, the contours of some hybrid models, including those advanced in the five contributions to this thematic issue, will be succinctly described. This is followed by a presentation of two quantitative studies of variable reduction processes in certain varieties of modern Dutch (Section 7). After a discussion of the possible relevance of some of the findings from the latter of the two studies, a few reflections regarding the possibilities and desiderata for further research conclude this contribution (Section 8).

2. Language variation

Diversity and variation occur both across and within languages. In this special issue no attention will be paid to language variation in the typological sense (cross-linguistic diversity or macro-variation), but rather in the sense of variation between (micro-variation) and within varieties of a given language. This distinction is sometimes referred to as one between inter- and intrasystemic variation, respectively; the latter type is sometimes called quantitative variation, i.e. Labov's (1972a:188) ''alternate ways of saying 'the same' thing''. Study after study has made clear that this type of language variation usually shows systematic, probabilistically constrained, quantitative regularities – hence 'orderly heterogeneity' (Weinreich et al., 1968).

In generative theory quantitative variation is traditionally referred to as optionality, although the notion of 'optionality' is sometimes also used to refer to choices which are indeed completely free, in the sense of linguistically entirely unconstrained (or not sufficiently understood). The label 'optional' or 'facultative', for example, has been used to refer to the fact that certain verbs (such as many equivalents of 'give') require two arguments, although there may be a third one, and also to cases of allophony which are not subject to complementary distribution (cf. Müller, 2003).

Both usage-based and rule-based paradigms conceive language as a cognitive object rather than a social one. They thus operate in a fundamentally different dimension than the sociolinguistic approach to language variation, since, as Labov (2010:7) put it, ''the central dogma of sociolinguistics is that the community is prior to the individual. [. . .] [L]anguage is seen as an abstract pattern located in the speech community and exterior to the individual. The human language faculty [. . .] is then viewed as the capacity to perceive, reproduce and employ this pattern.''

Apart from dialectology and sociolinguistics, until recently only few subdisciplines of linguistics were interested in inter- and intrasystemic variation. Differences between languages and especially between dialects or style levels were typically considered of marginal interest. This is rapidly changing. In the rule-based camp, there is a growing number of scholars who try to understand the smallest differences between dialects as manifestations of universal principles underlying the organization of language systems; the smallest difference (at the level of language as a system shared by the members of a community; cf. Chomsky's 1995 *E-language*) is thus explained on the basis of the highest common devisor (*I-language*, language as a cognitive commodity). Thus dialect features are sometimes explained as different instantiations of language universals or as instantiations of different language universals.

In the wake of developments in usage-based approaches to language, a subfield called Cognitive sociolinguistics has recently developed. It proposes a corpus-based method of describing and accounting for language variation which builds on the assumption that the clouds emanating from information stored in the mental lexicon contain, among other things, information regarding characteristics of both the speakers and the situations associated with specific variants. According to Hudson (2012, which contains some essential references) this approach to sociolinguistics differs from other approaches in that it focuses on individuals, their knowledge and behaviour, rather than on groups.

In this issue usage-based and rule-based paradigms of linguistic theory will be discussed from the point of view of variation in the sound components. The emphasis will be first and foremost on the answers that are typically given to two core questions regarding language variation: why does variation exist at all in a particular language system and how does it work?

Please cite this article in press as: Hinskens, F., et al., Grammar or lexicon. Or: Grammar and lexicon? Rule-based and usage-based approaches to phonological variation. Lingua (2014), http://dx.doi.org/10.1016/j.lingua.2014.01.005

3. Rule-based accounts

A useful stepping-stone for a concise comparison of rule-based and usage-based accounts is the question that is central in chapter 6 of Jackendoff's (2002) *Foundations of language*, viz. ''What aspects of an utterance *must* be stored in long-term memory, and what aspects *can* be constructed online in working memory?'' [emphasis in the source] (Jackendoff, 2002:152). The rule-based and usage-based answers to this question are very different. In rule-based (here: formal) linguistics the answer is: everything that is subject to grammatical or phonological regularities and therefore predictable (conditioned or not), e.g. the conjugation of weak verbs is not stored separately but is computed online instead. All idiosyncrasies, however, i.e. everything that is lexical or lexicalized, e.g. the conjugation of strong and irregular verbs, are memorized and stored – cf. Bloomfield's (1933) view of the lexicon as a list of exceptions – which is in line with De Saussure's claim that in general the connection between the form and the meaning of an item is arbitrary and unpredictable. This view is central in e.g. Pinker's (1999) words-and-rules model, also referred to as ''dual processing''. In the generative approach the regular, predictable properties of a language system are divided into general linguistic principles (''core grammar'', Universal Grammar or UG) which are assumed to be innate and those properties that have to be learned as they are language specific (''periphery'').[1]

In the course of its existence of over four decades, generative phonological thinking has been subject to many transformations and accelerations. In the approach that was inspired by Chomsky and Halle's *The Sound Pattern of English* (1968), the phonological part of grammar consisted of a small number of different modules. The interaction between these modules has been referred to as the modular feed forward model (Pierrehumbert, 2006:518), since it constituted a uni-directional process: lexicon > phonological encoding > phonetic implementation. The phonologically relevant part of a lexical item, the underlying form, which consisted of a sequence of phonemes which, in turn, consisted of lists of values of phonological features, were subject to one or more phonological rules. Under certain conditions, expressed in the rule's environment, these rules changed feature values, according to the following general format:

(1) $A \rightarrow B / C \_\_ D$

i.e. in the linguistic environment between C and D a phonological feature A is changed into B. For example, in the rule describing the process of syllable-final devoicing, A is the feature specification [-son] which targets obstruents, B is [-vce], meaning that obstruents become voiceless, C is not applicable, while D is]σ, indicating the right edge of a syllable (though not a unit in SPE). In this model, the phonological rules of a language were linearly ordered.

Let us now turn to sound change, i.e. diachronic variation. In the scenario of early (''linear'') generative phonology, causes of sound change and hence differences between an innovative and a conservative variety of a language may be due to either (a) the addition of a rule to one of the dialects, or (b) the loss of a rule from one of the dialects, or (c) a difference in the conditioning of a rule common to both systems, or (d) a difference in the ordering of rules common to both dialects. In all these cases, the rule machinery is the locus of inter-systemic variation. Whenever a rule applies optionally, it causes intra-systemic variation. When a rule loses its productivity and becomes lexicalized (i.e. lexically fixed), the underlying form is the locus of variation, which therefore represents a fifth scenario (e).

From the outset the principles underlying rule ordering have been a point of theoretical discussion, which, among other things, led to the development of Lexical Phonology. This designation indicates that the derivation of sound structure is no longer assumed to take place in one separate component of the grammar, but rather as divided across two components. In the earliest, lexical component, underlying representations may undergo morphological operations, which interact with lexical phonological rules (such as English trisyllabic shortening, i.e. shortening before two syllables in derived words, resulting in alternations such as op[ei]que vs. op[æ]city). Lexical phonological rules and morphological operations are subject to ''level ordering''. Lexical rules can be cyclic: they first apply to a smaller domain, then (typically after another morphological operation) to a larger domain, then again to a still larger domain, until the largest domain has been ''processed''. The output of the lexical component (or ''word grammar'') passes through the syntactic and post-lexical phonological components, resulting in the phonetic string.

Postlexical rules are independent of morphological structure; they share that property with *Lautwandel* in the Neogrammarian sense (cf. Hinskens, 1998). It is not uncommon for postlexical rules to be quantitatively variable (or ''optional'') and variable rules are typically postlexical in nature. Since all grammatical structure (both morphological and syntactic) which was erected in the course of the derivation is invisible to postlexical rules, their domain of application consists of prosodic units. Prosodic Theory treats these units as a hierarchy, which includes the syllable, the foot, the phonological word, the clitic group, the phonological phrase and the intonational phrase, with lower categories forming the constituents of the one that immediately dominates them, according to the so-called Strict Layer Hypothesis. The Strict

[1] This section contains a revised, extended and updated version of Section 3 in Hinskens (2011a).

Please cite this article in press as: Hinskens, F., et al., Grammar or lexicon. Or: Grammar and lexicon? Rule-based and usage-based approaches to phonological variation. Lingua (2014), http://dx.doi.org/10.1016/j.lingua.2014.01.005

Layer Hypothesis and the universal existence of the clitic group are not undisputed in the phonological literature (for discussion, see, recently, Grijzenhout and Kabak, 2009).

While Lexical Phonology is geared to the rule system and the interaction between the phonological system and other modules of the grammar, the nonlinear branch of generative phonology concentrates on aspects of phonological representation. There are two fields of interest, focusing on segment-internal and segment-external structure, respectively. Whereas segment-external aspects of sound structure are the subject of Metrical Phonology, which deals with issues such as syllable weight and related prosodic phenomena such as foot formation and stress (at the word-level), rhythm and intonation (at higher prosodic levels), segment-internal structure is central in Feature Geometry. The latter approach treats the universal inventory of phonological features as hierarchically ordered (Clements and Hume, 1995; Van Oostendorp and van de Weijer, 2005, eds.). This approach meets the objection from the sociolinguistic quarter that working with mutually unrelated distinctive features divides phonological space into as many independent dimensions, thus hampering the description of connected sound changes. For instance, the hierarchical organization of features in phonologically related subsets improves ''the possibility of describing in a coherent way a series of shifts moving around the periphery of the vowel trapezoid'' (Weinreich et al., 1968:149).

Lexical Phonology, Prosodic Theory, Metrical Phonology and Feature Geometry all have their own objects and since they supplement each other in several respects, they maintain orthogonal mutual relationships. Many of the insights of Lexical Phonology and Prosodic Theory have found their way into quantitative sociolinguistic studies of the word-final deletion of /t/ and /d/ (henceforth WFtD). This process applies variably, yet systematically (at least after obstruents), in varieties of English, Dutch and German.[2] The fact that compound words – which always consist of two or more phonological words – such as Dutch *lichtmast* (lit. 'light mast', 'lamppost') can undergo WFtD twice shows that the process targets phonological rather than grammatical words. The patterning of WFtD in ''several different varieties of North American English'' (Guy, 1991:4), and the dialect of Dutch spoken in Rimburg (Hinskens, 1992) suggests that this process applies both lexically and postlexically; in both systems, word-final coronal stops which are part of a monomorphemic word show the highest deletion rates, whereas /t/ and /d/ which have morphemic status show lower deletion rates, with the ones that are part of an irregular past tense (as in English *slept*) showing higher deletion rates than the fully regular, productive ones (as in English *worked*). In the Rimburg dialect of Dutch, WFtD has been lexicalized in several items, such as /kɔsə/ < /kɔstə/, 'to cost'. The Rimburg dialect, like neighbouring Limburg dialects, has a final [t] which has pronominal status. It occurs when in constituents of the type in (2a), in which the noun is grammatically neuter, the head is not expressed, yielding (2b):

(2a) $_{NP}$ [Det + Adnom + N]
e.g. ə lef kIŋk 'a sweet child'

(2b) $_{NP}$ [Det + Adnom-[t] + Ø]
e.g. ə lef-t 'a sweet [one]'[3]

In the Rimburg WFtD data, this particular [t], which must be added post-lexically, shows lower deletion rates than most morphemic [t]s (cf. Hinskens, 1992:244–246). These findings suggest that WFtD is an ''anywhere'' rule, which applies (optionally) wherever it can during the derivation – i.e. in the lexicon, morphology or syntax. Simplifying somewhat, one could say that [t]s which are part of a lexeme or irregular inflection form have three chances of being deleted (i.e. in each of the three main derivational stages), whereas among the derived [t]s the (regular) morphemic ones have two chances of being deleted and the pronominal ones only one. Both in the American data (Guy, 1991) and in the Rimburg data these differences are reflected in the actual deletion rates.

A comparison between the indigenous dialect with an ethnolectal variety and a ''daughter language'' of Dutch[4] reveals another interesting aspect of WFtD: in Jewish Dutch, WFtD was a variable and frequent process; in a few isolated cases, WFtD may have been lexicalized. In all other cases, final [t] was demonstrably present lexically or grammatically. In Jewish Dutch, WFtD did not appear to be sensitive to morphological structure, hence it applied merely postlexically.

In Afrikaans (as in Afro-American Vernacular English – Chuck Cairns, p.c.), on the other hand, WFtD has been lexicalized; a few isolated items, such as the numeral *acht*, 'eight', have retained their etymological final /t/ or /d/. The lexicalization of WFtD in Afrikaans has complicated the morphology, as in many cases the stop has become part of the plural ending (N) or inflectional suffix (adj), thus adding to the number of allomorphs, though on a lexically specific basis. Some nouns have two plural forms, one with and one without the etymological /t/ or /d/:

(3) kors: korste ~ korse 'crust(s)'

[2] Unlike English, Dutch and German have an automatic rule of final devoicing, as a result of which both /t/ and /d/ surface as [t] syllable-finally.
[3] Hinskens and Muysken (1986) and Hinskens (1992:179–181).
[4] Cf. Hinskens (2001:9–12) for sources, data and analyses.

Please cite this article in press as: Hinskens, F., et al., Grammar or lexicon. Or: Grammar and lexicon? Rule-based and usage-based approaches to phonological variation. Lingua (2014), http://dx.doi.org/10.1016/j.lingua.2014.01.005

Interestingly, in some nouns which do not have an etymological final [t] this alternation occurs too, e.g.:

(4) bos: bosse ∼ boste 'wood(s)'

Whether a noun does or does not have an underlying final /t/ or /d/ and whether it selects the plural allomorph with initial /t/ of /d/ is lexically stored (listed). This reflects scenario (e) above.

The comparison of WFtD in the Rimburg dialect, in Jewish Dutch and in Afrikaans shows that the same process can have a different rule-typological status in related language systems. A sound change can originate as a Neogrammarian sound change, viz. as an entirely productive and lexically exceptionless rule of phonetic implementation or as a postlexical process, and gradually percolate deeper into the language system. The older a process, the higher the chances that it becomes either grammatically conditioned or lexically stored. In the former case the process becomes part of the lexical component, where, as a lexical rule, it interacts with morphology. In the latter case the process is lexicalized, thus losing its productivity. A sound change which has been grammaticalised or lexicalized lives on either as a 'minor rule' (morphophonology) or as a correspondence rule, i.e. a device which allows lexical item-wise comparison between related varieties of a language, respectively. Postlexical processes and phonetic implementation on the one hand and lexical rules on the other share the characteristic that their application is entirely predictable on structural grounds; this is not the case for lexicalized sound change.[5] Before a sound change becomes lexicalized, it often lingers as a lexically diffuse sound change for a certain period of time. Rules which apply lexically diffusely are not conditioned by productive morphology. They have almost lost their productivity, although they are not entirely lexicalized yet. They eat their way through the lexicon in a word-by-word fashion; while being phonetically abrupt, lexically diffuse sound change is thus lexically gradual. Examples are the tensing and raising of /æ/ in certain British and American dialects of English and the palatalization of West-Germanic /u/ to /y/ in late Middle Dutch. (As borne out by the fact that there seems to be no single dialect with /y/ and /œy/ in the relevant lexical set, /y/ was exceptionlessly diphthongized in many dialects of early modern Dutch; see Kloeke, 1927).

Lexical Phonology predicts that this development (sometimes referred to as the 'life cycle' of a sound change) is directional in nature. In other words a development from lexicalized sound change through lexically diffuse to lexical to postlexical or phonetic change does not occur.

While the concepts of rule and derivation had faded into the background in nonlinear phonology in favour of representation (''if the representations are right, the rules will follow''), they were entirely abandoned in Optimality Theory (henceforth OT). In OT a set of constraints is assumed which determine the way in which the surface structure is allowed to deviate from lexically underlying representations. In principle all constraints are both

- universal, although their relative importance is language-specific. In this theory the constraints were originally conceived as the general linguistic principles ('core grammar'),[6] whereas the language specific ordering of the constraints (the 'periphery') must be learned, and
- ''soft'', as they can all be violated by candidate output forms in order to satisfy other, higher-ranked constraints.

The only generative capacity of the model resides in a function labelled GEN (''generator''), provided by UG. GEN projects an unlimited set of possible output candidates from a single lexical input form. It is up to the language learner to discover the underlying representations of the morphemes as well as the language-specific ranking of the universal constraints. All output candidates are rated according to their success in complying with the ranked set of constraints; the candidate that best satisfies the relevant high-ranked constraints is selected as the optimal one. This winning candidate is optimal, rather than perfect, since it will always violate certain lower-ranked constraints. In OT, analyses are typically summarized schematically in tables which are referred to as 'tableaux'. There are three types of constraints: Faithfulness constraints guard the phonological integrity of the lexical input form (and hence cognition), whereas markedness constraints serve phonological well-formedness. A third type of constraints shields the alignment of prosodic and grammatical structures. Examples of markedness constraints are NoCoDA, a constraint which disprefers syllables with a postvocalic consonant, and *COMPLEX^{COD}, a constraint which disprefers syllables with postvocalic consonant clusters (Kager, 1999:94, 97). A CC coda as in the English word *cost* is highly marked as it violates both constraints; NoCoDA is even violated twice.

[5] Kerswill (1987:25–26) points out that connected speech processes, ''can be involved in linguistic change (for example, German umlaut and Romance k-fronting before front vowels, giving Italian /a'mitSi/ (amici), are the result of 'fossilised' connected speech processes).''

[6] However, constraints have been proposed which are so specific that it may seem questionable if they are part of what is traditionally meant by ''core grammar''. Some scholars even posit word-specific constraints, cf. the discussion in Gabriel and Meisenburg (2009).

Please cite this article in press as: Hinskens, F., et al., Grammar or lexicon. Or: Grammar and lexicon? Rule-based and usage-based approaches to phonological variation. Lingua (2014), http://dx.doi.org/10.1016/j.lingua.2014.01.005

In OT analyses, variation resides in the constraint ranking. Variation *across* language varieties is usually described through subtle differences in constraint ranking between varieties; in this scenario (the ''multiple grammars'' model), all relevant constraints are completely ranked in competing grammars.[7] Every single variety is viewed as a categorical grammar of its own. Van Oostendorp (1997) accounts for style variation in French liaison, Dutch vowel reduction and Turkish vowel epenthesis in loan words, showing that as the style level becomes more elevated, the relevant faithfulness constraints are more highly ranked. Herrgen (2005) analyses cross-dialectal variation in WFtD in modern (southern and southwestern) dialects of German in the OT framework. Herrgen's data (from dialect atlases) concern two types of words with a sibilant + /t/ coda, viz. a monomorphemic noun and an inflected (2 sg pres) verb form for five groups of German dialects. Some dialects have [st], others [ʃt], yet others [s] (some of these dialects have [s] both in the verb and in the noun, some only in the verb), while [ʃ] only occurs in the inflected verb form. On the assumption that all dialects have underlying /st/, Herrgen analyses these five dialect groups with four constraints, which crucially differ in their ranking.

Although Lexical Phonology is derivational, its basic insights are not incompatible with OT. In Stratal OT (Bermúdez-Otero, 1999; Kiparsky, 2000), the output is generated and evaluated at three successive levels: stem-level, word-level, and phrase-level; the model thus combines parallel processing with derivation. The classical distinction between word phonology and phrase phonology can be implemented in OT in the form of Derivational Optimality Theory (DOT), as in Rubach (2000). In this approach the candidate output forms are evaluated in two steps, first at the word level, and subsequently at the post-lexical level.

Variation *within* varieties is typically analyzed as partial ordering of constraints. There are several variants of this type of analysis: the first one is based on the concept of floating constraints. Here, ''some particular constraints within a single grammar may be represented as falling anywhere within a designated range in the ranking hierarchy'' (Nagy and Reynolds, 1997:37), in other words, their position in the hierarchy is not entirely fixed. It would be relatively simple to recast Van Oostendorp's (1997) account of style variation in three distinct languages as subtly differing yet regularly related grammars into the floating constraints approach. Meanwhile floating constraints seem to have disappeared from the stage.

The second variant employs unranked constraints, in which several constraints coincide entirely, producing multiple optimal candidates as output, although not every logically possible ranking emerging from this scenario will necessarily yield another winner. This approach, of which several different versions have been developed and successfully implemented by Anttila (1997, 2002), enables predictions concerning the proportions of occurrence of the relevant variants. In Anttila's (2002) stratified grammar model constraints are mutually unordered within strata (sets of mutually unordered constraints); the fact that there can be unranked constraints in several different strata makes the number of possible grammars (i.e. constraint rankings) potentially very large; since here, too, not every grammar necessarily selects another type of candidate as optimal, clustering of optimal candidate types occurs; the clusters' proportions appear to be reflected in the actual quantitative variation; ''if a candidate wins in n tableaux and the total number of tableaux is t, then the candidate's probability of occurrence is n/t'', in the summary of Barbiers and Hermans (2007:4).

A closely related way to capture intrasystemic variation is by associating a relevant domain of the language to a different phonological grammar, i.e. a different constraint ranking. This concept is known as 'co-phonology' and it is also based on the insight that a single language system can have several co-existing distinct phonological systems. Each subsystem is indexed to a certain domain of the language and consists of a specific ranking of the universal constraints; variation between the domains results from different rankings of the constraints between co-phonologies. These distinct phonological systems are specific to domains such as the word, a given subset of the lexicon such as specific strata of non-native words, a register, a lexical class or a morphological category. Within a given co-phonology all constraints are fully general; since it builds on completely ranked, categorical grammars, this approach is akin to the ''multiple grammars'' model, briefly sketched above.

The third variant is fundamentally different from the other two. In the case of continuous ranking (Zubritskaya, 1997) and stochastic OT (Boersma and Hayes, 2001), the constraint ordering is not discrete, i.e. it is not based on integers, and hence the range of the intervals between neighbouring constraints can vary. On the ranking scale, every constraint occupies not just one single point, but rather a range of values, and as a consequence constraints can coincide partly or even entirely; the variation resides in the overlap of constraints.

This summarizes in our view the most important approaches to variation within OT. The same set of options would basically apply to Harmonic Grammar, in which constraints are assigned a specific weight, rather than an absolute position in the constraint ranking (Smolensky and Legendre, 2006; Pater, 2008, 2009, 2012). The implication here is that lower ranked constraints can 'team up' to match the violation of higher-ranked constraints. However, as far as we can see,

[7] In connection with 'causal principles' for language universals, Hawkins (1988:21) sketches this particular *avant la lettre* Optimality Theory-type approach to cross-dialectal variation in the scenario ''[. . .] two principles may be in conflict, and the resolution of the conflict may result in variation as languages opt for different solutions.''

Please cite this article in press as: Hinskens, F., et al., Grammar or lexicon. Or: Grammar and lexicon? Rule-based and usage-based approaches to phonological variation. Lingua (2014), http://dx.doi.org/10.1016/j.lingua.2014.01.005

this does not crucially affect our discussion of variation. A stochastic Harmonic Grammar model is discussed in Boersma and Pater (in press), for instance.

The relation between formal phonological theory and the study of language variation is far from unproblematic; cf. the tension between approaches to language as a cognitive object vs. a social object, touched upon in Section 2. Referring to 'autonomous phonology' as ''any approach to phonological investigation that assumes that the object of phonological enquiry can be studied in its own right, relatively independently of the study of factors such as the social context in which the speakers are located,'' Carr (2000:74, 84) claims that ''autonomous phonology is, in one sense, the best contender for a phonology as a properly scientific discipline. But [. . .] a fully autonomous phonology is unsustainable, since the data to be accounted for cannot be divorced from social context and are inherently variable.'' The latter aspects are to a large part matters of convention and there is a non-obvious ''relation between sociophonetic variation and UG as a natural object'' (96). Which role, then, does the 'natural' play in the 'conventional'?

Some quantitative sociolinguistic studies of intra-systemic phonetic and phonological variation explicitly adopt generative theories (linear, nonlinear and OT), although the focus is not on categorical, fully determined phenomena, but rather on variation and gradience, which are modelled through probabilistic quantification. In principle, language variation is explained internally as a result of the organization of the language system, which, in turn, is assumed to be subject to universal laws. Usually the theory serves to derive hypotheses concerning the nature and distribution of the variation; linguistic factor groups play a role as dimensions along which the variation patterns (typically in addition to or in interaction with extra-linguistic, usually macro- and micro-social, factors). With regard to WFtD, the grammatical status (either part of a lexeme or bound morpheme) of the stop appears to be an important dimension, and so does the place of articulation of the preceding obstruent. Typically, there is more deletion after a coronal (*in casu* sibilant) than after obstruents with other places of articulation (Guy and Boberg, 1997; cf. Van Hout, 1989:187). This finding is an instantiation of an alleged universal principle known as the Obligatory Contour Principle, a constraint prohibiting adjacent identical elements, i.e. in this case the features [coronal] and [-son] on two neighbouring consonants.

It appears that linguistic variables which occur in different languages are often influenced in the same way by the same or very similar linguistic factors (cf. Tagliamonte, 2011). This implies that the internal conditioning of language variation can contain important indications for possible universal constraints. This holds *a fortiori* for Bresnan et al's (2001) 'stochastic generalizations', i.e. generalizations which are categorical for some language(s) but probabilistic in others.

4. Usage-based accounts

From the point of view of usage-based approaches, the answer to Jackendoff's (2002:152) question ''What aspects of an utterance *must* be stored in long-term memory, and what aspects *can* be constructed online in working memory?'' will probably be: in principle everything is stored, whether regular or not, and whether predictable or not. Computation (the system of which is the research object of generative theory) plays a relatively small role. Storage takes place in the mental lexicon. In this approach the lexicon is not a list of exceptions, but rather a network of prototype-wise organized words, phrases and constructions, i.e. ''multi-word combinations whose properties cannot be fully accounted for compositionally'' (Booij, 2004:234). All types of regularities (and hence predictability and productivity), including grammatical structure, emerge bottom-up, climbing a ''ladder of abstractions'' (Pierrehumbert, 2003), from the information in the lexicon, through conventionalisation, which in turn results from repetition, hence from distributional frequency and frequency of usage. Bloomfield's view on the division of labour between lexicon, grammar and phonology, the essence of which was adopted by generative theory (sketched in Section 3 above; cf. Pinker's words-and-rules model), is called the 'rule/list fallacy' among usage-based linguists. The logical consequence of the usage-based view is that ''given that the set of utterances which any child hears in the course of language acquisition will be different from that of the next child, with different frequencies of, e.g. word variants, [. . .] every individual's grammar will be different at the level of phonetic implementation'' (Blevins, 2004:41). Language is ''an emergent system resulting from the general cognitive capacities of humans interacting with language substance over many instances of language use'' (Bybee, 2001:18). According to Goldberg (2006), this is the reason why most universals are only tendencies.[8]

This young paradigm, which, apart from usage-based theory, also includes *Exemplar Theory* (henceforth ET), *Construction Grammar* and *Cognitive Grammar*, represents an entirely different approach to phonology, morphology and syntax. These theories are inspired by connectionism, a school of thought in cognitive science which attempts to explain mental and/or behavioural phenomena as emergent processes of a self-organizing network of mutually connected units. ET was first introduced in psychology as a model of perception and categorisation (Nosofsky, 1986; Barbiers and Hermans (2007:2).

[8] This section contains a revised, extended and updated version of Section 4 in Hinskens (2011a).

Please cite this article in press as: Hinskens, F., et al., Grammar or lexicon. Or: Grammar and lexicon? Rule-based and usage-based approaches to phonological variation. Lingua (2014), http://dx.doi.org/10.1016/j.lingua.2014.01.005

Some central notions in usage-based theory may require further elucidation. First, the notion of connectionism needs some clarification. The prevailing variant of connectionism attempts to explain human intellectual skills with the aid of artificial neural networks, simplified models of the brain, which are composed of numerous units (representing neurons) which are interconnected. The weights which are attached to the individual connections indicate their relative strengths. In this system of weighted connections, certain patterns of neural activation can take shape as a result of which the activation of one particular unit is related with the activation of a number of other units. In connectionist models of language acquisition, an important question is how many constraints can be implemented into a network for the sake of the adequate construction of grammar on the basis of the input. The constraints are limitations or specifications of the type of associations the network can make (Van Geert, 2000:461–462).

The notion of self-organization can be illustrated with insights from the biological study of embryogenesis. In the past decades, it has become clear that genes are not building plans of the body, but merely specify the starting points of developmental processes. The specific shape of the body is determined by the interaction between all the parts that are involved in the process; this is a self-organizing process (cf. Van Geert, 2000:459).

The notion of emergent process, finally, refers to the emergence and subsequent development of (a) spontaneous order out of chaos – e.g. billions of neurons constitute one's consciousness and memory – and (b) new coherent patterns, structures and/or properties during the process of self-organization of a complex system. What is commonly referred to as speech organs are originally merely ''intended'' for eating, drinking and breathing. Speech is thus also an emergent property – although this view is probably not an essential part of any usage-based model.

In usage-based approaches to language, as theoretically laid out by Bybee (2001), among others, regularities of several types as well as processes of language variation and change are usually accounted for along quantitative lines, on the basis of distributional and usage frequencies (or type and token frequencies). Most linguists of this persuasion subscribe to the view that ''the cognitive and psychological processes and principles that govern language are not specific to language, but are in general the same as those that govern other aspects of human cognitive and social behavior''. Those facilities and skills include ''our enormous memory capacity, fine motor control, the ability to categorize experience, and the ability to make inferences'' (Bybee, 2001:17) as well as attention (Pierrehumbert, 2006:519). This is summarized in the thesis that language is grounded in domain-general cognitive processes (Bybee, 2010:1; cf. Diessel, 2011:831; see also Van de Weijer, this volume). Linguistic competence is a complex pattern of interactions between mental properties and capacities that are independently available. There is no mental blueprint which is specific for language (cf. the metaphor of the Invisible Hand Keller, 1994). Usage-based approaches are functionalist in nature. Extra-linguistic explanations are far from uncommon, although strictly speaking in this view the distinction between internal and extra-linguistic is not applicable, just like that between synchrony and diachrony (Bybee, 2010).

In view of the enormous importance of Bybee's (2001, 2006) model for the usage-based approach to variation and change in the sound components, some of its main aspects will be discussed here. Fundamental is the point of view that language is the bridge between phonetics and semantics, which together form the substance of language[9] and that its use (i.e. processing as well as the social and interactional uses) determines substance and language structure (2).[10]

Words, or more specifically lexemes, are lexically stored in associative networks; the associations are phonetically or semantically motivated. Some associations are stronger, others are weaker. ''Lexical organization provides generalizations and segmentation at various degrees of abstraction and generality [. . .] [S]torage in this model is highly redundant'' also because ''multimorpheme words are stored whole'' (7–8). Type frequency or family size is defined as ''the dictionary frequency of a particular pattern (e.g. a stress pattern, an affix, or a consonant cluster)'', in other words the lexical frequency of occurrence or co-occurrence[11] of a particular segment or sequence of segments, the number of words which have a similar phonological shape ('neighbours' – Luce and Pisoni, 1995); this number determines an item's 'lexical connections'; they constitute a 'schema' (cf. below). Token frequency is defined as ''the frequency of occurrence of a unit, usually a word''; this number determines the unit's lexical strength. Token frequency can also concern the collocations of individual items; the higher the token frequency of an item or pattern, the deeper its entrenchment. Given the prominent role played by frequency as the stem cells of grammar,[12] corpora are essential to usage-based and related approaches to language.

Schemas are ''non-process statements about stored items'' (22); there are at least four reasons (set out on p. 27) why a schema is not the same thing as a rule; one of these reasons is the fact that schemas can be gradient. As ''organizational

[9] Obviously, basic insights from Hjelmslev's *Glossematik* played an important role here, although references are lacking.
[10] In the remainder of this section, numbers between brackets refer to the page number in Bybee (2001).
[11] With all implied probability values, including, for example, ''transitional probabilities'' like ''probability of a C given a final V'' (Coetzee, 2008:250).
[12] As Jeroen van de Weijer (personal communication) points out, this comparison applies only partly, as frequency is a concept, whereas stem cells are items.

Please cite this article in press as: Hinskens, F., et al., Grammar or lexicon. Or: Grammar and lexicon? Rule-based and usage-based approaches to phonological variation. Lingua (2014), http://dx.doi.org/10.1016/j.lingua.2014.01.005

patterns across lexical items'' (28), schemas result from ''generalisations about linguistic units [. . .] by speakers as they categorize items for storage'' (22). Categorisation is based on analogy which, in turn, goes back to identity or similarity; on the basis of ''similarity'', gradience plays a role. Cf. ''One of the central issues in cognitive research on phonology and phonetics is categorization and gradience'' (Pierrehumbert, 2006:519). For instance, Ernestus and Baayen (2003) analyzed the underlying voicing/voicelessness of obstruents at the end of stems[13] of productive and fully regular Dutch verbs in connection with distributional frequencies; they traced subregularities which even enable language users to predict the voice specification of obstruents in nonsense words on the basis of phonologically similar morphemes. A similar study is Albright's (2003) experimental approach to the acceptability of conjugational variants of nonce infinitives in Italian, leading to the discovery of type frequency-based subregularities connected with particular phonological contexts, constituting 'islands of reliability', which serve as anchor points for the development of hypotheses concerning the subregularities at issue.

The lexicon does not contain stripped off phonemic representations of items, disrobed of all predictabilities. Abstract notions such as phonemes or segments, let alone distinctive features, do not exist in this framework, and there is no principled distinction between phonetics and phonology. The effects of phonological regularities are stored, and the regularities themselves emerge. Cf. ''phonological categories, as represented in the mind, are viewed as clusters of similar experiences'' (Pierrehumbert, 2006:519) and a language's phonological system, ''being an extraction from pronunciations of the words and phrases of the language'' (Phillips, 2006:3) is nothing more than an epiphenomenon. All encountered and stored realisations of an item ('exemplars') are substance-rich, full with fine (gradient, subcategorical) phonetic detail, and each single token is furnished with multi-modal episodic information about its phonetic and morphological context, semantic, pragmatic and social factors.[14] In these 'multiple trace models', the subtle characteristics of speakers' voices as well as the speakers' age, gender and dialect are supposed to be stored in association with the exemplars. All exemplars are organized in clouds, around a prototype and only ''rarely used, non-salient variants might fade from memory'' (52; cf. 6–7). Thus, old tokens can vanish (Barbiers and Hermans, 2007:3). On the level of the speech sounds, the prototype concept is reminiscent of Paul's (1880) '*Lautbild*', lit. ''sound imagine'', i.e. a mental representation of a given segment.

Prototypes can change with the actual tokens (or 'instances') of use, also on the level of the individual speaker, cf. ''probability distributions are acquired in great detail through experience [. . .] [and] continue to be updated in adult life'' (Pierrehumbert, 2006:523). Prototype change can consist of ''gradual phonetic change in categories, as well as in particular lexical units'' (p. 58). With Pagliuca and Mowrey (1987), Bybee assumes that most processes of sound change are articulatorily motivated; two types of gestural change can be distinguished: ''retimings [or compression] that cause overlap of gestures, and reductive changes that reduce the duration or magnitude of gestures'' (58). Gestures are series of actions performed by groups of muscles; individual gestures can include several articulators (cf. Browman and Goldstein, 1992). The magnitude of a gesture is reduced to zero in cases of deletion, and the retiming prevents a gap to be formed inside the utterance. Retiming (which, apart from deletion, also includes insertion and assimilation) is temporal, while reduction (which in the case of consonants involves lenition) affects the phonetic substance.[15] In these processes, repetition and thus token frequency (also locally, i.e. during a conversation – through priming or 'recency effects') plays an important role. There are two reasons for this. For one thing, the more frequently an item is used, the more strongly it is represented in the mental lexicon and hence the more easily accessible it is in online production. For another, speakers can afford to be more 'sloppy' in realizing them, because as a consequence of their stronger mental representation they are more easily predictable for the listener (Diessel, 2011:833). Of particular relevance is Gahl's (2008) study, which shows ''that homophone pairs in a corpus of spontaneous speech differ in duration, with high-frequency words being shorter than their lower-frequency homophone twins'' (Gahl, 2008:487), a finding, Gahl argues, which can only mean that it is lemma frequency rather than the frequency of underlying phonological forms which determines language production. This effect of token frequency cannot easily be reconciled with, let alone be incorporated in a rule-based model.

In the usage-based view on language, the fact that words with high token frequencies are subject to reduction effects does not contradict the observation (by Bybee, 2001, 2006:715, but also by scholars of older generations, such as Sturtevant) that, because of their deep entrenchment, high token frequency makes a form resistant to analogical change, which typically involves the morphology. This mechanism, labelled the Conserving Effect, underlies, for instance, the conjugation of many strong and irregular verbs, including English *be*, *have*, *come*, *do*. For Dutch, similar observations were made by Van Haeringen (1940).

[13] In syllable-final position, the voice contrast in obstruents is automatically neutralized in Dutch. Cf. footnote 1 above.

[14] However, according to Labov (2006:513), ''it is the stem that is socially evaluated, not the word as it is delivered with its inflectional affixes intact. In other words, social evaluation is performed on a lemmatized lexicon [. . .] [T]he social information is attached to a more abstract object than the remembered exemplars.''

[15] Cf. also Bybee (2006:714–715).

Please cite this article in press as: Hinskens, F., et al., Grammar or lexicon. Or: Grammar and lexicon? Rule-based and usage-based approaches to phonological variation. Lingua (2014), http://dx.doi.org/10.1016/j.lingua.2014.01.005

The role of type frequency in language change can be illustrated on the basis of Bybee's account of the gradual erosion of French liaison. This process is explained on the basis of the fact that the construction [det N adj]$_{plur}$ ''with an unmarked noun and an adjective'' has a higher type frequency ''which makes it more productive'' than that with a vowel-initial adjective, so that e.g. *des enfants [z] intelligents* [dɛz̪ɑ̃fɑ̃zɛ̃tɛliʒɑ̃] 'intelligent children', will gradually make way for [dɛz̪ɑ̃fɑ̃ɛ̃tɛliʒɑ̃]. ''Thus, it is not surprising that there is variation in the data'' between the occurrence and the non-occurrence of the liaison [z] (175). What is at work here is the 'rich-get-richer' phenomenon, also known as the 'gang effect' (McClelland and Rumelhart, 1981:304), based on similarity.

In this hardware-oriented view, the substance of phonological patterns and processes is phonetic in nature, more in particular motor-articulatory (gestures) and acoustic-perceptual representations (64). ''Phonological knowledge is procedural knowledge rather than static propositional knowledge'' (70) of discrete and abstract entities. Cf. Nathan's (2007:623) ''phonology is about entrenched motor skills.'' Formal universals result from ''common trajectories of change,[16] which themselves are fuelled by the human neuromotor ability to produce complex motor sequences, by storing and automating recurring stretches of behaviour and by the cognitive ability to categorize and organize stretches of linguistic experience in memory'' (215). Since language change is not only driven by general cognitive processes but also by specific social and cultural factors, universals are statistical and dynamic in nature (Bybee, 2010:201).

5. Pros and cons

Generally, one of the facts which incontrovertibly plead for Bybee's model is the cross-linguistic tendency that the more frequently words are used, the shorter they are on average (cf. Crystal, 1987:87). Cf. Zipf's (1935) observation that ''it seems reasonably clear that shorter words are distinctly more favoured in language than longer words.''

For the study of language variation usage-based approaches have attractive sides. First, variability is directly represented in memory in the shape of concrete exemplars, which are assumed to contain social-indexical information as well (cf. Docherty and Foulkes, 2000); many adherents of this approach use the label 'sociophonetics' to refer to Exemplar-Theoretical sociolinguistics. The link between ET and the sociolinguistic study of quantitative variation seems convincingly simple: in Pierrehumbert's view, an exemplar's token frequency determines its influence on perception/recognition and production/phonetic output, although ''the frequencies relating to the perceptual system are not necessarily the same as those relating to production'' (2006:524). In this way, usage-based models, which do not assume economical, canonical (hence 'abstract') lexical representations, have capacities which rule-based models lack. Most quantitative studies of language variation seem to show that internal conditions tend to have significant probabilistic effects. There is almost always a residue of variance which cannot be explained by any linguistic or extra-linguistic group of factors. An example is the reduction of unstressed vowels in Dutch; in a word such as *almanak* 'almanac', the /a/ in the second syllable is usually realized as schwa, but reduction to schwa is much less common for the /a/ in the second syllable of another integrated three-syllable loanword such as *karavaan* 'caravan', which is much less frequently used. One of the variants of the latter word has primary stress on the initial syllable. Both *almanak* and *karavaan* are nouns and have identical position of metrical stress (immediately following the syllable with word stress), and in both cases the stressed segments are identical as well. Moreover, in both cases the segments are in open syllable position. Usage-based models can in principle account for the fact that these items nevertheless differ in terms of the reducibility of the unstressed vowel – and for the phenomenon of lexically diffuse sound change more generally, which continues to be problematic for formal approaches.[17] Second, the model is not based on deterministic principles (in the way formal theories are), but rather probabilistic; as such it matches the nature of most documented instances of language variation. Third, just like adherents of usage-based approaches to language, many sociolinguists studying language variation reject the analytical distinction between diachrony and synchrony (applied by adherents of formal theories), as variation is often a synchronic reflection of an ongoing process of language change.

However, as Paul Boersma (personal communication) has pointed out, there is an annoying conceptual problem about the emergent nature of categories and generalizations, which paralyzes usage-based models. Enriching the exemplar cloud (from which all kinds of lexical, phonetic and grammatical categories emerge) with social categories potentially dilutes the cloud considerably, since it will have very many dimensions. Put differently, there will probably no longer be a coherent cloud, but rather a collection of isolated points in a multi-dimensional space. It will thus become very difficult to draw any generalisations. A related problem pertains specifically to (macro-, meso- and micro-) social categories. Since social categories are assumed to be just as emergent as linguistic ones and since the child acquiring its mother tongue is *tabula rasa* in every respect and is thus not assumed to have any ready-made social categories to attach to the exemplars it stores, the number of potentially relevant social categories will roughly be as large as the number of exemplars.

[16] A position which is also taken by Blevins (2004:23).
[17] Other weaknesses, some of which are connected with the analysis of variation, have been briefly pointed out above.

Please cite this article in press as: Hinskens, F., et al., Grammar or lexicon. Or: Grammar and lexicon? Rule-based and usage-based approaches to phonological variation. Lingua (2014), http://dx.doi.org/10.1016/j.lingua.2014.01.005

The space for the linguistic categories is enormous, but still considerably smaller (more finite, so to say) than that for social categories, resulting in potentially relevant social categories grounded in one single exemplar, so that generalizations across social aspects will be virtually impossible.

Linguistically, one of the general problems theories in which higher order units such as segments do not have an independent status are faced with concerns ''behaviours that children exhibit while acquiring language'' (Nathan, 2008:153). Phenomena of this type concern, for instance, systematic substitutions of particular (and typically relatively marked) segments, such as replacing the voiceless dental fricative /θ/ by labial [f], a simplification process which tends to be fully regular and exceptionless, despite the fact that the child has probably never been exposed to variants with [f], so no exemplars with /f/ have been stored.

A point of concern is also Bybee's conviction (2001:13, 26) that ''the productivity of a pattern, expressed in a schema, is largely [. . .] determined by its type frequency'', hence the higher the type frequency (and the stronger the similarity with stored items; see Bybee, 2010), the higher the productivity. The idea that productivity is gradient is also accepted in generative phonology (as shown, for instance, by the difference between postlexical and lexically diffuse sound change), but for the rest one could just as well maintain that type frequency results from productivity.

The statistical analyses of the effects of (token) frequency are sometimes surprisingly simplistic. Many researchers work with dichotomisations of this continuous variable, rather than with the raw numbers or (e.g. logarithmic) transformations thereof. Bybee (2000, 2001:112–113) demonstrates that there is a significant positive connection between WFtD in English regular past tense or past participle verb forms ending in *-ed* in non-prevocalic position on the one hand and token frequency on the other. In Bybee's analyses, the cut off-point between high and low frequency (i.e. the level where activation is supposed to gain full momentum) is set at 35 per million, while in the case of the lexical decision experiment with regular English past tense forms carried out by Alegre and Gordon (summarized in Bybee, 2001:111) the cut off-point is set at 6 per million. This makes one wonder about the outcomes of both studies if the cut off-points had been 6 per million in the Bybee study and 35 per million in the Alegre and Gordon case.[18] If, however, the different choices of cut-off points could independently be shown to be valid for each of the two types of data, the question arises why lexical effects for identical word sets can differ so drastically across task types.

As Walker (2012:407) puts it: ''if frequency has the consistent, monotonic effect predicted by a usage-based account, this effect may be masked by dividing the continuum of frequency into discrete categories''. Walker presents an advanced statistical analysis of variable WFtD in Toronto English in which the lexical items, including all their properties (such as frequency of usage, as a continuous variable), were treated as 'random effects', thus controlling for idiosyncratic lexical effects. In contrast with earlier multivariate analyses in which frequency of usage was treated as a nominal variable with arbitrary distinctions between levels ('high frequency' and 'low frequency'), no significant frequency effect emerges from Walker's statistically superior analysis.

On the other hand, Guy (2013) argues that, apart from continuous effects, discrete or threshold effects of token frequency are also conceivable – and in some cases do actually occur. In the latter scenario, in which frequency does not operate monotonically, there are no fine frequency gradations and it is conceivable that there is no independent frequency effect, but rather interaction effects between token frequency and lexical-phonological (Walker, 2012:412) or other internal variables, as in Erker and Guy's 2012 data regarding pro-drop, admittedly a syntactic phenomenon. For pro-drop in New York City Spanish (in the Otheguy-Zentella corpus), Erker and Guy found that the effect of the factor group present/preterite/imperfect is greater in more frequent forms (p. 544), while the effect of other factor groups, such as regular vs. irregular verbs, is neutralized (insignificant) in infrequent forms (pp. 542–543). Moreover, the effect of higher frequency is not constant: in some contexts, higher frequency favours more overt subject pronouns, while in others there are fewer subject pronouns. It is not clear whether similar complicated constellations of interaction effects can be expected outside of the syntactic modules.

In usage-based accounts, little attention is generally paid to the question where the relevant frequency counts come from, i.e. what they are actually based on. Are these available counts for large corpora? Or are they local corpus counts – as in the studies cited in the previous paragraph and in Erker and Guy (2012)? In the case of sociolinguistic variation, the corpus counts may even need to be carefully controlled for (linguistic and extra-linguistic) context and speech style.

With a few notable exceptions (Bybee, 2001, 2010; Docherty and Foulkes, 2000; Phillips, 2006), usage-based models and ET are relatively poor in terms of theoretical development; there are no handbooks concerning usage-based and/or ET approaches to phonology. Some of the related research seems to surf along on top of the wave of the 'digital humanities' (or 'computational humanities' or 'e-humanities') with its trend of sometimes theory-free data-mining of giant

[18] Cf. also the study of the deletion of medial /d/ = [ð] in New Mexican Spanish non-past participle forms, where the cut off-point is 123 on the total of 1040 tokens in the corpus of recordings (Bybee, 2001:148-149). In her study of vowel reduction in Dutch, Kloots (2005:135ff) distinguishes four arbitrary levels of token frequency. Clark (2008:262–265) confines her bivariate analysis of the use of the Scots monophthongal variant of English <ou> to items with a token frequency of >10 in her corpus.

Please cite this article in press as: Hinskens, F., et al., Grammar or lexicon. Or: Grammar and lexicon? Rule-based and usage-based approaches to phonological variation. Lingua (2014), http://dx.doi.org/10.1016/j.lingua.2014.01.005

amounts of data ('big data'). This specific approach often concerns the distillation of potentially meaningful patterns, rather than the testing of theories. The findings (which are not always genuinely new, although they are sometimes presented as such) are not always viewed from a wider perspective and sometimes explanations are far to seek.

A related aspect is that usage-based and ET studies tend to be slightly elusive; hard, directional predictions are sometimes absent, and in some cases explanations are post hoc. The findings are fairly heterogeneous and occasionally in apparent mutual contradiction, although in all fairness it should be pointed out that, to a certain extent, this holds for formal theories as well.

Interestingly, rule-based and usage-based theories of language are parallelled by two important approaches which are central in the modern natural sciences: on the one hand *reductionism*, i.e. the method which consists of reducing complicated phenomena to their elementary building blocks (e.g. atoms and subatomic particles; DNA), and on the other hand *emergence*, the insight that complicated phenomena can be more than the sum of their parts and can only be understood through the interactions of their constituent parts, such as self-reproducing life, developing out of the multitude of chemical reactions within the cells (Dijkgraaf and Fresco, 2008:16–17).

The discussion between rule-based and usage-based accounts of language concerns cognitive capacities (computation) vs. substance (storage). The debate is reminiscent of the debate regarding the role of nature (hereditary aptitude, innateness; genotype) vs. nurture (environment as the source of language input and training, repetition, and hence frequency; phenotype) in language acquisition. In the history of linguistic thinking, the discussion between rule-based and usage-based theories is parallelled by the debate between the first generation of Neogrammarians (grammatically blind and lexically exceptionless sound laws, the effects of which are merely obscured by analogy or borrowing) and the position of late nineteenth century mainstream dialectology, which explicitly takes idiosyncrasies of individual lexical items into account. One of the main representatives of this dialectological approach was Jaberg, whose dictum ''In Wirklichkeit hat jedes Wort seine besondere Geschichte'' (1908:6; 'in reality each word has its own particular history') has become famous. In his programmatic *Über die Lautgesetze: Gegen die Junggrammatiker* from 1885, Schuchardt claimed that frequently used words tend to be ahead in sound change. In purely synchronic linguistics the debate is relatively new.

Evidently, usage-based types of approaches to language differ in many respects from formal, rule-based approaches. The two paradigms have different views of language structure, the organization of grammar, modularity, linguistic competence, the role of psycholinguistics, and the division of labour between storage and computation in human cognition with respect to language production and processing, language acquisition as well as language variation and change. Hualde (2011:220) summarized one of the central differences between rule-based and usage-based approaches to phonological variation as ''phonemes or words as objects of phonological change''.

In addition there is at least one more central point of difference and this can be summarized in the notion of 'gradience', briefly touched upon in the previous section. Formal theories often presume categoriality, discreteness and at most binarity. A multi-valued feature such as sonority is (a) a borderline case and (b) not a primitive in any feature theory. Neogrammarian sound change is exceptionless (and often phonetically gradient); it can subsequently become prosodically, morphologically or lexically conditioned, but at least in the first two dimensions gradience as it can often be observed in the case of frequency effects does not play a role. By contrast, in usage-based approaches in principle all linguistic categories are conceived as gradient, fluid and organized around prototypes.

Analysing data concerning present-day type frequencies from six Bantu languages with five-vowel inventories and height harmony and six control languages with five-vowel inventories lacking harmony, Archangeli et al. (2012) compare predictions regarding vowel harmony derived from UG-flavoured phonological theory with what they refer to as Emergent Grammar, finding massive evidence in favour of the latter: (1) patterns (harmony outranking faithfulness or vice versa) in both types of systems tend to be loose, rather than (near-) categorical as predicted by UG theory, and (2) patterns tend to extend in both morphosyntactic and phonological domains, which is not predicted by UG theory. This leads the authors to plead for a more 'emergentist' approach to exceptions and statistical generalizations.

An entirely different message is put forward by Becker et al. (2011). On the basis of type frequencies, the authors present a range of statistical generalisations regarding Turkish laryngeal alternations in nouns marked with a possessive suffix. The authors conducted an experiment with novel words for which the subjects were forced to choose an allomorph. The subjects appeared to base their choices exclusively on phonologically (UG) motivated generalizations and disregard significant yet 'accidental' generalizations. The authors conclude that domain-general (i.e. not specifically linguistic) learning may well be successful, but in this case it is evidently supported by domain-specific (here: linguistic) filters (119–120). Obviously, there is a phonological bias in the data and not only frequencies are involved here. A comparable study (concerning voice alternations in the plural forms of mono- and polysyllabic nouns in English on the basis of experiments with nonce words) with similar outcomes is Becker et al. (2012).

While the findings of the study by Archangeli et al. yield positive evidence for usage-based approaches and negative evidence for rule-based models, the reversed picture emerges from the two studies by Becker et al. It is not immediately clear whether these contradictory findings are related to the fact that Archangeli et al. deal with productive phonology and

Please cite this article in press as: Hinskens, F., et al., Grammar or lexicon. Or: Grammar and lexicon? Rule-based and usage-based approaches to phonological variation. Lingua (2014), http://dx.doi.org/10.1016/j.lingua.2014.01.005

both studies by Becker et al. with productive morphophonology. Archangeli et al. mainly work with lexicographical data (from the Comparative Bantu OnLine Dictionary, CBOLD), whereas the two Becker et al. studies are based on experimental production data, but it is again unclear whether this methodological difference might play a role in the opposite findings.

6. Hybrid models. The contributions to this special issue

What usage-based and rule-based approaches of language have in common is, first, the fact that they were not specifically designed to handle language variation. Second, both types of model have pronounced universal ambitions. Third, both approaches recognize that a non-trivial learning theory is required; in either view, the challenge is to account for the acquisition of a system with infinite possibilities on the basis of finite bodies of evidence. All three properties follow from the fact that both paradigms look at language as a cognitive object.

Possibly on this basis and on the basis of the sense that the supporters of the opposite approach may have answers or at least methods to find answers that the own approach cannot provide, some scholars have tentatively come up with sketches of some of the contours of hybrid models. The available hybrid models are always clearly developed on the basis of either usage-based or rule-based models; for instance, this holds for Pierrehumbert (2002), McQueen et al. (2006) and Hay et al. (2013). Some proposals for hybrid models will be succinctly sketched in this section, along with the contributions to this volume, all of which contain building blocks for one or several bridges across the paradigmatic gap.

In this special issue, both Docherty and Foulkes and Ernestus explicitly plead, from an ET-position, for hybrid models. ***Docherty and Foulkes*** thoroughly discuss both manifest and latent possibilities in ET for dealing with the production, perception and acquisition of socially emblematic information alongside referential meaning codified in speech. Furthermore, the authors present a critical appraisal of ET, concentrating among other things on the foregrounding of the individual speaker and on the need to find the balance between top-down and bottom-up processing. In passing, Docherty and Foulkes point out that usage-based approaches may well be instrumental in substantiating the concept of 'salience', which plays a role in the sociolinguistic status of a phonetic or phonological phenomenon – as well as (one might add) in perception and language acquisition. The relationship between top-down and bottom-up (roughly: phonological and phonetic) processing also features prominently in the considerations with which ***Ernestus*** concludes her well documented contribution on a range of articulatory and perceptual (partially psycholinguistic) aspects of acoustic reduction in speech. Although, according to Ernestus, the mental lexicon contains representations for at least some reduced variants, the comprehension of reduced variants nevertheless demonstrably involves activation of the unreduced forms.

More and more often linguists working on usage-based models/ET argue that speech production and perception involve both abstract categories and word-specific, episodic memories. According to Kapatsinski (2013), findings from his experiments involving the learning of certain morphophonological regularities in artificial languages on the basis of type frequencies provide evidence for a synthesis of insights from both usage-based phonology and Harmonic Grammar.

In the OT-camp, too, a growing number of scholars plead for slightly hybrid models of phonology. Van de Weijer's (2012) model is ''based on the insight that selection (and not derivation) plays a role in both theories'' (p. 6), i.e. in OT and ET. In this model, representations are fully specified surface forms. In the spirit of stochastic OT and Harmonic Grammar, markedness and faithfulness constraints (the latter of which are all output–output in nature) have numeric values indicating their strengths; at the same time the candidates, which are not generated on the basis of an input form, are exemplars which are equally specified for strength in terms of frequency of usage. In numerically enriched OT tableaux the marginal totals for the strengths of both the relevant candidates (not token frequency) and the various constraints are multiplied, with constraints violation yielding negative numbers. The candidate with the highest sum is the winner. Constraints can have different strength values depending on the speech style.

Sloos (2013) also attempts to unite OT and ET. After having established at least four principled reasons ''to maintain a fully-fledged grammar in phonology'' (18–19), she develops her *E*xemplar-*P*rototype-*O*ptimality *T*heory. In this model, infrequently used words are represented in the mental lexicon in exemplars which are loosely connected with each other; this accounts for their exceptional behaviour. For frequently used words, however, a prototype automatically develops. ''The prototype forms the crucial link between the lexicon and the grammar'' (p. 22), which is constraint based and categorical. As in Van de Weijer's proposal, in EPOT the constraints are generalizations across lexical patterns. The communication between the grammar and the lexicon ''accounts for the fact that frequency effects appear to be relative regarding grammatical categories'' (p. 189).

In this volume, ***Van de Weijer*** claims that constraints are learned in the course of language acquisition. The author attempts to reconcile OT with usage-based models of grammar by positing that markedness and faithfulness constraints are not innate, but rather emerge during acquisition. On the assumption that OT is a valid theory, Van de Weijer claims that markedness constraints penalize structures that are statistically underrepresented in the child's primary linguistic data. Faithfulness constraints are claimed to emerge from patterns of association between exemplars in the lexicon. In short, both types of constraints are grounded in usage.

Please cite this article in press as: Hinskens, F., et al., Grammar or lexicon. Or: Grammar and lexicon? Rule-based and usage-based approaches to phonological variation. Lingua (2014), http://dx.doi.org/10.1016/j.lingua.2014.01.005

In his contribution to this special issue ***Van Oostendorp*** also argues from the perspective of a rule-based view on phonological competence. His main claim is that usage does play a role in determining non-absolute patterns, but this usage is outside of the grammar of cognition; it rather plays a role during the transmission of one generation to the next. The grammar itself can then still operate in a rule-based fashion, as modelled in classical OT. Lexicon optimization, the OT version of Kiparsky's (1968) 'alternation condition', which stipulated that underlying representations should not be overtly abstract, plays a central role in Van Oostendorp's proposal. Lexicon optimization is a means to derive an optimal input when several different inputs converge on a single output form. Of all possible inputs for a particular output, the one chosen is the form that is most harmonic, in that it incurs the least significant violations of the constraints (Prince and Smolensky, 1993:192). The optimization mechanism takes effect during acquisition and it is probably the mechanism underlying the lexicalization of a sound change which is ceasing to be productive – although the latter aspect is not discussed in Van Oostendorp's present paper.

As Coetzee (2008:248) puts it: ''much of recent linguistic research has focussed on evidence that formal grammar to the exclusion of usage frequencies influences language processing or vice versa. [. . .] What is largely lacking, however, is research that assumes that both of these influence processing and that asks how they interact.'' Citing Bybee's (2006:711) definition of grammar as ''the cognitive organization of one's experience with language'', and on the basis of the outcomes of his own psycholinguistic experiments, Coetzee points out that ''once this organization has happened, it gains a level of independence from this experience so that it can generalize into areas where the language learner has not had any experience'' (Coetzee, 2008:249). Or, as Carr (2000:101) put it: ''the acquisition of a phonology has a Rationalist rationale, but all of the capacities that are harnassed in the process of the acquisition of a phonology [. . .] are consistent with Empiricism.'' In investigations of first and second language acquisition (on phonological phenomena; see e.g. Nardy and Barbu, 2006; Rys, 2007; Rys et al., 2012), usage-related parameters often turn out to be relatively important, which is plausible against this background.

Bermúdez-Otero (2012) presents what he has coined Refined dual-route models as a parallel race model of processing, in line with Jackendoff's call for ''the necessity of a heterogeneous theory'' (Bermúdez-Otero, 2012:39). In this model the outcome of the 'race' between forms stored in the lexicon and computed forms from the grammar is determined by token frequency: the higher the token frequency the easier the retrieval of the stored form. In Bermúdez-Otero's account of English vowel reduction, schwas that cannot possibly develop because of cyclic stress in items such as *cons*[ə]*rvátion* or *tránsp*[ə]*rtátion* (variants of forms with a full vowel /ɜ́/ or /ɔ̀/ that is predicted on the basis of the English stress system) are frequency effects. The schwa variants, which are lexically diffuse, result from the fact that for these items the derivative in *-ation* is considerably more frequent in usage than the base (i.e. conse[ɜ́]rve and transp[ɔ́]rt) and the derivatives are listed non-analytically as a separate lexical entry. The probability for the 'problematic' variant to ''block the online computation of a competing form depends upon the ease with which the entry itself can be retrieved, which in turn partially depends on its resting activation, which in turn partially depends on its token frequency'' (Bermúdez-Otero, 2012:34).

Guy, in his contribution to this special issue, appears to remain agnostic with respect to the debate concerning usage-based vs. rule-based approaches (or, in Guy's words, ET and Mainstream Theory, respectively). He analyses specific lexical items which behave exceptionally with respect to WFtD in New Zealand English and final /s/ deletion in Salvadoran Spanish, which both show significantly higher reduction rates. Assuming that these items may have multiple lexical representations (viz. both reduced and unreduced underlying entries), Guy predicts that they show weaker surface effects of context on the rule than is the case for unexceptional, single-lexical entry words, *and* this appears indeed to be the case. Guy suggests to include frequency of usage as an independent variable (next to structural and extra-linguistic parameters) in multivariate analyses (i.e. Varbrul in case of the logistic regression type) of instances of language variation. With Pierrehumbert (2006:523), Guy concludes that a hybrid model is needed of variationist models ''augmented by incorporating quantified lexical representations, with additional lexical entries for words with clearly exceptional behaviour.''

7. Two studies from the domain of language variation and change

Finding indications for lexical effects in historical sound change on the basis of modern frequency data may be even more difficult than reconstructing astronomical phenomena that took place light-years away using data from cosmic background radiation. Therefore, it is probably unfair to draw conclusions regarding the relative success of usage-based proposals on the basis of the analysis of historical sound changes, since appropriate contemporary (preferably spoken and dialect-specific) corpora are typically lacking and therefore the relevant type and token frequencies will not be available. Instead, one should investigate sound change in progress.

With only few exceptions (such as Auer et al., 2011; Guy, this volume – also for more references; Walker, 2012), hardly any quantitative studies of phonological variation have tried to assess the relative impact of structural (and potentially universal) factors on the one hand and frequency-related factors on the other, on the basis of systematic and balanced

Please cite this article in press as: Hinskens, F., et al., Grammar or lexicon. Or: Grammar and lexicon? Rule-based and usage-based approaches to phonological variation. Lingua (2014), http://dx.doi.org/10.1016/j.lingua.2014.01.005

data for a specific variable phenomenon in a specific speech community. Such studies can in principle make it possible to compare the relative merits of rule-based and usage-based approaches.

Below, two recent studies on reductive sound change in specific varieties of modern Dutch in which structural factors and both type and token frequency were simultaneously tested will be briefly discussed. The first study reports on divergent, dialect-specific findings, none of which constitute positive evidence for usage-based models (Section 7.1). In the second study (a case study of one single speaker) both 'local' type and token frequencies and available frequencies from large corpora are put to work. This study reports on very divergent findings for variants which seem to result from two successive phases of one single process – with mixed and partially positive evidence for usage-based models (Section 7.2). The latter findings and their implications are related to one of the central issues in the ongoing theoretical debate (section 8).

7.1. Postvocalic /r/ in three groups of Dutch dialects

Several different groups of Dutch dialects have undergone (historical) deletion of postvocalic /r/ before coronal obstruents, as in the dialect variants of standard Dutch items such as those in (5):

(5)	eerst	iˑəʃ	'first'
	beurs	bʏˑəʃ	'wallet; stock market'
	kort	kɔ̰t	'short'
	woord	woːət	'word'
	koorts	koːərts	'fever'
	worst	wuːəʃ	'sausage'
	baard	baːt	'beard'

The second column in (5) contains broad phonetic transcriptions of the variants of the dialects of Waubach (Ubach over Worms), in the transition zone between Ripuarian and East-Limburg dialects of Dutch; some items (in this dialect e.g. the variant of standard Dutch *koorts*) are not r-less. Evidently, as far as the left-hand environment is concerned, the deletion occurred after both tense and lax vowels, after back and front vowels, after rounded and unrounded vowels. In many Limburg dialects of Dutch (including the subset of r-deleting ones), items such as *eerst*, *beurs* and *worst* have a high vowel; hence deletion occurred after high, mid and low vowels. As far as the right-hand environment is concerned, the deletion took place preceding both voiced (*baard*) and voiceless stops, preceding fricatives and preceding both single and complex codas. Usually, there are no indications of compensatory lengthening.

In modern standard Dutch, /r/ in items in which it precedes a coronal obstruent is fairly stable. The presence of absence of postvocalic /r/ is distinctive in minimal pairs such as those in (6):

(6)	kaars	-	kaas	'candle'; 'cheese'
	paars	-	Paas	'purple'; 'Easter'
	erwt	-	Ed	'pea'; 'Ed'
	kort	-	kot	'short', 'den, digs'

In order to investigate changes in the distribution and conditioning of the phenomenon in the speech community at large, a quantitative diachronic study was carried out. The method consisted of the real time comparison of data from two major diatopic, questionnaire-based fieldwork projects carried out in the twentieth century. The oldest one is the *Reeks Nederlandse Dialectatlassen* (RND), i.e. Series of Dutch Dialect Atlases. The data for this project were collected in fieldwork between 1925 (southwest) and the middle of the 1960s (north); they concern 1956 different local dialects and consist of dialect translations of 135 sentences plus isolated words and paradigms; the data are available in the form of narrow phonetic transcriptions. The later of the two studies is the Goeman-Taeldeman-van Reenen project (GTR). The data for this project were collected in fieldwork between around 1980 and 1995; they concern 613 different dialects of Dutch and Frisian and consist of dialect translations of 1854 words and 22 sentences. For this project, too, the data are available in the form of narrow phonetic transcriptions; they have been digitalized and the database is accessible through the website of the Meertens Instituut (http://www.meertens.knaw.nl/mand/database/).

There is some overlap and thus some limited direct comparability between RND and GTR as regards both local dialects and lexical items for which dialect variants were elicited. For the present study overlapping RND and GTR data were used for 50 different dialects which together represent three parts of the Dutch language area: the northeast (Lower Saxonian), the centre (Hollandic; western Brabantic), and the south-east (southeastern Brabantic; Limburgian). From the lexical items for which RND and GTR data overlap, nine were chosen for this study; all nine items are monomorphemic and monosyllabic. The choice of items was balanced for five phonological parameters. These pertain to the quality of the

Please cite this article in press as: Hinskens, F., et al., Grammar or lexicon. Or: Grammar and lexicon? Rule-based and usage-based approaches to phonological variation. Lingua (2014), http://dx.doi.org/10.1016/j.lingua.2014.01.005

Table 1
The main findings for the analyses of r-lessness per factor group.

	Preceding vowel			Following obstr.	
	Back–front	Low–nonlow	Rounded–unrounded	V–VV	C–CC
Correlation[a]	.77	.85	.82	.78	.71
Different level of r-lessness?[b]	Front > back	–	Unrounded > rounded	–	–
r-lessness per factor group significantly affected by[c]					
Time × phon.fact.	–	–	–	–	–
Time × phon.fact. × real. /r/	T	–	–	+	–
Time × phon.fact. × space	–	T	–	–	–

Legend: '–' = no significant effect, 'T' = trend towards significance ($.05 < p < .1$), '+' = significant effect ($p < .05$).

[a] All (paired samples) correlations $p = .000$.

[b] $t = 3.911$, df = 98, 2-tailed, $p = .000$; $t = -2.740$, df = 99, 2-tailed, $p = .007$.

[c] Outcomes of ANOVA's with repeated measures with phonological factor group as within subjects variable and time (RND-GTR), phonetic realization of /r/ and space (or region) as between subjects variables.

preceding vowel (viz. the place of articulation of the preceding vowel: back or front and low or non-low, as well as rounded vs. unrounded), the quantity of the preceding vowel (V – VV) and the number of following coda obstruents (C – CC).

The RND and GTR data for the realization of each of the 9 items in each of the 50 dialects were coded as '0' in case the /r/ is phonetically realized, even if only weakly, '1' in case both variants with and without phonetically realized /r/ were transcribed, and '2' for r-less variant(s). Moreover, the data were coded for the type of phonetic realization of /r/ in postvocalic position generally; in both sources two main types were distinguished, namely apical [r] and relatively velar or uvular [ʀ].

Some of the main findings are summarized in Table 1.

Overall, r-lessness turns out to be significantly affected by

- time (RND vs. GTR): to the extent that the data are reliable, it can be established that overall r-lessness is increasing;
- geographical space (northeast – centre – southeast): r-lessness appears to be most frequent in the southeastern and least frequent in the central dialect groups;
- time × space: in the northeastern and central dialects, r-lessness has increased, while it has decreased in the southeastern dialects.[19] Visually (Fig. 1):

  From the perspective of the Dutch language community at large, it can be established that in real time all three dialect groups are converging. This is brought about by the fact that r-lessness is decreasing in the southeastern dialects and growing in the northeastern and central dialects. It is obvious that in the dialects of the southeastern group r-lessness had attained a different status than in the other two dialect groups. In the southeastern dialects r-lessness has probably been lexicalized. The standard norm is r-full, so giving up r-lessness and restoring /r/ may well be a matter of accommodating to the standard language. Growing r-lessness in the northeastern and central dialects boils down to divergence away from the standard language. The main question is what underlies these different developments.

  Although the phonetic realization of /r/ as such does not play a statistically significant role, and despite the fact that the non-realization of postvocalic /r/ has not changed for the two major variants (time × phonetic realization of /r/), r-lessness is affected by
- space × phonetic realization of /r/: in the central and southeastern dialects, r-lessness indexes are significantly higher in case /r/ has an apical realization than in case it has a uvular realization (the northeastern dialects only have apical [r]). This finding indicates that r-deletion, apart from simplifying the cluster, may additionally be an instantiation of the Obligatory Contour Principle, i.e. the general constraint prohibiting adjacent identical elements, in this case the feature [coronal]. The fact that the r-lessness indexes are significantly higher following front vowels than following back vowels may be an additional indication that r-deletion is originally an OCP-effect; it has been proposed that front vowels and coronal consonants are both specified as [coronal] (e.g. Hume, 1992).To which extent do type and token frequencies contribute to these dialect group-specific developments? For this study, token frequencies were taken from the CELEX/INL 42 million word corpus of modern Dutch. The neighbours, on which type frequency counts were performed, were

[19] As to the three-way interaction effect involving tenseness: following lax vowels and in case /r/ is realized as [r], r-lessness increases strongly, in case /r/ is realized as [ʀ], r-lessness increases only weakly, following tense vowels and in case /r/ is realized as [r], r-lessness increases strongly, in case /r/ is realized as [ʀ], r-lessness decreases strongly ($F = 5.489$, df = 1, $p = .021$). So across the three dialect groups r-deletion or r-lessness is not (or at least not directly) determined by syllable weight.

Please cite this article in press as: Hinskens, F., et al., Grammar or lexicon. Or: Grammar and lexicon? Rule-based and usage-based approaches to phonological variation. Lingua (2014), http://dx.doi.org/10.1016/j.lingua.2014.01.005

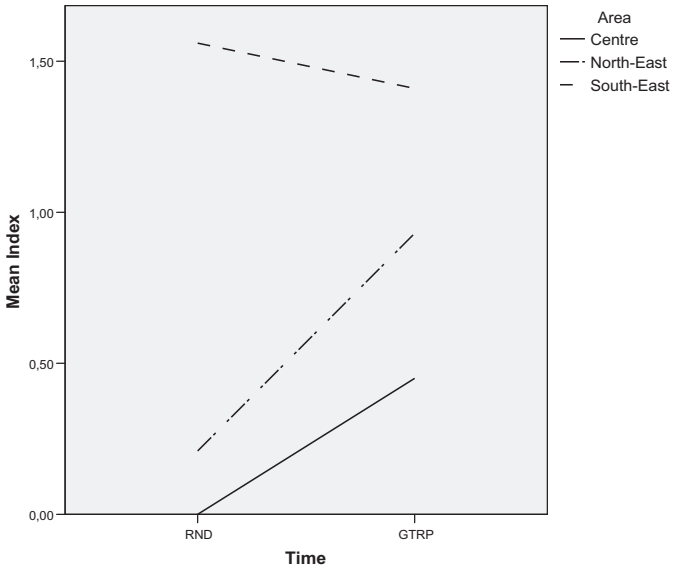


Fig. 1. Real time changes in the proportion of r-lessness in three Dutch dialect groups.

defined as all monosyllabic words ending in an /r/ followed by a tautomorphemic coronal obstruent; this group of words was assembled through Nieuwborg's (1978) retrograde dictionary of modern Dutch and counted 130 items. Breaking down the analyses for the dialect groups, it appears that

- in the northeastern dialects, frequency does not have a significant effect anywhere;
- in the central dialects, token frequency shows a significant, fairly strong, negative correlation with growing r-lessness ($r = -.722$, $p = .022$), while there is a significant, strong, positive correlation with type frequency for rounded vs. unrounded vowels ($r = .843$, $p = .004$). This means that the r-lessness indexes are higher after unrounded vowels than after round ones; cf. Fig. 2.
- in the southeastern dialects, there is a significant, strong, negative correlation with type frequency for back/front vowels ($r = -.799$, $p = .005$), as shown in Fig. 3.

Hence r-lessness indexes have increased after front vowels but decreased after back ones. Again, given the place of articulation of /r/ in many modern southeastern dialects, this drop may well be an OCP effect, i.e. a matter of phonology.

The fact that token frequency has either no (in the northeastern and southeastern dialects) or a negative (in the central

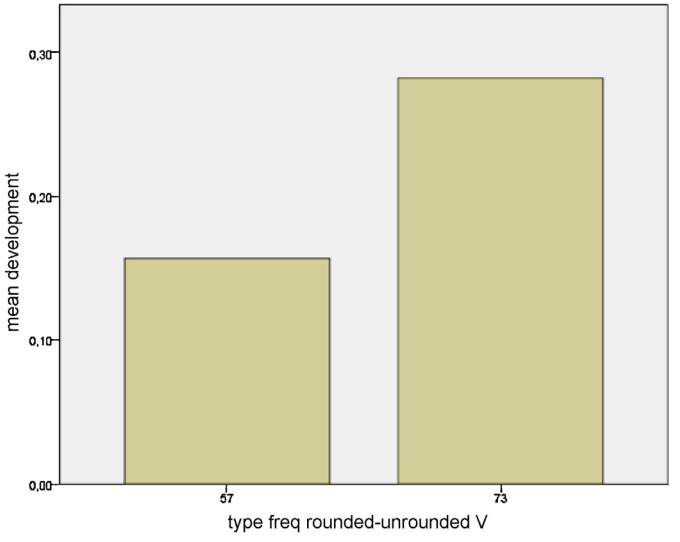


Fig. 2. The effect of rounded/unrounded type frequency on r-lessness in the central dialects.

Please cite this article in press as: Hinskens, F., et al., Grammar or lexicon. Or: Grammar and lexicon? Rule-based and usage-based approaches to phonological variation. Lingua (2014), http://dx.doi.org/10.1016/j.lingua.2014.01.005

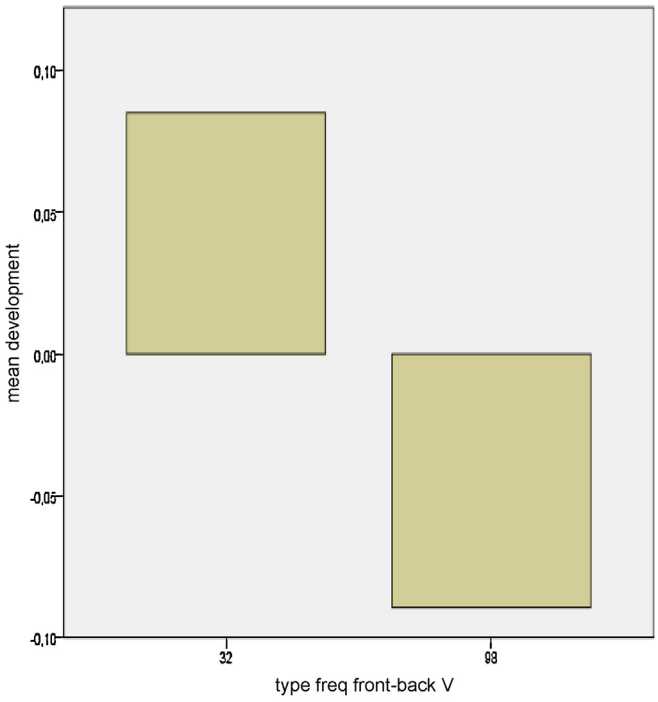


Fig. 3. The effect of front/back type frequency on r-lessness in the southeastern dialects.

dialects) relationship with r-lessness falsifies one of the central claims of usage based phonology, viz. the claim that high frequency of usage leads to reduction. The findings for the type frequency effects of the dimensions rounded vs. unrounded and front vs. back of the preceding vowel are in line with the theory, but the fact that there are so few of them (2 out of ($3 \times 5 =$) 15 cases) implies that here, too, the theory has only marginal explanatory power.

It appears, then, that in the northeastern and central dialects r-deletion is a productive, conditioned and phonologically (OCP) motivated sound change, which is causing these dialects to diverge from standard Dutch. In the southeastern dialect groups, in which r-deletion has presumably been lexicalized, the current restoration of /r/ as [r] or [ʀ] seems to be rather a matter of relexification of the relevant items with the lexical representations of the standard variants.

In short, the developments in these dialect groups are phonologically and/or sociolinguistically (on the basis of prestige) motivated, while lexical properties such as type and token frequencies hardly play a role. To the extent that they do, they appear to be dialect-specific. In one of the cases where they play a role, the direction of the effect is the opposite of what usage-based phonology would predict.[20]

7.2. The reduction and deletion of unstressed vowels in modern spoken Dutch – a case study

In modern Dutch, the reduction of unstressed vowels and schwa deletion is a variable process, partly determined by style level (the reduced and 'deletion' forms are characteristic of casual speech) and speech rate. The relationship with speech rate is not surprising, since speakers of Dutch produce between two and seven words per second (Ernestus, 2007). According to Hay and Sudbury (2005:811) style variation and token frequency are connected in ''that frequent words carry the brunt of the stylistic workload. While the production of infrequent words tends to remain fairly stable across different styles, frequent words display large amounts of phonetic variation across styles''.[21]

Processes of the reduction and deletion of unstressed vowels are not confined to particular varieties of Dutch and they do not concern specific vowels to the exclusion of others, although all authors agree that mid front unrounded /e/ is affected most and high round /y, u/ are affected least by reduction and deletion (cf. Booij, 1995; Geerts, 2008). Both processes are productive and hence not morphologised or lexicalized, although certain words are used more in reduced

[20] See Hinskens (2011b) for a much more detailed account; however, some of the usage-related findings are presented in more detail in the present contribution.

[21] This section contains a revised, abridged and updated version of Section 6 in Hinskens (2011a).

Please cite this article in press as: Hinskens, F., et al., Grammar or lexicon. Or: Grammar and lexicon? Rule-based and usage-based approaches to phonological variation. Lingua (2014), http://dx.doi.org/10.1016/j.lingua.2014.01.005

form than others. These cases concern lemmas rather than phonological forms (Gahl, 2008), as seen in the reduction and deletion in an item such as that in (7):

(7) natuurlijk > tuurlijk > tuuk

Reduction in this word only applies if the word is used as a discourse marker (the pragmatic value of which is comparable to 'of course'), but never if it is used as an adjective or adverb ('natural(ly)').

According to Booij (2008), vowel reduction in modern Dutch, where schwa is a phoneme, has both lexical and postlexical properties. Postlexically, every vowel without lexical stress can be reduced, e.g.:

(8) m[i]nuut ∼ m[ə]nuut 'minute'

On the other hand, vowels with lexical stress that are not stressed in a particular utterance can be phonetically implemented either full or reduced (Booij, 2008:502). At the same time the process has been lexicalized, producing items in which ''vowels in unstressed (non-word-final) syllables [. . .] have been replaced with schwa'' (Booij, 2008:502). The Dutch word for 'television' is therefore invariably pronounced with a schwa in the second syllable:

(9) tel[ə]visie ∼ *tel[e]visie

In Booij's view, Dutch vowel reduction is subject to lexical diffusion (cf. Booij, 2008:502). The author claims that ''this kind of reduction is boosted by high frequency [. . .] This is to be expected since vowel reduction reduces lexical contrasts, and thus impedes word recognition. High frequency, on the other hand, boosts recognition, and can thus compensate for the negative effects of vowel reduction'' (Booij, 2008:504).

A quantitative case study was carried out[22] to determine the reduction of vowel quality to schwa (Pagliuca and Mowrey's 1987 magnitude of gestures), and deletion of the vowel (presumably schwa) and sometimes of the entire rhyme. In Dutch these processes apply to items such as those in (10):

(10)

m[i]nuut	∼	m[ə]nuut			'minute'
fon[o]logie	∼	fon[ə]logie	∼	fonlogie	'phonology'
alm[a]nak	∼	alm[ə]nak			'almanac'

That is, it applies to unstressed vowels in word-initial, pretonic and posttonic position. The data for this quantitative case study were distilled from recordings of ten one-hour live radio shows with audience participation on a regional station in Amsterdam. The presenter (who will be referred to as AB, his initials) talks with many guests, some of whom are present in the studio and others who call in by telephone. The analyses were confined to the phone calls and there were two types: relatively serious conversations with experts (often academics) vs. more playful chats with listeners. See Hinskens (2011a) for a detailed account of the method and the analyses. Type frequency is the frequency of occurrence of the logical combinations of the values of the variables metrical position X hieght X tense/lax, on the understanding that Dutch all high vowels are phonological tense.

From the 1146 observations in the AB database, 688 show reduction of an unstressed vowel whereas no more than 40 show deletion. This is in line with findings from comparable studies. The main findings for reduction and deletion are presented in Tables 2a and 2b.[23]

The first remarkable finding is that the factor group 'style' does not have a significant effect on vowel reduction or deletion. A certain degree of stylistic differentiation was expected to occur in the radio host's language between the serious and the playful phone calls – but none was found, neither for reduction nor for deletion.

From all 14 factor groups that were studied no fewer than 10 appear to affect *reduction* significantly. Statistically, all five usage-related factors appear to be overruled by linguistic factors which are grammatical, metrical, segmental and prosodic in nature, respectively.

[22] By the authors together with the (then) students Marije Anker, Roos Appel, Sanne van der Schee and Marieke van der Worp.

[23] Nagelkerke R^2 is the proportion of the variance that is explained by the significant predictor(s); the value of this coefficient can range between 0 (no explanatory power) and 1.00 (all variance exhaustively explained by the predictor(s)). For a deeper understanding of the outcomes only −2 Log likelihood and the model χ^2 have some importance. −2 Log likelihood is a measure for the probability of the observed data, given the estimated effects (i.e. the weights) of the predictors; the higher this value, the smaller the fit between the data and the model (with fixed numbers of observations). Model χ^2 answers the question whether there are any significant differences between the calculated model and a model containing merely the constant in the equation.

Please cite this article in press as: Hinskens, F., et al., Grammar or lexicon. Or: Grammar and lexicon? Rule-based and usage-based approaches to phonological variation. Lingua (2014), http://dx.doi.org/10.1016/j.lingua.2014.01.005

Table 2a
Outcomes from the multivariate analyses (logistic regression). Criterion: reduction. Method: Backward Wald.

Predictor	B	df	Signif.
In stem	−.912	1	.000
In function word	−.668	1	.006
Word-initially	−.571	1	.000
In non-high vowels	.453	1	.048
In phonetically open Syllables	.334	1	.024
Token frequency AB	.015	1	.019
Token frequency CGN	.007	1	.027
Type frequency AB	.006	1	.015
Type frequency CGN	−.004	1	.033
Token frequency CELEX	.000	1	.002
−2 Log likelihood	1435.188		
Nagelkerke R^2	.121		
Model χ^2	107.030	11	.000

Table 2b
Outcomes from the multivariate analyses (logistic regression). Criterion: deletion. Method: Backward Wald.

Predictor		B	df	Signif
Constant		−3.745	1	.000
Token frequency AB		.035	1	.000
−2 Log likelihood	328.697			
Nagelkerke R^2	.061			
Model χ^2	18.302		1	.000

To the extent that there are frequency effects, the 'local' corpus counts outweigh counts on external large corpora (reduction); in the case of deletion, a local corpus frequency count is the only predictor which has a significant effect (deletion). Apart from this, the effects of the usage-related parameters on reduction are so weak that it make no sense discussing them any further. Also after a drastic reduction of the scale with the aid of a logarithmic transformation (as in the case of CELEX token frequency) no significant effects emerge.

With all factor groups studied, only slightly over 6% of the variance in *deletion* can be explained. It is noteworthy that in the final analysis there is only one single factor left with a significant effect, viz. the token frequency of words in the AB database. However, this effect is dwarfed by that of the constant in the regression equation.

These findings are tentative for several reasons, the first of which is the enormous difference in frequency of occurrence between reduction and deletion in the AB database, which, witness Kloots' (2005) findings (see above), does not seem to be untypical in this respect, however. A second reason why these conclusions can hardly be more than tentative is the fact that in both cases only small portions of the variance have been explained. Third, these conclusions only pertain to speaker AB. On the other hand, all investigators agreed that AB seems to be a fairly typical speaker without special characteristics.

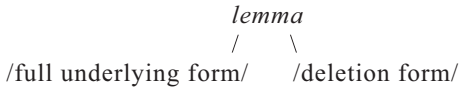
It is striking that, of all factors studied, reduction appears to be mainly determined by the linguistic ones, whereas all linguistic factors that were investigated turn out to have zero effect on deletion – which appears to be influenced exclusively by AB's own word frequencies. Putting it somewhat rashly: reduction, which in the database constitutes a very large majority of the non-full realisations (688/728) is rule-conditioned whereas deletion, a tiny minority of the non-full realisations (40/728), is usage-conditioned. It thus seems that the reduction and deletion of unstressed vowels (and in the case of deletion sometimes even of the entire rhyme), although seemingly phonetically related, are very different in their conditioning and may therefore well be unrelated. Very similar findings were reported by Kallen (2005:76) for the lenition and deletion of /t/ in Irish English as well as by Labov for the contraction and deletion of the copula in Black English Vernacular (or AAVE), which he describes as ''parallel but distinct rules'' which are ''alike in four variable constraints – all representing the influence of the grammatical environment'', while ''the variable phonetic constraints are diametrically opposed'' (Labov, 1972b:114).

Postlexical (or Neogrammarian) processes of sound change are blind to grammatical structure by definition. The fact that reduction of unstressed vowels appears to be conditioned mostly by grammatical factors (morphological status and function word vs. content word) thus indicates that from the point of view of rule-based approaches reduction cannot be a postlexical

Please cite this article in press as: Hinskens, F., et al., Grammar or lexicon. Or: Grammar and lexicon? Rule-based and usage-based approaches to phonological variation. Lingua (2014), http://dx.doi.org/10.1016/j.lingua.2014.01.005

process. While it is phonetically gradual, productive i.e. postlexical sound change is by definition exceptionless, i.e. lexically abrupt. It is therefore not conceivable that productive phenomena would be influenced (however marginally) by idiosyncrasies of individual lexical items, such as token frequency (in this case of lexical items or of vowel classes).[24] Since the type frequencies of productive processes are in principle infinite, it is equally inconceivable that they would be affected by type frequency (in this case of vowel neighbourhoods; cf. above). There are, in other words, several reasons to regard the reduction of unstressed vowels in Dutch as a lexical phenomenon. This holds *a fortiori* for the deletion of unstressed vowels. The online choice a speaker of Dutch makes for either the full form or the 'deletion variant' of an item does not appear to be a phonological let alone a phonetic choice, but rather a lexical one. The deletion variants are probably not 'derived' from the full underlying form; both forms appear to be available simultaneously, as in (11):

(11)



Booij posits the same organization for reduced forms: ''the fact that high frequency of use leads to a higher level of activation of reduced forms can only be accounted for if we assume that these reduced forms receive their own representation in memory'' (2004:227).

In these and similar cases, the output of a general, productive phonological process is pre-encoded by alterations to underlying representations and the existence of multiple lexical entries. To the extent that phonetic and postlexical processes are responsible for loanword adaption, the following observation by Smith is relevant. Discussing the historical phonology of Sranantongo and the adaptation of lexical items originally from Dutch and English, Smith (1987:31) claims that ''more frequently used items will tend to undergo a change first, less frequently used items will tend to undergo a change later, or even not at all. In other words a linguistic change will sometimes lose its force before it has applied to all the relevant forms.''

Booij (2012) discusses a range of cases of allomorphy in modern Dutch, arguing that most allomorphs have to be lexically listed. More generally, Booij claims that frequency effects can lead to the storage of regular forms once they attain unpredictable semantic or stylistic properties – not only after the disappearance of the relevant process, but even if the process is still productive.

Bybee does not subscribe to ''a strict modularization of phonology, grammar and lexicon, as proposed in generative theory'' (2001:97, 24), as she is convinced that the transitions between these components are gradient. In her (2002:271) and Pierrehumbert's (2006:521, 524) view, there are sound changes which are both phonetically and lexically gradient, since the effects of postlexical rules can be lexically stored.[25] Since the reduction of unstressed vowels in Dutch appears to lack several essential properties of postlexical processes, it cannot be such a phenomenon; of all factors studied, the deletion of unstressed vowels in Dutch appears to be only sensitive to a lexical property (token frequency), so this process is not a phonetically and lexically gradient sound change either. However, Rimburg WFtD (cf. Section 3) does represent a sound change of this type.

Reduction of unstressed vowels in Dutch may well be a lexically diffuse sound change. While being phonetically abrupt, lexically diffuse sound changes are by definition lexically gradual. Bybee (2001:58) points out that her model for sound change ''naturally provides an account of the pattern of lexical diffusion in which high-frequency items change more readily than low-frequency items.'' In his proposal for a formal account of lexically diffuse sound changes, Kiparsky (1995; cf. 1992:59) expressly built in the possibility that the token frequency of the relevant lexical items plays a determining role.

8. Outlook and issues for further research

1. If it is indeed the case that the effects of productive processes, such as phonetic and postlexical rules, can be lexically stored and if the number of stored items thus grows in a lexically diffuse fashion, then the underlying process must have an incremental component. Is this the mechanism that has been called the production/perception loop? This notion seems to refer to the fact that along with the number of changed (e.g. lenited) item-specific exemplars which are perceived and stored, the 'modus' in the mental lexicon, which will commonly be the target for speech production, gradually moves in the direction of the changed variant. Briefly: ''sound change affects stored representations incrementally each time a word is used'' (Bybee, 2000). How can other processes (than reduction) be excluded from this mechanism?
2. What exactly is the status of token and type frequency in usage-based models and which status could they have in rule-based theories? In models à la Bybee and ET, on the one hand, frequency as such does not exist as a primitive and the

[24] This also goes for the findings reported in Labov (2006:505 (Table 1) and p. 511).
[25] Bybee (2001:59, 60) and Booij (2008:502, 504).

Please cite this article in press as: Hinskens, F., et al., Grammar or lexicon. Or: Grammar and lexicon? Rule-based and usage-based approaches to phonological variation. Lingua (2014), http://dx.doi.org/10.1016/j.lingua.2014.01.005

relevant information is derived from the relevant exemplar clouds in the mental lexicon. In non-hybrid formal theory, on the other hand, frequency does not have any status at all – it is simply absent as it is considered irrelevant to the grammar. The question is where and how frequency in intermediate, hybrid models and theories is represented. (This is not the same question as how frequency functions in the analyses, which is clear for models such as Smolensky and Legendre (2006), Van de Weijer (2012), Sloos (2013) – cf. Sections 3 and 5.)

3. One of the main questions is whether the relevance of usage-based approaches to phonological variation is confined to reduction, lenition, assimilation, deletion and insertion (Bybee, 2001) or if they apply to any type of sound change, as Pierrehumbert (2002) argues. If Pierrehumbert is right, then any type of sound change (hence also regular vowel shift as well as processes of fortition such as e.g. diphthongisation; cf. Donegan, 1985:50; Balas, 2009:37–38), including phonetic or postlexical processes, is usage-driven, more in particular via significant, positive correlations with token frequency. This would cast another light on Clark's (2008:263–265) finding that the usage of the Scots monophthongal variant of English < ou > shows a significant *positive* correlation with token frequency, a finding which Clark attempts to explain as an instantiation of the Conserving Effect (against analogical sound change).[26]

   Evidence which goes diametrically against Pierrehumbert's claim comes from Hay (2013). Hay presents diachronic data regarding the chain shift of short front vowels in New Zealand English (ONZE) which almost systematically show that low-frequency items move significantly faster. Hay explains this fact with reference to the assumption that the iteration of a production/perception loop does not work in chain shifts: 'troublesome' tokens (i.e. tokens which are misunderstood or marked in some way) are less likely to be stored. In order to be stored, a token has to be identified, which is easier if (a) there is not much acoustic overlap with a phonologically similar item or (b) it is a high-frequency item. High-frequency items can therefore get away with being in the wrong place in vowel space more than low frequency items can.

   In Phillips' (2006) proposal, items that are used exceedingly often lead phonetic sound change (involving, for instance, reduction or assimilation), but relatively infrequently used items can lead gradient sound changes which require 'deeper' analysis from ''other components of the phonological system'' (Phillips, 2006:93–94) such as phonotactic constraints. For them to affect, in turn, the surface phonetics requires a level of lexical analysis on the part of the speakers. This model and especially the concept of 'deep analysis' seems hard to operationalize.

   Walker (2012:410) suggests that frequency (of usage) takes effect only in change in progress and not in stable variation, as in his data regarding WFtD in Toronto English.

   It is conceivable that systematically designed empirical research of phonological variation and change will reveal that usage-related factors play a role (a) *mainly* in the lexical component and (b) there and *only* there where grammar underdetermines usage. Anttila (2006:915–923) established that variation in the assibilation of *-ti* into *-si* in present-day Finnish occurs only in bimoraic verb stems (and never in monomoraic or trimoraic). The bimoraic forms with high token frequency show categorical assibilation, while those with low token frequency categorically resist assibilation; only the subset with intermediate token frequency values show variable assibilation. Under the same metrical conditions, apocope in nouns shows a significant positive correlation with the frequency of usage (931). In other words, usage-related parameters can be decisive where the choice of variants is not determined by the language structure, let alone by UG. Whenever the constellation of relevant phonetic and/or grammatical factors is fixed and a choice is still allowed, i.e. where variation is indeed free, type and token frequency may fine-tune the picture and thus shed further light on the conditioning of language variation and the directionality of language change. This view seems compatible with the findings from Becker et al.'s (2011, 2012) investigations, sketched in Section 5.

4. A concept which seems to be underexploited in usage-based models and ET is type frequency, which roughly corresponds to the incidence of an element or pattern in the lexicon of a given language system, i.e. the number of (free or bound) morphemes in which a given segment or sequence occurs (cf. Section 4). In Bybee's model (2001) type frequency is directly connected with productivity. It seems to be related with the notion of functional load. In structuralist thinking, this concept was used in connection with the number of items involved in specific phonological contrasts – or, in slightly more modern theories and more specifically, the number of minimal pairs for a given distinctive feature. According to Martinet (1955), the functional load of a phonological opposition affects the likelihood of a merger. Can rule-based and usage-based approaches to phonological variation and change meet with respect to type frequency? Which hypotheses would both approaches generate regarding mergers, and perhaps more generally regarding the stability of phonological contrasts?
5. Are speech production and processing in the matters discussed in this special issue truly different things, as suggested by Pierrehumbert (2006 – see Section 5)? If so, why? And if so, what good are frequency counts, which are, after all, merely based on production based?

[26] This analysis seems counter-intuitive, since monophthongization is a type of reduction.

Please cite this article in press as: Hinskens, F., et al., Grammar or lexicon. Or: Grammar and lexicon? Rule-based and usage-based approaches to phonological variation. Lingua (2014), http://dx.doi.org/10.1016/j.lingua.2014.01.005

6. What is the role played by orthography (which is supposed to be part of the non-linguistic information stored in the mental lexicon) in connection with the cognitive aspect of phonetic and phonological variation and the rate and nature of processes of sound change? Is there a systematic difference in this regard between language systems with and without a writing system (and, if so, what kind of writing system)? This question is also relevant to the standard vs. dialect distinction, as dialects are typically almost exclusively spoken and they usually do not have fixed spelling norms.
7. Which role do similarity and frequency play in cross-dialectal comprehension? Best et al. (2009) explore the concept of 'phonological constancy' in perception in early L1 acquisition. For this study, 15- and 19-month-old American toddlers were exposed to their own American dialect and to a dialect of Jamaican English. In an experiment in which the familiarity of words played an important role, the 19-month-olds appeared to abstract away from phonetic differences and to recognize words in either dialect, apparently having developed cross-dialect skills.
8. Are the relevant language systems mutually independent or interdependent in the mental organization of the linguistic competence of bilingual or bidialectal speakers? To which extent does the mental organization of the linguistic competence of bilingual or bidialectal speakers play a role in language change induced by language contact or dialect contact?

For questions 6 and 7, cf. Ernestus's new research project (2013:14–15).

Hopefully, a future special issue of *Lingua* will include contributions in which some or maybe even all of these questions are tackled.

Acknowledgements

Thanks to the reviewers for their very valuable remarks and suggestions on this paper. Only the authors are responsible for any shortcomings. To the extent to which Sections 3, 4, and (especially) 7.2 of the present contribution build on Hinskens (2011a) and the empirical study presented therein, thanks are due to Marije Anker, Roos Appel, Sanne van der Schee and Marieke van der Worp for their parts in the empirical excavation work, and to Radio Noord-Holland.

Generally, we should like to thank (1) the contributors to this special issue for their patience, trust and cooperativeness, (2) Peter Auer, Ricardo Bermúdez-Otero, Geert Booij, Michael Schäfer, Marjoleine Sloos and Wim Zonneveld for their thorough yet swift review of the various contributions, including the present introduction, and (3) *Lingua*'s Anikó Lipták and Johan Rooryck for their support.

References

Albright, A., 2003. Islands of reliability for regular morphology: evidence from Italian. Language 78 (4), 684–709.

Anttila, A., 1997. Deriving variation from grammar. In: Hinskens, F., van Hout, R., Wetzels, L. (Eds.), Variation, Change and Phonological Theory. Benjamins, Amsterdam/Philadelphia, pp. 35–68.

Anttila, A., 2002. Variation and phonological theory. In: Chambers, J., Trudgill, P., Schilling-Estes, N. (Eds.), The Handbook of Language Variation and Change. Blackwell, Malden, MA, pp. 206–243.

Anttila, A., 2006. Variation and opacity. Natural Language and Linguistic Theory 24, 893–944.

Archangeli, D., Mielke, J., Pulleyblank, D., 2012. Greater than noise: frequency effects in Bantu height harmony. In: Botma, B., Noske, R. (Eds.), Phonological Explorations. Empirical, Theoretical and Diachronic Issues. De Gruyter, Berlin, pp. 191–222.

Auer, P., Baumann, P., Schwarz, Chr., 2011. Vertical vs. horizontal change in the traditional dialects of southwest Germany: a quantitative approach. In: de Vogelaer, G., Heeringa, W. (Eds.), Talige en buitentalige factoren bij regiolectvorming = special issue of Taal en tongval 63 (1), 13–41.

Balas, A., 2009. Why can Poles perceive Sprite but not Coca-Cola? A natural phonological account. In: Boersma, P., Hamann, S. (Eds.), Phonology in Perception. De Gruyter Mouton, Berlin, pp. 25–53.

Barbiers, Sj., Hermans, B., 2007. Usage Based Approaches to Linguistics (Handout of a Talk Given at the Meertens Instituut, May 22, 2007).

Becker, M., Ketrez, N., Nevins, A., 2011. The surfeit of the stimulus: analytic biases filter lexical statistica in Turkish laryngeal alternations. Language 87 (1), 84–125.

Becker, M., Nevins, A., Levine, J., 2012. Asymmetries in generalizing alternations to and from initial syllables. Language 88 (2), 231–268.

Bermúdez-Otero, R., 1999. Constraint Interaction in Language Change: Quantity in English and Germanic. (Ph.D. dissertation) University of Manchester.

Bermúdez-Otero, R., 2012. The architecture of grammar and the division of labour in exponence. In: Trommer, J. (Ed.), The Morphology and Phonology of Exponence (Oxford Studies in Theoretical Linguistics 41). Oxford University Press, Oxford, pp. 8–83.

Best, C.T., Tyler, M., Gooding, T., Orlando, C., Quann, C., 2009. Development of phonological constancy. Toddlers perception of native- and Jamaican-accented words. Psychological Science 20 (5), 539–542.

Blevins, J., 2004. Evolutionary Phonology. The Emergence of Sound Patterns. Cambridge University Press, Cambridge.

Bloomfield, L., 1933. Language. Unwin, London (9th reprint 1969).

Boersma, P., Hayes, B., 2001. Empirical tests of the gradual learning algorithm. Linguistic Inquiry 32 (1), 45–86.

Boersma, P., Pater, J., 2014. Convergence properties of a gradual learning algorithm for Harmonic Grammar. In: McCarthy, J., Pater, J. (Eds.), Harmonic Grammar and Harmonic Serialism. Equinox Press, London (in press).

Please cite this article in press as: Hinskens, F., et al., Grammar or lexicon. Or: Grammar and lexicon? Rule-based and usage-based approaches to phonological variation. Lingua (2014), http://dx.doi.org/10.1016/j.lingua.2014.01.005

Booij, G., 1995. The Phonology of Dutch. Oxford University Press/Clarendon, Oxford.
Booij, G., 2004. Reflections on usage-based phonology. Review article of Joan Bybee, 2001, Phonology and language use. Studies in Language 28, 225–237.
Booij, G., 2008. Lexical storage and phonological change. In: Hanson, Chr., Inkelas, S. (Eds.), The Nature of the Word. Essays in Honour of Paul Kiparsky. MIT Press, Cambridge, MA, pp. 487–506.
Booij, G., 2012. Allomorphy and the architecture of grammar. In: Botma, B., Noske, R. (Eds.), Phonological Explorations. Empirical, Theoretical and Diachronic Issues. De Gruyter, Berlin, pp. 9–24.
Bresnan, J., Dingare, S., Manning, Chr., 2001. Soft constraints mirror hard constraints. Voice and person in English and Lummi. In: Butt, M., Holloway King, T. (Eds.), Proceedings of the LFG 01 Conference, Hong Kong. Online proceedings. Stanford (CSLI). http://www.stanford.edu/group/cslipublications/cslipublications/LFG/6/lfg01-toc.html (accessed Spring 2013).
Browman, C., Goldstein, L., 1992. Articulatory phonology: an overview. Phonetica 49, 155–180.
Bybee, J., 2000. Lexicalization of sound change and alternating environments. In: Broe, M., Pierrehumbert, J. (Eds.), Papers in Laboratory Phonology V: Acquisition and the Lexicon. Cambridge (CUP), pp. 250–268.
Bybee, J., 2001. Phonology and Language Use. Cambridge University Press, Cambridge.
Bybee, J., 2002. Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. Language Variation and Change 14, 261–290.
Bybee, J., 2006. From usage to grammar: the mind's response to repetition. Language 82 (4), 711–733.
Bybee, J., 2010. Language, Use and Cognition. CUP, Cambridge.
Carr, Ph., 2000. Scientific realism, sociophonetic variation, and innate endowments in phonology. In: Roberts, N., Carr, Ph., Docherty, G. (Eds.), Phonological Knowledge. Conceptual and Empirical Issues. OUP, Oxford67-14.
Chomsky, N., 1995. The Minimalist Program. MIT Press, Cambridge, MA.
Chomsky, N., Halle, M., 1968. The Sound Pattern of English. Harper & Row, New York.
Clark, L., 2008. Re-examining vocalic variation in Scottish English: a Cognitive Grammar approach. Language Variation and Change 20, 255–273.
Clements, G.N., Hume, E., 1995. The internal organization of speech sounds. In: Goldsmith, J. (Ed.), The Handbook of Phonological Theory. Blackwell, Cambridge, MA, pp. 245–306.
Coetzee, A., 2008. Grammaticality and ungrammaticality in phonology. Language 84 (2), 218–257.
Crystal, D., 1987. The Cambridge Encyclopedia of Language. CUP, Cambridge.
De Saussure, F., 1916. Cours de linguistique générale. Publié par Charles Bally et albert Séchehaye. Avec la collaboration de Albert Riedlinger. Édition critique préparée par Tullio de Mauro. Postface de Louis-Jean Calvet, 1972. Éditions Payot, Paris.
Diessel, H., 2011. Review article of Joan Bybee, 2010. Language 87 (4), 830–844.
Dijkgraaf, R., Fresco, L., 2008. Inleiding. In: Dijkgraaf, R., Fresco, L. (Eds.), De bètacanon. Wat iedereen moet weten van de natuurwetenschappen [. . .] Volkskrant en Meulenhoff, Amsterdam, pp. 11–19.
Docherty, G., Foulkes, P., 2000. Speaker, speech and knowledge of sounds. In: Burton-Roberts, N., Carr, Ph., Docherty, G. (Eds.), Phonological Knowledge: Conceptual and Empirical Issues. Oxford University Press, Oxford, pp. 105–130.
Donegan, P., 1985. On the Natural Phonology of Vowels. Garland Publishing, New York.
Erker, D., Guy, G., 2012. The role of lexical frequency in syntactic variability: variable subject personal pronoun expression in Spanish. Language 88 (3), 526–557.
Ernestus, M., 2007. Iedereen hoort dingen die er niet zijn. Akademie Nieuws 95 (December), 3–7.
Ernestus, M., 2013. Halve Woorden. Inaugural Address. Radboud Universiteit Nijmegen.
Ernestus, M., Baayen, H., 2003. Predicting the unpredictable: interpreting neutralized segments in Dutch. Language 79 (1), 5–38.
Gabriel, Chr., Meisenburg, T., 2009. Silent onsets? An optimality-theoretic approach to French h aspiré words. In: Féry, C., Kügler, F., van de Vijver, R. (Eds.), Variation and Gradience in Phonetics and Phonology. Mouton de Gruyter, Berlin, pp. 171–192.
Gahl, S., 2008. Time and thyme are not homophones. The effect of lemma frequency on word durations in spontaneous speech. Language 84 (3), 474–496.
Geerts, T., 2008. More About Less. Fast Speech Phonology; the Cases of French and Dutch. (Ph.D. dissertation) French Dept., Radboud Universiteit, Nijmegen.
Goldberg, A., 2006. Constructions at Work. The Nature of Generalization in Language. Oxford University Press, Oxford.
Grijzenhout, J., Kabak, B. (Eds.), 2009. Phonological Domains: Universals and Deviations. De Gruyter, Berlin.
Guy, G., 1991. Explanation in variable phonology: an exponential model of morphological constraints. Language Variation and Change 3, 1–22.
Guy, G., 2013. Lexical Frequency and Linguistic Variation. Department talk at Meertens Instituut (KNAW), Amsterdam 5 July 2013.
Guy, G., Boberg, Ch., 1997. Inherent variability and the obligatory contour principle. Language Variation and Change 9, 149–164.
Hawkins, J., 1988. Introduction. In: Hawkins, John (Ed.), Explaining Language Universals. Blackwell, Oxford, pp. 3–30.
Hay, J., 2013. Word memory and regular sound change. In: ICLaVE 7 plenary talk delivered at the 7th International Conference on Language Variation in Europe (ICLaVE), Trondheim, 26–28 June 2013.
Hay, J., Sudbury, A., 2005. How rhoticity became /r/-sandhi. Language 81 (4), 799–823.
Hay, J., Drager, K., Thomas, B., 2013. Using nonsense words to investigate vowel merger. English Language and Linguistics 17, 241–269.
Herrgen, J., 2005. Sprachgeographie und Optimalitätstheorie am Beispiel der t-Tilgung in Auslaut-Clustern des Deutschen. Zeitschrift für Dialektologie und Linguistik 72 (3), 278–317.
Hinskens, F., 1992. Dialect Levelling in Limburg. Structural and Sociolinguistic Aspects. (Ph.D. dissertation) Dept. of Linguistics and Dialectology, University of Nijmegen (revised and abridged version 1996, Niemeyer, Tübingen).
Hinskens, F., 1998. In: Ammon, Ulrich (Ed.), Variation Studies in Dialectology and Three Types of Sound Change. Linguistics of Variation (Sociolinguistica 12), pp. 155–193.
Hinskens, F., 2001. Die Niederländischen Sprachen. Sprachgeschichte, Sprachwandel und allgemeine Sprachtheorie. Inaugural Address. Electronic version: www.meertens.knaw.nl/medewerkers/frans.hinskens University of Leipzig.

Please cite this article in press as: Hinskens, F., et al., Grammar or lexicon. Or: Grammar and lexicon? Rule-based and usage-based approaches to phonological variation. Lingua (2014), http://dx.doi.org/10.1016/j.lingua.2014.01.005

Hinskens, F., 2011a. Lexicon, phonology and phonetics. Or: rule – based and usage – based approaches to phonological variation. In: Siemund, P. (Ed.), Linguistic Universals and Language Variation. Mouton de Gruyter, Berlijn, pp. 416–456.
Hinskens, F., 2011b. Koineization in the present-day Dutch dialect landscape: postvocalic /r/ and more. In: de Vogelaer, G., Heeringa, W. (Eds.), Talige en buitentalige factoren bij regiolectvorming = special issue Taal en tongval, pp. 99–126.
Hinskens, F., Muysken, P., 1986. Formele en functionele benaderingen van dialectale variatie: de flexie van het adjectief in het dialect van Ubach over Worms. In: Hoppenbrouwers, C., Schuurman, I., van Zonneveld, R., Zwarts, F. (Eds.), Syntaxis en Lexicon; Veertien Artikelen bij Gelegenheid van het Emeritaat van Albert Sassen. Foris, Dordrecht/Providence, pp. 13–24.
Hualde, J., 2011. Sound change. In: van Oostendorp, M., Ewen, C., Hume, E., Rice, K. (Eds.), The Blackwell Companion to Phonology, vol. 4. Blackwell, Malden, MA, pp. 2214–2235.
Hudson, R., 2012. Cognitive Sociolinguistics. Powerpoint for a Talk Given in Budapest, March 2012. In: www.phon.ucl.ac.uk/home/dick/talks/2012/socio.ppt (accessed October 2013).
Hume, E., 1992. Front vowels, coronal consonants and their interaction in non-linear phonology. (Ph.D. dissertation) Cornell University.
Jaberg, K., 1908. Sprachgeographie. Sauerlaender, Aarau.
Jackendoff, R., 2002. Foundations of Language. Brain, Meaning, Grammar, Evolution. Oxford University Press, Oxford.
Johnson, K., 1997. Speech perception without speaker normalization: an exemplar model. In: Johnson, K., Mullennix, J. (Eds.), Talker Variability in Speech Processing. Academic Press, San Diego, pp. 145–165.
Kager, R., 1999. Optimality Theory. Cambridge University Press, Cambridge.
Kallen, J., 2005. Internal and external factors in phonological convergence: the case of English /t/lenition. In: Auer, P., Hinskens, F., Kerswill, P. (Eds.), Dialect Change. Convergence and Divergence in European Languages. Cambridge University Press, Cambridge, pp. 51–80.
Kapatsinski, V., 2013. Conspiring to mean: experimental and computational evidence for a usage-based harmonic approach to morphophonology. Language 89 (1), 110–148.
Keller, R., 1994. Sprachwandel. Von der Unsichtbaren Hand in der Sprache. Francke/Uni-Taschenbücher, Tübingen.
Kerswill, P., 1987. Levels of linguistic variation in Durham. Journal of Linguistics 23, 25–49.
Kiparsky, P., 1968. Linguistic Universals and Linguistic Change. Foris, Dordrecht (reprinted in: Kiparsky, P., 1982, Explanation in Phonology).
Kiparsky, P., 1992. Analogy. In: Bright, W. (Ed.), International Encyclopedia of Linguistics, vol. 1. Oxford University Press, New York, pp. 56–61.
Kiparsky, P., 1995. The phonological basis of sound change. In: Goldsmith, J. (Ed.), The Handbook of Phonological Theory. Blackwell, Cambridge, MA, pp. 640–670.
Kiparsky, P., 2000. Opacity and Cyclicity. Linguistic Review 17, 351–367.
Kloeke, G., 1927. De Hollandsche expansie in de zestiende en zeventiende eeuw en haar weerspiegeling in de hedendaagsche Nederlandsche dialecten. Den Haag, Nijhoff.
Kloots, H., 2005. Vocaalreductie in het Standaardnederlands in Vlaanderen en Nederland (Vowel Reduction in Standard Dutch Spoken in Flanders and the Netherlands). (Ph.D. dissertation) University of Antwerp.
Labov, W., 1972a. Sociolinguistic Patterns. University of Pennsylvania Press, Philadelphia.
Labov, W., 1972b. Some principles of linguistic methodology. Language in Society 1, 97–120.
Labov, W., 2006. A sociolinguistic perspective on sociophonetic research. Journal of Phonetics 34, 500–515.
Labov, W., 2010. Principles of linguistic change. Cognitive and Cultural Factors, vol. 3. Wiley Blackwell, Oxford.
Langacker, R., 2008. Cognitive Grammar: A Basic Introduction. Oxford University Press, New York.
Luce, P., Pisoni, D., 1995. Recognizing spoken words. The neighborhood activation model. Ear and Hearing 19, 1–36.
Martinet, A., 1955. Économie des changements phonétiques: Traité de phonologie diachronique (Bibliotheca romanica, Series prima: Manualia et commentationes, No. 10). Éditions Francke, Berne.
McClelland, J., Rumelhart, D., 1981. An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. Psychological Review 88, 375–407.
McQueen, J., Cutler, A., Norris, D., 2006. Phonological abstraction in the mental lexicon. Cognitive Science 30, 1113–1126.
Müller, G., 2003. Optionality in optimality-theoretic syntax. In: Cheng, L., Sybesma, R. (Eds.), The Second Glot International State-of-the-Article Book. The Latest in Linguistics. Mouton de Gruyter, Berlin, pp. 289–321.
Nagy, N., Reynolds, W., 1997. Optimality theory and variable word-final deletion in Faetar. Language Variation and Change 9, 37–55.
Nardy, A., Barbu, S., 2006. Production and judgment in childhood: the case of liaison in French. In: Hinskens, F. (Ed.), Language Variation – European Perspectives. Benjamins, Amsterdam/Philadelphia, pp. 143–152.
Nathan, G., 2007. Phonology. In: Geeraerts, D., Cuycks, H. (Eds.), The Oxford Handbook of Cognitive Linguistics. OUP, Oxford, pp. 611–631.
Nathan, G., 2008. Phonology. A Cognitive Grammar Introduction. Benjamins, Amsterdam.
Nieuwborg, E., 1978. Retrograde woordenboek van de Nederlandse taal. Kluwer, Deventer (2e druk).
Nosofsky, R., 1986. Attention, similarity, and the identification–categorization relationship. Journal of Experimental Psychology: General 115 (1), 39–57.
Pagliuca, W., Mowrey, R., 1987. Articulatory evolution. In: Giacalone Ramat, A., Carruba, O., Bernini, G. (Eds.), Papers from the 7th International Conference on Historical Linguistics. Benjamins, Amsterdam/Philadelphia, pp. 459–472.
Pater, J., 2008. Gradual learning and convergence. Linguistic Inquiry 39 (2), 334–345.
Pater, J., 2009. Weighted constraints in generative linguistics. Cognitive Science 33, 999–1035.
Pater, J., 2012. Serial harmonic grammar and Berber syllabification. In: Borowsky, T., Kawahara, S., Shinya, T., Sugahara, M. (Eds.), Prosody Matters: Essays in Honor of Elisabeth O. Selkirk. Equinox Press, London, pp. 43–72.
Paul, H., 1880. Prinzipien der Sprachgeschichte. Max Niemeyer, Halle.
Phillips, B., 2006. Word Frequency and Lexical Diffusion. Palgrave Macmillan, New York.
Pierrehumbert, J., 2002. Word-specific phonetics. In: Gussenhoven, C., Warner, N. (Eds.), Laboratory Phonology VII. Mouton de Gruyter, Berlin, pp. 101–140.
Pierrehumbert, J., 2003. Probabilistic phonology: discrimination and robustness. In: Bod, R., Hay, J., Jannedy, S. (Eds.), Probabilistic Linguistics. MIT Press, Cambridge, MA, pp. 177–228.
Pierrehumbert, J., 2006. The next toolkit. Journal of Phonetics 34, 516–530.

Please cite this article in press as: Hinskens, F., et al., Grammar or lexicon. Or: Grammar and lexicon? Rule-based and usage-based approaches to phonological variation. Lingua (2014), http://dx.doi.org/10.1016/j.lingua.2014.01.005

Pinker, S., 1999. Words and Rules. The Ingredients of Language. Harper Collins, New York.
Prince, A., Smolensky, P., 1993. Optimality Theory: Constraint Interaction in Generative Grammar. Rutgers University Center for Cognitive Science Technical Report 2.
Rubach, J., 2000. Glide and glottal stop insertion in Slavic languages, a DOT analysis. Linguistic Inquiry 31, 271–317.
Rys, K., 2007. Linguistic and Non-Linguistic Factors in Secondary Dialect Acquisition by Children and Adolescents. (Ph.D. dissertation) University of Ghent.
Rys, K., Daelemans, W., De Pauw, G., Gillis, S., 2012. Leermechanismen in tweededialectverwerving. Tabu 40 (1–2), 59–88.
Schuchardt, H., 1885. Über die Lautgesetze: Gegen die Junggrammatiker. In: Leo, S. (Ed.), Hugo Schuchardt-Brevier. Ein Vademekum der allgemeinen Sprachwissenschaft. Niemeyer, Halle, pp. 51–87.
Sloos, M., 2013. Phonological Grammar and Frequency: An Integrated Approach: Evidence from German, Indonesian and Japanese. (Ph.D. thesis) Groningen University.
Smith, N., 1987. The Genesis of the Creole Languages of Surinam. (Ph.D. thesis) University of Amsterdam.
Smolensky, P., Legendre, G., 2006. The Harmonic Mind: From Neural Computation to Optimality-Theoretic Grammar. MIT Press, Cambridge, MA.
Tagliamonte, S., 2011. Variation as a window on universals. In: Siemund, P. (Ed.), Linguistic Universals and Language Variation. Mouton de Gruyter, Berlijn, pp. 128–168.
Van de Weijer, J., 2012. Grammar as Selection: Combining Optimality Theory and Exemplar Theory. Kougaku, Nagoya.
Van Geert, P., 2000. Theorieën over taalverwerving (Theories on language acquisition). In: Gillis, S., Schaerlaekens, A.M. (Eds.), Kindertaalverwerving. Een handboek voor het Nederlands [. . .] Nijhoff, Groningen, pp. 435–472.
Van Haeringen, C., 1940. De taaie levenskracht van het sterke werkwoord. De Nieuwe Taalgids 34, 241–255.
Van Hout, R., 1989. De structuur van taalvariatie. Een sociolinguïstisch onderzoek naar het stadsdialect van Nijmegen. (Ph.D. dissertation) University of Nijmegen.
Van Oostendorp, M., 1997. Style levels in conflict resolution. In: Hinskens, F., van Hout, R., Wetzels, L. (Eds.), Variation, Change and Phonological Theory. Benjamins, Amsterdam/Philadelphia, pp. 207–229.
Van Oostendorp, M., van de Weijer, J. (Eds.), 2005. The Internal Organization of Phonological Segments. Mouton de Gruyter, Berlin.
Von Goethe, J.W., 1826. Maximen und Reflexionen. Aus Kunst und Altertum, Fünfter Band drittes Heft. Nach den Handschriften des Goethe-und Schiller Archivs. Verlag der Goethe-Gesellschaft, Weimar, pp. 1907.
Walker, J., 2012. Form, function, and frequency in phonological variation. Language Variation and Change 24 (3), 397–415.
Weinreich, U., Labov, W., Herzog, M., 1968. Empirical foundations for a theory of language change. In: Lehmann, W., Malkiel, Y. (Eds.), Directions for Historical Linguistics; A Symposium. University of Texas Press, Austin, pp. 95–189.
Zipf, G., 1935. The Psycho-Biology of Language. Houghton Mifflin, Boston.
Zubritskaya, K., 1997. Mechanism of sound change in optimality theory. Language Variation and Change 9, 121–148.

Frans Hinskens[a,b,*]
[a]*Meertens Instituut (KNAW), Netherlands*
[b]*VU University Amsterdam, Netherlands*

Ben Hermans
Meertens Instituut (KNAW), Netherlands

Marc van Oostendorp[a,b]
[a]*Meertens Instituut (KNAW), Netherlands*
[b]*Leiden University, Netherlands*

*Corresponding author at: Meertens Instituut (KNAW), Netherlands.
E-mail address: frans.hinskens@meertens.knaw.nl (F. Hinskens)

25 November 2013