

The role of morphology in fine phonetic detail: The case of Dutch *-igheid*

Mark Pluymaekers^{*}; Mirjam Ernestus[#]; Harald Baayen[○]; Geert Booij[‡]

^{*} Radboud University Nijmegen, Mark.Pluymaekers@mpi.nl; [#] MPI for Psycholinguistics, Mirjam.Ernestus@mpi.nl;

[○] Radboud University Nijmegen, Harald.Baayen@mpi.nl; [‡] Leiden University, G.E.Booij@let.leidenuniv.nl

Résumé: Cette étude montre que la structure morphologique d'un mot influence la réalisation acoustique de ses affixes. Dans un corpus de Néerlandais parlé, l'agglomérat consonantique [xh] du suffixe -igheid est prononcé d'une façon plus courte si les deux consonants sont séparés par une coupe morphologique. Cette observation peut être expliquée dans une approche de la morphologie basée sur la théorie informationnelle selon laquelle l'information d'un suffixe dépend de la densité du paradigme morphologique.

The pronunciation of words and affixes is characterized by immense intra- and inter-speaker variation. In the current study, we investigated whether some of this variation can be accounted for by morphological structure. Our focus was on the Dutch suffix *-igheid* (*/əxhEiɪt/*), which occurs in different types of words. Three morphological types were considered, which differ in whether the base word and the form ending in *-ig* occur in isolation (see Table 1).

Type	Boundary location	Example	Translation
+igheid	Before <i>-igheid</i>	vast + igheid	'security'
+heid	Before <i>-heid</i>	zuinig + heid	'thriftiness'
+ig+heid	Either side of <i>-ig</i>	baz + ig + heid	'bossiness'

Table 1: Types of words containing *-igheid*, including examples and the locations of morphological boundaries

Words of the +igheid type are necessarily formed by adding *-igheid* to the base word, given that there is no existing form ending in *-ig*. The same derivation is likely to take place in +ig+heid words, albeit for a different reason. Hay (2003) found that derived words that are more frequent than their bases are more likely to be accessed as a whole. Since for the majority of +ig+heid words, the base word is more frequent than the *-ig* form, it can reasonably be assumed that in these words, *-igheid* tends to be added directly to the base. For words of the +heid type, such a derivation is impossible, as only the form ending in *-ig* is an existing Dutch word.

This study investigated whether the differences in morphological structure outlined above are reflected in the phonetic implementation of the phoneme cluster */xh/*. Two hypotheses were distinguished that make different predictions with regard to the realization of this cluster. The first hypothesis, henceforth referred to as the Prosodic Structure hypothesis, focuses on the implications of morphological structure for prosodic structure. In +igheid and +ig+heid words, the suffix as a whole is added to the base. As a result, the */h/* no longer occurs at the beginning of a prosodic domain and might be deleted, leading to cluster simplification (cf. Booij, in preparation). No simplification is expected in +heid words, since the */h/* serves as a morphological boundary marker there.

The second hypothesis, henceforth referred to as the Information-Based hypothesis, is concerned with the informativeness of the */xh/* cluster given the other words in the morphological paradigm. This approach finds its roots in Word-and-Paradigm morphology, where the focus is on paradigmatic rather than syntagmatic relations between words. In +heid words, the */xh/* cluster is relatively uninformative, as there are only a few competitors in the morphological paradigm from which the word needs to be distinguished. This is mainly due to the absence of a base form without *-ig*: Since *zuin* does not exist, the morphological paradigm of *zuinigheid* is rather sparse, containing only words that are derived from *zuinig*. Because of the

low informativeness of the /xh/ cluster, the Information-Based hypothesis would predict that it is shortened, given that less informative linguistic units are generally found to be more reduced than more informative linguistic units (e.g., Lieberman, 1963; Jurafsky, 2001; Aylett & Turk, 2004). In +igheid and +ig+heid words, on the other hand, the cluster is relatively informative, as the morphological paradigm of the base word contains more lexical competitors from which the *-igheid* word needs to be distinguished. As a result, simplification of the cluster could hinder word recognition and would therefore not be expected here. It should be noted that the Information-Based hypothesis makes predictions that are exactly opposite to those made by the Prosodic Structure hypothesis. The two hypotheses were pitted against each other in a corpus study.

The materials were taken from the subcorpus 'Library for the blind' of the Corpus of Spoken Dutch (Oostdijk, 2000). All 432 occurrences of *-igheid* in the subcorpus were analyzed using Automatic Speech Recognition (ASR) technology. Since it is extremely difficult to determine whether or not an /h/ is realized after a fricative, the ASR only measured the duration of the /xh/ cluster on the basis of a pre-determined phonemic transcription. The ASR was pre-tested on this task and found sufficiently reliable when compared with human transcribers. A linear regression model was fitted to predict the duration of the cluster. In addition to morphological type, variables such as speech rate (in syllables per second), speaker characteristics, and the position of the word in the utterance were taken into account in the analysis.

The duration of the /xh/ cluster was found to be shorter in +heid words than in +igheid and +ig+heid words. This finding supports the Information-Based hypothesis, which predicts cluster simplification in +heid words because the cluster is less informative there. It could be argued that the Prosodic Structure hypothesis has not been falsified by this observation, as its predictions mainly concern the presence or absence of /h/. However, it is difficult to conceive how the deletion of /h/ in +igheid and +ig+heid words, which was predicted by the Prosodic Structure hypothesis, could have led to a longer duration of the cluster as a whole in exactly these words.

We conclude that some of the variation in the acoustic realization of *-igheid* can indeed be accounted for by morphological structure. Since morphological structure depends on the language under investigation, this variation is language-specific and cannot be ascribed to low-level motor processes. More importantly, our results suggest that morphological effects on acoustic realizations need not be mediated by phonology. A prosodic account of *-igheid* words made predictions that were opposite to the effects that were observed, which suggests that the effects of morphological structure on fine phonetic detail are more direct than has thus far been assumed. Finally, these findings also lend support to accounts of morphological structure that stress the paradigmatic relations between words rather than the syntagmatic relations within words (e.g., Blevins, 2003).

References

- AYLETT, M. & TURK, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prominence, and duration in spontaneous speech, *Language and Speech*, 47, 31-56.
- BLEVINS, J.P. (2003). Stems and paradigms, *Language*, 79, 737-767.
- BOOIJ, G.E. (in preparation). Construction morphology.
- HAY, J. (2003). *Causes and consequences of word structure*, Routledge, New York & London.
- JURAFSKY, D., BELL, A., GREGORY, M., & RAYMOND, W.D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In Bybee, J. & Hopper, P. (eds), *Frequency and the emergence of linguistic structure*, 229-254, John Benjamins, Amsterdam.
- LIEBERMAN, P. (1963). Some effects of the semantic and grammatical context on the production and perception of speech, *Language and Speech*, 6, 172-175.
- OOSTDIJK, N. (2000). The Spoken Dutch Corpus Project, *The ELRA Newsletter*, 5, 4-8.