Rules and exceptions in Dutch word stress

Anneke Neijt and Vincent J. van Heuven

1. Basic characteristics of Dutch word stress and examples

Dutch word stress has been the subject of extensive investigation during the last ten years, cf. Neijt and Zonneveld (1982), van der Hulst (1984), Langeweg (1988), Kager (1989) and Trommelen and Zonneveld (1989). The various authors describe Dutch stress in terms of a few main rules, supplemented by subrules and exception lists. The exact choice of the main rules has been subject to debate. Evidence for the main rules has been sought in various domains: stress shifts in loan words, stress patterns in newly-formed words, brand names, acronyms, child language, or speech errors, but the most common criterion in the references above has been lexical frequency of stress patterns. In this paper we shall confront three closely related accounts of Dutch word stress with lexical data in order to evaluate their predictions. The three turn out to be descriptively equivalent, but the survey of data suggests that a more restrictive theory is possible, in which phonological and morphological subregularities are accounted for outside of the general framework.

The following characteristics of Dutch stress seem to be generally agreed upon (for a detailed survey and references cf. Kager 1989):

- (1) Generally assumed characteristics of Dutch word stress:
 - a quantity sensitive foot structure, e.g. no heavy (h) or superheavy (sh) syllables as weak nodes of feet;
 - b syllables form bounded feet;
 - c direction: from right to left;
 - d left dominant foot structure (in metrical trees: strong-weak assignment);
 - e right dominant word structure (in metrical trees: weak-strong assignment, and right branching structures).

Foot structure (1a) depends on segmental distinctions of syllable rimes. The relevant types of syllables are the following:

This paper could not have been written without the help of Maarten Hijzelendoorn, who wrote the software to run the searches on the RUL-corpus. We have profited from comments by Harry van der Hulst, René Kager, Jacques Koreman and Anneke Nunn.

- Superlight syllables (sl), sometimes called schwallables (Kager 1989), typically have rimes with schwa, optionally followed by consonants.
- Light syllables (1) have rimes with a long vowel (VV) only.
- Heavy syllables (h) have rimes with a short vowel plus one consonant only (VC). In Dutch a short vowel is obligatorily followed by a consonant in its rime; a single intervocalic consonant is ambisyllabic after a short vowel (van der Hulst 1985).
- Superheavy syllables (sh) have rimes with (i) long vowels plus at least one consonant (VVC), and (ii) short vowels plus at least two consonants (VCC). We extend this category to (iii) diphthongs (ViVj) or loan vowels such as /E:/ and /O:/ (migraine, zone) followed by zero or more consonants, since these cases, too, are almost invariably stressed. Superheavy syllables with at least one consonant occur word finally only (Trommelen 1983).

The motivation of this typology can be found in e.g. stress assignment: superlight syllables never bear main stress, whereas superheavy syllables in final position generally do. The regular position of main stress in words ending in light or heavy syllables depends on the preceding syllable type; the distinction between light and heavy syllables is motivated by the fact that antepenultimate stress is possible if the penultimate syllable is light, but not if it is heavy. The following examples illustrate the rules of (1) and the distinctions proposed thus far.

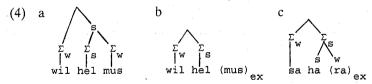
The syllable inventory has been used to further analyze two corpora: one is a collection of words (van der Hulst and Langeweg 1984), the H&L-list henceforth. It contains ca. 4,000 polysyllabic words from van Dale's dictionary, which are morphologically simplex or have an opaque morphological structure. The second is the RUL-list, a morpheme lexicon that was originally developed at Leyden University (RUL) for automatic text-to-speech conversion (Heemskerk and van Heuven 1992). This list contains ca. 7,000 polysyllabic words and suffixes. It specifies information on, among other things, deep phonology, syllable boundaries and stress position, on the basis of which classification could take place automatically, cf. the examples in (3) (* represents main stress).

(3) Example analyses of morphemes in the RUL-list

syllable types	orthography	phonemic transcription			
[h,l]	mama	[mA-m*a]			
[h,sl,l,l]	ballerina	[bA-l@-r*i-na]			
[l,sh,l,l]	hydraulica	[hi-dr*W-li-ka]			
[h,sl,sh,l,l]	banderillero	[bAn-d@-ril-j*e-ro]			
[1,1,1]	-atica	[*a-ti-ka]			

Observe that syllables with a short vowel are analyzed as heavy, whether followed by a consonant in the superficial syllabification or not; we assume ambisyllabicity for the onset consonant of the next syllable (see above). The items are subdivided into groups defined by syllable structure, word length (disyllabic words are separated from words containing a larger number of syllables), and stress position. The numerical outcome of analyzing both corpora is given in Table 1 (next page).

As will be clear from Table 1, theories of word stress in Dutch should be capable of accounting for the fact that stress on the penultimate is the regular case in words ending in two light or two heavy syllables. This is difficult to incorporate in the rules, since heavy syllables form a foot of their own. Three solutions have been proposed in the literature. The first, illustrated in (4a), is based on the Lexical Category Prominence Rule, the LCPR: word trees are labeled w-s only if the right node branches (cf. van der Hulst 1984). A foot consisting of one heavy syllable is non-branching, and therefore labeled 'weak'. This way, main stress will be on the heavy syllable preceding a final heavy syllable. The second solution, cf. (4b), yields the same result by extrametricality of final heavy syllables (cf. Lahiri and Koreman 1988) and a straightforward w-s-assignment of word trees. The third proposal is based on extrametricality of light and heavy syllables after foot formation has applied (cf. Kager 1989, Trommelen and Zonneveld 1989). This 'late' extrametricality brings about prefinal stress for heavy syllables as in (4b), but its effect is annihilated in light ones, since foot structure remains intact, cf. (4c).



The three proposals need different ways of handling exceptions, cf. (5). Note that not all of these are mentioned in the publications cited above, but they seem to be the 'natural' possibilities for generating exceptional patterns:

Table 1. Frequency distribution of primary stress in words of two and words of more than two syllables

	H&L-list			RUL-list			
Word type	APU	PEN	FIN	APU	PEN	FIN	
l,sl h,sl sh,sl		(a) (a) (a)	(a) (a) (a)		501 719 178	0 0 0	
s1,1 1,1 h,1 sh,1		(a) 203 211 2	(a) 55 26 0		0 265 265 35	8 97 54 3	
sl,h l,h h,h sh,h		(a) 187 145 4	(a) 135 92 3		0 178 142 14	18 132 117 11	
sl,sh l,sh h,sh sh,sh		(a) 65 69 5	(a) 434 354 24		0 66 52 13	66 582 487 49	
sl,sl l,sl h,sl sh,sl	(a) 28 1 0	(a) 358 127 39	(a) 0 0 0	21 20 3 1	0 485 267 78	0 0 0	
sl,1 l,1 h,1 sh,1	13 149 0 0	0 219 149 0	9 80 29 0	20 152 1 0	0 228 135 8	42 160 46 1	
sl,h l,h h,h sh,h	17 199 0 0	0 63 38 0	15 58 11 0	21 185 5 0	0 26 33 0	29 74 11 0	
sl,sh l,sh h,sh sh,sh	20 30 2 0	0 0 1 0	79 496 58 1	17 10 0 0	0 31 5 2	173 806 104 7	

Legend

APU, PEN, FIN: main stress on the antepenultimate, the penultimate

or the final syllable.

sl, l, h, sh: superlight, light, heavy and superheavy syllables, preceded by dots in words of more than two syllables.

(a): the H&L-list does not include such word types.

Table 2. Comparison of LCPR, early extrametricality of heavy syllables (E-ex), and late extrametricality (L-ex)

		ntepenultimate		Penultimate Stress		Final Stress			
Word type	LCPR	E-ex	L-ex	LCPR	E-ex	L-ex	LCPR	E-ex	L-ex
,sl				R	R	R	*	*	*
sl,l 1,l h,l sh,l				* R R R	* R R R	* R R R	R +br +br +br	R fF fF fF	R fF,-ex fF,-ex fF,-ex
sl,h l,h h,h sh,h				* R R R	* R R R	* R R R	R +br +br +br	R -ex -ex -ex	R -ex -ex -ex
sl,sh l,sh h,sh sh,sh				* -br -br -br	* +ex +ex +ex	* +ex +ex +ex	R R R R	R R R	R R R R
sl,sl l,sl h,sl sh,sl	R * *	R +ex *	R * *	* R R R	* R R R	* R R R	* * *	* * *	* * *
sl,1 1,1 h,1 sh,1	R fF *	+ex +ex *	R fF *	* R R R	* R R R	* R R R	+br +br +br +br	R fF fF fF	-ex fF,-ex fF,-ex fF,-ex
sl,h l,h h,h sh,h	R R *	R R *	R R *	* pF R R	* pF R R	* pF R R	+br +br +br +br	-ex -ex -ex -ex	-ex -ex -ex -ex
sl,sh l,sh h,sh sh,sh	-br -br *	+ex +ex *	+ex +ex *	* pF,-br -br -br	* pF,+ex +ex +ex	* c pF,+ex +ex +ex	R R R	R R R R	R R R R

Legend

sl, l, h, sh: superlight, light, heavy and superheavy syllables, preceded by dots in words of more than two syllables.

R, *: predicted regular classes and systematic gaps.

t ex: exceptional syllable (non-)extrametricality. t br: exceptional (non-)branchingness.

pF, fF: lexical prefinal or final foot.

- (5) a Exceptions to the LCPR-approach:
 - exceptional branchingness (+br) which feeds the LCPR;
 - exceptional non-branchingness (-br) which bleeds the LCPR.
 - b Exceptions to a theory with early extrametricality of heavy syllables:
 - heavy syllables are exceptions to extrametricality (-ex);
 - non-heavy syllables are lexically extrametrical (+ex).
 - c Exceptions to a theory with late extrametricality of light and heavy syllables:
 - light and heavy syllables are exceptions to extrametricality (-ex);
 - superheavy syllables are lexically extrametrical (+ex).

In order to concentrate on essential differences between the proposals, all three are allowed the possibility of a lexical final or prefinal foot (abbreviated fF and pF, respectively), a possibility proposed by Kager (1989) and Trommelen and Zonneveld (1989). An overview of the range of the three theories is presented in Table 2 (previous page), in which R stands for 'regular', i.e. the position generated regularly, without idiosyncratic behavior of syllables. Asterisks indicate positions predicted by the theories to be systematic gaps. Table 2 shows that it is difficult to evaluate the theories on the basis of disyllabic words, since the same predictions are made by each theory, both in terms of regular stress positions and of systematic gaps. As for words with a larger number of syllables, several points need discussion: the different predictions of regularity and systematic gaps, examples of words that should be systematically absent and examples of words predicted to occur, but small in number, differences in degree of exceptionality, and overgeneration.

2. Predicted regularity

The three theories are similar in their prediction of what is the regular case, except for words with a final light syllable preceded by a superlight one. The LCPR and late extrametricality theories claim that antepenultimate stress is regular in such words, cf. (6), whereas the theory of early extrametricality predicts final stress to be regular, cf. (7) (the RUL-list is used as a basis, the tag "H&L" is used for additional H&L-words).

(6) Antepenultimate stress in words ending in [...,sl,l]:
etcetera, cinema, kinema, cholera, camera, opera, kiekeboe, numero,
hetero, rotary, society, algebra, gerbera, selderie, kaketoe, mistletoe,
barbecue, marteko, privacy, penalty, royalty, Albeda (H&L), Alkema
(H&L), Altena (H&L), Hengelo (H&L).

(7) Final stress on words ending in [...,sl,l]:
procede, hotelier, kukeleku, tournedos, decolleté, sodeju, avenue,
maintenee, craquelé, nondeju, parvenu, entredeux, filatelie, anemie
(H&L), maroquinerie, coterie, etc. (30 words on -ie in RUL-list).

Both types exist in equal proportions, so that quantitative criteria cannot be used to decide which is the regular group. Comparison of the groups reveals that (6) is less exotic, and more compound-like (e.g. kiekeboe is clearly a compound; the endings -da, -ma and -lo typically occur in names, cf. Roorda, Scheltema, Venlo). Some of the obvious subregularities, e.g. -era, -ero and -ty in (6) and the final /e, y, i/ in (7), are no doubt related to a different etymology of the groups: English, Latin or Greek in (6) versus French or Greek in (7).

3. Systematic gaps

In a theory with regular early extrametricality of heavy syllables, the notion of extrametricality is defined in its original sense: the extrametrical syllable is a separate domain to the rules of foot formation. Since heavy syllables are not allowed to be weak nodes of feet, the foot structure of preceding syllables is unaffected by early extrametricality of heavy syllables. In this regard, the theory is similar to the LCPR-theory (which has no extrametricality at all) and the theory which uses extrametricality after foot formation has applied (late extrametricality). However, when light or superlight syllables are extrametrical, early extrametricality will affect the foot structure of preceding syllables. Therefore, unlike both alternative theories, the theory of early extrametricality predicts the possibility of antepenultimate stress for words ending in a light syllable followed by a superlight syllable. The words attested are presented in (8):

- (8) Antepenultimate stress in words ending in [...,l,sl]:²
 - a metropolis, trivialiter, syfilis, notulen, andoren, Nijmegen (H&L);
 - b weduwe, Veluwe, Betuwe (H&L);
 - c Brazilië (plus 22 additional names of countries on -ië in the H&L-list), chemicaliën, genitaliën, saturnaliën, neuriën, Scandinaviër, Australiër, sauriër, carrier, terriër, Helvetiër, Serviër, cheviot.

According to the RUL-list, some words on -is are pronounced with a schwa (metropolis and syfilis), but others with /I/ (dualis, cf. (16) below). Here, the database may be inconsistent. Most words on -en in (8) can be considered as inflected forms. This reduces the number of words, but has no bearing on the issue discussed.

Presumably, words of type (8b) and (8c) need a different analysis (cf. Trommelen 1983, van der Hulst 1984, Kager and Zonneveld 1986). We assume that word internal syllables with a high vowel directly in front of the vowel of the next syllable are stress rejecting (the exceptions are small in number, cf. messías, María).

Next, observe the mismatches between theory and corpus, i.e., words predicted not to occur and predicted words that occur in small numbers only. First consider the set of words that should not occur at all:

- (9) Antepenultimate stress in:
 - a [...,h,sl]: umpire, terminal, establishment;
 - b [...,sh,sl]: landauer;
 - c [...,h,l]: graffiti;
 - d [...,h,h]: talisman, rombombom, sanhedrin, triathlon, badminton.

Some words are rare (e.g. sanhedrin, /sAn-hE-drIn/), and the number of such words is small, although more examples than those enumerated in (9) (viz. names) are known from the literature. Moreover, the phonemic transcription of some of these words may be questioned (/t*Ur-mI-n@l/ and /gr*E-fI-ti/ may rather be analyzed as /t*Ur-mi-n@l/ and /grE-f*I-ti/) and some words could receive their stress pattern by being analyzed as compounds. The existence of (9) therefore cannot be used as an argument for rejecting any of the three theories.

The set of rare but predicted words is the following:

(10) Antepenultimate stress in [...,l,sh] (31 irregular forms, 13 of which are proper names):

(in RUL:) tomahawk, samovaar, samowaar,³ deficit, olifant, horizont, crucifix, katapult, mocassin, bungalow, uniform, (additional words in H&L:) ablatief, locatief, vocatief, hospitaal, Casimir, Elzevier, Olivier, Oedipus, heliport, leukoplast, Beatrix, Biotex, Conimex, Dulcolax, Duralex, Isolex, Moulinex, Odorex, tubifex, Beatriis.

(11) Penultimate stress in:

a [...,sh,l] (8 regular forms): cigarillo, tequila, resumptie, resorptie, secundo, guerilla, absorptie, andijvie;

³ Orthographic variants are deleted from the RUL-list, unless they correspond with a difference in pronunciation, as seems to be the case for samovaar and samowaar.

- b [...,h,sh] (5 irregular forms): appendix, cyrillisch, lucullisch, Kretenzisch, Kaapverdisch;
- c [...,sh,sh] (2 irregular forms): caoutchouc, hydraulisch.

(12) Final stress in:

- a [...,sh,l] (1 irregular form): impromptu;
- b [...,h,h] (13 irregular forms): violoncel, stewardess, monoftong, apostil, medaillon, carillon, franskiljon, bataljon, postiljon, compagnon, castagnet, bombardon (H&L), tarantel (H&L);
- c [...sh,sh] (8 regular forms): hovaardij, polyptiek, liaison, asymptoot, augustijn, artistiek, peremptoir, conjunctuur (H&L).

In Dutch, word internal rimes are superheavy by exception only, which explains the small numbers in (11a/12a) and (11c/12c). The small numbers elsewhere indicate that, perhaps, the theories proposed are too permissive, and that all such cases, as those in (9), should be considered systematically absent. In most cases the existence of these classes of words relates to their exceptional phonology or morphology: frozen compounds or blends such as uniform and heliport in (10), -x and -tief in (10), -x and -isch in (11b), and -on in (12b).

4. Degrees of exceptionality

As is shown by Table 2, each of the three stress placement accounts for Dutch has several exception features at its disposal. It seems reasonable to assume that words whose stress pattern can be generated by an appeal to a single exception feature are in some meaningful sense less irregular than words whose stress placement involves multiple exception features (cf. Kager 1989). We shall test this assumption against our frequency data on stress patterns. Most irregular classes of words are accounted for by one idiosyncracy. All theories need two exception features in words such as (13) with penultimate stress on a light syllable followed by a superheavy syllable:

(13) Penultimate stress in [...,l,sh]: tragikomisch, macaronisch, hypothetisch, etc. (28 words on -isch); exequatur, approbatur, imprimatur.

These words receive a prefinal foot lexically, and an extrametrical or non-branching final foot. Less exceptional is, according to the theories, antepenultimate stress in such words, cf. group (10) discussed above. Both (13) and (10) show phonological and morphological subregularities, however, which suggests that their existence may be explained differently and that the theories wrongly predict such word types.

Another instance of degree of exceptionality is predicted by late extrametricality for words ending in two light syllables (see Kager 1989 for an extensive discussion of this). The theory predicts (14) to be less exceptional than (15), a prediction borne out by our frequency data.

(14) Antepenultimate stress in [...,l,l]:

- a quadragesima, duodecimo, majolica, alinea, apocope, paprika, dominee, aloe, piccolo, risico, etc. (87 in RUL-list);
- b cafetaria, lobelia, continuo, papoea, legio, audio (62 words with a prefinal high vowel immediately followed by a vowel in RUL-list; these forms are probably not [...,l,l] but [...,sl,l]).

(15) Final stress in [...,l,l]:

- a onomatopee, ericacee, theodicee, introducé, debouché, farmacopee, communiqué, individu, ambigu, continu, chocola, lavabo, rococo, kariboe, acajou, etc. (47 in RUL-list);
- b chromolithografie, oleografie, ideologie, analogie, charivari, etc. (127 -ie-forms).

Moreover, (14) seems to be the more usual pattern than (15).

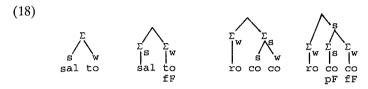
The final instance of degree of exceptionality can be found in words ending in a light syllable followed by a heavy syllable. If the use of a prefinal lexical foot is considered more exceptional (because it allows irregularity to occur word internally) than the use of exceptional branchingness or exceptional non-extrametricality, the words in (16) should be more exceptional than the words in (17):

- (16) Penultimate stress in [...,l,h]: carbolineum, conopeum, protozoön, mausoleum, nasigoreng, ultimatum, alligator, gradatim, privatim, senator, equator, ricinus, papyrus, desideratum, vademecum, leviathan, gladiator, atheneum, dualis, coadjutor, mercator, spectator, mecenas, spondeus, trocheus, etc. (26 RUL-forms and 63 H&L-forms).
- (17) Final stress in [...,l,h]: stafylokok, kameleon, macadam, toreador, samoerai, parasol, bajonet, fontanel, paralellogram, elektricien, etc. (74 RUL-forms and 58 H&L-forms).

Here the difference between the two categories seems related to whether they are early loans (16) or more recent ones (17). In general, contrary to the predictions by the theories, the early loans seem less irregular.

5. Overgeneration

On the basis of comparison of (3) and (5) one may conclude that the theories generate all occurring word types and only these. These surveys, however, do not uncover the possibility of assigning different analyses to non-ambiguous words. Such cases exist, for instance in words ending in two light syllables, cf. (18):



To predict which structure is correct, two strategies are available. First, an additional device can be invented such that analyses without idiosyncratic features are favored over those that include such features. Second, the notion of idiosyncratically marking non-final syllables could be excluded in principle. Other ways of generating words such as (13) and (16) then need to be found. Consonant extrametricality is a good alternative: it is more restrictive than prefinal feet in that it does not allow the variation of (18), and predicts the availability of penultimate stress to words such as (13) with a final VVC-syllable preceded by a light one, but not to words with VCC in the final syllable.

6. Discussion

In this paper we compared three theories, for this purpose described uniformly in terms of currently accepted general features, on the basis of two corpora: the H&L-list and the RUL-list. This comparison shows the three variants to be empirically equivalent to a large extent. The conclusion must be that only theoretical considerations can decide which theory is to be preferred. Extrametricality is to be preferred over the LCPR, since additional evidence for the LCPR is hard to find, whereas additional evidence for extrametricality can be found in other types of extrametricality such as consonant extrametricality. Moreover, use of the LCPR excludes extrametricality, since a combination of both would predict there to be words with stress on the pre-antepenultimate syllable. Such words are included in the lists, cf. (19), but are rightly excluded theoretically, since alternative analyses are available:

- (19) Words with stress at the preantepenultimate syllable:
 - a linguistic terms on -tief: vocativus, accusatief, diminutief, adverbatief (H&L), imperatief (H&L), indicatief (H&L), infinitief (H&L), nominatief (H&L);
 - b frozen compounds (?): kronometer, monoxide, synthesizer, disagio, hemicyclus, oversized, horticultuur;
 - c complex proper names (H&L): Amerongen, Kemenade, Scheveningen.

The survey of data above, in which primarily the exceptional classes were discussed, shows that the three proposals are too permissive. For most of the sets of words discussed, obvious subregularities hold, based on phonological or morphological characteristics. Many of these subregularities have been noted in the literature, but have not led to a more restrictive formulation of rules and exception devices, such that some classes are excluded in principle (as is the class exemplified in (19)).

At several points in our survey, the opposition of early versus recent loans or degree of naturalness was mentioned, cf. (6)-(7), (14)-(15) and (16)-(17). Such parallels are accounted for in a non-uniform way by the three proposals reviewed here, and progress is to be expected in this area, perhaps related to above-mentioned different approach of exceptions.

References

Heemskerk, J.S.M. and V.J. van Heuven (1992) 'MORPA, a morpheme lexicon based morphological parser', in V.J. van Heuven and L.C.W. Pols, eds., *Analysis and synthesis of speech, strategic research toward high-quality text-to-speech generation*, Mouton De Gruyter, Berlin (in press)

Hulst, H.G. van der, and S.J. Langeweg (1984) Nederlandse klemtoon; ongelede woorden, INL Working Papers.

Hulst, H.G. van der (1984) Syllable Structure and Stress in Dutch, Foris, Dordrecht.

Hulst, H. van der (1985) 'Ambisyllabicity in Dutch', in H. Bennis and F. Beukema, eds., Linguistics in the Netherlands, Foris, Dordrecht, 57-66.

Kager, R.W.J. (1989) A Metrical Theory of Stress and Destressing in English and Dutch, Foris, Dordrecht.

Kager, R. and W. Zonneveld (1986) 'Schwa, Syllables, and Extrametricality in Dutch', The Linguistic Review 5, 197-221.

Langeweg, S.J. (1988) The stress system of Dutch, PhD dissertation, Leyden University.

Lahiri, A. and J. Koreman (1988) 'Syllable weight and quantity in Dutch', Proceedings of the 7th West Coast Conference on Formal Linguistics, Stanford Linguistic Association.

Neijt, A. and W. Zonneveld (1982) 'Metrische fonologie: de representatie van klemtoon in Nederlandse monomorfematische woorden', De Nieuwe Taalgids 75, 527-547.

Trommelen, M.T.G. (1983) The Syllable in Dutch, Foris, Dordrecht.

Trommelen, M. and W. Zonneveld (1989) Klemtoon en metrische fonologie, Coutinho, Muiderberg.