The Dutch plural landscape

Judith van Wijk
Utrecht University / UIL-OTS

1. Introduction

This paper aims to test claims about Dutch pluralisation found in the literature against data on Dutch nouns and their plurals in a corpus from Van Dale.[1] With the information gained from this research I hope to get as complete a picture of pluralisation in Dutch as possible, and find out how strong the claims that are often made actually are on a quantitave basis. This information is vital for the development of test items for the *wug-tests* (Berko, 1958) I intend to do at a later stage. The nonce words that children will be tested on in these *wug-tests* should be representative of the Dutch plural landscape to mirror the input children receive. I will do these experiments to test the *words-and-rules* theory, as proposed by Pinker (1999). This theory, which I will explain in greater detail below, claims the use of a default affix for regular plurals. Dutch, however, has two plural affixes, neither of which is the default by well-established criteria, and could therefore be a problem for the theory. If the theory is universal, though, children are expected to show default use, even if adults do not. They will expect a situation with a single default. Default use will then manifest itself in overgeneralisations of one, not both, of the plural affixes and in productive use of their chosen default in the *wug-test*.

The *words-and-rules* theory claims there are separate components for regulars and irregulars in the language faculty. The theory states that irregulars are stored in the mental lexicon, while rules are used to form regulars. The most important rival of the dual theory is the connectionist model (Rummelhart & McClelland, 1986). According to this model both regulars and irregulars are stored in the mental lexicon. In the connectionist model, as well as in the dual model, the mental lexicon works as a pattern associator, organising the stored items according to patterns. In the dual model, however, when a stem is not found in the lexicon a rule applies by default. This rule attaches an affix to the stem to form, for instance, the past tense or the plural. Examples of an affix used by default, a default affix, are the English plural -*s* in *houses* and the German plural -*s* in *Autos*.

How do we know which affix is used by default? Looking at the English example

mentioned above the answer may seem obvious: the most frequently used affix is the default affix. The answer is more complicated however, and German pluralization shows us why. The German plural affix *-s* is one of eight possible plural forms. Even though *-s* is used in a minority of cases it is the default affix. The crucial criterion for defaulthood is the diversity of types an affix attaches to (Pinker & Prince, 1991). A default poses no restrictions on the stem, except for the category it belongs to, e.g. English plural *-s* can attach to any word as long as it is a noun. This ensures that all singular nouns that are not stored in the lexicon can be handled by the rule.

Accordingly, the affix that is considered to be the default can be determined with the help of diversity-related criteria that are characteristic of the default affix; it attaches to new words, loanwords, acronyms, and names. The new words criterium is especially important, since these words have not been heared before and can therefore not be stored in the mental lexicon. It is because of these criteria that we know *-s* is the default plural affix in both English and German. It attaches to new words, e.g. *wugs* (Eng.), loanwords, e.g. *Kiosks, Hits* (G.), acronyms, e.g. *LKWs* (trucks) (G.), and names, e.g. *Johnsons* (Eng.).

Much research has been done on this subject, mostly for German and English. Dutch pluralisation may shed some new light on the subject as it has two plural affixes, *-s* (/s/) and *-en* (/ən/), each of which is dominant within its own phonologically defined domain. The relevant phonological features are stress and sonorancy of the final consonant. Besides being dominant within their own domains, they both seem to obey the criteria for the default. They attach to:

(1)   new words → *kepsen, flonkers* (Snow et al., 1980)
       loanwords → *fezzen, ciabatta's*
       acronyms → *HATten, PABOs* (Baayen et al., to appear)
       names → *de van Wijken, de van Dijks*

If it turns out that Dutch has two default affixes, however, this is problematic for the dual model. The central criterion for the default in this model is the ability to attach to a stem without morphological or phonological conditions. For adult Dutch such conditions are relevant if not necessary.

The fact that adult Dutch seems problematic for the dual model does not mean the theory cannot be universal. If it is universal children acquiring Dutch will attempt to establish a default no matter what the input is. Acquisition data of Dutch pluralisation, potentially, have important consequences for the dual model. The first possibility is that Dutch children show no evidence for a default at all, which would be evidence against the dual model. It would imply that either the theory is language specific, or that it is incorrect all together. The second and third possibility both confirm the theory, but in different ways. If children do acquire a default overgeneralisations are expected to be found. Consistently using one of the two affixes for all plurals of nonce words is strong evidence in favour of the theory. The

third option, two defaults in children's productions, would also be a confirmation of the dual model, but would result in a different kind of overgeneralisation. The child will then distinguish two domains, one for each affix. Each domain contains irregulars that the child will overgeneralise. In this case the definition of 'default affix' as it is used now in the theory has to be adepted.

In this paper I will first establish what has been said about Dutch pluralisation in the literature, and then check these claims against data from a corpus from Van Dale. Finally I will discuss what the results of the comparison mean for the development of test items for my experiments.

2.   Dutch pluralisation in the literature

Several sources give a description of Dutch pluralisation, the most important of which are Van Haeringen (1947), De Haas & Trommelen (1993), and Haesereyn et al. (1997).

All sources agree that stress is an essential factor in the pluralisation process. Van Haeringen calls it the rhythmic factor. This factor states that Dutch strives for a strong-weak pattern at the end of a word. Therefore a word with a stressed final syllable will prefer *-en* (/ən/) over *-s* (/s/), while a word with an unstressed final syllable will prefer *-s.*

(2)   tafel-s   stoel-en
       s   w   s   w
      *tafel-en   *stoel-s
       s w w   s

The rhythmic factor is not the only factor involved in pluralisation; other important phonological factors are sonorancy and number of consonants in the final coda. These factors show their relevance in the form of preferences. The preferences that (sometimes) go against the rhythmic factor are summarised in (3).

(3)

-en	*-s*
final obstruent (*arend*)[2]	
final diphthong (*juffrouw*)	final back vowel (*niveau*)
final long vowel + sonorant (*etmaal*)	polysyll. with final front vowel (*paraplu*)
both *-s* and *-en* possible	
simplex words in *ə* (*pagode*)	

As it says in (3), words ending in *ə* can take both *-en* and *-s* as plural affix, but De Haas & Trommelen state that such words show a clear preference for *-s.* Also, the

number of words ending in *ə* that exclusively take *-en* decreases, whereas the number of words ending in *ə* that exclusively take *-s* is increasing.

The literature is divided on the subject of words ending in the front vowel *-ee* (/eː/). According to Haesereyn et al. *-ee* should be grouped with diphthongs because it behaves similarly. In that case words ending in *-ee* should prefer *-en* as can be seen in (3). Van Haeringen also states that both *-ee* and *-eu* (/ø/) are like diphthongs and prefer *-en* as plural affix. De Haas & Trommelen, on the other hand, distinguish between monosyllabic words in *-ee* and polysyllabic words in *-ee*. The monosyllabic ones take *-en* for the plural, in accordance with the rhythmic factor, while the polysyllabic ones take *-s*.

Besides the rhythmic factor and other phonological factors there are also non-phonological factors that lead to the preference of a certain affix. Loanwords always take *-s*, possibly because the plural of the word in its original language takes *-s* and has also been borrowed. The presence of a suffix can also lead to a preference for a certain plural affix. For instance, *-ier*, as in *winkelier* (/wɪŋkəlir/, shopkeeper) always takes *-s*, but *-ing*, as in *lading* (/ladɪŋ/, load), always *-en*. The preferences for one of the affixes with these words are striking because their stress patterns would predict the opposite affix of the one they pick. These words go against the rhythmic factor as the result of the strong influence of their suffixes.

To find out how strong the claims in the literature are I will consider data on affix distribution from the Van Dale corpus.

3.   Results Van Dale corpus

3.1   Material

The corpus of data from Van Dale[3,4] that was used for this research consists of separate lists for nouns taking *-s* (/s/) in the plural and nouns taking *-en* (/ən/) in the plural. Within these lists the words are ordered according to number of syllables. The sublists that arise as a result are ordered for stress pattern, and within these sublists nouns are ordered according to their final sound.

Token frequencies of both singulars and plurals are based on the contents of five Dutch newspapers over the course of 2½ years. Although token frequencies give a good indication of the frequency of words, it has to be kept in mind that there is some 'pollution' in the figures. Though not often, in some cases the frequency is affected by the fact that a verb or an adverb exists that is written exactly the same as the plural of a noun. Type frequency of classes was derived by counting the words in a class, token frequency by adding up their individual frequencies.

The results presented in this paper are based on analyses of the monosyllables, bisyllables and trisyllables in the corpus. Together they contain 12,715 plurals.

In my analysis of the data I have not distinguished between loanwords and native words, because my main interest is to get an idea of the possible input children base their knowledge on. Children do not know whether a word is a loanword or not so they will not treat them differently from other words. The only loanwords that are treated differently are those with a singular form in -*s* and a plural that takes -*s* as well (a *ə* is inserted between the singular and plural -*s*). There is a small number of words in the list that behaves like this, an example being *lunch-lunches*. I have not classified these words as taking -*s*, but as taking a different, *ander*, plural affix.

The lists also contain compounds. As with loanwords there is no way of knowing for sure if children know these words to be compounds, and even if they do taking them out of the corpus would distort the picture of the plural landscape. Children do not disregard them, so neither will I.

3.2 Results

The information I hope to gain from this analysis is a quantification of the relative strength of factors identified in the literature, as well as, possibly, an identification of new factors. Zooming in on the plural landscape, dividing the corpus into smaller parts on the basis of different features, a more refined picture gradually emerges. The features considered here are stress, number of syllables, sonorancy and the manner of articulation of the final consonant. Special attention will be paid to words ending in a vowel, since the literature is divided on this subject.

Combining all monosyllables, bisyllables, and trisyllables in the corpus to determine the overall distribution of the affixes we find that 50.8% of types and 73.5% of tokens takes -*en* as plural affix. However, frequency is not an overriding factor for the default (Clahsen et al., 1992), and the phonological diversity of bases of affixation is more important. This general picture of Dutch pluralisation is not detailed enough to provide the information necessary for the development of well-formed wug-items.

According to the literature one of the most important factors in Dutch pluralisation is the rhythmic factor. As can be seen in (4)[5] the data from Van Dale confirm this.

(4)   Percentages of words taking -*en* according to stress pattern

final	penultimate	antepenultimate
type: 4851/6037 (80.4%) token: 2717452/2915846 (93.2%)	type: 1025/5094 (20.1%) token: 152967/885262 (17.3%)	type: 587/1571 (37.4%) token: 36927/152923 (24.2%)

A majority of words with final stress takes -*en* as the plural affix. The reverse is true for words with penultimate stress; here a majority takes -*s*. For words with antepenultimate stress the preference for one of the two affixes is less strong. A majority takes -*s* as plural affix, which seems to be the most logical choice since it does not add another weak syllable. A large minority, however, does add that weak syllable, which is made possible by the fact that these words often have an extra stress on the final syllable.

Sonorancy of the final consonant is another feature that is prominent in the literature. A sonorant at the end of an unstressed final syllable is an environment that favours the use of -*s* for the plural. An obstruent as the final consonant of a word, regardless of the stress pattern, is an indication of -*en* being used for the plural. Leaving out stress for the moment, and considering only the final segment the following distribution is found in the data:

(5) Percentages of words taking -*en* according to the final segment

obstruent	sonorant	vowel
type: 4039/4774 (84.6%)	type: 2432/5446 (44.7%)	type: 364/2892 (12.6%)
token: 1882740/1944937 (96.8%)	token: 1016210/1570652 (64.7%)	token: 27439/459151 (6.0%)

Words that do not end in a consonant prefer -*s* over -*en* according to the literature, and as can be seen in (5) this claim is supported by the Van Dale data.

To zoom in closer on the pluralisation 'rules' the features stress pattern and sonorancy of the final consonant will be combined. In doing so the results become more polarised.

(6) Percentages of words taking -*en* when stress and sonorancy of the final consonant are considered

	final	penultimate	antepenultimate
obstruent	type: 3418/3773 (90.6%) token: 1794069/1842434 (97.4%)	type: 373/659 (56.6%) token: 68046/80130 (84.9%)	type: 248/342 (61.2%) token: 20625/22373 (92.2%)
sonorant	type: 1465/1961 (74.7%) token: 924418/1061622 (87.1%)	type: 634/2587 (24.5%) token: 75491/405379 (8.6%)	type: 333/898 (37.1%) token: 16301/103651 (15.7%)
vowel	type: 340/698 (48.7%) token: 18008/32119 (56.1%)	type: 18/1863 (1.0%) token: 9430/400133 (2.4%)	type: 6/331 (1.8%) token: 1/26899 (0.01%)

This table shows that stress is more important than sonorancy, but that both features are essential for pluralisation. The strong influence of stress can be seen most clearly in the frequencies of plurals whose bases end in a sonorant. When sonorancy is the only feature under consideration around 50% of words with a final sonorant take -*en*. When stress is also taken into account the results become more polarised.

3.2.1 *Final sonorants*

By far the strongest preference found in the data that can be seen in the table is that words with an unstressed final syllable and ending in a vowel take -*s* as plural affix. I will return to vowel final words later, and first zoom in on words with a final sonorant by looking at the vowel preceding the sonorant.

(7)   Percentages of words ending in a vowel and a sonorant taking -*en*

	final	penultimate	antepenultimate
long vowel /a, e, i, o, u, y/	type: 1008/1257 (80.2%) token: 574748/640272 (89.8%)	type: 16/35 (45.7%) token: 3995/6628 (60.3%)	type: 32/150 (21.3%) token: 778/12819 (6.1%)
short vowel /ɑ, ɛ, ɪ, ɔ, y/	type: 394/583 (87.6%) token: 315834/375826 (84.0%)	type: 614/982 (62.5%) token: 69556/98034 (71.0%)	type: 298/377 (79.0%) token: 15516/18125 (85.6%)
ə	n.a. n.a.	type: 1/1513 (0.06%) token: 3/297829 (0.001%)	type: 1/350 (0.3%) token: 0/72227 (0.0%)
loan vowel /ɑː, ɛː, ɔ, yː, iː, uː/	type: 6/52 (11.5%) token: 15357/21699 (70.8%)	type: 0/26 (0.0%) token: 0/312 (0.0%)	type: 0/15 (0.0%) token: 0/252 (0.0%)

It is clear from the table that words in which a ə precedes the final sonorant behave differently from words with a short vowel. The claim that these words have a tendency to take -*s* as plural affix is strongly confirmed. However, according to the literature those words that end in a short vowel followed by a sonorant and that do not have stress on the final syllable have a preference for -*s* as well, but this is not confirmed by the data. Instead a considerable majority prefers -*en*. This is partly due to the presence of words ending in -*ing* (/ɪŋ/), a suffix that always takes -*en*, in this group.

It is claimed in the literature that words ending in a long vowel and a sonorant have a tendency to take -*en* over -*s*, but from (7) it is clear that this tendency decreases when the stress moves away from the end of the word. The reason that these words seem to prefer -*s* over -*en* can be found in a 'sub-rule' of Dutch pluralisation which states that personal names take -*s* as plural affix. One of the

suffixes that creates such names is -*aar* (/ar/), as in *wandelaar* (/wandəlar/, walker). Words of this type are very common in Dutch.

The factor manner of articulation of the final consonant is not mentioned in the literature, but I did consider it in my analysis. I found that in general plosives and fricatives behave alike, however when the number of syllables is taken into account a deviant category is found amongst bisyllabic words with penultimate stress. Here words with a final fricative generally show a preference for -*en* as plural affix, which goes against the rhythmic factor. The source of this deviation is the fricative -*s*. Practically all words ending in -*s* take -*en* in the plural to ensure a difference between singular and plural, which -*s* would not give. A small number of words in this category, all loanwords, take -*es* in the plural. These are all plurals that are taken from the language the word was originally borrowed from.

A second group of words that was found to behave different from the others are words with a final liquid. When these words have an unstressed final syllable they have a very strong preference for -*s* over -*en*. The reason is again the 'sub-rule' concerning personal names mentioned above, now applying to the person affixes -*ier* (/ir/) and -*eur* (/ør/). Another cause for the high percentage of -*s* with liquid final words is that there are large amounts of words in Dutch that end in -*əl* or -*ər*, like *kabel* (cable) and *loper* (masterkey), which, as was shown in (7), practically always select -*s* for the plural.

3.2.2 *Vowels*

As opposing views on vowel final words were found in the literature, this section will take a closer look at those words. Dividing the vowels into diphthongs, back vowels, and front vowels shows there is dichotomy between diphthongs and monophthongs.

(8)   Percentage of vowel final words taking -*en*

	final	penultimate	antepenultimate
diphthong /au, ɛi, œy/	type: 225/232 (97.0%) token: 11937/11941 (99.96%)	type: 8/10 (80.0%) token: 870/873 (99.7%)	type: 6/6 (100.0%) token: 1/1 (100.0%)
back vowel /a, o, u, ɑ/	type: 1/91 (1.0%) token: 1/2784 (0.03%)	type: 0/493 (0.0%) token: 0/24855 (0.0%)	type: 0/102 (0.0%) token: 0/11345 (0.0%)
front vowel /e, i, ø, y, ɛ/	type: 107/294 (36.4%) token: 5472/15516 (35.3%)	type: 8/470 (1.7%) token: 1507/88151 (1.7%)	type: 0/31 (0.0%) token: 0/494 (0.0%)

The data strongly confirm the claims about monophthongs made in the literature; there is a strong preference for -*s* as plural affix. The only group of monophthongs

that deviates from the others is that of words with a final front vowel and a stressed final syllable. This is the result of high numbers of words ending in *-i* (/i/). When *-i* is stressed it prefers *-en* over *-s*. In the literature these words are claimed to have a tendency to take *-en*, however, according to the data a majority prefers *-s*.

In the literature several interpretations of words with final *-ee* could be found. Van Haeringen en Hasereyn both grouped *-ee* with diphthongs indicating that a majority of words with final *-ee* should take *-en* as plural affix. De Haas & Trommelen, on the other hand, distinguished between polysyllables and monosyllables with final *-ee*. Polysyllables are expected to take *-s*, while monosyllables are expected to take *-en*. The data from Van Dale confirm this last interpretation. Overall 85–100% of types and 99–100% of tokens in the Van Dale corpus take *-s* as plural affix when *-ee* is the final segment of a polysyllabic word. These high percentages can be explained by the large number of loanwords from French that end in *-ee*, e.g. *logee–logees* (guest(s)), but that does not change the fact that children may deduce from their input that most words ending in *-ee* take *-s* in the plural. Monosyllabic words ending in *-ee* are divided when it comes to picking a plural affix. Of the types a slight majority takes *-s*. This contradicts the literature. However, the token frequency of words taking *-en* is much higher, which is in accordance with the literature. This shows that the analysis of De Haas & Trommelen is correct, and that the tendency of polysyllables in *-ee* to take *-s* is very strong.

Finally something has to be said about ə, as there are a lot of words in Dutch that end in this vowel. Since ə only occurs in unstressed syllables the figures concerning final ə are exclusively based on polysyllabic words with penultimate or antepenultimate stress. Of all the words that end in a vowel 42.6% of types and 67.8% of tokens of *-s* plurals come from words ending in ə. For *-en* plurals these figures are much lower; 0.5% of types and 25.7% of tokens come from words ending in ə. This is partly due to the presence of diminutives which exclusively take *-s*, but it can also be the result of a phenomenon already mentioned in the literature; the group of words ending in ə that exclusively take *-s* is increasing, whereas those exclusively taking *-en* is becoming smaller.

4.   Discussion

The data from the Van Dale corpus have confirmed that it is unlikely Dutch has a unique default. They both attach to a great diversity of types, but they both impose quite a few restrictions on the stems they attach to as well. Even if adult Dutch does not have a clear default, children are still expected to search for a default if the dual model is universal. By means of *wug-tests* I will try to find out if they do indeed look for a default.

The analysis discussed in this paper has provided an indication of which type of words have a strong preference for a certain affix and in which cases the preference is less strong. This information is a prerequisite for the development of items for a *wug-test* on pluralisation. Environments with a strong tendency towards a certain affix are less interesting as test items than those in which both affixes occur frequently. If children do indeed overgeneralise one of the two affixes they are more likely to do this in areas where both affixes occur in adult Dutch. Therefore more information can be gained from a test item like *bapoel* (/bˈapul/), final long vowel and sonorant with penultimate stress, than from a test item like *spootje* (/spˈotjə/), ə final. Other interesting areas for recruitment of test items are those areas where both type and token frequency are quite low. If words do not occur often in the input, it is more likely that overgeneralisations occur because high frequency words could be stored in the mental lexicon.

Notes

1. The corpus used for this research was made available to the author by Van Dale Data bv. The corpus is a selection of nouns and their plurals taken from Van Dale Woordkenmerkenbestand. The nouns/plurals that were selected are the ones that can be found in *Van Dale Groot woordenboek hedendaags Nederlands*. A more detailed description of the corpus will be given later in this paper.

2. Pronunciation and translation of Dutch words in this table:

arend	/ˈarənt/	eagle
juffrouw	/jˈʊffrau/	Miss
niveau	/nivˈo/	level
etmaal	/ˈɛtmaal/	24 hours
paraplu	/paraplˈy/	umbrella
pagode	/paɣˈodə/	pagoda

3. Since an adult corpus can give a more general picture of pluralisation, and because data on child-directed speech from CHILDES are mainly on speech directed at very young children, the Van Dale corpus was chosen over CHILDES for this research.

4. After writing the article it was discovered there are three gaps in the data from Van Dale. The gaps concern plurals taking -(*e*)*n* for (a) words ending in -ə, (b) words ending in -ə+sonorant (*appelen* (apples)), and (c) words ending in -*f* or -*s* that do not have softening in the plural, e.g. *fotografen* (photographers). Type/token for (a) 565/818167, (b) 99/35955, (c) 62/1408.

5. All tables give the percentages of words taking -*en*. Percentages of words taking -*s* can easily be computed, as they are the only other affix. In the rare event of the occurrence of a different (*ander*) affix the percentage will be mentioned separately, if relevant.

The absolute numbers are also given; they should be read as follows: number of words taking -*en*/total number of words in this class.

A Chi-square test was performed for all distributions, which has shown that all distributions are significant.

References

Baayen, H., Schreuder, R., DeJong, N., & Krott, A. to appear. Dutch inflection: the rules that prove the exception. In *Storage and computation in the language faculty*, eds. S. Nooteboom, F. Weerman and F. Wijnen.

Berko, J. 1958. The child's learning of English morphology. *Word* 14: 150–177.

Clahsen, H., Rothweiler, M., Wöst, A., & Marcus, G. 1992. Regular and irregular inflection in the acquisition of German noun plurals. *Cognition* 45: 225–255.

de Haas, W., & Trommelen, M. 1993. *Morfologisch handboek van het Nederlands: een overzicht van de woordvorming*. 's-Gravenhage: SDU.

Haesereyn, W., Romijn, K., Geerts, G., de Rooij, J., & van den Toorn, M.C. 1997. *Algemene Nederlands Spraakkunst*. Groningen: Martinus Nijhoff.

Pinker, S., & Prince, A. 1991. Regular and irregular morphology and the psychological status of rules of grammar. In *The reality of linguistic rules (1994)*, eds. S.B. Lima, R.L. Corrigan and G.K. Iverson. Amsterdam: John Benjamins.

Pinker, S. 1999. *Words and rules: the ingredients of language*. New York: Harper Collins.

Rummelhart, D., & McClelland, J.L. 1986. On learning the past tense of English verbs. Implicit rules or parallel distributed processing? In *Parallel distributed processing: Explorations in the microstructure of cognition*, eds. J. McClelland, D. Rumelhart and PDP Research Group. Cambridge, Massachussetts: MIT Press.

Snow, C.E., Smith, N.S., & Hoefnagel-Höhle, M. 1980. The acquisition of some Dutch morphological rules. *Journal of Child Language* 7: 539–553.

van Haeringen, C.B. 1947. De meervoudsvorming in het Nederlands. In *Neerlandica: verspreide opstellen*, ed. C.B. van Haeringen, 186–209. 's-Gravenhage.

Van Dale Woordkenmerkenbestand, Van Dale Data, Utrecht.