# Where is the Dutch Stress System?

Björn Köhnlein and Marc van Oostendorp

June 24, 2016

## 1 Introduction

One of the problems in work on the typology of stress is that the data are so difficult to pinpoint even for an individual language. Can we rely on the ears of native speakers? On those of fieldworkers? Do we need some kind of instrumental validation (and if yes, what should we measure)? Can we only determine the stress pattern after some analysis (and if yes, what analytical technique should we use)? How do we deal with exceptions? And how do we decide that some facts are exceptions in the first place?

One guideline in this might be to first decide what exactly it is we decide to find. Are we interested in generalizations about the stress patterns in corpora? Or are we rather interested in the knowledge of stress in individual speakers? And in the latter case: how do we decide that such speakers do have certain knowledge?

One might surmise that the problem at least is easily solved for a language like Dutch, for which we have a respectable literature (Trommelen and Zonneveld, 1989; Kager, 1989; van der Hulst, 1984; van Oostendorp, 1997; Linke, 2016; Gussenhoven, 2009; Langeweg, 1988). It is our goal in this paper to show that even for such a language, this evidence is debatable. As such, this paper should be seen as a sequel to van Oostendorp (2012). That paper makes the rather radical claim that there is little evidence for Dutch stress to be either quantity-sensitive or subject to a three-syllable window, as is usually assumed in the literature. van Oostendorp claims instead that these properties are diachronic consequences of the fact that the relevant loanwords which are considered to be evidence are loanwords, mostly from Romance languages which *do* have these properties. Such evidence — static distribution of lexical items in the lexicon — cannot be seen as an indication of the knowledge of the system by an individual speaker. If it is the goal of

linguistic theory to characterize the latter, we should rather look for other means to get this information.

There are many problems with taking actual words as the basis for our study of knowledge. The set of words in a given language is influenced by many factors which might not all be part of the speakers' "knowledge", but rather the result of contingencies in the history of the language involved. We know from the study of (mostly) phonotactics that not every statistic generalisation about the actual lexicon is also reflected in the behaviour of speakers.

But although van Oostendorp thus ends with a call for experimentation, the paper is mainly intended as an overview article of the existing literature, hence it does not provide experimental evidence itself. In recent years, however, we have been able to put forward new evidence, and test the predictions of existing theories against new data, from two on-line experiments: one on biblical names (1,774 participants, 2,000 names; conducted 2012), and one on nonsense words (1,631 participants, 461 words; conducted 2013-14).

In both cases, particpants were shown a list of written words — a selection of all items — and asked to mark which syllable they would stress. The first list is a representative sample of all names in the Bible (which includes about 4,000) names, except those that are commonly known (*Adam, Moses*, etc.) It turned out that most of these names do indeed behave like nonsense words; they are unknown even to many orthodox Christians, because they are mostly hapaxes, appearing in lists of names. (Some informants thought we tried to fool them, because they did not recognize most names in spite of the fact that they assumed to have a thorough knowledge of the Bible.)

At the same time, the participants were aware of the fact that they were doing a test on biblical names, and furthermore the names were recognizable as of Greek, Aramaic, Hebrew, Latin (etc.) origin. For this reason we also did an experiment with carefully designed nonsense words in which all the relevant kinds of words were present. In particular, we had 3- and 4-syllabic words with all possible combinations of tense and lax vowels and diphthongs in all syllables; since lax vowels in Dutch only occur in closed syllables, this means also all combinations of syllable structure were taken into account. (The difference between tense and lax vowels is reflected also in standard Dutch orthography; typically, tense vowels are represented as long.)

In this contribution, we check these data against the main generalisations from the literature on Dutch stress. We take these to be the following:

- Dutch has a three-sylable window at the end of the word.
- Default stress is on the penultimate syllable, modulo the fact that:
- Stress is (partly) quantity-sensitive.
- Dutch stress is partly lexically specified, although the 'functional load' of the contrast is not very large (there is at most a handful of minimal pairs).

In the following sections, we briefly discuss the results for each of these. We will show that speakers of Dutch do indeed seem to have knowledge of at least some of these generalisations. In the last section we then turn to the intriguing question *why* people have such knowledge and what we learn about human language from this observation.

## 2 Three-syllable window

Usually, Dutch stress is supposed to be subjected to a three-syllable window at the end of the word: stress is on one of the last three syllables, and never on antepenultimate syllables. The evidence for this is that words like *[mákaroni] are absent from Dutch; Gussenhoven 2014, for instance, claims that it is impossible to have exceptions to the relevant constraint.

Yet there are several problems with this assumption. First, the only evidence we have is 'static' evidence. Morphologically caused alternations, such as we have in Modern Greek, are missing. In the latter language, *góndola* 'gondola' has stress on the antepenultimate syllable. If we add a proclitic, as in *gondolá-mu* 'my gondola', the stress shifts because it would otherwise fall outside of the window. Such kinds of 'dynamic' alternations are missing in Dutch.

Static evidence comes from loanwords, from mostly Indo-European languages which themselves respect the three-syllable window; and since the position of stress is very often borrowed faithfully, we do not have to see the three-syllable window as active in the case of borrowing either. Furthermore, as we will see below, Dutch *does* show a preference for antepenultimate stress: in some cases, stress shifts to that position. However, this is not evidence for a three-syllable window either, as such shifts also occur when the original stress already is within the window.

Finally, there are at least two known classes of exceptions, viz. a class of grammatical terms (*infinitief, accusatief, nominatief,...*), and secondly a number of place names *Scheveningen, Wageningen, ...* (Note however that Köhnlein (2015) provides a more detailed analysis of stress patterns in place names, where it is argued that many of them are morphologically complex.) All of these are words of 4 syllables with stress on the first syllable. Furthermore, CELEX, provides us with some more indivual examples *reutemeteut, tingeltangel, erlenmeyer, ammelaken*, although it should be noted that the first two words have some kind of reduplicative structure and the latter look like personal names and like compounds.

Experimental evidence also do not really point in the direction of a three-syllable window. In our own experiment, the following is the distribution of stress assigned to four-syllable words by our participants:
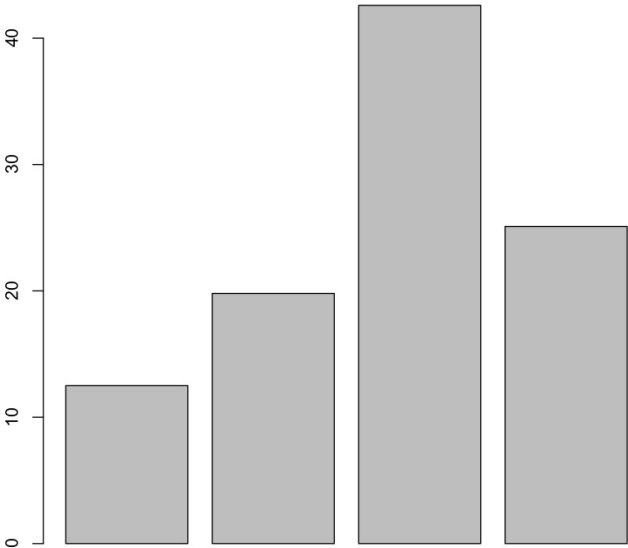
(1)

There is clearly less stress on the preantepenultimate syllable than on any other syllable, yet in 10% of the cases, participants chose antepenultimate stress when confronted with a word of four syllables.

An even more dramatic result was found by Nouveau (1994) in a study where three-year old children where asked to repeat a nonsense word (presented to them as the name of an animal) and error rates were recorded:

(2)

|           | PAPU | APU | PU | U  |
|-----------|------|-----|----|----|
| karabilo  | 65   | 30  | 20 | 65 |
| monitaron | 50   | 35  | 45 | 55 |

Clearly, children made more errors when stress was put on the preantepenultimate syllable than when it was put on some other syllable — except, interestingly, the ultimate. Typically, an error would be that stress was shifted on some other syllable. However, the error rate was far from being 100%. We may conclude that preantepenultimate stress is (heavily) marked, but not that it is absent.

One problem in our methodology, as well as that of Nouveau (1994) is that there is always the option option of (fake) compounds: maybe the participants did not analyse the word as monomorphemic, but as morphologically complex. Here is for instance the word where the largest number of participants (29.8%) chose stress on the initial syllable:

(3) *zeiboltiroon*

It is possible that forms such as these are analysed as being compounded of two nonsense words, such as *zeibol+tiroon*, with compound stress on the first element, as is usual in Dutch (nominal, prosodic) compounds (Trommelen and Zonneveld, 1989; Booij, 1999). This however makes the whole idea of a window virtually untestable, at least using the methodology of nonsense words: if speakers assume that there monomorphemic words with more than three syllables are absent, or very rare, and tend to assign a compound analysis to such words, why would they posit some constraint on stress placement in such words?

We may thus conclude that there is no evidence for a three-syllable window in Dutch beyond the fact that many Dutch words happen not to have stress on the PAPU or an even earlier syllable. Note that final stress is almost as marked as PAPU stress.

## 3    Penultimate default

The following generalisation we turn to is that of the penultimate default. 'Default' stress in Dutch is supposed to be on the penultimate syllable by every contemporary author I am aware of. Furthermore, antepenultimate stress is supposed to be more marked in most types of words, and ultimate stress by most authors is considered more marked still. The typical evidence for this comes from lexical distribution (most words have stress on penultimate stress, and much fewer on the antepenultimate or ultimate syllable), but also from regularizations in everyday speech or even in language change (*normáliter* 'normally' is sometimes pronounced as *normalíter*, and mistakes do not happen the other way around).

Our experiment unequivocally confirms this generalisation. For instance, we find the following for words ending in three light syllables:
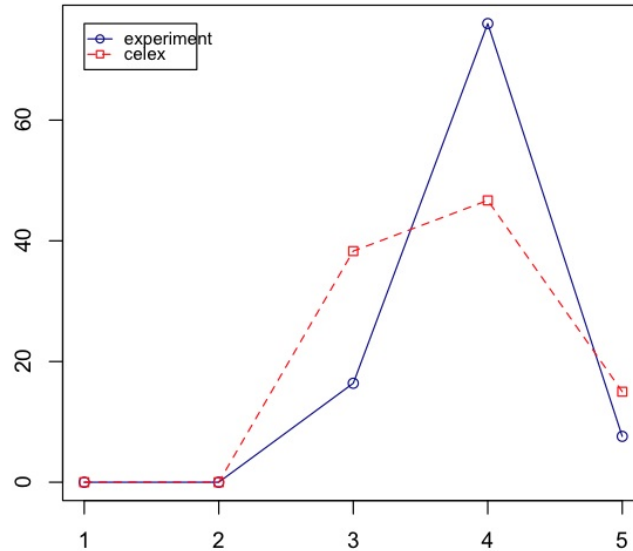
(4)

The blue line gives stress in the experiment; the red line stress on trisyllabic words with only light syllables in the CELEX database. For maximal comparability with longer words, 3 denotes the antepenultimate syllable, 4 the penultimate and 5 the last syllable of the word.

Clearly, there is a very strong preference in the experiment with nonsense words to choose the penultimate syllabe. The preference is visible in the existing wordbase in CELEX, although it is much less clear: in particular, the many words (word types) with antepenultimate stress do not translate directly into speaker judgements.

## 4   Weight

We now turn to the issue of syllable weight. Following the literature, we probably have to make (at least) a four way distinction of syllable type:

(5)   a.  Schwa
      b.  Light (open, tense vowel)
      c.  Heavy (lax vowel and one conosonant or diphthong)
      d.  Superheavy (tense vowel plus one consonant, lax vowel plus two consonants)

The sensitivity to schwa goes uncontested. First, syllables with word stress never have schwa[1]

---

[1]One old point of contention here, though, is based on the observation that schwa and [œ] are in complementary distribution, since the latter never occurs in unstressed position.

But this is not the only way in which schwa interacts with stress. Another generalisation is that stress immediately precedes the schwa syllable, at least when the latter is final (van der Hulst, 1984). In other words, the tendency towards penultimate stress would be even stronger in this case. On the other hand, this generalisation is not without exceptions. Those in (6a) are due to van der Hulst (1984); those in (6b) are from Celex.

(6)   a.   *Manchester*, *armoede* 'poverty', *Nijmegen*, *weduwe* 'widow', *oorkonde* 'certificate', *Betuwe*, *Veluwe*

   b.   *maluwe* 'malva', *clitoris* 'id.', *syfilis* 'word', *cannabis* 'id.', *sauriër* 'id.'

None of these examples are unproblematic. *Manchester*, *Veluwe*, *Betuwe* and *Nijmegen* are place names (the former is apparently also used for a type of textile, but we are not sure that we have every encountered it in that meaning); the words *oorkonde, armoede* and *sauriër* look like suffixed words (ending in *-de* and *-er* respectively); the words *Betuwe, Veluwe, weduwe, maluwe* all end in *-uwe*, where the *u* itself could be analysed as an underlying schwa. And as for the words in *-is*, I am not sure why CELEX assumes that the vowel in these cases is a schwa rather than an [ɪ].

There thus do not seem to be any unproblematic counterexamples to the claim that schwa is 'prestressing'. We have a third, even stronger piece of evidence for the relation between schwa and stress: (optional) reduction of unstressed vowels. Unlike in English, unstressed vowels can be unreduced; but in informal registers, reduction is at least possible, and in many cases also required. Reduction therefore seems a 'real' fact of the phonology of the language.

Given the methodology of our experiment, we could not test this particular property of Dutch further, as Dutch orthography misses a special symbol for schwa: it is usually spelled with *e*, but so are [e] and [ɛ], and we do not know which phonetic interpretations informants use when they see a particular letter.

Given the virtually uncontested nature of the effects of schwa, searching for experimental evidence also does not seem to be all that urgent. That is not the case, however, for the other end of the weight scale, the so-called superheavy syllables. The claim here is that final superheavy syllables always attract stress.

---

In a structuralist analysis, this may lead one to assume that schwa and [œ] are allophones of the same phoneme. Opponents of this idea have pointed to the pair *katterig* 'headachy' and *katterug* 'back of a cat', but these words have a different morphological structure. Both are derived from the noun *kat* 'cat', the former by affixation, and the second by compounding with *rug* 'back'. It is therefore possible to claim that the second form, but not the first, has compound stress.

An alternative claim would be that [e] and schwa are in complementary distribution, as the former shows a very strong tendency to reduce when it occurs in unstressed position, but this has never been worked out, as far as we know.

There is a list of exceptions (*kroepoek* 'krupuk, Indonesian prawn cracker', *lichaam* 'body', *arbeid* 'labour', *vampier* 'vampyre', *nonsens* 'nonsense', *ambacht* 'craft', *asbest* 'asbestos'. These are all bisyllabic word with stress on the first syllable although the second syllable is superheavy. It has been proposed to analyse these as 'prosodic compounds' consisting of two cranberry morphemes, but one can of course wonder whether that doesn't make any claim about the weight of superheavy syllables unfalsifiable.

In our experiments, it turns out that the tendency to stress the final syllable is seen mostly in the way it counterbalances the preference for a trochee. If the last syllable is light, there is a very strong preference for stress on the penultimate (as we already know), but this preference disappears if the final syllable is not light. The distribution then becomes 50-50. Yet it is not clear that superheavy syllables have a much stronger effect than heavies. The following data are from bisyllabic words in our corpus:

(7)

|       | Syllable    | Stress |
|-------|-------------|--------|
| X-SH  | ultimate    | 45.4   |
|       | penultimate | 54.6   |
| X-H   | ultimate    | 42.5   |
|       | penultimate | 57.5   |
| X-A   | ultimate    | 5.9    |
|       | penultimate | 94.1   |

(bisyllabic words; SH=superheavyy, A=light, H=heavy)

Actually, in this case the effect is much stronger if we look at existing (trisyllabic) words in CELEX.
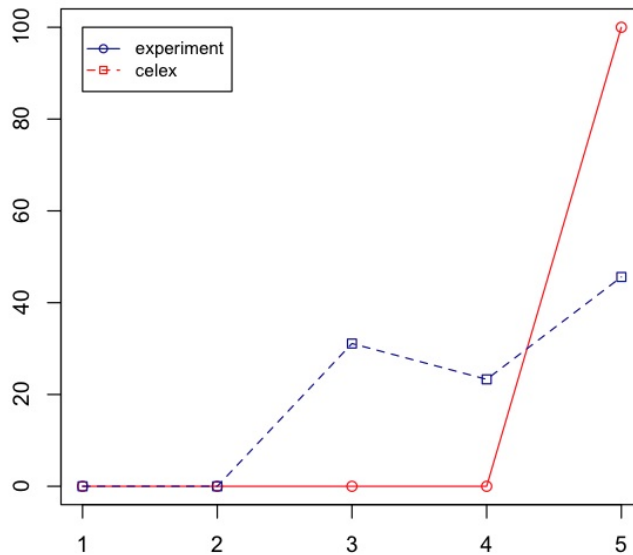
(8)



(Words of type heavy-light-superheavy; 19 examples in CELEX)

We thus found little evidence for a difference between heavy and superheavy syllables; but the results above do seem to indicate that light syllables are different than heavies or superheavies. In stress-based analyses of Dutch prosody, this difference has played an important role. Analyses differ in various ways but it seems reasonable to summarise that we have count as heavy all closed syllables or diphtongs, and as light all open syllables, and furthermore that the following holds:

(9)  When the penultimate syllable is heavy, stress cannot be on the ante-penultimate.

In Celex we find the following exceptions to this generalization:

(10)  'VC. VC. VC          *mangistan, badminton*
      'VC. VC. VV          *chimpanzee*
      'VV.VC.VV            *shoarma*
      (VC).'VV.VC.VC   *sanniyasin*

Note: The latter two examples seem rather questionable; we have not found people that share the stress judgements on these examples. That in turn implies that if we take the three counterexamples seriously, the relevant generalisation should also refer to the weight of the antepenultimate syllable. There have been no relevant proposals which can deal with this; but on the other hand, the number of these counterexamples is of course vanishingly small.

We can test more weight-related generalizations from the literature. Here is a list:

(11)  a. **Main generalization**: Only the weight of the last two syllables count.

b. If the penult is closed, it will be stressed

c. If the ultimate is closed and the penult is open, the antepenultimate will have stress

d. If the ultimate is open and the penult is open, the penult will have stress

We have already observed above that what we called the 'main generalization' is not straightforwardly confirmed; the structure of the antepenultimate syllable seems relevant:
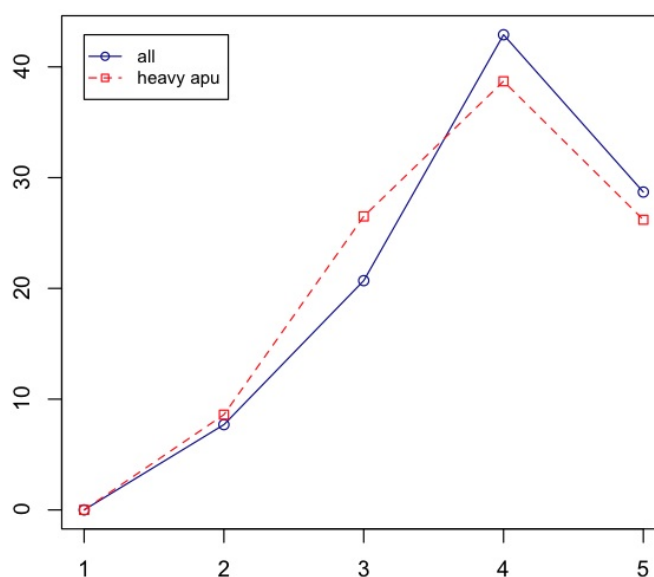
(12)



Weight effects thus are *not* confined to the last two syllables. Interestingly, we do on the other hand *not* find any evidence for the relevance of weight on the preantepenultimate syllable:
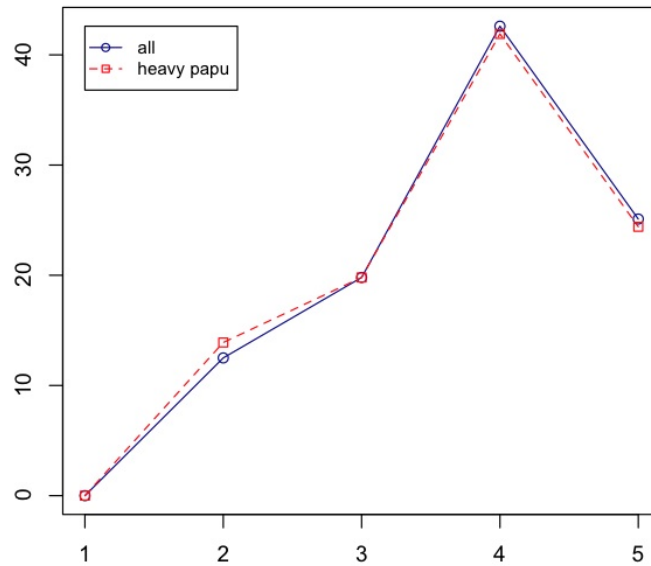
(13)

There is no real difference between the red line (words with a heavy prean-
tepenultimate) and the blue line (words where this syllable is light). There
actually is thus some indication for a kind of three-syllable window effect in
the sense that only the weight of the last three syllables are computed.
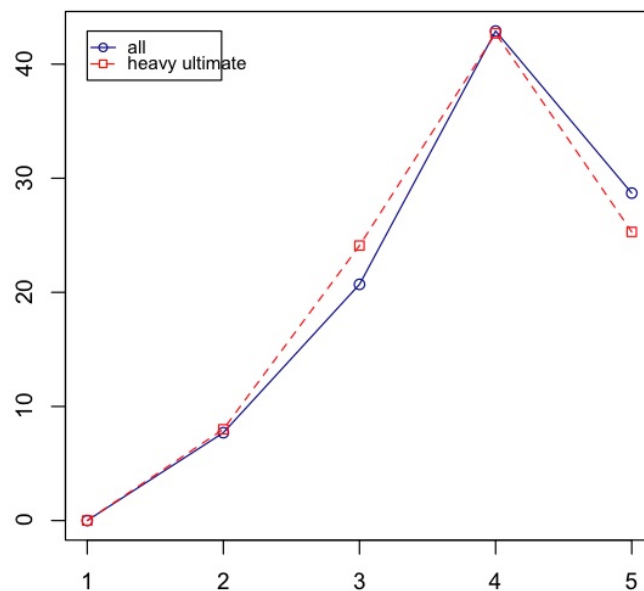
Let us now turn to the three other generalisations:

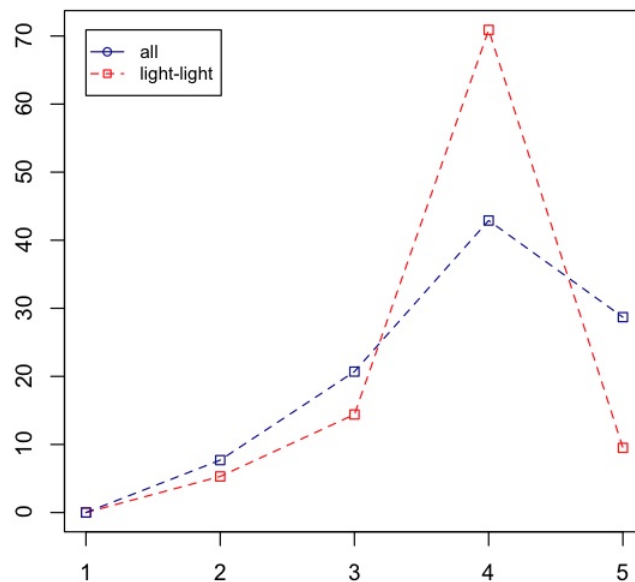(14)   If penult is closed, it will be stressed
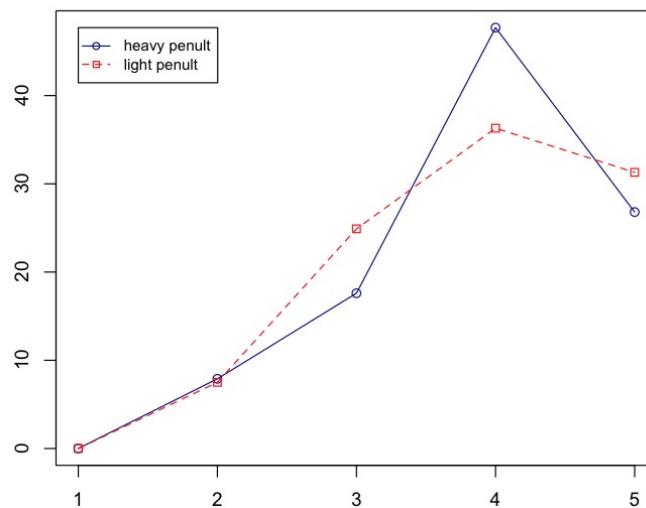


This seems to hold (as a tendency).

(15)   If the ultimate is closed and the penult is open, the antepenultimate
       will have stress

This seems to be a weak tendency at best: there is just a little bit more stress in these cases.

(16) If the ultimate is open and the penult is open, the penult will have stress



This is a rather strong effect, as we have actually already seen. This is the

domain where the penultimate default can play its role, not disturbed in any way by any kind of weight consideration.

## 5 Discussion

The data from the experiments thus seem to show that speakers have access to the penultimate default and (more subtly) to weight considerations, although for the latter our data do not seem to fit any particular theory extremely well. There is no strong evidence for the three syllable window or for some of the more precise predictions of stress-theories. However, none of the factors on their own seem to make absolute predictions as to where stress goes. Furthermore, some of the more precise predictions have not be confirmed, such as that stress would be on superheavy syllables, or can only fall on one of the last three syllables of the underived word.

In any case, van Oostendorp (2012)'s challenge seems to have been met: there is evidence for speaker's knowledge for some regularity in Dutch stress — the system is not purely lexical. The important question now is: what kind of knowledge are we actually talking about? What is this knowledge that speakers have? And what is its purpose? Why do speakers possess this kind of model.

In unpublished work, Cei and Hayes (2012) look at data from the (somewhat similar) system of Italian. They propose a system of stochastic OT which models the data fairly well: a set of constraints with some probability attached to them gives us the kind of variability we see in our data as well. The question now is: why would people possess this type of knowledge, stochastic or otherwise?

Grammatical knowledge typically helps the learner create new forms; but that does not apply here: our experiment creates a very exceptional situation, viz. one of nonsense words which are only seen and not spoken. This is the kind of situation in which one would be forced to 'guess' the stress of a word, but there is little reason to assume that that is a very natural kind of situation. We could assume that we are learning how to distribute stress over unknown forms, but that just does not seem to be a very useful linguistic skill to have.

A classical OT answer to the question why we have such knowledge is that this is an Emergence of the Unmarked (TETU) Effect: all constraints are available in all grammars, so also in grammars where they are masked by faithfulness and therefore do not really show their face. In other grammars they may be useful  if stress is fully predictable we do not have to store it, or it can function as a parsing cue , and the fact that they show up also in this case is just a side effect. Maybe some of the variability can then be attributed to the fact that there is little evidence to set the rankings properly  they have to be ranked in some way, but since there is little evidence as to their precise ranking, speakers may choose different solutions for this.

However, such an account is incompatible with the general approach of Cei and Hayes (2012), which is based on the assumption that constraints are not universal, but made up on the go. Given the limited usefulness of such constraints under language use, the question then stays why such constraints are introduced by the learner at all. We actually see the results presented here as well as those reported by Cei and Hayes (2012) as problematic for any kind of emergentist approach: it is not clear why such knowledge would 'emerge' if it does not really do much. Under any kind of nativist approach — be it OT or a parameter account — the fact that people know these things is easier to understand: it may be of no use to them, but it is for people with slightly different rankings or parameter settings.

One could also entertain the possibility that maybe this somewhat helps the listener parse the incoming string, but given the variability of stress position, stress is not a very good boundary marker in Dutch (especially given that words which are longer than three syllables seem quite exceptional). Also, the interaction with syllable weight does not really help: the fact e.g. that stress tends to be more on the antepenultimate syllable when the ultimate is heavy would help parsing only in a complicated way.

Given the fact that these choices for buiding constraints are stochastic, the question also arises, why are there absolutely no lexical words in Dutch (or Italian) with no stress? Or with more than one main (non-compound) main stress? Gussenhoven (2014) for instance says that the constraint 'that ensures that all words have main stress' is undominated, and that he does not consider it further; but he does not go into the interesting observation that this constraint seems 'undominated' in all known languages with stress. On the other hand, there are languages in which lexical stress does not play a role, so where the constraint is at the bottom of the hierarchy. Why is it never in an intermediary position? It looks as if stress is a real 'macroparameter' which can be switched on or off; it is not 'violable constraint'. Again, we think that this is more of an indication for some form of nativism than for emergentism.

## Bibliography

Booij, G., 1999. Morpheme structure constraints and the phonotactics of dutch. In: Hulst, H. v., Ritter, N. (Eds.), The Syllable. Views and Facts. Mouton de Gruyter, pp. 53–68.

Cei, E., Hayes, B., 2012. Multiple factors affecting italian word stress. Handout March 7, 2012.

Gussenhoven, C., 2009. Vowel duration, syllable quantity and stress in dutch. In: The nature of the word. Essays in honor of Paul Kiparsky. MIT Press, pp. 181–198.

Gussenhoven, C., 2014. Possible and impossible exceptions in dutch word stress. In: Hulst, H. v. (Ed.), Word Stress. Theoretical and Typological Issues. Cambridge University Press, pp. 276–296.

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

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

Köhnlein, B., 2015. The morphological structure of complex place names: the case of dutch. The Journal of Comparative Germanic Linguistics 18, 183–212.

Langeweg, S., 1988. The stress system of Dutch. Ph.D. thesis, University of Leiden.

Linke, K., 2016. Word stress . URL `http://www.taalportaal.org/taalportaal/topic/pid/topic-14020545852112030`, retrieved February 04, 2016 from http://www.taalportaal.org/taalportaal/topic/pid/topic-14020545852112030.

Nouveau, D., 1994. Language Acquisition, Metrical theory, and Optimality: A Case Study of Dutch Word Stress. Ph.D. thesis, University of Utrecht.

van Oostendorp, M., 1997. Lexicale variatie in optimaliteitstheorie. Nederlandse Taalkunde 2, 133–154.

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

van Oostendorp, M., 2012. Quantity and the three-syllable window in dutch word stress. Language and Linguistics Compass 6, 343–358.