

University of Groningen

Phonological grammar and frequency

Sloos, Marjoleine

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:
2013

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Sloos, M. (2013). *Phonological grammar and frequency: an integrated approach*. s.n.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Chapter 8

Discussion & Conclusion

In the light of the division of labour between generative and usage-based approaches in phonology, and the more recent call for hybrid modelling, the goal of this thesis was to investigate the relation between the phonological grammar and the lexicon. The investigation consisted of the collection of more relevant data and on the basis of these findings a further step in phonological hybrid modelling was taken. The relation between the grammar and the lexicon was investigated by examining casestudies in which an interaction between grammar and frequency was attested. Such interactions were found in the variation of the pronunciation of Standard German long vowel <ä>, Dutch loanword integration in Indonesian, and sequential voicing (rendaku) in Japanese. The theoretical framework of this thesis is a combination of Exemplar Theory and Optimality Theory, incorporating the notion of the prototype: the resulting model is referred to as Exemplar Prototype Optimality Theory (EPOT).

This discussion summarizes the results on the basis of previous findings about frequency effects in the literature and the hypotheses as defined in chapter 1 in §8.1. Further, in §8.2, the modelling of the data in EPOT is discussed. The final conclusion is provided in §8.3.

8.1 Results

The findings of this thesis build on prior frequency studies, which have shown that two basic frequency effects can be distinguished in language variation and change, namely frequency effects in analogy and frequency effects in reduction, as defined as follows:

(1) *Type I frequency effects (analogy)*

Frequency of occurrence correlates with analogical change such that HF words are less likely to undergo analogical change and LF words are more likely to undergo analogical change.

(2) *Type II frequency effects (reduction)*

Frequency of occurrence correlates with reduction such that HF words are more likely to undergo reduction and LF words are less likely to undergo reduction.

In general, frequency studies suggest that frequency effects, the differences in pronunciation of similar words which depend only on differences in frequency of occurrence, affect the whole lexicon. On the other hand, grammar is supposed to be blind for frequency. As far as I know, possible interactions between frequency and grammar had not been investigated systematically before. Sporadically, however, some interactions were reported, like Bybee (2002) and Coetzee & Pater (2008) who found that English -t/-d deletion occurs more often in nouns, adjectives, past tense verbs, and monomorphemic words. Through an in-depth statistical analysis of new data, I investigated the exact nature of the interaction between frequency effects and grammar, to shed light on the relation between grammar and lexicon, the content of the input and how the input is derived.

In this thesis, I investigated synchronic categorical variation in Japanese rendaku, synchronic gradient variation in the pronunciation of the long vowel <ä> in the Alemannic variety of Standard German, diachronic categorical variation in coalescence in Dutch loanword integration in Indonesian, and diachronic gradient change in the pronunciation of the long vowel <ä> in Standard German. We found an interaction between frequency and grammar in all four cases. The novel findings can be summarized as follows.

(3) *Type III frequency effects (opaque structure)*

In chapter 4, we investigated Japanese *rendaku*, which showed that compounds with roots that are lowly frequent in isolation, are less likely to undergo the morphophonological rule of *rendaku*. *Rendaku*, voicing of the initial obstruent of the right-hand member of a compound, is a phonological rule which has a number of exceptions. Whereas usually HF words behave idiosyncratically regarding the grammar, in *rendaku* we found that the exceptions are LF words. Similarly, in chapter 5, we found that in the Alemannic variety of Standard German, the dialectal [e:] as pronunciation for long vowel <ä> is substituted for [ɛ:], catalysed by pre-r vowel lowering. HF words in non pre-r contexts which are the result of umlaut, are also lowered, but LF words do not follow this rule. Subsequently, in chapter 6, the integration of Dutch loanwords in Indonesian, clearly shows that HF frequency words change first, which is probably not a reduction effect. Finally, in chapter 7, a three-level frequency effect, of low-, mid-, and high-frequency words, was observed in the change of the long vowel <ä> in Low-Saxon. I suggest that Type III frequency effects are default frequency effects, which occur in the absence of Type I and Type II frequency effects. Type III frequency effects may also combine with Type I frequency, leading to a three-level distinction of HF, MF, and LF words (like in chapter 7 on the long vowel <ä> in Standard German). Type III frequency effects seem to occur when the structure (lexical or grammatical) of a word is opaque.

(4) *Frequency effects are sensitive to the phonological or morphological context*

Frequency effects interact with grammar such that, depending on the context, frequency effects may be blocked or different frequency types may co-occur. In Dutch loanwords in Indonesian, we found that Type I frequency effects are blocked in non-native phonological contexts. In Standard German long vowel <ä> lowering, Type I frequency effects occur in non pre-r context and Type II frequency effects occur in pre-r context.

(5) *Frequency is relative depending on the grammatical context*

Within a particular case of variation, we find the same frequency type within subcategories. Words with a frequency F may in context α may behave as a relatively LF word, and (other) words with frequency F in context β may behave as relatively HF words. Thus, frequency is a relative notion. This was illustrated for coalescence in Dutch loanword integration in Indonesian, where words with a particular frequency value behave as LF words among /t/-initial words, but as HF words among /p/-initial words. I would suggest that future frequency studies carefully investigate whether relative frequency effects occur and, if so, whether it is useful for quantitative studies and probabilistic modelling, to use relative frequency values, rather than absolute frequency values.

In the light of these results, let us evaluate the hypotheses that were formulated in §1.4, repeated below for convenience.

(6) *Hypothesis I*

Frequency effects within a particular variation pattern occur in particular grammatical contexts and are blocked in other grammatical contexts.

Frequency effects may be blocked in a particular grammatical context. Dutch loanwords in Indonesian show a Type III frequency effect (opaque structure), in which HF words change first, except for words that begin with a loan phoneme /f/ or with a consonant cluster. This effect is, however, parasitic upon the fact that words with borrowed structures do not integrate into the native grammar. Words that have non-native structure will hardly be integrated, so there is no variation and hence no frequency effects occur.

(7) *Hypothesis II*

Frequency effects within a particular variation pattern occur in all grammatical contexts, but they are sensitive to the grammatical difference between these contexts.

Dutch loanwords in Indonesian may undergo coalescence, but the proportion of coalesced forms depends on the initial consonant of the stem (which is part of the input for coalescence) and a Type III frequency effect occurs in different subcategories slightly different, on the basis of the stem-initial consonant. So Hypothesis II is confirmed.

(8) *Hypothesis III*

Frequency effects are independent of the grammar.

In none of the studies, frequency effects were independent of the grammar, as explained above. In all cases that were investigated in this thesis, variation and change in Standard German long vowel <ä>, Japanese rendaku, and Dutch loanwords in Indonesian, frequency effects were sensitive to the grammar in one way or another. In Standard German long vowel <ä>, pre-r context or non pre-r context determine which frequency type occurs; in the Alemannic variety, we saw an interaction between frequency and umlaut; in Japanese, an interaction between frequency and rendaku has been showed; in Dutch loanwords in Indonesian, an interaction between frequency and coalescence occurred. So Hypothesis III is rejected. But the fact that we did not find grammar-independent frequency effects does not mean, of course, that they do not exist. What the investigation shows is that future variation studies in linguistics should be carried out with more attention for the co-occurrence of grammatical factors and frequency effects.

(9) *Hypothesis IV*

In a particular process of language variation and change, either Type I frequency or Type II frequency applies.

Surprisingly, we found that in German long vowel <ä> lowering, a Type I frequency effect occurred in non pre-r context, and a Type II frequency effect occurred in pre-r context. We also found that it is possible that in non pre-r context, a Type III frequency effect (opaque structure) occurs and in pre-r context, a Type II frequency effect occurs. So Hypothesis IV must be rejected.

Besides these frequency-grammar interactions, we found two other significant facts about language processing. The first is that native coders of linguistic data may be susceptible to halo effects or coder bias. This is an important topic for further investigation, which could have repercussions in all fields of linguistic research that make use of coding or transcription: clinical linguistics, dialectology, fieldwork, second language acquisition, and sociolinguistics. A thorough investigation of this topic should clarify what the scope of coder bias is, to what extent coders may be biased by accent, how coder bias can be avoided, whether experience of the coders diminishes coder bias, whether coders can be trained to avoid coder bias, and, if so, how. The comparison between acoustic and auditory analyses seems a suitable way to investigate coder bias, but more coders have to be investigated, in a more systematic manner.

Furthermore, we found that in the reversal of a merger, orthography may serve as a reference point. This implies that orthography should be represented in the lexicon as well. However, orthography seems to play a role only in case of competition of different forms and in the absence of other decisive cues for the pronunciation. The way orthographical representation is stored must therefore differ from other lexical information. This lexical representation of the phonology-orthography relation is left for future research.

8.2 Exemplar-Prototype-Optimality-Theory (EPOT)

We saw that frequency effects may interact in different ways with grammar. Further, we also found that words that are very infrequent behave differently from words that have a “normal” or high frequency. The results were modelled in Exemplar-Prototype-Optimality-Theory (EPOT): a combined model with a lexicon that can account for frequency effects and with a fully fledged grammar. The lexicon is modelled such that perception of each linguistic item is stored as an exemplar and categorized on the basis of (dis-)similarity between the exemplars, where similarity depends on categorical perception and meaning. Exemplars are stored at different levels, linguistic and non-linguistic, and thus represent fine phonetic detail, sociolinguistic information, and also orthographical information (chapter 2). So far, the EPOT lexicon is identical to ET. However, I suggest that only very infrequent words, or newly perceived words, are stored exclusively in an exemplar-based manner, since when more exemplars are stored, automatically an abstraction is made: a prototype. The prototype may be the “mean” or the “weighted mean” of the category, but this is not necessarily always the case. Psychological experiments have shown that prototype construction also relies on personal relevance and motivation (Sherman 1996). For linguistics, I suggest that, besides frequency, saliency, recency, and new innovative forms that are related to social economic status are factors that could contribute to the construction of a prototype, which is in line with cognitive phonology. Prototypes can also be formed on the basis of analogical networks, in which case they are comparable with underlying forms in generative grammar. The prototypes differ from the usual concept of underlying forms, however, in that they may contain fine phonetic detail and that more prototypes of the same word may be formed. In EPOT, prototypes form the connection between lexicon and grammar, since, on the one hand, they are an abstraction over lexical storage and, on the other hand, they serve as input for the grammar. Fine phonetic detail thus enters the grammar in the input. In chapter 7, I showed that a restriction on GEN, which requires that candidates differ only minimally, but still categorically, can assure that gradiency is retained in the output, which captures the frequency effect in production. Crucially, the grammar itself remains categorical and free from frequency values.

EPOT can account for stable variation as well as diachronic change, or in other words, EPOT functions as a model of the individual language user as well as the language community. Stable variation and loanword integration can be modelled in EPOT as follows. Since LF words are supposed to have opaque morphological, morphophonological, or lexical structure, which makes them susceptible for neighbourhood frequency effects, their linguistic output may vary. Words with higher frequency have stronger prototypes, which behave in a usual, grammatical, way. Diachronic change is accounted for in EPOT in the following way. The initial and final stages of change, as well as the phonological processes during change, are modelled in constraint-based grammar. Like the prototypes, which are abstractions over lexical categories, I also regard constraints as generalizations but over the whole lexicon rather than over exemplars of a single word. Between the initial and final stages of language change, lexical reorganization may occur: each perceived instance is stored as an exemplar and due to changes in category storage, the prototype of the category may change over time. In EPOT, these changes are modelled in the lexical component.

In §1.3.3, I introduced “Grammar as Selection” (van de Weijer (2012)) on which EPOT builds. I outlined some potential problems of GS: lack of empirical evidence, unclear account of stylistic variation, frequency assignment to groups of exemplars rather than an account of word frequency, and possibly overrepresentation of frequency. This thesis provided empirical data, and showed that they can easily be modelled in EPOT. Stylistic variation was not accounted for, but works similar as the account for the German data: different prototypes of a single word may exist and occur in the input. Word frequency is accounted for by exemplar modelling, which has a direct impact on the form of the prototype, so groups of exemplars do neither occur as input nor as candidates in the grammar. Frequency is represented by the exemplar storage but not as constraint weight, so that no overrepresentation of frequency occurs.

8.3 Conclusion

In conclusion, this thesis contributed to hybrid modelling in phonology by data which show a clear interaction between frequency and grammar. There appear to be different frequency-grammar interactions: frequency effects are sensitive to the grammatical context or may be blocked in a certain grammatical context. We unexpectedly found a third frequency type. We also found that orthographical information should be incorporated in the lexical representation. Further, I suggested that lexical storage involves exemplar-based storage as well as prototype storage. The first occurs in new, very infrequent words, the latter occurs in more frequent words. These results were modelled in a combined Exemplar-Prototype-Optimality-Theory. In future research, this model should be tested on the basis of language acquisition, loanword phonology, and other areas of phonology. EPOT can also easily be extended with a probabilistic component.

Further research is also needed to examine Type III frequency effects (opaque structure) and frequency-grammar interactions in more detail. I hypothesize that Type III frequency effects might generally occur in loanword integration, language acquisition (van de Weijer & Sloos, forthcoming), and stable variation (where it might co-occur with Type I frequency effects). In addition, it should be investigated whether there is neurolinguistic evidence for an exemplar-only based storage for newly acquired and extremely infrequent words, on the one hand, and prototype storage for more frequent words, on the other hand. I hope to have shown in this dissertation that frequency and phonological grammar should be studied simultaneously in order to get a full picture of phonological processing.