PREDICTING THE UNPREDICTABLE: INTERPRETING NEUTRALIZED SEGMENTS IN DUTCH

MIRJAM ERNESTUS R. HARALD BAAYEN

University of Nijmegen and *Max Planck Institute for Psycholinguistics*

Among the most fascinating data for phonology are those showing how speakers incorporate new words and foreign words into their language system, since these data provide cues to the actual principles underlying language. In this article, we address how speakers deal with neutralized obstruents in new words. We formulate four hypotheses and test them on the basis of Dutch word-final obstruents, which are neutral for [voice]. Our experiments show that speakers predict the characteristics of neutralized segments on the basis of phonologically similar morphemes stored in the mental lexicon. This effect of the similar morphemes can be modeled in several ways. We compare five models, among them STOCHASTIC OPTIMALITY THEORY and ANALOGICAL MODELING OF LANGUAGE; all perform approximately equally well, but they differ in their complexity, with analogical modeling of language providing the most economical explanation.*

1. INTRODUCTION. The way language-users deal with foreign borrowings or words that they have never heard before may show us what knowledge language-users actually have about the distribution of sounds in their language, and how they organize and use this knowledge (see e.g. Hyman 1970, Hooper 1976:10, Derwing 1980, Gussenhoven & Jacobs 1998:38). For instance, adaptations of foreign borrowings into different languages often show (depending on sociocultural factors such as attitudes toward borrowing) that borrowings may be changed in the receiving language so that they conform to the phoneme inventory of that language, which suggests that languages have restricted phoneme inventories not because they happen to have words with no other phonemes, but because of highly ranked wellformedness constraints (Jacobs & Gussenhoven 2000).

We used experimental neologisms to come to grips with final devoicing in Dutch, the phenomenon that underlyingly voiced and voiceless obstruents are realized identically in syllable-final positions. Experimental neologisms can provide new insights into the knowledge speakers use when dealing with neutralized obstruents of which the underlying [voice]-specification is unavailable to them. We carried out an experiment in Dutch that allows us to answer questions such as the following. Do listeners know when hearing a segment in a neutralizing position in an unknown word that this segment may be realized differently in a non-neutralizing position, that is, that its underlying representation may be different from its surface representation? If, for instance, both /t/ and /d/ are realized as [t] at the end of words in a language, and listeners hear the new word [fat], do they know that the [t] is possibly [d] before affixes? And if listeners know this and are forced to determine the realization of the segment in non-neutralizing positions, is their choice for one of the possible realizations random? Is it based on the relative phonological strengths of the corresponding phonemes in the language, that is, on the extent to which these phonemes are resistant to assimilation processes and on the ease with which they can be phonetically realized? Or is their choice based on the

* We would like to thank Paul Boersma, Geert Booij, Wolfgang Dressler, Pim Mak, Rob Schreuder, and Royal Skousen for stimulating discussion and comments, and the editors Brian Joseph and Keith Johnson and the two anonymous referees for their careful and constructive criticism.

characteristics of phonologically or phonetically similar words in the lexicon? The answers to these questions take us beyond the traditional wisdom that underlying [voice] specifications are idiosyncratic only, and call for models that can deal with the probabilistic aspects of lexical structure.

2. NEUTRALIZATION OF [voice] IN DUTCH. In Dutch, the feature [voice] is distinctive for obstruents that precede a vowel-initial suffix (except *-achtig*), the past-tense morpheme, or vowels that belong to the same morpheme as the obstruents. For instance, the words *verwijden* ([vɛrʋɛidən]) and *verwijten* ([vɛrʋɛitən]) only differ in the [voice] specification of their coronal stop preceding the infinitive suffix *-en* ([-ən]), but differ in meaning (see 1a). The [+voice] specification of the [d] of *verwijden* seems to be an idiosyncratic property of this lemma. Similarly, the [−voice] specification of the [t] of *verwijten* seems to be lexically marked and unpredictable.

In word-final position, [voice] is nondistinctive. The realization of obstruents in this position as voiced or voiceless does not depend on their realization before vowel-initial suffixes, that is, on their underlying [voice] specification, but mainly on the type of following segment. They are generally realized as voiced before voiced stops (traditionally seen as the result of final devoicing and regressive voice assimilation, e.g. Booij 1995:58) and as voiceless before all other types of segments and before phrase boundaries (traditionally seen as the result of final devoicing only, e.g. Booij 1995:22). This is illustrated in 1b–d, which show that both *verwijd* (/vɛrʋɛid/) and *verwijt* (/vɛrʋɛit/) are generally realized with [d] before [b] (1b), and that they are both generally realized with [t] before [n] (1c) and before phrase boundaries (1d). The forms [vɛrʋɛit] and [vɛrʋɛid] are allomorphs of both *verwijd* and *verwijt*.

(1) a.	verwijden	[vɛrʋɛidən]	'widen-INF'
	verwijten	[vɛrʋɛitən]	'reproach-INF'
b.	verwijd bijna	[vɛrʋɛid bɛina:]	'widen almost'
	verwijt bijna	[vɛrʋɛid bɛina:]	'reproach almost'
c.	verwijd niet	[vɛrʋɛit nit]	'widen not'
	verwijt niet	[vɛrʋɛit nit]	'reproach not'
d.	verwijd	[vɛrʋɛit]	'widen'
	verwijt	[vɛrʋɛit]	'reproach'

It has generally been assumed that underlyingly voiced and underlyingly voiceless neutralized obstruents in Dutch are completely identical to each other with respect to all their acoustic characteristics. This assumption is supported in phonetic studies by Jongman, Sereno, Raaijmakers, and Lahiri (1992) and Baumann (1995). It is partly contradicted by Warner, Jongman, Sereno, and Kemps (2001), whose study suggests that if speakers have to read aloud pairs of words that differ only in the underlying [voice] specification of the neutralized obstruent, some of them tend to realize the neutralized obstruents that are underlyingly voiced with slightly more acoustic characteristics of voiced obstruents than the neutralized obstruents that are underlyingly voiceless.

3. HYPOTHESES. Speakers know the underlying characteristics of a certain neutralized segment only if they happen to know the realization of that segment in non-neutralizing positions. Thus, speakers of Dutch know that the final obstruent of *verwijd* is underlyingly /d/, instead of /t/, only if they know the realization of, for instance, *verwijden* [vɛrʋɛidən]. If speakers do not know relevant realizations, as in the case of unknown words, they have to guess the underlying characteristics of the neutralized segment. They can follow any of the following four strategies. Each strategy implies a different organization of the grammar.

Hypothesis 1. Speakers ignore the neutralization and assume that the relevant segment has underlyingly exactly the characteristics with which it was realized in the neutralizing position.

Hypothesis 2. Speakers randomly assign one of the possible underlying representations to the neutralized segment. They recognize that there is neutralization but do not undo it in a principled way.

Hypothesis 3. Speakers tend to choose that phoneme as the underlying representation that has the strongest position in the phonology of the language. They recognize that there is neutralization and base their choice for the underlying representation on phonology.

Hypothesis 4. Speakers tend to choose that phoneme as the underlying representation that makes the morpheme resemble similar morphemes in the lexicon. That is, they are more likely to choose a given underlying representation when there are more similar words in the lexicon sharing this underlying representation. Speakers recognize that there is neutralization and base their choice for the underlying representation on the distribution of the underlying representations among existing morphemes, serving as exemplars.

The correct hypothesis can be ascertained on the basis of obstruents neutralized for [voice] in Dutch, since the four hypotheses make clearly different predictions with respect to the phonological interpretation of these segments in unknown words.

According to hypothesis 1, the underlying [voice] specification that is assigned to the neutralized obstruent should depend on its actual realization. Recall that the realization of an obstruent neutralized for [voice] in Dutch depends mainly on the type of following segment. The hypothesis therefore implies that the interpretation of a neutralized obstruent in Dutch depends on the type of following segment. If the neutralized obstruent is not followed by a voiced stop, it is generally realized as voiceless, and, according to hypothesis 1, it should be interpreted as underlyingly voiceless. The pseudo-word [dɛrʋa:t], for instance, should be analyzed as /dɛrʋa:t/. But if the neutralized obstruent is followed by a voiced stop, it is generally realized as voiced, and should be interpreted as underlyingly voiced. The first word in the phrase [dɛrʋa:d bɛina] should be interpreted as /dɛrʋa:d/.

According to hypothesis 2, it should be completely unpredictable whether a certain speaker assigns an underlying [+voice] or [−voice] specification to a certain neutralized obstruent in Dutch. Every neutralized obstruent should be assigned an underlying [+voice] specification as often as an underlying [−voice] one.

Hypothesis 3 predicts that neutralized obstruents should be more often assigned an underlying [+voice] specification if they are higher in the following hierarchy.

(2) Hierarchy A: strength of voiced obstruents in Dutch phonology
bilabial stop > alveolar stop > alveolar fricative > labiodental fricative > velar fricative.

Hierarchy A summarizes the strength of voiced obstruents in Dutch phonology. It positions the voiced fricatives lower than the voiced stops since all underlyingly voiced fricatives in morpheme-initial position are invariably realized as voiceless after obstruents (see 3a), whereas underlyingly voiced stops in initial position, at least those belonging to content words, are never realized as voiceless (see 3b; see Ernestus 2000:50).

(3) a.	groep	[ɣrup]	'group'	werkgroep	[ʋɛrkxrup]	'working group'
	vlucht	[vlɵxt]	'flight'	dagvlucht	[dɑxflɵxt]	'day flight'
	zuster	[zɵstər]	'sister'	dagzuster	[dɑxsɵstər]	'day sister'
b.	dag	[dɑx]	'day'	werkdag	[ʋɛrgdɑx]	'working day'
	boot	[bo:t]	'boat'	dagboot	[dɑɣbo:t]	'day boat'

Among the stops, the bilabial one is ranked highest in hierarchy A, because, while initial /d/s can be realized as voiceless after obstruents if they belong to certain function words, initial /b/s never are (e.g. Zonneveld 1983:306; see 4). Apparently, /b/ has a stronger position in Dutch phonology than /d/.

(4) dan	[dɑn]	'then'	loop dan	[lo:bdɑn], [lo:ptɑn]	'walk then'
ben	[bɛn]	'am'	ik ben	[ɪgbɛn], *[ɪkpɛn]	'I am'

The relative strengths of the voiced fricatives in Dutch phonology emerge from the effectiveness of their voiced/voiceless opposition in the different varieties of Dutch. First, the voiced/voiceless opposition is effective in fewer varieties of Dutch for the velar fricative than for the labiodental fricative, and in fewer varieties for the labiodental fricative than for the alveolar fricative. Thus, the opposition between /x/ and /ɣ/ is maintained only in the varieties of Dutch spoken in the Southern parts of the Netherlands and in Flanders, whereas the opposition between /s/ and /z/ is maintained in all varieties, except those spoken in the north and northwest of the Netherlands. If a variety does not maintain the opposition for a fricative, it lacks the voiced variant (Collins & Mees 1981:159, Gussenhoven & Bremmer 1983:57, Slis & van Heugten 1989, van Reenen & Wattel 1992). Second, those varieties of Dutch maintaining the opposition for all fricatives preserve the distinction in a larger number of phonological contexts for the alveolar fricative (in all contexts except in obstruent clusters) than for the labiodental one (in all contexts except phrase initial and in obstruent clusters), and for the velar fricative in even fewer contexts (only intervocalically, Gussenhoven & Bremmer 1983). Apparently, /z/ is phonologically a stronger phoneme than /v/, which is a stronger phoneme than /ɣ/, as encoded in hierarchy A.

Note that hierarchy A not only reflects the relative strength of voiced obstruents for Dutch but also their strength in the languages of the world. Ohala (1983:201) noted that for the 706 languages surveyed by Ruhlen (1975), there is more than twice the probability of voicing being present on stops than on fricatives (Ladefoged & Maddieson 1996:176), showing that also in the general hierarchy voiced stops are ordered above voiced fricatives. And in this hierarchy, bilabial stops are ranked highest, since they are more often voiced than stops of other places of articulation, probably due to articulatory, acoustic, and auditory factors (Ohala 1983:195). Hierarchy A reflects the general unmarkedness hierarchy of voiced obstruents.

Finally, turning to hypothesis 4, we find that it predicts that the mirror image of hierarchy A should hold, and that neutralized obstruents should more often be assigned an underlying [+voice] specification if they are high in the following hierarchy.

(5) Hierarchy B: underlyingly voiced morpheme-final obstruents in the Dutch lexicon
velar fricative > labiodental fricative > alveolar fricative > alveolar stop > bilabial stop.

Hierarchy B is based on all 1697 words attested in the Dutch section of the CELEX lexical database (Baayen et al. 1995) that consist of a nominal, verbal, or adjectival base morpheme ending in an obstruent of which both the voiced and voiceless variants are phonemes in Dutch, and that is followed by the comparative suffix ([-ər]), the infinitive suffix ([-ən]), or the plural suffix [-ən]. The final obstruents of these nominal, verbal, and adjectival base morphemes are not in neutralizing positions for [voice], since they are followed by vowel-initial suffixes. They can be in neutralizing positions, however, which is the case when the morphemes are used in isolation. Inspection of the [voice] specification of these obstruents therefore provides information on the underlying [voice] specification of neutralized obstruents in Dutch. The probability that a neutralized obstruent is underlyingly voiced depends on its place and manner of articulation, as reflected by hierarchy B. This is shown in Table 1, which lists the absolute numbers and percentages of morphemes with underlyingly voiced and voiceless final obstruents, subcategorized for the types of these obstruents. In this table and in the following ones, we use the symbols P for bilabial stops, T for alveolar stops, S for alveolar fricatives, F for labiodental fricatives, and X for velar fricatives, both the voiced and the voiceless variants. In Table 1, the percentage of morphemes with underlyingly voiced final obstruents increases systematically from 9% to 97% as we proceed from those with bilabial stops down to those with velar fricatives, as is reflected by hierarchy B. A chi-squared test confirms that we are not observing a random pattern ($\chi^2(4) = 415.1$, $p < 0.001$), and separate chi-squared tests on pairs of adjacent obstruents in the hierarchy are all significant as well ($p < 0.01$ after Bonferroni-adjustment).

OBSTRUENT	VOICED		VOICELESS		TOTAL	
	#	%	#	%	#	%
P	20	9	210	91	230	100
T	177	25	542	75	719	100
S	151	33	300	66	451	100
F	116	70	50	30	166	100
X	127	97	4	3	131	100

TABLE 1. Morphemes ending in underlyingly voiced or voiceless final obstruents in the CELEX data set, by type of obstruent.

Hypothesis 4 embodies a further prediction, namely that speakers of Dutch exploit all correlations between the characteristics of the morphemes present in their lexicons and the underlying [voice] specifications of the final obstruents of these morphemes. When interpreting a neutralized segment, they might not restrict themselves to the correlation between the underlying [voice] specification of the neutralized obstruent and its type, which is the correlation reflected by hierarchy B. If there are also correlations between other characteristics of the morphemes and the underlying [voice] specifications of the neutralized obstruents, speakers might exploit these as well. For instance, if there is a correlation between the quality of a vowel and the underlying [voice] specification of the following obstruent, such that high vowels are typically followed by underlyingly voiced obstruents and low vowels by underlyingly voiceless obstruents, speakers may tend to assign an underlying [+voice] specification to neutralized obstruents following high vowels and an underlying [−voice] specification to neutralized obstruents following low vowels.

To ascertain whether such correlations are present in Dutch, we analyzed the 1697 words in our data set with a k-nearest neighbors algorithm using information gain

PHONOLOGICAL PROPERTY	INFORMATION GAIN
Quality of onset of final syllable (–, b, j, x, kr, . . .)	0.054
Quality of vowel of final syllable (u, a:, ɔ, . . .)	0.095
Presence and quality of consonant preceding the final obstruent (–, l, n, s, k, . . .)	0.103
Type of final obstruent (P, T, S, F, X)	0.187
Number of syllables in the word (1, 2, 3, 4)	0.017
Position of stress in the word (antepenult, penult, final)	0.001

TABLE 2. Information gains for phonological properties of morphemes as predictors of underlying [voice] specification of final obstruents.

weighting as available in TiMBL (Daelemans et al. 2002). For each phonological property listed in Table 2, we inspected the information gain. The information gain measures the extent to which one's uncertainty (entropy) about the underlying [voice] specification of a final obstruent decreases when one is told the value of that phonological property. That is, the information gains tell us how informative phonological properties are for deciding on the [voice] specification of final obstruents. We found that the quality of the vowel in the final syllable has a high information gain (0.095) compared to the information gain associated with the position of stress (0.001). This implies that the quality of the vowel is relevant, or that its quantity is relevant, since there is a fixed relation between vowel quality and quantity in Dutch. In addition, we found a similarly high information gain (0.103) for the quality of the consonant, if present, preceding the final obstruent. The highest information gain of all emerged for the type of the final obstruent (0.187). When we allow TiMBL to take into account only the correlation between the underlying [voice] specification and the type of the neutralized obstruent, that is, hierarchy B, it classifies the underlying [voice] specification of the neutralized obstruent of an existing word, given all other words in the lexicon (using leave one out validation), correctly in 76.3% of cases. Its classification accuracy increases to 82.6% when we allow TiMBL to also take into account the quality of the vowel in the final syllable and the presence and quality of the consonant preceding the final obstruent. This is a statistically significant improvement ($\chi^2(1) = 20.29$, $p < 0.001$, with continuity correction). Providing even more information to the analysis, such as the onset of the final syllable (information gain weight 0.054), the position of stress in the word (information gain weight 0.001), or the number of syllables in the word (information gain 0.017), does not improve classification accuracy.

We also analyzed our data set with a CLASSIFICATION AND REGRESSION TREE ANALYSIS (CART, Breiman et al. 1984). This technique divided the data into groups of morphemes characterized by their final rhymes, and as similar as possible with respect to the underlying [voice] specification of their final obstruent. CART created eleven groups, which are characterized in 6, together with the percentage of morphemes in each group underlyingly ending in a voiced obstruent. The misclassification rate of this CART analysis for existing words is 16.8%; that is, 16.8% of the words have a different underlying voice specification than the majority of the words in their group. This is a nonsignificant improvement with respect to the 17.4% misclassification rate of TiMBL ($\chi^2(1) = 1.13$, $p = 0.133$, with continuity correction). The division of the data shows again that the quality of the final vowel, the presence and quality of an extra consonant in the final rhyme, and the type of the final obstruent correlate to the underlying [voice] specification of the final obstruent.

(6) Groups created by CART

Morphemes ending in	Percentage of voiced obstruents
1. {ɛi, ɑu, œy, a:, e:, o:, ø:, i, u}{–, j, l, m, n, r}P	0
2. {ɛi, ɑu, œy, a:, e:, o:, ø:, i, u}{–, j, l, m, n, r}T	37.2
3. {ɛi, ɑu, œy, a:, e:, o:, ø:, i, u}{–, j, l, m, n, r}S	76.5
4. {f, k, p, s, t, x}{P, T, S}	1.9
5. {ɑ, ɛ, ɪ, ɔ, ɵ, y}{–, m, r}{P, T, S}	13.5
6. {ɑ, ɛ, ɪ, ɔ, ɵ, y}{l, n}{P, T, S}	35.7
7. {ɛi, ɑu, a:, e:, o:, ø:, y}{–, j, l, r, m, n}{F, X}	99.2
8. {i, u}{–, m} F	77.8
9. {ɑ, ɛ, ɪ, ɔ, ɵ}{–, m} F	9.1
10. {ɑ, ɛ, ɪ, ɔ, ɵ, i, u}{l, r} F	87.5
11. {ɑ, ɛ, ɪ, ɔ, ɵ, i, u}{–, j, l, r, m, n} X	95.3

In contrast to TiMBL, CART builds a classificatory partition for this data set that is well interpretable phonologically. This partition suggests that long vowels preceding final obstruents tend to favor underlyingly voiced specifications for these obstruents (compare in 6 group 3 with group 6, and group 7 with groups 9 and 10; the only exception is provided by the bilabial stops in group 1). Thus, in accordance with Booij 1999, we see that fricatives are generally underlyingly voiced after long vowels (groups 3 and 7). In addition, we find that preceding obstruents tend to favor underlyingly voiceless specifications (compare group 4, morphemes with additional obstruents in the final rhyme, with all other groups, in which the final obstruents are not preceded by other obstruents).

Finally, the effect of the quantity of the vowel in the final syllable, the presence and sonority of an extra consonant in the final rhyme, and the type of the final obstruent on the underlying [voice] specification of this final obstruent also appears from Tables 3 and 4. These tables list the numbers of morphemes with underlyingly voiced and voiceless final obstruents, cross-classified for type of obstruent and the preceding type of segment. Table 3 lists the numbers for morphemes ending in syllables with long vowels, while Table 4 lists those for morphemes ending in syllables with short vowels. The figures in Table 3 are in general higher than the corresponding figures in Table 4, indicating an effect of vowel quantity. In addition, we see that the percentage of underlyingly voiced obstruents is the lowest in the right most columns, suggesting an effect of the type of segment preceding the final obstruent. Finally, we see again an effect of the type of final obstruent, as we proceed from the top of the tables to the bottom.

OBSTRUENT TYPE	NO PRECEDING CONSONANT				PRECEDING SONORANT				PRECEDING OBSTRUENT			
	VOICED		VOICELESS		VOICED		VOICELESS		VOICED		VOICELESS	
	#	%	#	%	#	%	#	%	#	%	#	%
P	0	0	52	100	–		–		–		–	
T	55	32	116	68	22	71	9	29	4	17	20	83
S	78	82	17	18	14	67	7	33	0	0	11	100
F	62	100	0	0	1	100	0	0	–		–	
X	65	98	1	2	–		–		–		–	
	$\chi^2(4) = 231.1, p < 0.001$				$\chi^2(2) = 0.551, p = 0.759$				$\chi^2(1) = 0.751, p = 0.386$			

TABLE 3. Morphemes ending in underlyingly voiced or voiceless obstruent in the CELEX data set, by type of obstruent and type of preceding segment (all morphemes end in syllables with long vowels).

OBSTRUENT TYPE	NO PRECEDING CONSONANT				PRECEDING SONORANT				PRECEDING OBSTRUENT			
	VOICED		VOICELESS		VOICED		VOICELESS		VOICED		VOICELESS	
	#	%	#	%	#	%	#	%	#	%	#	%
P	20	22	72	78	0	0	60	100	0	0	6	100
T	21	16	113	84	56	35	103	65	1	1	124	99
S	3	3	115	97	31	41	45	59	0	0	85	100
F	4	10	38	90	28	82	6	18	–		–	
X	32	91	3	9	19	100	0	0	–		–	
	$\chi^2(4) = 143.9, p < 0.001$				$\chi^2(4) = 96.4, p < 0.001$							

TABLE 4. Morphemes ending in underlyingly voiced or voiceless obstruent in the CELEX data set, by type of obstruent and type of preceding segment (all morphemes end in syllables with short vowels).

The two tables do not present exactly the same quantitative pattern. Consider, for instance, the columns that tabulate the cases in which the obstruent immediately follows the vowel. In the case of a long vowel, the percentage of underlyingly voiced obstruents largely follows hierarchy B. However, in the case of a short vowel, hierarchy B is not followed. Clearly, the quantity of the vowel interacts with hierarchy B. The role of vowel quantity is supported by a GLM (generalized linear model) analysis with a logit link function of the counts of voiced and voiceless alveolar obstruents (T and S) in Tables 3 and 4, the only obstruents for which we have data for all six conditions. Both vowel quantity and the kind of preceding segment emerge from this analysis as significant ($F(1,7) = 15.7$, $p = 0.005$ for vowel quantity; $F(2,7) = 14.6$, $p = 0.003$ for the preceding consonant). No reliable difference between T and S could be observed ($F(1,7) < 1$).

Tables 3 and 4 do not cover words with [i, u, y] in their final syllables. These vowels behave like long vowels in phonology (e.g. like long vowels they need not be followed by coda consonants), but are acoustically as short as short vowels (e.g. Moulton 1962, Booij 1995:5). Since the vowels [i, u, y] are phonologically long but phonetically short, they cannot be univocally classified with the long or short vowels. In 6, these vowels pattern sometimes with long vowels and sometimes with short vowels. Table 5 provides a more detailed comparison of words with [i, u, y] with words with long vowels and with words with short vowels. A GLM analysis with a logit link function of the counts of underlyingly voiced and voiceless obstruents for the words with short vowels and [i, u, y] shows a statistically (nearly) significant effect for vowel type ($F(1,8) = 5.17$, $p = 0.053$). Final obstruents are more often voiced after [i, u, y] than after short vowels. Interestingly, a GLM analysis with a logit function of the counts for the words with

CODA	[i, u, y]				LONG VOWELS				SHORT VOWELS			
	VOICED		VOICELESS		VOICED		VOICELESS		VOICED		VOICELESS	
	#	%	#	%	#	%	#	%	#	%	#	%
- P	0	0	20	100	0	0	52	100	20	22	72	78
- T	13	24	41	76	55	32	116	68	21	15	113	85
Son T	5	50	5	50	22	71	9	29	56	35	103	65
Obstr T	0	0	11	100	4	17	20	83	1	1	124	99
- S	24	71	10	29	78	82	17	18	3	3	115	97
Son S	1	25	3	75	14	67	7	33	31	41	45	59
Obstr S	0	0	7	100	0	0	11	100	0	0	85	100
- F	21	78	6	22	62	100	0	0	4	10	38	90
- X	11	100	0	0	65	98	1	2	32	91	3	9

Son = sonorant consonant Obstr = obstruent

TABLE 5. Underlyingly voiced and voiceless obstruents for morphemes containing [i, u, y] and for corresponding morphemes with long and short vowels, by type of final coda.

[i, u, y] and for the words with long vowels shows a statistically significant effect of vowel type as well ($F(1,8) = 8.92$, $p = 0.017$). Obstruents following [i, u, y] are underlyingly voiced less often than obstruents following long vowels. The vowels [i, u, y] behave like neither the phonologically and phonetically long vowels nor like the phonologically and phonetically short vowels with respect to their effect on the [voice] specification of final obstruents. They behave like vowels in between.

In conclusion, we have used nonparametric analogical models like TiMBL and CART to reveal the regularities present among existing words in the lexicon. We have shown that the underlying [voice] specification of final obstruents is predictable to a large extent on the basis of the type of this obstruent and the types of the preceding segments. This finding is surprising, since the [voice] specification of the neutralized obstruent is distinctive, and precisely therefore generally tacitly assumed to be unpredictable.

If hypothesis 4 is correct, we expect that the probability with which speakers choose an underlying [+voice] specification for the final obstruent of a new word mirrors the percentage of words underlyingly ending in voiced obstruents among the words similar to the new word. We expect, for instance, that speakers assign an underlying [+voice] specification to the final obstruent of [plɔs] with a probability of approximately 13.5%, since this word falls in group 5 as defined by CART (see 6). In addition, we expect that speakers assign a [+voice] specification to the final obstruent of [bɛlf] with a probability of approximately 87.5%, since this word belongs to group 10.

In order to test the four hypotheses about the interpretation of neutralized obstruents in unknown words, we carried out a production experiment with phonotactically legal pseudo-words in Dutch.

4. A PRODUCTION EXPERIMENT. Speakers of Dutch were presented with pseudo-words. Their task was to interpret the final, neutralized, obstruents of these words as underlyingly voiced or voiceless.

PROCEDURE. Participants listened through closed head-phones (Sony MDR-55) to phrases consisting of the subject pronoun *ik* [ɪk] 'I' and a nonexisting verb in the first person singular present tense, obeying the phonotactic restrictions of Dutch. Examples of these phrases are [ɪk tif] *ik tief*, [ɪk dɛnt] *ik dent*, and [ɪk dɑup] *ik daup*. The final obstruents of the verb forms are word final and therefore phonologically neutralized for [voice]. The participants' task was to write down as accurately as possible the past-tense forms of the verb forms. In Dutch, past tense is created by suffixing *-te* [tə] or *-de* [də] to the verb stem. The suffix *-te* is used if the final obstruent is underlyingly voiceless; *-de* is used if this obstruent is underlyingly voiced. Thus, the choice of the suffix reveals the participants' phonological interpretation of the presented neutralized obstruents. A participant who creates the past-tense form *tiefte* has interpreted the final obstruent of [tif] as underlyingly voiceless, and a participant who writes down *tiefde* has interpreted the final obstruent as underlyingly voiced.

The participants were asked to write down the full past-tense forms, instead of just *de* or *te*, so that we could ascertain that they had understood the rhymes of the pseudo-words as intended. We need this information since hypothesis 4 predicts that not only the characteristics of the final obstruents themselves but also the qualities of the preceding vowel and consonant, if present, affect the participants' interpretation of the obstruents.

The experiment was self-paced. Participants were presented with a new phrase only after they had indicated that they were ready by pushing a button.

MATERIALS. The participants were presented with 192 nonexisting monosyllabic verb forms representing nearly all possible rhymes in Dutch. Not included in the experiment were very low-frequency rhymes, phonotactically abnormal rhymes, and rhymes that are almost always created by a suffix. The forms presented in the experiment are listed in the Appendix.

The phrases were recorded by a female speaker of a variety of Dutch maintaining the voiced/voiceless opposition for all fricatives and stops. Except for the twenty pseudo-verbs ending in a velar fricative, all final obstruents were spelled with a voiceless grapheme in the list of phrases that the speaker read aloud for the recording. Since the pseudo-verbs were phrase final, their final obstruents were realized as voiceless. (For detailed information on the phonetic characteristics of the materials, the role of the orthographical representation of the final obstruents, and the problem of incomplete neutralization, see Ernestus & Baayen 2002). The phrases were recorded on a DAT (BASF master 94) in a soundproof room on a portable DAT-recorder Aiwa HD S100 and a Sony microphone ECM MS957. The recordings were stored as .wav files (sample rate: 48 KHz) on a computer by means of the speech analysis package Praat (Boersma 1996). The phrases were presented in one of three random orders to the participants, with four intervening breaks. The actual test phrases were preceded by eleven practice phrases with existing verbs, and twenty practice phrases with pseudo-verbs.

PARTICIPANTS. Twenty-eight native speakers of Dutch, students at Nijmegen University, participated in the experiment. They were paid for their participation.

RESULTS AND DISCUSSION. In the great majority of trials, participants wrote past-tense forms ending in *-de* or *-te*. Occasionally, however, they produced other, unexpected past-tense forms. Some of these forms were created by vowel-alternation (mostly by one participant, who is responsible for 28 out of 34 such responses). For instance, *ties* was produced as the past-tense form for [tas] *taas*, and *bast* as the past-tense form for [bɪst] *bist*. These past-tense forms reveal nothing of the participants' interpretations of the presented neutralized obstruents as underlyingly voiced or voiceless, and were therefore not taken into account in the analysis.

We also discarded responses ending in *-bte*, *-pde*, *-dte*, *-tde*, *-zte*, *-vte*, *-gte*, and *-chde*, which are illegal according to Dutch orthography. In these orthographic transcriptions, the grapheme representing the stem-final obstruent indicates a voice specification that is opposite to the specification indicated by the form of the past-tense suffix. As a consequence, we do not know what the intended underlying [voice] specification is. Possibly, the participants were themselves undecided. Our decision to leave out these past-tense forms does not affect the final pattern of results. We also discarded past-tense forms the stems of which do not completely correspond to the stems of the presented stimuli. For instance, we disregarded *bunsde* as the past-tense form for [bɵnt] *bunt*, and *duitte* as the past-tense form for [dyt] *duut*. These particular forms are probably not the past-tense forms of the pseudo-verbs we presented, but the past-tense forms of slightly different pseudo-verbs which the participants thought they had heard. A rough analysis of the data with and without these past-tense forms shows that the decision to leave them out does not affect the final pattern of results. The remaining numbers of past-tense forms ending in *-de* and *-te* for each pseudo-word can be found in the Appendix. The intersubject agreement about the choice of the past tense forms as measured by Cohen's κ was 0.294 ($Z = 99, 91$, $p < 0.001$).

Tables 6 and 7 list the counts of these responses ending in *-de* or *-te*, subcategorized by the type of the stem-final obstruent of the presented form and by the type of the

OBSTRUENT TYPE	NO PRECEDING CONSONANT				PRECEDING SONORANT				PRECEDING OBSTRUENT			
	de		*te*		*de*		*te*		*de*		*te*	
	#	%	#	%	#	%	#	%	#	%	#	%
P	4	3	124	97	–		–		–		–	
T	7	4	168	96	8	11	68	89	36	14	227	86
S	98	55	79	45	25	48	27	52	39	15	221	85
F	88	64	49	36	–		–		–		–	
X	128	81	30	19	–		–		–		–	
	$\chi^2(4) = 322.3, p < 0.001$				$\chi^2(1) = 20.8, p < 0.001$				$\chi^2(1) = 0.092, p = 0.762$			

TABLE 6. Responses ending in *-de* and *-te*, by type of stem-final obstruent and type of preceding segment (pseudo-words with long vowels).

preceding segment. Table 6 shows the pseudo-verbs with long vowels, Table 7 the pseudo-verbs with short vowels. The words with [i, u, y] are not included. We return to them later.

The data in Tables 6 and 7 allow us to evaluate the four hypotheses formulated in the preceding section. First, consider hypothesis 1, according to which speakers should ignore neutralization and assign that underlying [voice] specification to a neutralized obstruent which is in accordance with its realization. Given that 24% of the pseudo-verbs were interpreted as having an underlyingly voiced obstruent, even though they were all realized with voiceless obstruents, this hypothesis is obviously wrong.

OBSTRUENT TYPE	NO PRECEDING CONSONANT				PRECEDING SONORANT				PRECEDING OBSTRUENT			
	de		*te*		*de*		*te*		*de*		*te*	
	#	%	#	%	#	%	#	%	#	%	#	%
P	11	12	83	88	3	1	200	99	1	1	78	99
T	9	7	127	93	35	13	241	87	26	6	446	94
S	20	15	112	85	84	38	137	62	32	8	392	92
F	33	25	101	75	151	53	132	47	–		–	
X	66	65	36	35	148	90	17	10	–		–	
	$\chi^2(4) = 131.4, p < 0.001$				$\chi^2(4) = 410.5, p < 0.001$				$\chi^2(2) = 5.1, p = 0.078$			

TABLE 7. Responses ending in *-de* and *-te*, by type of stem-final obstruent and type of preceding segment (pseudo-words with short vowels).

Hypothesis 2 is clearly incorrect as well. The structure in Tables 6 and 7 shows that speakers do not randomly assign one of the possible underlying representations to neutralized segments. Consider, for instance, the past-tense forms created for the twenty verbs ending in a velar fricative. The majority of the past-tense forms created for seventeen of these verbs end in the suffix *-de*, showing that the probability of finding *-de* after the velar fricative is greater than 0.5 (two-tailed sign-test, $n = 20$, $T = 14$, $p = 0.010$).

The experimental results also falsify hypothesis 3. The percentage of neutralized obstruents interpreted as underlyingly voiced increases in Table 6, which lists the results for pseudo-words with long vowels, as we move down from the bilabial stops to the velar fricatives. This pattern is fully reliable for the contexts in which there is no preceding consonant, as well as in the context in which the final obstruent is preceded by a sonorant consonant (see the chi-squared tests, also listed in this table). Turning to Table 7, which lists the results for pseudo-verbs with short vowels, we find the same pattern for neutralized obstruents following sonorant consonants and obstruents. The data clearly do not reflect the phonological hierarchy A. Speakers do not interpret the underlying [voice] specification of neutralized obstruents on the basis of the relative phonological strengths of the voiced variants of these obstruents.

We now turn to hypothesis 4. This hypothesis is supported by the data. As we just mentioned, the percentage of neutralized obstruents interpreted as underlyingly voiced increases as we move down from the bilabial stop to the velar fricative for the pseudo-words with long vowels (Table 6, data block 1), and for the pseudo-words ending in a short vowel, a consonant, and the final obstruent (Table 7, data block 2). For both these data sets, the probability of obtaining the observed percentage-based ordering of $P < T < S < F < X$ out of all possible orderings is 0.008. The observed orderings are in accordance with hierarchy B, and support hypothesis 4.

Also the obstruents directly preceded by a short vowel (Table 7) provide some support for hypothesis 4. When we proceed from the bilabial stops to the velar fricatives directly following short vowels (Table 7, data block 1), we observe a pattern that differs somewhat from hierarchy B, in that there are more (but not significantly more) *-de* responses to words ending in P than to words ending in T. The existing words listed in Table 4 (data block 1) also show a divergence from hierarchy B, here not only with respect to P but also with respect to T. The existing words show a more pronounced divergence from hierarchy B, in that P is more often underlyingly voiced than S ($\chi^2(1) = 17.61$, $p < 0.0001$) for existing words, but not for the experimental data ($p > 0.5$). In other words, the experimental pattern holds the middle ground between hierarchy B and the pattern displayed by the existing words. This suggests that the interpretation of words with short vowels is not only affected by the existing words with short vowels, but also by the other existing words that do follow hierarchy B, for instance, the words with [i, u, y] (see Table 5), exactly as is predicted by the CART analysis (see groups 5, 6, 10, and 11 in 6).

The data further support hypothesis 4 since they reflect similar correlations between the properties of the final rhymes of the words and the underlying [voice] specification of the final obstruents as we found in the CELEX data set. This is shown in Figure 1, which gives both the percentages of words underlyingly ending in a voiced obstruent in the lexical data set (left panels) and in the experimental results (the percentages of past-tense forms created with the suffix *-de* in the experiment, right panels). In the upper panels, the words are broken down by the type of their final obstruent. Note that the percentage of underlying voicing ranks the obstruents in the CELEX data (upper left panel) in the same way as in the experimental data (upper right panel). The probability of this same ranking under change conditions is 0.008 (1/5!), indicating that we are dealing with a significantly similar pattern. In the middle panels, the words are broken down by the quantity of their final vowel. We observe a similar decrease in percentage of underlying voicing when going from the long to the short vowels, but the words with phonologically long but phonetically short vowels [iuy] pattern differently in CELEX (middle left panel) and the experiment (middle right panel). In the lower panels, the words are broken down by the type of segment preceding the final obstruent. Again, the patterns in the lower left and lower right panels are similar.[1]

Final support for hypothesis 4 comes from Figure 2, which plots for each of the 192 experimental words the observed percentage of past-tense forms created with *-de* and

[1] GLM analyses with a logit link function of the counts of voiced and voiceless obstruents in CELEX and in the experiment revealed significant main effects of data source and type of rhyme (obstruent in the upper panels of Fig. 1, vowel in the middle panels, and segment preceding the final obstruent in the bottom panels), as well as interactions between the two ($p < 0.0001$ in all cases). The main effect of data source points to a bias for voiceless obstruents in the experiment. Inspection of the coefficients of the model (using contrast coding) shows that, except for the analysis of the vowel data, the effect of the interaction is small compared to the main effect of rhyme. This confirms that the main pattern in the experimental data mirrors that in CELEX.

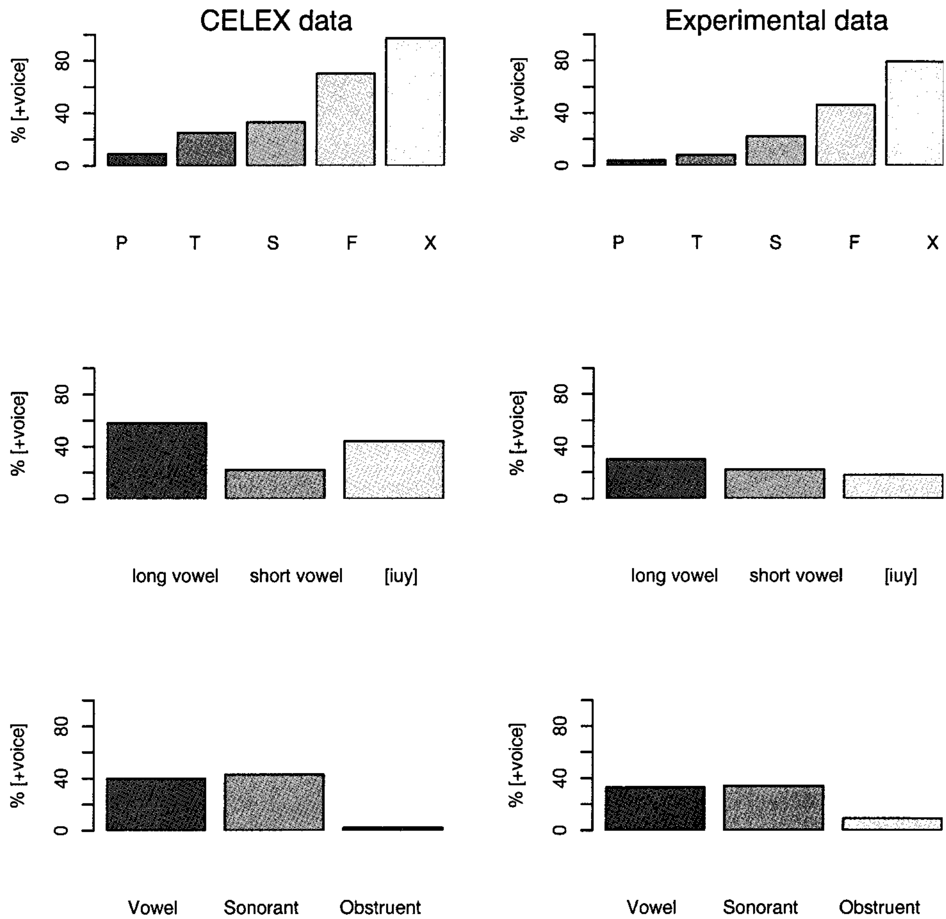

FIGURE 1. Words underlyingly ending in a voiced obstruent in CELEX data set and in experimental data. Broken down by type of final obstruent (upper panels), quantity of final vowel (middle panels), and type of the segment preceding final obstruent (lower panels).

the percentage of past-tense forms with *-de* predicted by the CART analysis, that is, the probability that the final obstruent is interpreted as underlyingly voiced (see 6). The solid line in the figure represents a nonparametric scatterplot smoother (Cleveland 1979). This smoother shows a good correlation between the observed and predicted percentages of voiced obstruents ($r_s = 0.50$, $S = 588680$, $p < 0.001$). In addition, it reveals an overall bias for a voiceless interpretation of the neutralized obstruent. Without the bias, the smoother would have been approximately identical to the dashed line, which represents the line $y = x$. The bias for a voiceless interpretation of the obstruents in the experiment corresponds to the fact that the majority of the final obstruents in the lexical database (65%) are underlyingly voiceless.

We now turn to the experimental words with the vowels [i, u, y], which we have not yet considered. Recall that these vowels are special in that they typically behave like long vowels in phonology, but as short vowels in phonetics. They behave neither like phonologically and phonetically short vowels nor like phonologically and phonetically long vowels with respect to their effects on the underlying [voice] specification

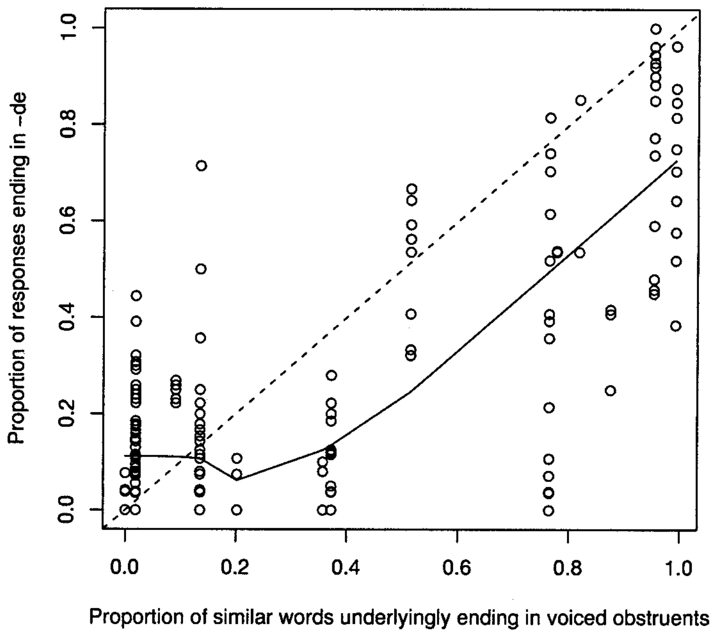

FIGURE 2. Percentage of voicing as a function of analogical support for the 192 pseudo-words. Solid line represents a nonparametric regression line; dashed line represents the line $y = x$.

of final obstruents of existing words (see §3). Table 8 lists the numbers and percentages of responses ending in *-te* and *-de* for the experimental words containing [i, u, y], broken down by the type of coda (type of final obstruent and presence/absence and sonority of a preceding consonant). In addition, the table lists the responses for the corresponding types of words with short and long vowels. The table shows that also the obstruents following [i, u, y] follow hierarchy B ($\chi^2(4) = 82.2587$, $p < 0.001$). A GLM analysis with a logit link function of all data in the table with type of coda and type of vowel as independent variables and excluding the interaction of these two factors from the model definition confirms that type of vowel affected the participants' responses ($F(2,14) = 8.38$, $p = 0.004$). More interestingly, a subanalysis of the words

CODA	[i, u, y]				LONG VOWELS				SHORT VOWELS			
	de		*te*		*de*		*te*		*de*		*te*	
	#	%	#	%	#	%	#	%	#	%	#	%
- P	1	2	48	98	4	3	124	97	11	12	83	88
- T	4	5	76	95	7	4	168	96	9	7	127	93
Son T	5	19	22	81	8	11	68	89	35	13	241	87
Obstr T	19	4	243	96	36	14	227	86	26	6	446	94
- S	22	40	33	60	98	55	79	65	20	15	112	85
Obstr S	24	9	243	91	39	15	221	85	32	8	392	92
- F	39	49	41	51	88	64	49	36	33	25	101	75
- X	29	64	16	36	128	81	30	19	66	65	36	35

Son = sonorant consonant Obstr = obstruent

TABLE 8. Responses ending in *-de* and *-te* for words containing [i, u, y] and corresponding types of words with long and short vowels (broken down by types of segments in rhymes).

with long vowels and with [i, u, y] only also shows a reliable effect of type of vowel ($F(1,7) = 20.05$, $p = 0.003$), whereas a subanalysis of the words with short vowels and the words with [i, u, y] does not ($F(1,7) = 3.59$, $p = 0.100$). We also conducted a second analysis of the data, in which we included the interaction term. Including the interaction term leads to a saturated model that overfits the data (all residuals are zero), which is a disadvantage compared to the previous analysis. But this second analysis makes it possible to trace more details of the structure in the data set. The second analysis revealed not only significant main effects of coda and vowel but also an interaction. A subanalysis of the data excluding the short vowels revealed main effects of coda and vowel and no interaction of coda by vowel. Words with long vowels elicited more *-de* responses than did the words with [i, u, y]. A subanalysis of the data excluding the long vowels revealed both main effects and the interaction to be significant. Although in general there is not much of a difference between words with [i, u, y] and words with short vowels, notably short vowels directly followed by P elicited more *-de* responses than did [i, u, y]. Considered jointly, these analyses suggest that words with [i, u, y] patterned predominantly with the words with short vowels. Apparently, the quantitative similarity structure which the participants relied on appears to be based at least in part on phonetics.

In summary, the results of our experiment provide several cues to how listeners deal with neutralization. Speakers of Dutch do not ignore the [voice] neutralization of final obstruents (contra hypothesis 1), and undo the neutralization in a principled way (contra hypothesis 2). They agree about the underlying [voice] specification of final obstruents in pseudo-words, which does not follow from theories in which [voice] specifications are specified in the lexicon, without additional ad hoc redundancy rules. The speakers do not base their prediction on hierarchy A, the hierarchy of the phonological strength of voiced obstruents in Dutch (contra hypothesis 3). Instead, they rely on their knowledge of the underlying representations of similar words. They use these words as exemplars, which is in accordance with hypothesis 4.

5. MODELING. We now consider several formal and quantitative models that account for the observed effects of the similarity structure in the speakers' mental lexicon. We compare both the performance and complexity of these models. We first discuss three models that derive rules of some kind for predicting the [voice] specification of final obstruents. These rules are derived from the data, and once formulated, they can be applied to new forms without consulting the original data. We then turn to two analogical models, which differ from these rule models in that the individual instances in the data keep playing a crucial role even after the model has been built.

The first rule model we consider is a stochastic version of optimality theory (OT). Optimality theory (Prince & Smolensky 1993, McCarthy & Prince 1993a,b, 1995, and related work) assumes that there are constraints on the possible characteristics of word forms in a language. These constraints are stored separately from the words in the speakers' grammar. They are ranked in a hierarchy, and only those word forms that optimally satisfy the hierarchy in force for that language are possible.

The effects of the similarity structure in the speakers' mental lexicon revealed in this article can be accounted for within optimality theory by positing constraints that reflect this similarity structure, and by assuming that these constraints affect the underlying representations of words, instead of their surface representations as they normally do. A minimal set of constraints is given in 7.

(7)	*P[+voice]	Bilabial stops are not underlyingly voiced.
	*T[+voice]	Alveolar stops are not underlyingly voiced.
	*S[+voice]	Alveolar fricatives are not underlyingly voiced.
	*F[−voice]	Labiodental fricatives are not underlyingly voiceless.
	*X[−voice]	Velar fricatives are not underlyingly voiceless.
	*V:O[−voice]	Obstruents are not underlyingly voiceless after long vowels.
	*iuyO[−voice]	Obstruents are not underlyingly voiceless after the vowels [i, u, y].
	*VO[+voice]	Obstruents are not underlyingly voiced after short vowels.
	*SonO[−voice]	Obstruents are not underlyingly voiceless after sonorant consonants.
	*OO[+voice]	Obstruents are not underlyingly voiced after other obstruents.

The first five constraints account for the effects of the type of obstruent on the underlying [voice] specification. The constraints *V:O[−voice], *iuyO[−voice], *VO [+voice] account for the effects of the type of the preceding vowel, while *SonO [−voice] and *OO[+voice] account for the effects of the degree of sonority of a preceding consonant.

For some words, some of these constraints are in conflict, such as *V:O[−voice] and *T[+voice] for words ending in a long vowel and an alveolar stop. The former constraint forbids an underlyingly voiceless specification for the final obstruent, whereas the latter forbids an underlyingly voiced specification. The variation present in the existing words shows that there must be variation in which constraint is the most important. Apparently, the positions of the constraints in the hierarchy are variable: Sometimes they dominate conflicting constraints, and sometimes they are subordinated to these constraints. In other words, the constraint evaluation must be stochastic, as proposed by Boersma (1998). Boersma defines the position of a constraint in the hierarchy at a certain time as in 8 (Boersma 1998:331)

(8) position in hierarchy = average position + RankingSpreading * z

with z a Gaussian random variable with mean 0 and standard deviation 1.

The average ranking of the ten constraints which are relevant for our data can be determined by means of Boersma's GRADUAL LEARNING ALGORITHM (Boersma 1998: 273, Boersma & Hayes 2001), as implemented in Praat (Boersma 1996). This algorithm assumes that initially all constraints prohibiting certain characteristics, like the constraints formulated above, are ordered at one given position in the hierarchy. Every time a word form that is not optimal according to the reigning stochastic hierarchy is attested, the constraints that are offended by this word form are moved down. In addition, the constraints that are offended by the forms that, according to this stochastic hierarchy, are more correct than the attested word form are moved up. We determined the average ranking of the ten constraints by training the gradual learning algorithm with the CELEX data described in §3. We presented the algorithm with these data one hundred thousand times. The RankingSpreading in equation 8 was set to 2. This setting ensures that only constraints that are at a maximal average distance of 10 units are regularly in an inverse domination relation. The starting positions of the constraints were set at 100 units.

RANKING	POSITION	CONSTRAINT
1	− 161.286	*X[− voice]
2	− 230.270	*OO[+ voice]
3	− 503.969	*P[+ voice]
4	− 507.022	*T[+ voice]
5	− 507.353	*V:O[− voice]
6	− 507.963	*iuyO[− voice]
7	− 508.604	*SonO[− voice]
8	− 508.627	*S[+ voice]
9	− 1286.323	*VO[+ voice]
10	− 1287.326	*F[− voice]

TABLE 9. Average rankings and positions of the 10 constraints.

The resulting average hierarchy of the constraints is listed in Table 9. The table shows that the constraints *P[+ voice], *T[+ voice], *V:O[− voice], *iuyO [− voice], *SonO[− voice], and *S[+ voice] hardly differ in their average position. Since the positions of the constraints in the hierarchy are stochastic, the relative ranking of these six constraints is variable. Thus, *T[+ voice] is predicted to be sometimes ranked higher than *V:O[− voice], and sometimes lower. As a consequence, a word such as [dɑut] would sometimes be interpreted as having an underlyingly voiceless final obstruent (see 9), and sometimes as having an underlyingly voiced one (see 10).

(9) Interpretation of [dɑut] when *T[+voice] dominates *V:O[−voice]

[dɑut]	*T[+voice]	*V:O[−voice]
☞ /dɑut/		*
/dɑud/	*!	

(10) Interpretation of [dɑut] when *V:O[−voice] dominates *T[−voice]

[dɑut]	*V:O[−voice]	*T[+voice]
/dɑut/	*!	
☞ /dɑud/		*

We evaluated this grammar by comparing its predictions for the pseudo-words in our experiment with the participants' responses. Since the positions of the constraints are stochastic, and the output of the grammar is consequently variable, we considered the average prediction for every pseudo-word. We generated these average predictions by having the grammar predict the responses for each pseudo-word one hundred thousand times. We calculated for each pseudo-word the proportion of cases in which the final obstruent was assigned an underlyingly voiceless representation, and compared this proportion with the proportion of participants who had interpreted the final obstruent as underlyingly voiceless. Our results are presented in the upper left panel of Figure 3 (Fig. 3A), which is a scatterplot of the observed proportion of *-te* responses (horizontal axis) and the expected proportion of *-te* responses (vertical axis) for every word in the experiment. The solid line represents a nonparametric smoother (Cleveland 1979). The plot shows a strong correlation between the predicted and observed proportions ($r = 0.83$, $t(190) = 20.4161$, $p < 0.001$; $r_s = 0.67$, $S = 384968$, $p < 0.001$).

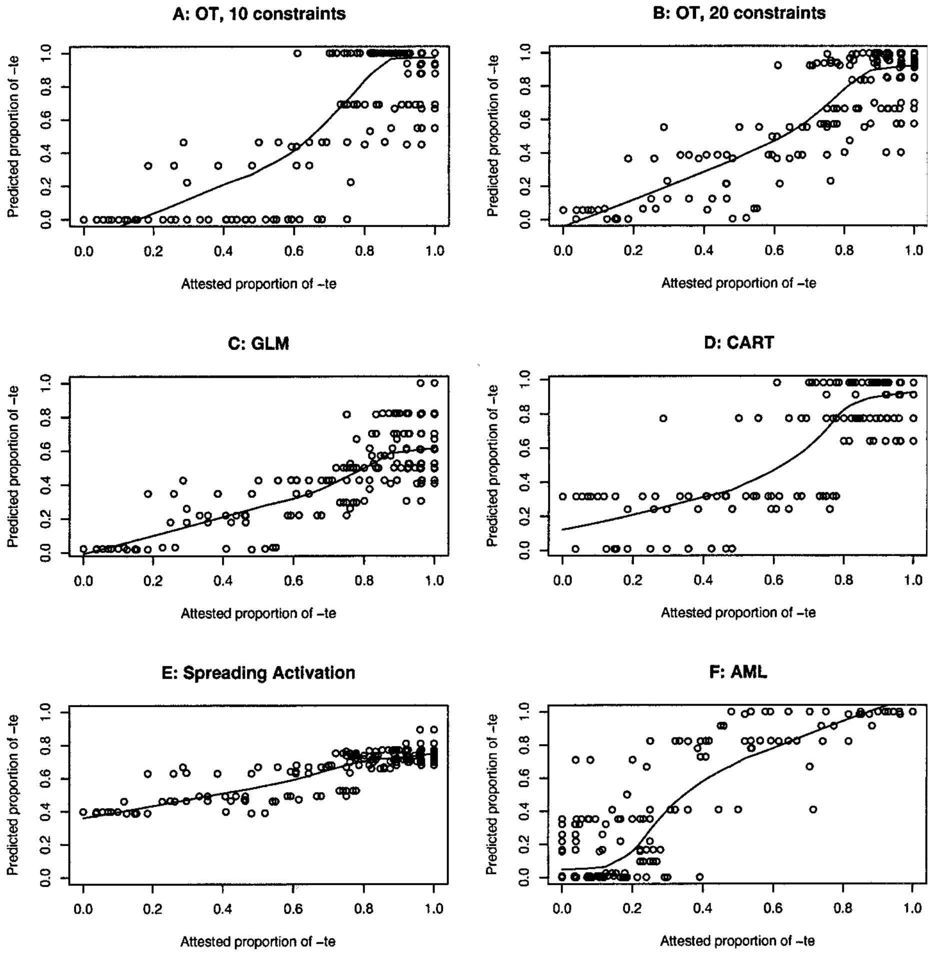

FIGURE 3. Proportions of created past-tense forms with *-te* as a function of proportions predicted by the models. Solid lines represent nonparametric smoothers of the data (Cleveland 1979).

The model clearly captures a substantial part of the structure in the experimental data. This conclusion is supported by the fact that in 87.0% of cases the majority of the participants agreed with the majority prediction of the grammar. We can conclude that an optimality theory grammar with ten constraints provides a good account of our data.

This optimality theory account faces one problem, however. Though it successfully predicts the participants' majority choices, it does not account properly for the fact that the participants are often not unanimous in their choices. For many instances, the model predicts zero or one hundred percent *-te*, while our participants showed a graded behavior, that is, in the center of Fig. 3A, a great many points lie either on the bottom line or at the top line of the graph. The grammar predicts deterministic behavior where it should not do so. It does so if either a voiced or a voiceless interpretation of the final obstruent does not violate any constraint.

This problem is eliminated when we add the mirror-constraints of the ten constraints already in our hierarchy. For instance, the constraint *T[+voice] has *T[−voice] as its mirror constraint. Both the voiced and voiceless interpretation of the final obstruent of every word now violates several constraints, allowing variation to emerge for most cases.

We determined the positions of the twenty constraints with the gradual learning algorithm, again setting RankingSpreading = 2, and again presenting the CELEX data one hundred thousand times. The resulting average hierarchy is listed in Table 10. The ten constraints of the previous analysis are all ranked above 100, the starting positions of the constraints. They should not be violated by the optimal forms. Their mirror images, which are often violated, are all ordered below 100, in exactly the opposite order.

RANKING	POSITION	CONSTRAINT
1	105.985	*X[−voice]
2	105.463	*OO[+voice]
3	104.299	*P[+voice]
4	101.930	*T[+voice]
5	101.735	*SonO[−voice]
6	101.699	*VO[+voice]
7	101.452	*V:O[−voice]
8	101.239	*F[−voice]
9	100.633	*iuyO[−voice]
10	100.609	*S[+voice]
11	99.391	*S[−voice]
12	99.367	*iuyO[+voice]
13	98.761	*F[+voice]
14	98.548	*V:O[+voice]
15	98.301	*VO[−voice]
16	98.265	*SonO[+voice]
17	98.070	*T[−voice]
18	95.701	*P[−voice]
19	94.537	*OO[−voice]
20	94.015	*X[+voice]

TABLE 10. Average rankings and positions of the 20 constraints.

As before, we evaluated the resulting grammar by comparing its predictions for the words in the experiment with the participants' responses. We again generated the average predictions by having the grammar predict the underlying [voice] specification for the experimental words one hundred thousand times. We found a strong correlation of the expected with the observed probabilities ($r = 0.85$, $t(190) = 21.8394$, $p < 0.001$; $r_s = 0.72$, $S = 335915$, $p < 0.001$). This is shown in Fig. 3B, which plots the predicted proportions of *-te* against the observed proportions of *-te*. In addition, we found that the majority prediction for a given pseudo-word is identical to the participants' majority response in 88.5% of the cases. This performance does not differ significantly from the performance of the grammar with only ten constraints (both for the correlations, $p > 0.05$ calculated using Fisher's Z-transformation [Woods et al. 1986:165] and for the majority choice congruence, Fisher's exact test two-tailed $p > 0.20$). The grammar with the mirror constraints is, nevertheless, a qualitative improvement with respect to the grammar without the mirror constraints, as can be seen from Figs. 3A and 3B. The grammar with twenty constraints predicts *-te* in either zero or one hundred percent of cases for fewer words than the grammar with ten constraints does, and provides a better qualitative fit to the experimental data.

The constraints discussed above are actually conditional statements. They state that if some condition is met, for instance, if the rhyme of a word contains a sonorant consonant, the final obstruent should be underlyingly voiced or voiceless. In optimality theory, conditional statements (henceforth conditionals) are strictly ordered in a hierarchy, and the interpretation of a final obstruent at a given moment is in fact determined by only one conditional, namely the one among the relevant ones that happens to be the highest in the hierarchy at that moment.

A second kind of model to consider is one in which the conditionals are not ordered in a strict hierarchy, but in which all relevant conditionals simultaneously contribute to the probability of a certain output, every time an output has to be produced. Since some conditionals are often more important than others, the contributions of the conditionals can be weighted. The model starts from a default type of word, in this case the type of word with the highest probability to end in an underlyingly voiceless obstruent. Such a word ends in a short vowel, has an extra obstruent in the rhyme, and has a bilabial stop. If a word deviates from this default type, the relevant active conditionals reduce by a certain amount the probability that the final obstruent of the word is interpreted as underlyingly voiceless. Mathematically, this type of model belongs to the class of generalized linear models (GLM), to which the VARBRUL technique used in sociolinguistics (e.g. Sankoff 1987) belongs as well. We estimated the parameters of this model on the basis of the CELEX data set discussed in §3, using the loglinear equation (11), with υ defined as in 12.

(11) $$\mu = \frac{e^{\upsilon}}{1.35}$$

(12) $$\begin{array}{ccccccc} \upsilon = \text{default} & - & V & - & C & - & O \\ 0.30 & & \begin{Bmatrix} 0.28\ \mathrm{I}_{[\mathrm{iyu}]} \\ 0.50\ \mathrm{I}_{[\mathrm{long}]} \end{Bmatrix} & & \begin{Bmatrix} 0.20\ \mathrm{I}_{[\mathrm{None}]} \\ 0.50\ \mathrm{I}_{[\mathrm{Sonorant}]} \end{Bmatrix} & & \begin{Bmatrix} 0.19\ \mathrm{I}_{[\mathrm{T}]} \\ 0.35\ \mathrm{I}_{[\mathrm{S}]} \\ 1.02\ \mathrm{I}_{[\mathrm{F}]} \\ 3.29\ \mathrm{I}_{[\mathrm{X}]} \end{Bmatrix} \end{array}$$

The variable μ denotes the probability that a word underlyingly ends in a voiceless obstruent. The variable υ has the value of 0.30 when the word is of the default type. If the word is not of the default type, it meets some of the specifications indicated in the subscripts of the conditional operators I. These conditional operators then evaluate to 1, and the preceding factors are substracted from the default value. The conditionals that are not met by the word evaluate to 0, and the preceding weights are not substracted. To give an example, in the case of [pulf], υ equals 0.30 (default value) -0.28 (since the final vowel is [u], and $\mathrm{I}_{[\mathrm{iyu}]}$ is consequently 1) -0.50 (since the final rhyme contains a sonorant consonant) -1.02 (since the word ends in a labiodental fricative) $= -1.52$, and μ equals 0.16, which implies that the final obstruent of [pulf] is predicted to be interpreted as underlyingly voiced in (100% − 16%) 84% of cases.

This model also appears to be a good predictor of our experimental results, as is shown in Fig. 3C, which plots μ against the proportion of observed past-tense forms created with *-te* for all words in the experiment. The model requires nine parameters, and achieves a Pearson correlation of $r = 0.79$ ($t(190) = 17.8985$, $p < 0.001$) between the predicted and observed proportions of past-tense forms with *-te*, a Spearman correlation of $r_s = 0.71$ ($S = 339278$, $p < 0.001$), and a correctly predicted majority choice score of 72.4%. The correlations do not differ significantly from those observed for the stochastic optimality theory models ($p > 0.05$, calculated using Fisher's Z-transformations). The majority choice congruence, however, is significantly worse (Fisher's exact test two-tailed, $p < 0.001$). This suggests that a model with weighted, simultaneously applying conditionals is observationally less adequate than the OT models. Note, however, that it is also less complex than the OT grammars in the number of conditionals that it requires to fit the data. For a discussion of the relation between stochastic OT and loglinear models, the reader is referred to Manning 2003.

We now turn to a third type of model, a model based on CART (Breiman et al. 1984), in which the conditionals are unweighted but partially ordered. Recall that we also used a CART model in the analyses of the CELEX data and the experimental results in §§3–4. When analyzing the data in these sections, we considered all vowels separately, and all extra consonants in the final rhyme separately, instead of grouping them into the classes long vowels, short vowels, [iuy], sonorant consonants, and obstruents. In order to compare the performance of the CART model with the performance of the other models discussed in this section, we also determined its performance while making use of these segment classes. An analysis of the CELEX data set using the segment classes resulted in a pruned tree of six nodes, which can be found in Figure 4. This tree shows that there is a large difference between the underlying [voice] specification of labiodental and velar fricatives (F, X) on the one hand and bilabial stops, alveolar stops, and alveolar fricatives (P, T, S) on the other hand (split at the top of the figure). The former obstruents (left head branch of the tree) have a smaller probability of being underlyingly voiceless. Those in syllables with long vowels are underlyingly voiceless in 1% of the cases, while those in syllables with short vowels or [i, u, y] are

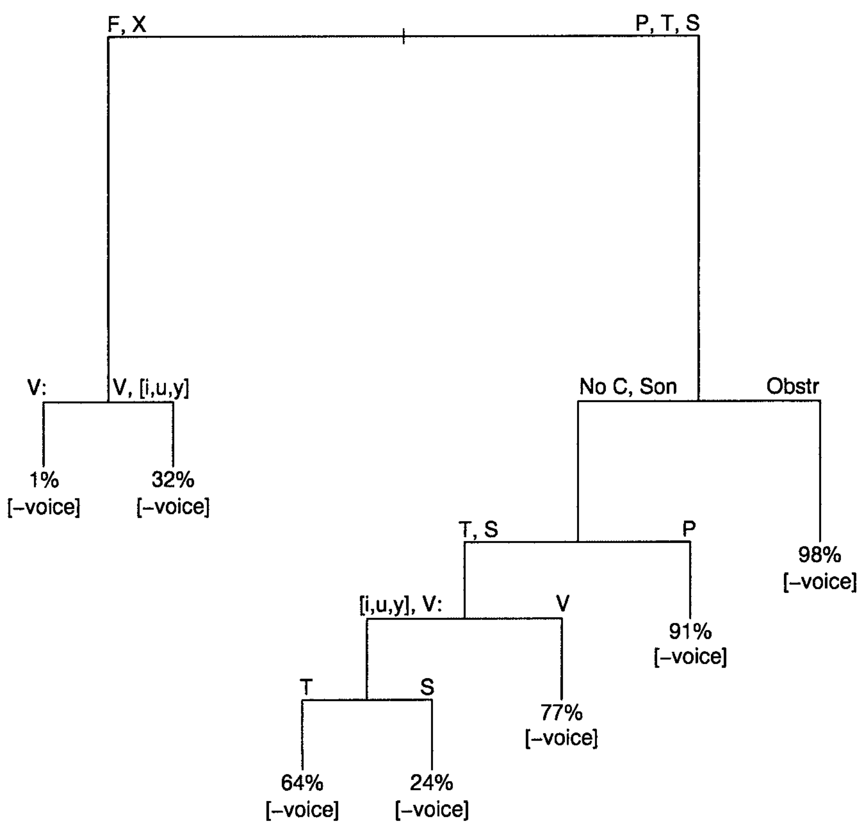

FIGURE 4. CART analysis of CELEX data, showing predicted percentages of voiceless specifications. V: indicates long vowel, V a short vowel, Son a sonorant consonant before the obstruent, Obstr a nonsonorant consonant before the obstruent, and No C indicates no extra consonant in the final rhyme.

underlyingly voiceless in 32% of cases. Bilabial stops, alveolar stops, and alveolar fricatives (right head branch of the tree) are underlyingly voiceless in 98% of cases if they follow another obstruent. If they do not follow an obstruent, bilabial stops are underlying voiceless in 91% of cases, while the probability with which the alveolar obstruents are underlyingly voiceless is affected by the type of the preceding vowel. In syllables with short vowels, they are underlyingly voiceless in 77% of cases, while in syllables with [i, u, y] or long vowels, alveolar stops are underlyingly voiceless in 64% of cases, and alveolar fricatives in 24% of cases. The predicted proportions of underlyingly voiceless obstruents made by this tree correlate well with the observed proportions ($r = 0.80$, $t(190) = 18.4041$, $p < 0.001$; $r_s = 0.66$, $S = 398926$, $p < 0.001$; see Fig. 3D), and the majority choice of the participants is predicted correctly in 88.5% of the cases. The performance of this model does not differ significantly from the performance of the preceding OT grammars. This implies that a grammar with partially ordered conditionals should be taken seriously as an alternative to the traditional, strictly hierarchical OT models.

We now turn to the two analogical models, in which the individual exemplars keep playing a crucial role even after the model has been built.

The first analogical model is a SPREADING ACTIVATION model along the lines of Schreuder & Baayen 1995 and Krott et al. 2001. It can be viewed as a variant of the IB1-IG algorithm of TiMBL (Daelemans et al. 2002). As in the model proposed in Bybee 1985, words are activated in the mental lexicon together with phonologically, phonetically, morphologically, or semantically related words. In the model, this phenomenon of lexical co-activation is captured by allowing activation to spread to other words along lines of form and meaning similarity. Figure 5 illustrates the kind of network architecture required to model the present data for the case that the listener is presented with the nonexisting word [da:rt].

There are three layers in the model. The first layer contains the features specifying the rhyme structure of the target word, which in this case has a long vowel, a sonorant preceding the final obstruent, and a final /T/. Each feature value is connected with the words in the lexicon sharing this feature value, only a small subset of which is shown in Fig. 5. In turn, these words are connected with the [voice] specification of their final obstruent, which is either voiced or voiceless.

Activation flows from the rhyme features via the lexical exemplars to the [voice] specifications. Crucially, the activation flow is modulated by the weights on the connections between the rhyme features and the exemplars. The feature vowel quantity has outgoing weights w_1, the feature type of preceding segment has outgoing weights w_2, and the final obstruent feature has outgoing weights w_3. Thanks to these weights, exemplars such as /pɪnt/, /lɪst/, /zʋa:rd/, /be:md/, /ko:rd/, /le:z/, /tʋa:lf/, and /prɪns/ are activated to different degrees, depending on their degree of similarity with the target word and on the relevance of the shared features as coded by the weights. The word /zʋa:rd/, for instance, is activated proportionally to $w_1 + w_2 + w_3$, while /lɪst/ receives activation proportionally to only w_3. If these two words were the only exemplars in the lexicon co-activated by the target [da:rt], the probability that the final obstruent of this word would be interpreted as voiced would be $(w_1 + w_2 + w_3)/(w_1 + w_2 + w_3 + w_3)$.

We determined the weights of this model given our CELEX data set in two steps. We first calculated the information gain values (see e.g. Daelemans et al. 2002) for the three features, which gave us the initial estimates of the weights. We then applied Nelder and Mead's simplex optimization procedure (1965) to obtain the final estimates,

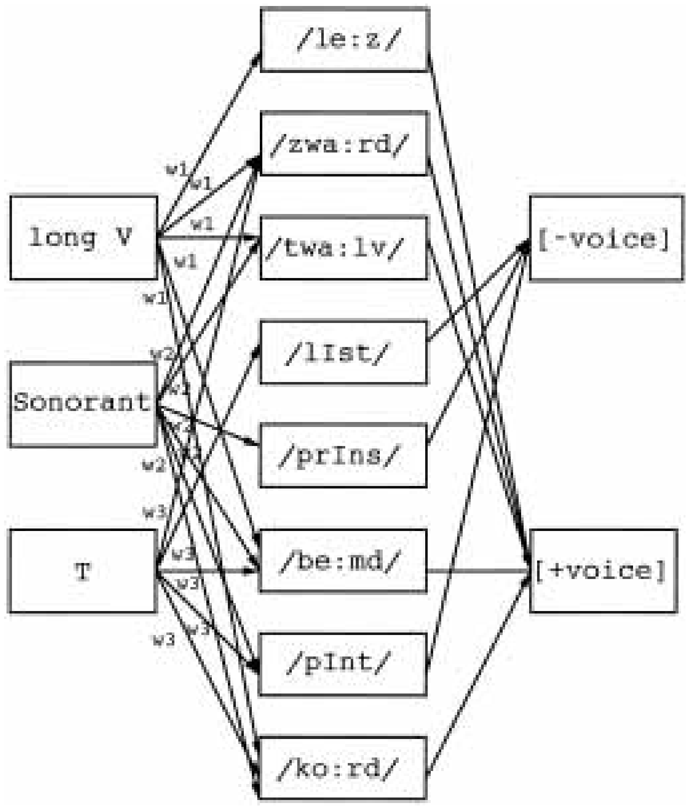

FIGURE 5. Spreading Activation model.

0.0689, 0.1711, and 0.7965 respectively. Applied to the experimental data, we obtained a Pearson correlation of $r = 0.85$ ($t(190) = 22.2395$, $p < 0.001$), and a Spearman correlation of $r_s = 0.76$ ($S = 283010$, $p < 0.001$) between the predicted and observed responses ending in *-te*; see Fig. 3E. The model correctly predicts the participants' majority choice in 91.2% of cases. Performance accuracy is again similar to that of the other models discussed in this section (with the exception of the GLM model, which performs slightly worse with respect to the majority choice congruence). It is remarkable (in light of Occam's razor) that the model that receives the highest scores of all models requires only three parameters, one for each of the three features: The more free parameters a model has, the easier it is to fit to the data, and the less surprising high accuracy scores are. Note that these parameters play a role similar to the conditional statements in the preceding models, in that they regulate the analogical force of the exemplars depending on which features the exemplars share with the target word. In other words, the weights embody the rule part of the model, the part of the model that has to be calculated over the exemplars in the lexicon before the model can be applied. At the same time, the exemplars in the lexicon also play a crucial role when the model is actually applied.

The final model, ANALOGICAL MODELING OF LANGUAGE (AML, Skousen 1989, 1993), dispenses with any a priori calculations over the exemplars in the lexicon. This model bases its predictions on the exemplars in what is called the analogical set, which contains the nearest neighbors as well as homogeneously behaving exemplars at greater distances. Crucially, the analogical set contains those words that share subsets of feature-values with the target word and that evidence the same kind of behavior with respect to the [voice] specification of the final obstruent. Leaving the technical details of AML aside, it is important to realize that AML calculates the analogical set for a given target word on the fly, potentially on the basis of all exemplars in the lexicon, which makes it a computationally intensive model (but see Skousen 2000).

Applied to the 1697 words in our CELEX data set, we observed a high correlation between the predictions of AML and the experimental observations ($r = 0.85$, $t(190) = 21.8384$, $p < 0.001$; $r_s = 0.72$, $S = 326062$, $p < 0.001$). AML correctly predicts the participants' majority choice in 89.6% of cases. This performance of AML is similar to the performance of the models discussed above, again with the exception of the GLM model, which performs slightly worse. Fig. 3F plots the predicted probabilities against the observed proportions of voiceless interpretations. It is remarkable that a model without any free parameters achieves this high level of performance.

This AML analysis is based on the assumption that only the type of the final vowel, the type of segment preceding the final obstruent, and the type of the final obstruent itself are relevant for the interpretation of this final obstruent as underlyingly voiced or voiceless. This is, in fact, against the philosophy of AML, which assumes that all characteristics of a word, even the most unlikely ones, may affect linguistic interpretation. According to the philosophy of AML, we should not use segment classes, but consider the segments separately, as we did in §3 and §4. Moreover, we should also take the onsets of the words into account, since the experimental words differ in their onsets. We also computed the probabilities assigned by AML that the experimental words are interpreted as underlyingly ending in a voiceless obstruent, while considering the segments separately, and taking the onsets into account. The resulting probabilities correlate with the observed proportions as well as the probabilities computed in the previous AML analysis ($r = 0.78$, $t(190) = 17.2232$, $p < 0.001$; $r_s = 0.66$, $S = 401009$, $p < 0.001$) and therefore do not form an improvement. Moreover, the analysis also correctly predicts the participants' majority choice in approximately the same percentage of cases (86%). We conclude that for our data the performance of AML is independent of the precise form of the lexical representations.

This brings us to the end of our discussion of models accounting for the observed relation between the characteristics of the words in the lexicon and the speakers' interpretation of new words. Table 11 recapitulates the performance of the models by listing the Pearson and Spearman correlations between the predicted and observed proportion of voicing for each model, and the percentages of words for which the models correctly predict the participants' majority choice. The column labeled Complexity summarizes in a crude way the complexity of the various models in terms of the numbers of constraints, coefficients, weights, decision nodes, and supracontexts. Not taken into account, for example, is the complexity of the theory of a priori assumptions about the

METHOD	COMPLEXITY	r	r_s	MAJORITY CHOICE CONGR.
OT	10 constraints	0.83	0.67	87
OT	2*10 constraints	0.85	0.72	89
GLM	9 coefficients	0.79	0.71	72
CART	6 decision nodes	0.80	0.66	89
Spreading Activation	3 weights	0.85	0.76	91
AML	8 supracontexts	0.85	0.72	90

TABLE 11. Performance of different models. (All correlations are statistically significant: $p < 0.001$).

possible parameter space, the theoretical complexity of the learning process, the actual computation time required for training and fitting the models (which was huge for stochastic OT), and the computation time required for processing a data point (relatively long for AML). We see that the models are well matched in their performance: their predictions correlate approximately equally well with our experimental results, and they correctly predict the participants' majority choice in approximately the same percentage of cases, the only exception being the GLM model.

None of the models discussed so far take the frequencies of occurrence of the exemplars into account. They all hypothesize that existing words of a high frequency are as relevant as existing words of a low frequency for the interpretation of new words. To test this hypothesis, we calculated the probabilities on underlyingly voiceless obstruents under the assumption that the relevance of each word in the CELEX data set is proportional to its frequency of occurrence. We chose CART as our model since the relevance of the frequencies of words in the CELEX data set can easily be set in this model by means of weights. The resulting probabilities correlate well with the observed probabilities ($r = 0.68$, $t(190) = 12.8929$, $p < 0.001$; $r_s = 0.63$, $S = 438427$, $p < 0.001$), but, crucially, the correlation coefficient $r = 0.68$ is significantly smaller than the $r = 0.80$ obtained for the CART model without frequency weights ($p < 0.05$, calculated using Fisher's Z-transformation). Moreover, the model with frequency weights predicts the participants' majority choice in only 68.2% of cases, which is also significantly smaller than the majority choice congruence for the CART model without weights, which is 89% (Fisher's exact test, two-tailed: $p < 0.001$). Similarly degraded results were obtained when token frequencies were included in the spreading activation model. These results suggest that the token frequencies of the exemplars are irrelevant for the prediction of the underlying [voice] specification of final obstruents in Dutch.

6. General discussion. This article addresses the question of how speakers interpret neutralized segments, focusing on final devoicing in Dutch. The received wisdom about final devoicing is that the underlying voice specification of final obstruents is an idiosyncratic lexical property of words with final obstruents. Against this background, we formulated four hypotheses. Hypothesis 1 predicts that a listener interprets a neutralized obstruent as underlyingly voiced if it happens to be realized as voiced and as underlyingly voiceless otherwise. Hypothesis 2 predicts random interpretation with voiced interpretations in half of the cases. Hypothesis 3 predicts more voiced interpretations when the voiced variant of the obstruent is less marked in phonology. Hypothesis 4 states that speakers are more likely to choose that phoneme as the underlying representation when there are more phonologically/phonetically similar morphemes in the lexicon sharing that phoneme. Our inspection of the Dutch lexicon revealed that the underlying [voice] specification of a neutralized obstruent is correlated with the place and manner of articulation of this obstruent and with the characteristics of the preceding segments. If speakers are more likely to choose that underlying [voice] specification for the neutralized obstruent that is shared by more existing similar morphemes, they consequently should more often choose the specification that is predicted by the place and manner of articulation of the final obstruent and the characteristics of the preceding segments.

We tested these four hypotheses by means of a production experiment. Speakers of Dutch were presented with pseudo-words ending in obstruents neutralized for [voice], and had to interpret these neutralized obstruents as underlyingly voiced or voiceless. We found that the percentage of participants interpreting the neutralized obstruent of a given word as underlyingly voiced correlates strongly with the percentage of phonologically/phonetically similar words in the mental lexicon ending in an underlyingly voiced obstruent. For instance, the word [mɑrx] was interpreted as ending in an underlying voiced /ɣ/ by all participants, which correlates with the fact that nearly all final velar fricatives are underlyingly voiced. The word [te:s] was interpreted as underlyingly ending in a voiced /z/ by 74.1% of the participants, which correlates well with the fact that 76.5% of the words ending in a phonologically long vowel and an alveolar fricative end in /z/. Apparently, speakers of Dutch make use of the phonological/phonetic similarity patterns in the lexicon and interpret new words in such a way that they conform to these similarity patterns. Hypothesis 4 is correct.

The main lines of this phonological similarity structure in the lexicon can be summarized as follows. First, long vowels favor voiced final obstruents more than short vowels. This makes sense, as in general vowels before voiced obstruents tend to be longer than vowels preceding voiceless obstruents (see e.g. Slis & Cohen 1969 and the references given there). Second, sonorant segments preceding the final obstruent, both vowels and sonorant consonants, favor voicing compared to obstruents preceding the final obstruent. This also makes sense, as morpheme-internal obstruent clusters tend to be voiceless in Dutch (Zonneveld 1983). Third, the type of the final obstruent itself is an important predictor of voicing. Surprisingly, a mirror image of the hierarchy of phonological strength emerges from the data, especially for words with a long vowel. For instance, although /b/ is phonologically stronger than /p/ (Ohala 1983:195, Gussenhoven & Jacobs 1998:34), there are only a few words with final /b/ and many words with final /p/. Conversely, at the other end of the hierarchy, even though /ɣ/ is phonologically weaker than /x/ (Ladefoged & Maddieson 1996:176, Ohala 1983:201), there are only a few words ending in /x/ and many words ending in /ɣ/. A possible functional explanation might be that the underlying [voice] specification of obstruents in Dutch optimizes language comprehension. Given that /p/ has a nonnegligible probability to be realized as [b], [p] has a higher cue validity: if [p] is perceived, there is no ambiguity, whereas a perceived [b] can be either /p/ or /b/. Similarly, given that /ɣ/ may be realized as [x], it has the higher cue validity. If [ɣ] is heard, it must represent /ɣ/, while a perceived [x] may represent both /x/ and /ɣ/. If this explanation is on the right track, Dutch shows a preference for words to use final obstruents with a high cue validity. More research is clearly required here.

In §4 we discussed several formal and quantitative models that account for the observed effects of the statistical distributional structure in the lexicon on the participants' interpretation of the final obstruents. We first considered two optimality theory grammars. We found that the grammar assuming twenty constraints (ten constraints expressing the correlations observed in the lexicon, plus their mirror constraints) accounts well for the data, including the variation in the data. Crucially, the ranking of the constraints is assumed to result from a stochastic function, allowing it to vary in time. We discussed models assuming unordered and weighted (GLM), or partially ordered (CART), conditionals. These models require fewer parameters than the optimality theory grammars; that is, they are less complex but perform approximately as well as the optimality theory grammars. We then discussed a psycholinguistic spreading activation model derived from a computational model of systematic analogy (TiMBL) that also provides an excellent fit to the data with just three parameters. Finally, we discussed AML, a parameter-free model of analogy that also performs very well. All models score about equally well with our experimental data, with the GLM model as the exception when the majority choice congruence is considered.

Our results have several important implications. First, our data show that the underlying [voice] specification of final obstruents in Dutch is predictable to a far greater extent than has generally been assumed. It is predictable not only for linguists having computerized statistical techniques at their disposal, but also for naive speakers, since they use this predictability in language production.

Second, we see that the predictability is based on the similarity structure in the lexicon. Depending on the preferred model, speakers base their choice between an underlying voiced or voiceless specification on separately stored generalizations/conditionals reflecting the characteristics of the similar words in the lexicon (the optimality theory account, CART, and GLM), or they base their choice directly on the phonologically/phonetically similar words, which then act as exemplars (TiMBL, the spreading activation model, AML).

Third, the finding that speakers can predict the underlying [voice] specification of final obstruents may explain the fact that although the underlying [voice] specification of morpheme-final obstruents is neutralized in a large percentage of cases, speakers are seldom in doubt about the correct underlying [voice] specification of the final obstruent for a given morpheme. If a speaker forgets the underlying [voice] specification of a final obstruent, he or she can deduce it from the other morphemes in the lexicon. In fact, we think that the structure in the lexicon may well be instrumental in maintaining the underlying voiced/voiceless distinction for morpheme-final obstruents in Dutch.

Finally, our discussion of quantitative models provides an overview of the range of possibilities for handling nondeterministic morphophonological data. We have shown that Boersma's stochastic optimality theory yields a predictive accuracy that is as good as that of standard statistical techniques like GLM and CART. Note, however, that stochastic optimality theory also comes with a high cost in terms of theoretical complexity, since it must assume a large number of constraints, two for each relevant feature-value. The spreading activation model performs equally well with just three parameters, one for each feature, and analogical modeling of language yields very accurate predictions as well with no free parameters at all.

A frequent argument against analogical models is that many entries in the lexicon have to be accessed. This would make such models uneconomical and unnecessarily complex for real-time processing compared to, for instance, stochastic OT. One possible reply is that processing complexity within sequential processing systems, such as stochastic OT with a sequential evaluation of ranked constraints, differs from processing complexity in massively parallel processing systems as used by AML and the human brain. More importantly, the psycholinguistic evidence points unambiguously to substantive co-activation of lexical candidates in the mental lexicon (see e.g. Collins & Loftus 1975, Pisoni et al. 1985, Allopenna et al. 1998, De Jong et al. 2000).

We think that phonological theory has thus far underestimated the usefulness and power of the analogical approaches as a means for coming to grips with noncategorical phenomena in phonology. Other successful applications of analogical models have been reported by, for example, Daelemans et al. (1994), Eddington (2000a,b), and Krott et al. (2001). However, Albright and Hayes (2001) report an analogical model for English past-tense forms with less than optimal performance, a model along the lines of Nosofsky's (1990) generalized context model, implemented following Nakisa, Plunkett, and Hahn (2000). Crucially, the analogical model studied by Albright and Hayes assigns equal weight to all dimensions of similarity, so that, for instance, a difference between two segments in the onset is as important as that same difference in the final position of the coda. The resulting variegated similarity has various undesirable properties, as documented by Albright and Hayes. It is also highly undesirable from the perspective of machine learning (Daelemans et al. 2002). In the analogical models considered here, variegated similarity does not arise (CART, spreading activation), or it arises only under very special circumstances (AML; see Skousen 1989).

Interestingly, Albright and Hayes observed the effects of sets of phonologically similar words for regular pseudo-verbs in English, which they describe in terms of rules detailing environments with high degrees of support for a regular outcome. Their evidence for islands of reliability for regular verbs, which supports the role for local generalizations and constellations as argued for by Janda & Joseph 1986 and Joseph 1997, is in line with the kind of probabilistic phenomena reported here for Dutch. As we have shown, such probabilistic data can be understood surprisingly well with more sophisticated models of analogy.

Further support for our results comes from an experiment in which speakers of Dutch were exposed to existing words and appeared to use the statistical phonological/phonetic structure in the lexicon that we revealed in this article. Ernestus and Baayen (2001) presented participants with existing verbs and, again, the task was to decide whether the past-tense forms of these verbs are created by adding *-te* or *-de* to the verbal stem. The experimental data show that speakers of Dutch not only base their choice between *-te* and *-de* on the simple rule stating that *-te* follows underlyingly voiceless obstruents and *-de* underlyingly voiced ones. In case of low-frequency words, speakers tend to choose *-te* if the majority of phonologically/phonetically similar words end in a voiceless obstruent and *-de* if the majority of similar words end in a voiced obstruent. As a consequence, they create violations of the simple rule in the case of verbs that happen to end in an obstruent with an underlying [voice] specification that is different from the [voice] specification of the final obstruents of the majority of the phonologically/phonetically similar verbs in the lexicon.

In another experiment, Ernestus and Mak (2002) asked speakers of Dutch to read sentences containing past-tense forms, some of which ended in the 'incorrect' past-tense suffix. Using the self-paced reading paradigm, they found that an incorrect past-tense form leads to longer reading times, especially when the inappropriate suffix is not supported by the phonologically similar words in the lexicon. The statistical structure in the lexicon governing the [voice] specification of final obstruents emerges also for comprehension.

Finally, Ernestus and Baayen (2002) report replication of the experimental results presented here even when some final obstruents are realized as slightly voiced. Recall that Warner et al. 2001 reported incomplete neutralization for word-final obstruents. Ernestus and Baayen also observed that underlyingly voiced obstruents in word-final position can be articulated with residual characteristics of voicing, primarily realized by shorter burst durations. They observed that words ending in such slightly voiced plosives elicited more *-de* than the same words ending in completely voiceless plosives. Importantly, however, by far the strongest predictor of the past-tense allomorph again was the set of phonologically similar words.

Summing up, what we have shown is that the underlying [voice] specification of final obstruents in Dutch words, instead of being an arbitrary lexical property of these words, is lexically structured. This structure can be captured by various quantitative models, ranging from stochastic OT to AML. The predictions of these models for novel experimental words are in line with the choice patterns of the participants: both the participants and the models show that the unpredictable can be predicted after all.

APPENDIX: EXPERIMENTAL PSEUDO-WORDS AND RESPONSES

PSEUDO-WORD		RESPONSES			
		-te		-de	
		#	%	#	%
taap	[ta:p]	23	95.8	1	4.2
deep	[de:p]	25	96.2	1	3.8
poop	[po:p]	24	100.0	0	0.0
tuip	[tœyp]	24	92.3	2	7.7
dijp	[dɛip]	28	100.0	0	0.0
dap	[dɑp]	18	75.0	6	25.0
tep	[tɛp]	21	100.0	0	0.0
mip	[mɪp]	24	96.0	1	4.0
gop	[xɔp]	20	83.3	4	16.7
talp	[tɑlp]	19	100.0	0	0.0
telp	[tɛlp]	27	100.0	0	0.0
dulp	[dɵlp]	25	96.2	1	3.8
pamp	[pɑmp]	22	100.0	0	0.0
bemp	[bɛmp]	25	100.0	0	0.0
kimp	[kɪmp]	28	100.0	0	0.0
tomp	[tɔmp]	23	92.0	2	8.0
berp	[bɛrp]	26	100.0	0	0.0
torp	[tɔrp]	25	100.0	0	0.0
tesp	[tɛsp]	26	100.0	0	0.0
dosp	[dɔsp]	24	96.0	1	4.0
tusp	[tɵsp]	28	100.0	0	0.0
fiep	[fip]	23	95.0	1	4.0
boep	[bup]	25	100.0	0	0.0
paat	[pa:t]	21	87.5	3	12.5
feet	[fe:t]	25	96.2	1	3.8
foot	[fo:t]	28	100.0	0	0.0
beut	[bø:t]	25	96.1	1	3.9
fuit	[fœyt]	26	96.3	1	3.7
kijt	[kɛit]	24	100.0	0	0.0
daut	[dɑut]	19	95.0	1	5.0
daant	[da:nt]	20	80.0	5	20.0
faart	[fa:rt]	25	96.1	1	3.9
foort	[fo:rt]	23	92.0	2	8.0
teecht	[te:xt]	21	77.8	6	22.2
puicht	[pœyxt]	19	76.0	6	24.0
bijcht	[bɛixt]	20	74.1	7	25.9
peekt	[pe:kt]	26	100.0	0	0.0
kuikt	[kœykt]	27	100.0	0	0.0
pijkt	[pɛikt]	27	100.0	0	0.0
duipt	[dœypt]	25	96.2	1	3.8
keest	[ke:st]	23	88.5	3	11.5
buist	[bœyst]	18	72.0	7	28.0
doost	[do:st]	21	77.8	6	22.2
slat	[slɑt]	27	100.0	0	0.0
ket	[kɛt]	26	100.0	0	0.0
jit	[jɪt]	28	100.0	0	0.0
fot	[fɔt]	21	77.8	6	22.2
but	[bɵt]	25	89.3	3	10.7
pilt	[pɪlt]	17	94.4	1	5.6
fult	[fɵlt]	25	92.6	2	7.4
dant	[dɑnt]	27	100.0	0	0.0
dent	[dɛnt]	23	95.8	1	4.2
dint	[dɪnt]	20	80.0	5	20.0
jont	[jɔnt]	21	75.0	7	25.0
bunt	[bɵnt]	20	83.3	4	16.7
fart	[fɑrt]	19	90.5	2	9.5
dert	[dɛrt]	10	76.9	3	23.1

APPENDIX: EXPERIMENTAL PSEUDO-WORDS AND RESPONSES (CONTINUED)

PSEUDO-WORD		RESPONSES			
		-te		*-de*	
		#	%	#	%
nort	[nɔrt]	21	84.0	4	16.0
burt	[bɵrt]	14	77.8	4	22.2
dift	[dɪft]	28	100.0	0	0.0
koft	[kɔft]	27	96.4	1	3.6
buft	[bɵft]	25	92.6	2	7.4
tacht	[tɑxt]	24	88.9	3	11.1
pecht	[pɛxt]	28	100.0	0	0.0
bicht	[bɪxt]	23	85.2	4	14.8
docht	[dɔxt]	26	100.0	0	0.0
nucht	[nɵxt]	20	87.0	3	13.0
rakt	[rɑkt]	25	100.0	0	0.0
fekt	[fɛkt]	27	100.0	0	0.0
dokt	[dɔkt]	27	100.0	0	0.0
kelt	[kɛlt]	24	92.3	2	7.7
fipt	[fɪpt]	26	100.0	0	0.0
bopt	[bɔpt]	18	90.0	2	10.0
dast	[dɑst]	25	92.6	2	7.4
dest	[dɛst]	25	89.3	3	10.7
bist	[bɪst]	23	88.5	3	11.5
bost	[bɔst]	26	96.3	1	3.7
pust	[pɵst]	23	92.0	2	8.0
fiet	[fit]	28	100.0	0	0.0
noet	[nut]	23	88.5	3	11.5
bijs	[bɛis]	17	60.7	11	39.3
baus	[bɑus]	9	64.3	5	35.7
meus	[møs]	18	64.3	10	35.7
kijns	[kɛins]	8	29.6	19	70.4
taars	[ta:rs]	19	76.0	6	24.0
kuifs	[kœyfs]	17	70.8	7	29.2
bijfs	[bɛifs]	22	81.5	5	18.5
kuichs	[kœyxs]	14	70.0	6	30.0
pijchs	[pɛixs]	14	60.9	9	39.1
duiks	[dœyks]	25	89.3	3	10.7
pijks	[pɛiks]	27	96.4	1	3.6
teeps	[te:ps]	27	100.0	0	0.0
kuips	[kœyps]	22	78.6	6	21.4
tijps	[tɛips]	26	96.3	1	3.7
faats	[fa:ts]	27	96.4	1	3.6
fas	[fɑs]	22	84.6	4	15.4
tes	[tɛs]	24	85.7	4	14.3
tis	[tɪs]	21	87.5	3	12.5
gos	[ɣɔs]	22	84.6	4	15.4
fus	[fɵs]	23	82.1	5	17.9
dals	[dɑls]	19	67.9	9	32.1
duut	[dyt]	25	96.2	1	3.8
poert	[purt]	22	81.5	5	18.5
poecht	[puxt]	23	92.0	2	8.0
tiekt	[tikt]	27	100.0	0	0.0
toekt	[tukt]	26	100.0	0	0.0
doept	[dupt]	25	100.0	0	0.0
diest	[dist]	22	81.0	5	18.0
boest	[bust]	22	88.0	3	12.0
taas	[ta:s]	10	38.5	16	61.5
tees	[te:s]	7	25.9	20	74.1
foos	[fo:s]	5	18.5	22	81.5
duis	[dœys]	13	48.1	14	51.9
bels	[bɛls]	15	55.6	12	44.4

APPENDIX: EXPERIMENTAL PSEUDO-WORDS AND RESPONSES (CONTINUED)

PSEUDO-WORD		RESPONSES			
		-te		*-de*	
		#	%	#	%
pans	[pɑns]	18	64.3	10	35.7
kens	[kɛns]	14	50.0	14	50.0
tons	[tɔns]	8	28.6	20	71.4
dars	[dɑrs]	21	75.0	7	25.0
bers	[bɛrs]	18	69.2	8	30.8
kors	[kɔrs]	24	85.7	4	14.3
tafs	[tɑfs]	26	92.9	2	7.1
bifs	[bɪfs]	20	83.3	4	16.7
tofs	[tɔfs]	26	96.3	1	3.7
pufs	[pɵfs]	23	82.1	5	17.9
dechs	[dɛxs]	23	88.5	3	11.5
kichs	[kɪxs]	28	100.0	0	0.0
tochs	[tɔxs]	21	91.3	2	8.7
tuchs	[tɵxs]	26	92.9	2	7.1
baks	[bɑks]	25	89.3	3	10.7
tiks	[tiks]	25	89.3	3	10.7
poks	[pɔks]	25	92.6	2	7.4
dats	[dɑts]	24	100.0	0	0.0
bets	[bɛts]	24	100.0	0	0.0
kits	[kɪts]	26	96.3	1	3.7
dots	[dɔts]	25	89.3	3	10.7
kuts	[kɵts]	25	96.1	1	3.9
dies	[dis]	17	60.7	11	39.3
toes	[tus]	16	59.3	11	40.7
tiefs	[tifs]	26	92.9	2	7.1
toefs	[tufs]	25	89.3	3	10.7
kiechs	[kixs]	19	82.6	4	17.4
boechs	[buxs]	19	76.0	6	24.0
tieks	[tiks]	26	92.9	2	7.1
toeks	[tuks]	25	96.1	1	3.9
kieps	[kips]	25	89.3	3	10.7
doeps	[dups]	26	96.3	1	3.7
kiets	[kits]	27	100.0	0	0.0
boets	[buts]	25	92.6	2	7.4
taaf	[ta:f]	10	35.7	18	64.3
deef	[de:f]	8	29.6	19	70.4
puif	[pœyf]	13	46.4	15	53.6
tijf	[tɛif]	11	42.3	15	57.7
boof	[bo:f]	7	25.0	21	75.0
taf	[tɑf]	21	75.0	7	25.0
tef	[tɛf]	20	74.1	7	25.9
bif	[bɪf]	19	73.1	7	26.9
chof	[xɔf]	21	77.8	6	22.2
chuf	[xɵf]	20	76.9	6	23.1
talf	[tɑlf]	13	46.4	15	53.6
belf	[bɛlf]	10	35.7	18	64.3
bolf	[bɔlf]	9	33.3	18	66.7
bamf	[bɑmf]	16	59.3	11	40.7
pemf	[pɛmf]	18	75.0	6	25.0
tumf	[tɵmf]	14	58.3	10	41.7
tarf	[tɑrf]	19	67.9	9	32.1
merf	[mɛrf]	18	66.7	9	33.3
perf	[pɛrf]	7	43.7	9	56.3
porf	[pɔrf]	11	40.7	16	59.3
purf	[pɵrf]	16	59.3	11	40.7
tief	[tif]	12	46.2	14	53.8
doef	[duf]	13	46.4	15	53.6

APPENDIX: EXPERIMENTAL PSEUDO-WORDS AND RESPONSES (CONTINUED)

PSEUDO-WORD		RESPONSES			
		-te		*-de*	
		#	%	#	%
kuuf	[kyf]	16	61.5	10	38.5
paag	[pa:x]	4	15.4	22	84.6
teeg	[te:x]	1	3.7	26	96.3
foog	[fo:x]	3	12.5	21	87.5
puig	[pœyx]	4	14.8	23	85.2
bijg	[bɛix]	5	18.5	22	81.5
keug	[køx]	13	48.2	14	51.9
pag	[pɑx]	5	26.3	14	73.7
beg	[bɛx]	5	22.7	17	77.3
fig	[fɪx]	2	11.8	15	88.2
bog	[bɔx]	11	55.0	9	45.0
tug	[tɵx]	13	54.2	11	45.8
dalg	[dɑlx]	1	5.6	17	94.4
pelg	[pɛlx]	2	8.0	23	92.0
bulg	[bɵlx]	1	3.8	25	96.2
marg	[mɑrx]	0	0.0	26	100.0
perg	[pɛrx]	2	7.1	26	92.9
kirg	[kɪrx]	9	40.9	13	59.1
dorg	[dɔrx]	2	10.0	18	90.0
kieg	[kix]	13	52.0	12	48.0
koeg	[kux]	3	15.0	17	85.0

REFERENCES

ALBRIGHT, ADAM, and BRUCE HAYES. 2001. Rules vs. analogy in English past tenses: A computational/experimental study. Los Angeles: UCLA, MS.

ALLOPENNA, PAUL D.; JAMES S. MAGNUSON; and MICHAEL K. TANENHAUS. 1998. Tracking the timecourse of spoken word recognition using eye movements: Evidence for continuous mapping models. Journal of Memory and Language 38.419–39.

BAAYEN, R. HARALD; RICHARD PIEPENBROCK; and LEON GULIKERS. 1995. The CELEX lexical database (CD-ROM). Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

BAUMANN, MONIKA. 1995. The production of syllables in connected speech. Nijmegen: Benda Drukkers.

BOERSMA, PAUL. 1996. Praat: Doing phonetics by computer. Amsterdam: University of Amsterdam, MS.

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: Oxford University Press.

BOOIJ, GEERT. 1999. The syllable, views and facts. Morpheme structure constraints and the phonotactics of Dutch. Studies in generative grammar 45, ed. by Harry van der Hulst and Nancy A. Ritter, 53–68. Berlin: Mouton de Gruyter.

BREIMAN, LEO; JEROME H. FRIEDMAN; RICHARD A. OLSHEN; and CHARLES J. STONE. 1984. Classification and regression trees. Belmont, CA: Wadsworth International Group.

BYBEE, JOAN. 1985. Morphology: A study of the relation between meaning and form. Amsterdam: John Benjamins.

CLEVELAND, WILLIAM S. 1979. Robust locally weighted regression and smoothing scatterplots. Journal of the American Statistical Association 74.829–36.

COLLINS, ALLAN M., and ELIZABETH F. LOFTUS. 1975. A spreading-activation theory of semantic processing. Psychological Review 82.407–28.

COLLINS, BEVERLY, and INGER MEES. 1981. The sounds of English and Dutch. Leiden: Leiden University Press.

DAELEMANS, WALTER; STEVEN GILLIS; and GERT DURIEUX. 1994. The acquisition of stress, a data-oriented approach. Computational Linguistics 20.421–51.

Daelemans, Walter; Jakub Zavrel; Ko van der Sloot; and Antal van den Bosch. 2002. TiMBL: Tilburg memory based learner reference guide, version 4.2. Tilburg: Computational Linguistics Tilburg University.

De Jong, Nivja H.; Robert Schreuder; and R. Harald Baayen. 2000. The morphological family size effect and morphology. Language and Cognitive Processes 15.329–65.

Derwing, Bruce L. 1980. English pluralization: A testing ground for rule evaluation. Experimental linguistics, integration of theories and applications, ed. by Gary D. Prideaux, Bruce L. Derwing, and William J. Baker, 81–120. Ghent: E. Story-Scientia.

Eddington, David. 2000a. Analogy and the dual-route model of morphology. Lingua 110.281–98.

Eddington, David. 2000b. Spanish stress assignment within the analogical modeling of language. Language 76.92–109.

Ernestus, Mirjam. 2000. Voice assimilation and segment reduction in casual Dutch: A corpus-based approach of the phonology-phonetics interface. Utrecht: LOT.

Ernestus, Mirjam, and Harald Baayen. 2001. Choosing between the Dutch past-tense suffixes *-te* and *-de*. Linguistics in the Netherlands 2001, ed. by Ton van der Wouden and Helen de Hoop, 81–93. Amsterdam: John Benjamins.

Ernestus, Mirjam, and Harald Baayen. 2002. The functionality of incomplete neutralization in Dutch: The case of past-tense formation. Nijmegen: University of Nijmegen and Max Planck Institute for Psycholinguistics, ms.

Ernestus, Mirjam, and Willem Marinus Mak. 2002. Analogical effects in the reading of Dutch verb forms. Nijmegen: University of Nijmegen, ms.

Gussenhoven, C., and R. H. Bremmer, Jr. 1983. Voiced fricatives in Dutch: Sources and present-day usage. North-Western European Language Evolution 2.55–71.

Gussenhoven, Carlos, and Haike Jacobs. 1998. Understanding phonology. London: Arnold, Hodder Headline Group.

Hooper, Joan B. 1976. An introduction to natural generative phonology. New York: Academic Press.

Hyman, Larry. 1970. The role of borrowing in the justification of phonological grammars. Studies in African Linguistics 1.1–48.

Jacobs, Haike, and Carlos Gussenhoven. 2000. Loan phonology: Perception, salience, the lexicon and OT. Optimality theory: Phonology, syntax, and acquisition, ed. by Joost Dekkers, Frank van der Leeuw, and Jeroen van de Weijer, 192–210. Oxford: Oxford University Press.

Janda, Richard D., and Brian D. Joseph. 1986. One rule or many? Sanskrit reduplication as fragmented affixation. Studies on Language Change (Ohio State University Working Papers in Linguistics 34) 84–107.

Jongman, Allard; Joan A. Sereno; Marianne Raaijmakers; and Aditi Lahiri. 1992. The phonological representation of [voice] in speech perception. Language and Speech 35.137–52.

Joseph, Brian D. 1997. How general are our generalizations? What speakers actually know and what they actually do. ESCOL '96. Proceedings of the thirteenth Eastern States Conference on Linguistics, ed. by Anthony D. Green and Virginia Motopanyane, 148–60. Ithaca: Cascadilla Press.

Krott, Andrea; R. Harald Baayen; and Robert Schreuder. 2001. Analogy in morphology: Modeling the choice of linking morphemes in Dutch. Linguistics 39.51–93.

Ladefoged, Peter, and Ian Maddieson. 1996. The sounds of the world's languages. Oxford: Blackwell.

Manning, Christopher D. 2003. Probabilistic approaches to syntax. Probability theory in linguistics, ed. by Rens Bod, Jennifer Hay, and Stefanie Jannedy. Cambridge, MA: MIT Press, in press.

McCarthy, John, and Alan Prince. 1993a. Prosodic morphology 1: Constraint interaction and satisfaction. Amherst, MA and New Brunswick, NJ: University of Massachusetts and Rutgers University, ms.

McCarthy, John, and Alan Prince. 1993b. Generalized alignment. Yearbook of Morphology 1993.79–153.

McCarthy, John, and Alan Prince. 1995. Faithfulness and reduplicative identity. Papers in optimality theory, ed. by Jill L. Beckman, Laura Walsh Dickey, and Suzanne Ur-

banczyk, 249–384. (University of Massachusetts Occasional Papers in Linguistics 18.) Amherst, MA: Graduate Linguistic Student Association.
MOULTON, WILLIAM G. 1962. The vowels of Dutch: Phonetic and distributional classes. Lingua 11.294–312.
NAKISA, RAMIN; KIM PLUNKETT; and ULRIKE HAHN. 2000. A cross-linguistic comparison of single and dual-route models of inflectional morphology. Models of language acquisition: Inductive and deductive approaches, ed. by Peter Broeder and Jaap Murre, 201–21. Oxford: Oxford University Press.
NELDER, J. A., and R. MEAD. 1965. A simplex method for function minimization. Computer Journal 7.308–13.
NOSOFSKY, ROBERT M. 1990. Relations between exemplar similarity and likelihood models of classification. Journal of Mathematical Psychology 34.393–418.
OHALA, JOHN J. 1983. The origin of sound patterns in vocal tract constraints. The production of speech, ed. by Peter F. MacNeilage, 189–216. Berlin: Springer.
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. Optimality theory: Constraint interaction in generative grammar. Boulder, CO and New Brunswick, NJ: Rutgers University and University of Colorado at Boulder, MS.
RUHLEN, MERITT. 1975. A guide to the languages of the world. Stanford: Stanford University, MS.
SANKOFF, DAVID. 1987. Variable rules. Sociolinguistics: An international handbook of the science of language and society, vol. 1, ed. by Ulrich Ammon, Norbert Dittmar, and Klaus J. Mattheier, 984–97. Berlin: Walter de Gruyter.
SCHREUDER, ROBERT, and R. HARALD BAAYEN. 1995. Modeling morphological processing. Morphological aspects of language processing, ed. by Laurie B. Feldman, 131–54. Hillsdale, NJ: Lawrence Erlbaum.
SKOUSEN, ROYAL. 1989. Analogical modeling of language. Dordrecht: Kluwer.
SKOUSEN, ROYAL. 1993. Analogy and structure. Dordrecht: Kluwer.
SKOUSEN, ROYAL. 2000. Analogical modeling and quantum computing. Los Alamos National Laboratory. Online: http://arXiv.org.
SLIS, I. H., and A. COHEN. 1969. On the complex regulating the voiced-voiceless distinction 2. Language and Speech 12.80–102,137–55.
SLIS, I. H., and M. VAN HEUGTEN. 1989. Voiced-voiceless distinction in Dutch fricatives. Linguistics in the Netherlands 1989, ed. by Hans Bennis and Ans van Kemenade, 123–32. Dordrecht: Foris.
VAN REENEN, PIETER, and EVERT WATTEL. 1992. De uitspraak van /s/ en /z/ voor klinker in het Nederlands: Zes eeuwen variatie. De Binnenbouw van het Nederlands: een bundel artikelen voor Piet Paardekoper, ed. by Hans Bennis and Jan W. de Vries, 291–306. Dordrecht: ICG Publications.
WARNER, NATASHA; ALLARD JONGMAN; JOAN SERENO; and RACHÈL KEMPS. 2001. Sub-phonemic durational differences in production and perception. Nijmegen: Max Planck Institute for Psycholinguistics, MS.
WOODS, ARTHUR E.; PAUL FLETCHER; and ANTHONY HUGHES. 1986. Statistics in language studies. Cambridge: Cambridge University Press.
ZONNEVELD, WIM. 1983. Lexical and phonological properties of Dutch voicing assimilation. Sound structures: Studies for Antonie Cohen, ed. by Marcel van den Broecke, Vincent van Heuven, and Wim Zonneveld, 101–35. Dordrecht: Foris.

Max Planck Institute
for Psycholinguistics
Wundtlaan 1
6525 XD Nijmegen
The Netherlands
[mirjam.ernestus@mpi.nl]
[Harald.Baayen@mpi.nl]

[Received 5 September 2001;
accepted 7 August 2002]