

Chapter 3

DUTCH INFLECTION: THE RULES THAT PROVE THE EXCEPTION

Harald Baayen,
Robert Schreuder,
Nivja de Jong
and Andrea Krott

Abstract This chapter addresses the balance of storage and computation in the mental lexicon for fully regular and productive inflectional processes in Dutch. We present evidence that both regular inflected nouns and regular inflected verbs show clear and robust effects of storage, but that at the same time on-line parsing also plays a role. We argue that the balance of storage and computation cannot be predicted on the basis of economy of linguistic description. Instead, a range of cognitive and linguistic factors are crucial determinants.

Keywords: Morphological processing, frequency effects, dual route modelling, inflection, regularity, economy of description, defaults.

1. Introduction

This chapter addresses the issue of the balance of storage and computation for regular inflected words in Dutch in language comprehension. Traditionally, this question has not been an issue in linguistics. Ever since Bloomfield, 1933, the lexicon has been the repository of idiosyncracies. Listing in the lexicon is restricted to monomorphemic words and to complex words that are not predictable by rule, whether at the level of form or at the level of meaning. For linguistics as a descriptive discipline, this approach is justified, as the listing of complex words sharing features that can be predicted by rule implies the loss of generalizations.

In computational linguistics, practical considerations have led researchers to store the shorter complex words in their parsing programs in order to minimize the costs of having to analyze — on-line — such often highly

ambiguous and frequent forms. In recent computational approaches such as data-oriented parsing (Bod, 1998) and lazy learning (Daelemans et al., 1999), extensive storage of regular words and multi-word structures is used to optimize on-line performance. In these approaches, the data are the primary source of information, and no attempt is made to extract explicitly from the data rules for analysing unseen constructions.

In psycholinguistics, the complete range of possible trade-offs between storage and computation has found its defenders. Butterworth, 1983 was the first to formulate the hypothesis that complex words in a language such as English are extensively stored in the mental lexicon, and that parsing is a back-up analogical procedure used only for the rare instances in which novel forms are encountered. Connectionist models of the kind advocated by Seidenberg, 1987 are implementations of this position using a framework in which distributed data storage takes place without any role for explicit morphological rules. Effects of morphological structure observed in on-line processing experiments are, in this approach, an epiphenomenon of statistical regularities between input and output patterns. The opposite extreme has been defended by Pinker, 1991; Marcus et al., 1995; Clahsen et al., 1997 and Clahsen, 1999. Their approach assumes that Bloomfield's position is an adequate characterization of the balance of storage and computation in human language processing. An intermediate position has led to the formulation of a parallel dual route model (Schreuder and Baayen, 1995), in which rules and memory-based retrieval operate in parallel.

The aim of the present chapter is to address the question of the balance of storage and computation for regular inflected words. Authors such as Pinker, Marcus, Clahsen, and also Jaeger et al., 1996 have argued that regular and irregular inflected complex words are processed differently by the brain, irregulars being handled by memory-based retrieval, and regulars being handled by rule (but see Plunkett and Juola, 1999, and Seidenberg and Hoeffner, 1998, for alternative interpretations of the data). One of the arguments used by Pinker and Clahsen concerns the occurrence of frequency effects for complex words, which they argue to be restricted to irregular complex words.

There are two distinguishable token frequency effects. The Surface Frequency Effect concerns the finding that the frequency of occurrence of a complex word such as *walks* can be an important predictor of response latencies in a variety of word recognition tasks. The Base Frequency Effect concerns the observation that the frequency of the lexeme *walk*, i.e., the summed token frequency of *walk* and all its inflectional variants, can also be an important predictor of such response latencies. The effects of Surface Frequency and Base Frequency are independent, as shown by

Taft, 1979 and Burani et al., 1984 for English and Italian inflected words. The existence of a Surface Frequency Effect is generally interpreted as evidence for storage of the inflected word as a whole at some level of representation. The existence of a Base Frequency Effect indicates that at some point during lexical access the base word itself has been accessed. The claim advanced by Pinker and Clahsen is that for regular inflected words only Base Frequency Effects should be observed and no Surface Frequency Effects.

Our research, however, has revealed very reliable Surface Frequency Effects for regular inflected words. For Dutch, Italian, and English, regular noun plurals show substantial Surface Frequency Effects (Baayen et al., 1997a; 1997b; Sereno and Jongman, 1997; and Allegre and Gordon, 1999). In Finnish, such an effect has been reported for regular partitive noun plurals, even though Finnish is a language with a very rich morphology (Bertram et al., 2000b). In Dutch, furthermore, fully regular comparatives have been found to show Surface Frequency Effects (Bertram et al., 2000c). These results call into question the idea that the economy of linguistic description, which requires maximization of rule-based description and minimization of list-based description, can be mapped directly onto the domain of lexical processing in the mental lexicon. In the cognitive domain, it is not a priori self-evident that the costs of processing by rule are less than the costs of storage (Frauenfelder and Schreuder, 1992). The storage capacity of the human brain is enormous (Landauer, 1986), and it is well known from various cognitive domains (e.g., hand reaches, Rosenbaum et al., 1992, and arithmetic, Rickard et al., 1994) that storage may take over when computations become complex and time-consuming. In the domain of morphological processing, storage might likewise be a useful mechanism complementing rule-based access.

Moreover, from a more general linguistic point of view, storage for regular noun plurals as observed by Baayen et al., 1997b is not that surprising either. Tiersma, 1982 has pointed out that high-frequency plural forms are what he called locally unmarked. Normally, plurals are formally and semantically marked with respect to their singulars, as in the case of *nose* versus *noses*. On the other hand, plurals that are much more frequently used than their corresponding singulars tend to be semantically unmarked. For instance, *eyes* and *feet* are unmarked in the sense that they represent the default occurrence of the objects they refer to. Tiersma shows that locally unmarked plural nouns serve as attractors in language change, which implies that such nouns must have been stored in the mental lexicon. Cross-linguistically, it is precisely the locally unmarked plural forms that emerge as simplex forms in languages

such as Bari (Eastern Nilotic), which uses a singulative suffix to derive the singular form of a monomorphemic semantic plural noun (e.g., *kuru*, 'worms', *kuru-töt*, 'worm', Dimmendaal, 1987).

Clahsen et al., 1997, however, have suggested another reason why effects of storage might have been observed for Dutch noun plurals, namely, their supposed irregularity. Clahsen and his colleagues argue that a more subtle notion of regularity is called for than the one used in traditional grammars (Clahsen et al., 1997, p.208). In order to ascertain whether the plural noun suffixes of Dutch fall into what they call the regular cluster, the properties of Dutch noun plurals would have to be checked against a reference list of linguistic criteria for regularity as proposed by Marcus et al., 1995, p.197. In this approach, the traditional notion of regularity has been replaced by the notion of default. Default forms are said to be free of morphological restrictions and lexical governance, they are claimed to attract members of non-default classes in historical change. Loan words and marginal forms such as acronyms and abbreviations are also analyzed to fall into default classes.

Other scholars have likewise questioned the regularity of the Dutch plural suffixes (Gordon, 1985; Pinker and Prince, 1988; 1991), on the basis of the occurrence of Dutch noun plurals in compounds, and on the basis of the plurals being rival affixes.

In what follows, we will examine this important alternative explanation for why storage effects in lexical processing might arise: irregularity. If it is indeed the case that some form of irregularity is the driving force behind any storage in the mental lexicon, then irregularity would overrule more general cognitive considerations concerning the balance of storage and computation in general. Before assigning primacy to irregularity as the driving force behind lexical storage, we should consider in detail whether it is linguistically justified to describe Dutch noun pluralization as in some sense irregular. In section 2, we therefore examine noun plural formation in Dutch with respect to the notion of default developed by Marcus et al., 1995. We will show that the system of noun pluralization is regular, and that native speakers of Dutch coin new noun plurals exactly as predicted. In section 3, we will proceed to show that Surface Frequency Effects may also appear for fully regular verbal inflections which can only be described as default suffixes. At the same time, we will show that these Surface Frequency effects can occur simultaneously with Base Frequency effects, suggesting that storage and computation are not mutually exclusive. In the General Discussion, we will integrate our findings from the regular nominal and regular verbal inflectional domains, and we will trace some of the resulting theoretical consequences.

2. Dutch noun plurals

In this section, we first present an outline of the rules governing the distribution of the plural suffixes in Dutch. We then proceed to show that speakers of Dutch apply these rules consistently without problems to phonotactically legal pseudowords, which illustrates that pluralization is indeed rule-governed.

2.1. The rules

There are three plural suffixes in Dutch: *-en*, *-s*, and *-eren*. The latter suffix occurs only in 15 nouns, and need not concern us here.

There are five factors that determine the distribution of *-en* and *-s*, both of which are fully productive. The main factor is the phonological structure of the base word: *-s* appears after unstressed syllables, and *-en* appears after stressed syllables. From the phonological point of view, *-s* and *-en* are in complementary distribution, and can be considered synchronically as suppletive forms. There are some further phonological restrictions: Words ending in a full stressed vowel often select *-s* rather than *-en*, which may be due to avoid a sequence of two non-homorganic vowels: *pa's*, 'dads'. Furthermore, if a word already ends in an *-s*, the appropriate suffix is *-en*: *cactussen*.

A second important factor is morphological in nature. Some suffixes always require *-en*, e.g., *-ing*, '*-ing*', other suffixes only take *-s*, e.g., the diminutive suffix *-tje*, while yet others allow both, e.g., *-isme*, '*-ism*'.

Third, loan words select *-s*, e.g., *stations*, *trams*, *jeeps*.

Fourth, the *-s* tends to be selected for persons as opposed to things, especially in the case of homonyms: *portiers*, 'doorkeepers' versus *portieren*, 'car doors'.

Fifth, both suffixes are possible for nouns ending in a schwa (*kade*: *kades*, *kaden*, 'quay'); more formal registers prefer the suffix *-en*.

The existence of choice for words ending in a schwa can be understood in terms of the general preference for disyllabic trochees in Germanic (see Booij, 1998, for an account in terms of optimality theory). Both *kades* and *kaden* as plural forms of *kade* satisfy this rhythmic constraint (Van Haeringen, 1947) of Dutch. There are words for which the first factor (the rhythmic principle) clashes with the fourth factor, which specifies *-s* as the choice for person names. In such cases, both suffixes can be used: *directeur*, 'director', has as plural forms both *directeurs* (by factor 4) and *directeuren* (by factor 1).

Finally, there are also some exceptions to these rules, such as *wonder*, 'miracle', which has as its plural *wonderen* instead of *wonders*. In a

<http://www.springer.com/978-1-4020-0526-8>

Storage and Computation in the Language Faculty
Nooteboom, S.G.; Weerman, F.; Wijnen, F.N.K. (Eds.)
2002, 356 p. 3 illus., Hardcover
ISBN: 978-1-4020-0526-8