

Statistically gradient generalizations for contrastive phonological features¹

MIRJAM ERNESTUS

Abstract

In mainstream phonology, contrastive properties, like stem-final voicing, are simply listed in the lexicon. This article reviews experimental evidence that such contrastive properties may be predictable to some degree and that the relevant statistically gradient generalizations form an inherent part of the grammar. The evidence comes from the underlying voice specification of stem-final obstruents in Dutch. Contrary to received wisdom, this voice specification is partly predictable from the obstruent's manner and place of articulation and from the phonological properties of the preceding segments. The degree of predictability, which depends on the exact contents of the lexicon, directs speakers' guesses of underlying voice specifications. Moreover, existing words that disobey the generalizations are disadvantaged by being recognized and produced more slowly and less accurately, also under natural conditions. We discuss how these observations can be accounted for in two types of different approaches to grammar, Stochastic Optimality Theory and exemplar-based modeling.

1. Introduction

Native speakers are able to coin new morphologically complex words that sound well-formed and are easily understood by their interlocutors. Such new words reflect the form, form-meaning, and form-function regularities among the words in the lexicon. According to traditional linguistic theory, the language-

1. I would like to thank Harald Baayen, Susanne Gahl, Mark Pluymaekers, Alan Yu, and two anonymous reviewers for their valuable comments on earlier versions of this article.

learning child translates these lexical regularities into abstract rules (SPE: Chomsky and Halle 1968) or constraint hierarchies (OT: e.g., Prince and Smolensky 1993/2004; Tesar and Smolensky 1998, 2000; Kager 1999) and it is these rules or hierarchies that speakers use when producing new words. In exemplar-based models, in contrast, speakers are assumed to coin new words in analogy with sets of existing words. As a consequence, lexical regularities are predicted to be statistically gradient, that is, to range from absolute exceptionless generalizations to weak trends which only hold for a few words (e.g., Bybee 2001; Skousen 2002).

Many form-function and form-meaning regularities are indeed not absolute. Thus, whereas English past-tense forms typically consist of the verb stem and the suffix *-ed*, some past-tense forms just contain a different vowel than the verb stem (e.g., *sing – sang*, *hold – held*). Irregulars of a low frequency of use may be regularized in course of time, which shows the force of the lexical regularities (cf. Stemberger 1988). Conversely, groups of irregulars may also function as attractors and irregularize regular forms (e.g., Stemberger 1988; Bybee 2001; Albright and Hayes 2003; Pierrehumbert 2006). Such analogical effects, driven by just a small number of words, are expected within exemplar-based approaches.

Form regularities appear to contrast with form-function and form-meaning regularities in that they tend to be absolute. For instance, the regularity that obstruents are voiceless in word-final position (final devoicing) is without exceptions in languages such as Dutch, German, and Polish, even in the non-native vocabulary. As form regularities are typically absolute, they only concern non-contrastive properties. According to the traditional view, contrastive properties are simply listed in the lexicon and do not generalize to other words. Within exemplar-based models, in contrast, analogical effects may arise from statistically gradient generalizations over these contrastive features.

This article reviews a series of experimental studies investigating gradient generalizations over the underlying voice specifications of stem-final obstruents in Dutch. These specifications are contrastive, as is illustrated by the words [vɛrvɛidən] *verwijden* ‘to widen’ and [vɛrvɛitən] *verwijten* ‘to reproach’, which differ phonologically only in the [voice]-specification of the stem-final alveolar plosive preceding the infinitive suffix [ən] *en*. The contrast is neutralized in syllable coda position, where all obstruents are realized as voiceless (Booij 1995). Thus, in isolation, the stems of both *verwijden* and *verwijten* are realized as [vɛrvɛit].

We investigated whether the underlying voice specifications of stem-final obstruents in Dutch may be predictable, even though they are contrastive (Section 2). Subsequently, we examined whether the gradient regularities that we found may affect linguistic behavior (Sections 3 to 7). In Section 8, we discuss possible sources of the attested regularities and describe how our findings can

Table 1. The eleven phonological gangs and their support for underlying voicing, defined as the proportion of words underlyingly ending in voiced obstruents (Ernestus and Baayen 2003). The segments between the first pair of brackets represent the possible vowels of the final rhymes, while the segments between the second pair represent the possible pre-final consonants. A “-” represents the possibility of the absence of a pre-final consonant. The final pair of brackets enclose the possible final obstruents.

Phonological gangs	Support for voicing
1. { ϵ i, au, α y, a:, e:, o:, \emptyset :, i, u} {-, j, l, m, n, r} {p}	.000
2. { ϵ i, au, α y, a:, e:, o:, \emptyset :, i, u} {-, j, l, m, n, r} {t}	.372
3. { ϵ i, au, α y, a:, e:, o:, \emptyset :, i, u} {-, j, l, m, n, r} {s}	.755
4. { ϵ i, au, α y, a:, e:, o:, \emptyset :, i, u, a, ϵ , i, ɔ , y, y} {f, k, p, s, t, x} {p, t, s}	.019
5. {a, ϵ , i, ɔ , y, y} {-, m, r} {p, t, s}	.135
6. {a, ϵ , i, ɔ , y, y} {l, n} {p, t, s}	.357
7. { ϵ i, au, a:, e:, o:, \emptyset :, y} {-, j, l, r, m, n} {f, x}	.992
8. {i, u} {-, m} {f}	.778
9. {a, ϵ , i, ɔ , y} {-, m} {f}	.081
10. {a, ϵ , i, ɔ , y, i, u} {l, r} {f}	.775
11. {a, ϵ , i, ɔ , y, i, u} {-, j, l, r, m, n} {x}	.953

be accounted for in two different approaches of grammar, Stochastic Optimality Theory and exemplar-based models.

2. Predictability of underlying voicing

We investigated the predictability of the underlying voice specification of Dutch stem-final obstruents in Ernestus and Baayen (2003). From the CELEX lexical database (Baayen et al. 1995), we collected all 1697 monomorphemic nouns, adjectives, and verbs that end in an obstruent of which both the voiced and voiceless variants represent Dutch phonemes (/b, p/, /d, t/, /v, f/, /z, s/, /ʒ, x/). A k-nearest neighbors algorithm using information gain weighting as available in the Tilburg Memory Based Learner (Daelemans et al. 2004) showed that the underlying voice specification of a stem-final obstruent correlates with the type of this obstruent (bilabial plosive, alveolar plosive, labiodental fricative, alveolar fricative, velar fricative), with the quality of the preceding consonant, if present, and with the quality of the preceding vowel. The quality of the onset of the final syllable, the position of main stress in the word, and word class emerged as irrelevant.

A Classification Tree analysis (Breiman et al. 1984) supported these findings. It classified the words into phonological gangs, based on their final

rhymes as pronounced in word-final position (i.e., with a voiceless final obstruent). Rhymes were grouped together if they showed similar probabilities for the final obstruent to be underlyingly voiced. Table 1 defines the resulting eleven phonological gangs in terms of their final rhymes and lists for each gang the proportion of words underlyingly ending in voiced obstruents. Importantly, each gang consists of final rhymes that are phonologically similar. For instance, all words in Gang 1 end in a phonologically long vowel, optionally followed by a sonorant consonant, and a bilabial plosive.

In some gangs, nearly all words underlyingly end in either voiced (Gangs 7, 11) or voiceless (Gangs 1, 4, 9) obstruents. For most words belonging to these gangs, the underlying voice specification is predictable given the voice specification for the majority of words in the gang. In other gangs, the variation among the words is greater (e.g., Gangs 2, 3, 6). Note, however, that the proportion of underlyingly voiced obstruents in a gang varies between 0.000 and 0.372 and between 0.755 and 0.992. In other words, all phonological gangs show a clear preference for either underlyingly voiceless or voiced final obstruents. The gangs thus represent statistically gradient generalizations on the underlying voice specification of stem-final obstruents in Dutch, a contrastive phonological property.

Henceforth, we will refer to the proportion of stems that underlyingly end in a voiced obstruent as the gang's analogical support for underlying voicing. The analogical support for an underlyingly voiceless specification equals the remaining proportion (i.e., $1 -$ analogical support for underlying voicing).

Naturally, the number of gangs and therefore also the analogical support for voicing in each gang depends on the statistical technique used to determine the gangs. The technique that we used here, a classification tree with pruning, is known to be conservative.

3. Past-tense formation for pseudowords

The crucial question in phonological and psycholinguistic theorizing is whether linguistically naïve speakers of Dutch have access to the gradient generalizations on underlying voicing. If so, they may use these generalizations for guessing the underlying voice specifications of new words.

We created 192 phonotactically legal pseudo verb stems, representing the eleven phonological gangs in Table 1, and presented the stems auditorily to 28 native speakers of Dutch (Ernestus and Baayen 2003). The stem-final obstruents were realized as voiceless, as they were word-final (final devoicing). Hence, our participants could neither retrieve the underlying voice specification of the stem-final obstruent from the word's lexical representation nor deduce it from the acoustic signal.

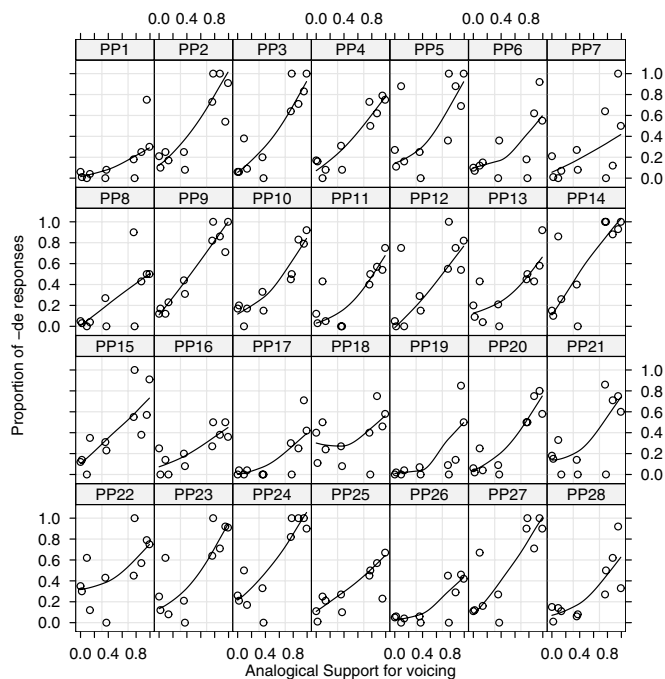


Figure 1. The correlation between analogical support for voicing and percentage of past-tense forms created with *de* for each participant (PP).

The participants nevertheless needed the underlying specification since they were asked to create past-tense forms for the pseudoverbs. In Dutch, regular past-tense forms are created by suffixing the allomorphs /tə/ *te* or /də/ *de* to the verb stem. According to received wisdom, the choice between the two allomorphs is completely regular and depends exclusively on the underlying voice specification of the stem-final segment (e.g., Booij 1995). The standard allomorph is /tə/ *te* if the stem-final segment is underlyingly voiceless, otherwise /də/ *de* is selected. Thus, the standard past-tense allomorph is *te* for *stap* /stap/ ‘step’ (/staptə/ *stapte*) and *de* for *krab* /krab/ ‘scratch’ (/krabdə/ *krabde*).

As predicted, the participants based their choices between the past-tense allomorphs *de* and *te* on the lexical regularities for underlying voicing. The percentage of participants choosing *de* for a given pseudoverb correlated positively with the analogical support for voicing in the pseudoverb’s phonological gang ($r_s = 0.50$): The higher the analogical support for voicing for a given verb, the more participants chose *de*. Clearly, speakers of Dutch have access to the gradient regularities on underlying voicing.

In this initial study, we focused on the average percentages of participants choosing *de* for a given pseudoverb. Importantly, the analogical support for underlying voicing was also predictive at the level of the individual participant, as illustrated in Figure 1. This figure contains a subplot for each participant. Every dot in a subplot represents a phonological gang. The x-axis plots the analogical support for underlying voicing, while the y-axis shows the proportion of verbs for which the participant selected *de*. The lines in the subplots, representing a non-parametric scatterplot smoother (Cleveland 1979), show that there was a positive correlation between the proportion of *de* responses and the analogical support for voicing for all participants. Every participant chose *de* for some verbs and *te* for other verbs belonging to the same phonological gang, with the probability of *de* correlating with the gang's analogical support for underlying voicing.

A multi-level model (e.g., Quené and Van den Bergh 2004) with the percentages of *de* responses given by each participant for each phonological gang (i.e., each dot in Figure 1) as dependent variable, the analogical support for underlying voicing as predictor, and participant as random factor, revealed that the participants did not differ in their sensitivity to the analogical support for voicing ($p > 0.1$). The correlation between the probability of a *de* response and the analogical support for voicing appeared equally strong for all participants. All speakers showed probabilistic behavior that reflected the statistical gradience of the regularities on underlying voicing.

The question arises whether participants also show this behavior if the acoustic realizations themselves contain information on the underlying voice specification of the final obstruent. In Dutch, the neutralization of underlyingly voiced and voiceless obstruents in word-final position is often incomplete, as the underlyingly voiced obstruents tend to have more acoustic characteristics of voiced obstruents than underlyingly voiceless obstruents. That is, underlyingly voiced obstruents tend to be slightly voiced (e.g., Warner et al. 2004; Ernestus and Baayen in press). We presented two additional groups of 28 participants with the same 192 pseudoverbs as in the initial study, but some of the final obstruents were realized as slightly voiced (Ernestus and Baayen 2006). All participants relied in their selection of the past-tense allomorph on the subtle acoustic cues to underlying voicing: They were more likely to choose *de* if the final obstruent was slightly voiced. For instance, the words ending in a sonorant and a plosive received *de* responses in 7.2% of the trials in the initial study and in 13.7% of the trials in this follow-up study, in which the plosives were slightly voiced, when mixed with words ending in completely voiceless obstruents. Nevertheless, by far the most important predictor for the participants' choices was the analogical support for voicing, which was almost as predictive in this follow-up study as in the initial study. Participants relied heavily on the statistically gradient generalizations for voicing even though they could obtain

the information that they needed from the acoustic signal. Hence, these generalizations represent more than a last resort: Speakers use these generalizations also when other sources of information are available.

4. Restoration of underlying voicing

In order to understand the relevance of the statistically gradient generalizations in everyday speech, we have to investigate whether they also affect the processing of existing words. In a first step, we investigated whether speakers use the generalizations if they feel uncertain about the voice specifications of existing words (De Schryver et al. 2005).

The native Dutch lexicon contains three pairs of fricatives: /f, v/, /s, z/, and /x, ɣ/. For speakers from the West of the Netherlands, however, the opposition between the voiced and voiceless fricatives is disappearing, leaving only the voiceless fricatives. This increases the uncertainty among these speakers about the standard pronunciation and spelling of words.

We presented 31 university students from the West of the Netherlands and 31 from the Dutch speaking area of Belgium (Flanders) with two orthographic representations of 98 infinitives consisting of a fricative-final stem and the suffix *en*. One orthographic representation was standard, while in the other representation the underlyingly voiced fricative was transcribed as underlyingly voiceless, or vice versa (non-standard). For instance, participants read both standard *kruisen* and non-standard *kruizen* ‘to cross’. They were asked to indicate which orthographic transcription was standard.

As expected, the students from the West of the Netherlands chose the non-standard representations more often (in 21.6% of the trials) than the students from Flanders (8.2%). Crucially, the probability of the non-standard representation was higher for verbs receiving substantial analogical support for the non-standard voice specification of the fricative. For instance, the verb *kruisen* (/krøysen/) falls in phonological gang 3 (see Table 1), in which the analogical support for /z/ is 0.755, and 7 out of the 31 participants from the West of the Netherlands opted for non-standard *kruizen*. Speakers use the predictability of underlying voicing also for existing words.

5. Past-tense formation for existing verbs

In a next step, we investigated whether the gradient generalizations on voicing also affect the processing of existing words for which speakers know the underlying voice representations.

We presented 40 Dutch university students from the East of the Netherlands with the plural present tense forms of 176 common, regular, existing verbs

(Ernestus and Baayen 2004). In the plural present tense form, the verb-stem is followed by the suffix /ən/ (e.g., [krabən] *krabben* ‘scratch’). The stem-final obstruent surfaces with its underlying voice specification and the participants were thus literally told what the underlying voice specification was. We asked the participants to select for each verb as fast as possible the standard, prescribed, past-tense allomorph. Since our participants were highly educated, they were well trained on past-tense formation. Moreover, they had ample experience with reading and writing and it may therefore be assumed that their mental lexicons contain lexical representations for (most of) the standard past-tense forms. In sum, we did not expect these participants to produce non-standard forms.

The participants chose the standard allomorph in almost 95 % of the trials. Eighty-two verbs elicited non-standard responses from 1 to 24 participants. Participants created non-standard forms more often and reacted more slowly for verbs with low frequency past-tense forms. More importantly, they did so too for verbs with greater analogical support for the non-standard voice specification of the stem-final obstruent. For instance, whereas the morphological rule prescribes *de* for the verb /krab/, the lexical regularities favor an underlyingly voiceless specification for the final obstruent (Gang 5), thereby supporting *te*. Accordingly, almost half (19) of the participants chose *te* and both the standard and the non-standard responses were relatively slow.

These results have been replicated with a participant group consisting of ten native speakers of Dutch and nine native speakers of languages as diverse as Russian and Farsi, who had been speaking Dutch only for a few years. (This experiment has not been reported elsewhere.) The native and non-native speakers were matched for age and education level. All participants attended high school at the time of the experiment.

The data were analyzed by means of stepwise multi-level models with speaker and verb as crossed grouping factors (e.g., Bates and Sarkar 2005). A generalized linear model of the responses (standard versus non-standard) showed no statistically significant differences between the native and the non-native speakers of Dutch. The probability of the non-standard allomorph decreased with the frequency of the past-tense form ($\beta = -0.072, F(1, 3334) = 6.83, p < 0.01$) and was especially low for stems ending in alveolar plosives (Type of obstruent: $F(1, 3334) = 32.75, p < 0.0001$). More importantly, the probability of the non-standard allomorph increased with the analogical support for the non-standard voice specification of the stem-final obstruent ($\beta = 0.013, F(1, 3334) = 49.52, p < 0.0001$).

As the response latencies were skewed, we applied a logarithmic transformation and discarded response latencies longer than 3000 ms. The native and non-native speakers did not differ significantly from each other in their reaction times. All participants reacted faster on verbs with stems ending in alveolar

plosives (Type of Obstruent: $F(1, 3217) = 17.87, p < 0.0001$) and they chose the standard allomorph faster than the non-standard allomorph ($\beta = 0.0038, F(1, 3217) = 72.56, p < 0.0001$). Furthermore, the higher the frequency of the past-tense form, the faster the participants chose the standard allomorph (Frequency: $\beta = 0.007, F(1, 3217) = 23.35, p < 0.0001$; interaction Frequency and Standard: $\beta = -0.024, F(1, 3217) = 18.84, p < 0.0001$). Crucially, participants reacted more slowly on verbs with more support for the non-standard voice specification of the stem-final obstruent ($\beta = -0.0005, F(1, 3127) = 10.07, p < 0.005$), independently of whether they chose the standard or non-standard allomorph.

In a third experiment with existing verbs (Ernestus and Baayen 2001), 28 university students listened to the stems, instead of the plural present tense forms, of the verbs. As the stem-final obstruents were word-final, they were all realized as voiceless. Also in this experiment, the probability of the non-standard past-tense allomorph was greater for verbs with more analogical support for the non-standard voice specification of the stem-final obstruent.

In all three experiments, verb forms that do not conform to the gradient regularities on underlying voicing were disadvantaged by being produced more slowly and less accurately. This suggests a strong effect of these gradient generalizations on everyday speech processing. However, as the participants were put under time pressure, they may have used resources that they do not use under natural conditions. In order to ascertain whether the gradient regularities on underlying voicing indeed play a role in normal processing, we therefore also conducted two corpus studies.

One study (Ernestus and Mak 2005) was based on a search of the world wide web. It showed that the non-standard past-tense allomorph is chosen significantly more often for verbs with an analogical support of at least 0.5 for the non-standard allomorph (in 13% of tokens) than for verbs with smaller analogical support for this allomorph (only in 1% of tokens). In the second study, based on the Spoken Dutch Corpus (Oostdijk 2000), we found that words with great analogical support for the non-standard allomorph are also regularly pronounced with this allomorph (Ernestus and Baayen 2004).

We conclude that the statistically gradient generalizations on underlying voice specifications affect language processing also under natural conditions. These gradient generalizations must therefore form an inherent part of grammar.

6. Reading past-tense forms

So far, we focused on language production. The gradient generalizations may, however, also play a role in language comprehension. For instance, listeners

and readers may process non-standard forms faster if they are in line with the generalizations.

Ernestus and Mak (2005) tested this hypothesis for reading by contrasting 14 past-tense forms that receive substantial analogical support for the underlying voice specification of the stem-final obstruent (congruent verbs) with 14 past-tense forms receiving only little analogical support (incongruent verbs). The forms of the congruent and incongruent verbs were matched for their length in letters and their lexical log frequency. Each verb form was embedded in a simple sentence, in which it was spelled with the prescribed allomorph (e.g., *krabde*) for half of the participants and with the non-standard allomorph (e.g., *krabte*) for the other half. All participants read both sentences with standard and with non-standard past-tense forms. The vast majority of sentences in the experiment were fillers, consisting only of correctly spelled words.

Participants read the sentences from a computer screen in a self-paced reading experiment. The words of a sentence were presented one by one. As soon as a participant finished reading a word, he or she pressed a button and the next word became visible. We measured the time span between the successive button presses, that is, the time a participant needed to read each word in the sentence. Every sentence was followed by a question, which stimulated the participants to focus on content instead of form.

Fifty-two native speakers of Dutch participated in the experiment. Their reading times were 21 ms longer for the non-standard than for corresponding standard past-tense forms and 41 ms longer for words directly following non-standard versus standard forms, but only if the verb was congruent ($p < 0.01$). If the verb was incongruent, the non-standard past-tense forms did not slow reading ($p > 0.1$): Participants read such forms on average 13 ms faster than the corresponding standard forms and the following word 21 ms more slowly.

From this, we might conclude that the gradient regularities for voicing directly affect word comprehension. An alternative explanation is possible, however. We saw in the production experiments and in the corpus studies that non-standard past-tense forms are relatively frequent for incongruent verbs. The non-standard forms of such verbs may therefore be stored in the reader's mental lexicon. Congruent verbs, in contrast, are seldom produced with non-standard allomorphs, and only their standard past-tense forms are probably lexically represented. This implies that the reading of a non-standard past-tense form may simply involve lexical retrieval for incongruent verbs, whereas readers have to overcome the inappropriate allomorph for congruent verbs. Irrespective of which explanation is correct, the data bear witness to the role of gradient regularities in normal language processing.

7. The role of vocabulary size

According to the traditional view, generalizations, once translated into rules or constraints, function independently of the exemplars in the lexicon (e.g., Chomsky and Halle 1968). Exemplar-based models, in contrast, assume that generalizations are computed on-line and, hence, remain sensitive to exactly which words are stored in the mental lexicon (Bybee 2001; Skousen 2002). We investigated the role of the precise contents of the lexicon by studying past-tense formation in language acquisition, that is, by primary school children of different grades (Mak and Ernestus 2005).

As children grow older, their general spelling skills improve, and therefore we may predict that they perform better on both congruent and incongruent verbs. However, children acquire relatively more congruent than incongruent words. Many incongruent words are highly frequent (De Schryver et al. 2005) and acquired early. For instance, the four incongruent words for Gang 11 are *lach* 'laugh', *kuch* 'cough', *juich* 'cheer', and *poch* 'boast'. Many of the congruent words, in contrast, are of a low frequency and acquired late, such as *pleeg* 'commit' and *sarcofaag* 'sarcophagus'. This suggests that the generalizations on underlying voicing are more pronounced, and incongruent verbs are more exceptional, for speakers with greater lexicons. We therefore predicted that the children's improvements would be significantly greater for congruent than for incongruent verbs.

We selected the past-tense forms of 11 congruent and 11 incongruent verbs that were of a similar log frequency of occurrence. We incorporated the verb forms in a story and asked 288 children to complete the forms by adding *de* or *te* to the verb stems. In the Netherlands, children begin to learn to read and write in Grade 3, while they are explicitly taught the spelling of past-tense forms in Grade 7. The children in our experiment came from Grades 5 to 8, from four different schools. They knew all words in the story. As predicted, the percentage of standard forms created for the congruent verbs increased steadily with Grade (from 65% in Grade 5 to 82% in Grade 8), whereas there was no statistically significant improvement for the incongruent verbs (from 61% to 64% standard responses).

A high percentage of the children were non-native speakers of Dutch. Interestingly, these non-native children outperformed the children with Dutch as their native language for incongruent verbs (67% versus 62% standard responses, a statistically significant difference), whereas the two groups of children did not differ with respect to the congruent verbs (77% and 78%, respectively). As native speakers typically have larger vocabularies than non-native speakers, also this finding is in line with the hypothesis that incongruent verbs are relatively more problematic, the larger the speaker's vocabulary. We conclude that the gradient generalizations on underlying voicing, which

affect language processing, remain sensitive to the exact contents of the lexicon.

8. General discussion

This article reviewed experimental evidence showing that statistically gradient generalizations are not restricted to the domain of morphology, but play an active role in phonology as well. Gradient generalizations can be formulated on the underlying voice specifications of stem-final obstruents in Dutch, even though these voice specifications are contrastive. The gradient generalizations thoroughly affect linguistic behavior.

The underlying voice specifications of Dutch final obstruents are predictable (to some extent) given the obstruent's place and manner of articulation and the phonological properties of the preceding segments (Table 1). It is unsurprising that the relevant properties are those of the obstruent itself and of segments close to the obstruent. Most non-contrastive phonological properties are predictable given other properties of the segment, such as voicing for nasal vowels and backness for round vowels (e.g., Gussenhoven and Jacobs 1998:30). Furthermore, nearly all dependencies between segments are between immediately adjacent segments (e.g., assimilation) or between segments that are only separated by segments for which the phonological property at issue is irrelevant (e.g., vowel harmony).

In addition, the nature of the correlations between the underlying voice specification of the final obstruent and the properties of the preceding segments are as expected, given the aerodynamics of speech. We found that phonologically long vowels favor underlyingly voiced obstruents (compare, e.g., Gang 7 with Gangs 9–11), and this corresponds with the observation that vowels tend to be acoustically longer before voiced than before voiceless obstruents (e.g., Slis and Cohen 1969). Furthermore, we found that preceding obstruents favor final obstruents to be underlyingly voiceless (Gang 4), which is in accordance with the fact that obstruents in clusters are difficult to be realized as voiced (e.g., Ernestus 2000).

In contrast, the correlations between the underlying voice specification of an obstruent and its place and manner of articulation are the opposite of what would be expected given the aerodynamics of speech and the general trends among languages. First, we observed that in Dutch plosives are less often underlyingly voiced than fricatives (e.g., compare Gangs 1, 2, with Gang 3), even though in the languages of the world voicing is twice as common on plosives than on fricatives (Ohala 1983). Second, we found that underlying voicing is disfavored most for bilabial plosives (Gang 1), whereas in the languages of the world, bilabial plosives are typically voiced, probably due to articulatory,

acoustic, and auditory factors (Ohala 1983). Finally, we documented that the velar fricative is nearly almost underlyingly voiced (Gang 11), whereas the realization of voicing on this fricative is hard (Gussenhoven and Jacobs 1998: 34). A possible functional explanation for the unexpected correlations might be that underlying voice specifications in Dutch optimize language comprehension: Obstruents tend to have underlying voice specifications that are difficult to realize, as these specifications have higher cue validities. For instance, if [ɣ] is perceived, there is no ambiguity, whereas a perceived [x] can be underlyingly /x/ and /ɣ/.

The statistically gradient generalizations affect linguistic behavior in that they direct speakers' guesses of underlying voice specifications. The probability that a speaker assumes the final obstruent of a new word to be underlyingly voiced correlates with the probability that an existing word of the same phonological type underlyingly ends in a voiced obstruent. Speakers thus do not consistently choose the underlying specification with the highest probability, but rather apply probability matching. In this, their behavior is similar to that of goldfish, rats, and pigeons, among others. These animals show probability matching when they have to choose between sources of food with unequal probabilities of indeed providing food or with different amounts of food, which results in an optimal distribution of animals over sources (e.g., Labov 1994).

The gradient generalizations also affect the processing of existing words: Existing words that disobey the generalizations are recognized and produced more slowly and less accurately, even under natural conditions. Speakers bring existing words in line with the gradient generalizations by changing the voice specification of the stem-final obstruent, rather than by adjusting some other phonological property of the word. For instance, they change /dʏb/ 'waver' into /dyp/ instead of /dʏɣ/. The reason for this may be that the voice specifications of obstruents play a minor role in word comprehension in Dutch. They are less relevant than place and manner of articulation (Ernestus and Mak 2004), probably due to the many phonological processes that affect them (final devoicing, regressive voice assimilation, and progressive assimilation).

Given that the generalizations thoroughly affect linguistic behavior, they have to be incorporated in the grammar. The grammar should account for our findings that generalizations induce probabilistic behavior, that they affect the accuracy and the time required for lexical processing, and that they are continuously updated, as witnessed by the data from the primary school children.

Roughly, two types of models can handle the data. One type, Stochastic Optimality Theory (SOT, Boersma 1998), incorporates the traditional view that lexical regularities are stored as abstract generalizations. The other type of models is exemplar-based and includes connectionist networks with sufficient memory to store both generalizations and exemplars.

SOT can easily incorporate the generalizations on underlying voice specifications if we assume constraints not only on non-contrastive but also on contrastive phonological properties. The generalizations on underlying voicing may be captured by constraints such as “Labiodental fricatives are not underlyingly voiceless” and “Fricatives are voiced after long vowels”. In SOT, constraints vary in their relative positions, which may result in stochastic behavior. An account within SOT may thus obtain a correlation of at least 0.85 with the average responses obtained in our first experiment with pseudoverbs (Section 3), a correlation that is as high as those obtained by other, rule-based and analogy-based, models, as shown in Ernestus and Baayen (2003).

Boersma’s Gradual Learning Algorithm for SOT (Boersma 1998: 263; Boersma and Hayes 2001) may account for the continuous updating of the constraint hierarchy. This algorithm assumes that the exact average positions of constraints are adjusted after every new word. Thus, if a child learns the words *pleeg* /pley/ ‘commit’ and *sarcofaag* /sakrofaɣ/ ‘sarcophagus’, the constraint or complex of constraints favoring velar fricatives to be underlyingly voiced is promoted.

Finally, SOT may account for our observation that speakers create non-standard past-tense forms even for verbs of which they know the underlying voice specifications. For this, we have to allow the constraints on underlying voice specifications to outrank the Input-Output Faithfulness constraints for stems in past-tense forms.

It is less clear how SOT would account for our finding that speakers produce both standard and non-standard past-tense forms more slowly for verbs with greater analogical support for the non-standard allomorph. Boersma (personal communication) suggests that participants in production experiments may compute the optimal output as often as necessary to reach a certain degree of certainty. They would reach this degree faster if the faithfulness constraints and the constraints on underlyingly voicing agree (congruent verbs) than if they disagree (incongruent forms). This account implies that the difference in processing time between congruent and incongruent words is restricted to laboratory conditions, a hypothesis worth testing in further research.

Exemplar-based approaches differ from generative accounts in that form, form-function, and form-meaning regularities are not translated into abstract rules or constraints that operate independently from the words stored in the lexicon (e.g., Bybee 2001; Skousen 2002). Instead, these accounts take as their starting point the often replicated finding that the semantic, phonological, morphological, and orthographic neighbors of a word affect the ease and speed with which this word is produced and understood (e.g., Pisoni et al. 1985; Schreuder and Baayen 1997). The lexical processing of a word implies the activation of its neighbors.

The activated neighbors may affect the processing of a new word (e.g., a borrowing or a pseudoword). If some property is shared by all neighbors (absolute regularity), the new word may be forced to incorporate this property as well. If a property represents a gradient regularity, the neighbors that follow the regularity and those that do not are in direct competition. This competition may result in pronunciation variation and long processing times. The strength of the competition is affected by which words are exactly present in the mental lexicon.

Similar analogical effects arise for existing words. Some activated neighbors may support a given property of a word, while others do not. For instance, some activated phonological neighbors may support the underlying voice specification of the stem-final obstruent of the word, whereas others may support the opposite, non-standard, specification. This results in competition between the two specifications. If the specifications differ only little from each other in the amount of support that they receive from the activated neighbors and from the lexical representation of the word itself, competition is strong. Speakers show variation in their productions and their production times are long.

Likewise, our findings can be captured by connectionist models that allow for the storage of both exemplars and generalizations. Moscoso del Prado Martín et al. (2004) presented several models with all Dutch words incorporated in CELEX, such that the models acquired the systematicities of Dutch orthography. In the next training phase, the models received as input the present tense forms of regular and irregular verbs, while their output consisted of the corresponding past-tense forms. In the actual test phase for past-tense formation, the models with relatively small memories showed performances for the pseudowords similar to that of human participants (see Section 3), while their performance for existing regular verbs was substantially worse. In contrast, the model with a sufficient large memory to store both individual items and regularities could not be distinguished from human participants, neither for pseudoverbs nor for existing verbs. These results strongly suggest that the grammar is exemplar-based.

Several researchers have studied the role of statistically gradient generalizations in the domain of morphology. The present article revealed that gradient generalizations are also important in phonology: Gradient generalizations on phonological contrastive properties affect linguistic behavior. This supports a view in which an important role is reserved for gradient regularities in all subdomains of grammar.

References

- Albright, Adam and Bruce Hayes (2003). Rules vs. analogy in English past-tenses: A computational/experimental study. *Cognition* 90: 119–161.
- Baayen, R. Harald, Richard Piepenbrock, and Leon Gulikers (1995). *The CELEX Lexical Database (CD-ROM)*. Philadelphia, PA.: Linguistic Data Consortium, University of Pennsylvania.
- Bates, Douglas M. and Deepayan Sarkar (2005). *The lme4 Library*. On-line available: <http://lib.stat.cmu.edu/R/CRAN/>.
- Boersma, Paul (1998). *Functional Phonology*. The Hague: Holland Academic Graphics.
- Boersma, Paul and Bruce Hayes (2001). Empirical tests of the gradual learning algorithm. *Linguistic Inquiry* 32: 45–86.
- Booij, Geert (1995). *The Phonology of Dutch*. Oxford: Clarendon Press.
- Breiman, Leo, Jerome H. Friedman, Richard A. Olshen, and Charles J. Stone (1984). *Classification and Regression Trees*. Belmont, California: Wadsworth International Group.
- Bybee, Joan (2001). *Phonology and Language Use*. Cambridge: Cambridge University Press.
- Chomsky, Noam and Morris Halle (1968). *The Sound Pattern of English*. New York: Harper and Row.
- Cleveland, William S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association* 74: 829–836.
- Daelemans, Walter, Jakub Zavrel, Ko Van der Sloot, and Antal Van den Bosch (2004). *TiMBL: Tilburg Memory Based Learner, version 5.1. Reference Guide*. Tilburg University.
- De Schryver, Johan, Mirjam Ernestus, Anneke Neijt, and Pol Ghesquière (2005). Analogy, frequency and sound change: The case of Dutch devoicing. Manuscript, Institute for Higher Education in the Sciences and the Arts, Brussels.
- Ernestus, Mirjam (2000). *Voice Assimilation and Segment Reduction in Casual Dutch. A Corpus-Based Study of the Phonology-Phonetics Interface*. Utrecht: LOT.
- Ernestus, Mirjam and R. Harald Baayen (2001). Choosing between the Dutch past-tense suffixes *-te* and *-de*. In *Linguistics in the Netherlands (2001)*, Ton van der Wouden and Helen de Hoop (eds.), 81–93. Amsterdam: Benjamins.
- (2003). Predicting the unpredictable: Interpreting neutralized segments in Dutch. *Language* 79: 5–38.
- (2004). Analogical effects in regular past-tense production in Dutch. *Linguistics* 42: 873–903.
- (2006). The functionality of incomplete neutralization in Dutch: The case of past-tense formation. In *Laboratory Phonology 8*, Louis Goldstein, Douglas H. Whalen, and Catherine Best (eds.), 27–49. Berlin: Mouton de Gruyter.
- (in press). Intraparadigmatic effects on the perception of voice. In *Voicing in Dutch*, Erik Jan van der Torre and Jeroen Van de Weijer (eds.). Amsterdam: Benjamins.
- Ernestus, Mirjam and Willem Marinus Mak (2004). Distinctive phonological features differ in relevance for both spoken and written word recognition. *Brain and Language* 90: 378–392.
- (in press). Analogical effects in reading Dutch verb forms. *Memory and Cognition* 33 (7): 1160–1173.
- Gussenhoven, Carlos and Haike Jacobs (1998). *Understanding Phonology*. London: Arnold.
- Labov, William (1994). *Principles of Linguistic Change. Internal Factors*. Oxford UK and Cambridge USA: Blackwell.
- Kager, René (1999). *Optimality Theory*. Cambridge: Cambridge University Press.
- Mak, Willem Marinus and Mirjam Ernestus (2005). The impact of increasing vocabulary size on analogical effects in spelling. Ms., University of Utrecht.
- Moscato del Prado Martín, Fermín, Mirjam Ernestus, and R. Harald Baayen (2004). Do type and token effects reflect different mechanisms: Connectionist modeling of Dutch past-tense formation and final devoicing. *Brain and Language* 90: 287–298.
- Ohalá, John, J. (1983). The origin of sound patterns in vocal tract constraints. In *The Production of Speech*, Peter F. MacNeilage (ed.), 189–216. New York: Springer-Verlag.

- Oostdijk, Nelleke (2000). The Spoken Dutch Corpus Project. *The ELRA Newsletter* 5: 4–8.
- Pierrehumbert, Janet B. (2006). An unnatural process. In *Laboratory Phonology 8*. Louis Goldstein, Douglas H. Whalen, and Catherine Best (eds.), 81–106. Berlin: Mouton de Gruyter.
- Pisoni, David B., Howard C. Nusbaum, Paul A. Luce, and Louisa M. Slowiaczek (1985). Speech perception, word recognition and the structure of the lexicon. *Speech Communication* 4: 75–95.
- Prince, Alan and Paul Smolensky (1993/2004). *Optimality Theory: Constraint Interaction in Generative Grammar*. Oxford UK and Cambridge USA: Blackwell.
- Quené, Hugo and Huub Van den Bergh (2004). On multi-level modeling of data from repeated measures designs: a tutorial. *Speech Communication* 43: 103–121.
- Schreuder, Rob and Harald R. Baayen (1997). How complex simplex words can be. *Journal of Memory and Language* 37: 118–139.
- Skousen, Royal (2002). *Analogical Modeling*. Amsterdam: Benjamins.
- Slis, Iman H. and Cohen, Antonie (1969). On the complex regulating the voiced-voiceless distinction ii. *Language and Speech* 12: 80–102, 137–155.
- Stemberger, Joseph P. (1988). Morphology in language production with special reference to connectionism. In *The Handbook of Morphology*, Andrew Spencer and Arnold M. Zwicky (eds.), 428–452. Oxford: Blackwell Publishers.
- Tesar, Bruce and Smolensky, Paul (1998). Learnability in Optimality Theory. *Linguistic Inquiry* 29: 229–268.
- (2000). *Learnability in Optimality Theory*. Cambridge, Massachusetts: The MIT Press.
- Warner, Natasha, Allard Jongman, Joan Sereno and Rachèl Kemps (2004). Incomplete neutralization and other sub-phonemic durational differences in production and perception: Evidence from Dutch. *Journal of Phonetics* 32: 251–276.