Word stress assignment in German, English and Dutch: Quantity-sensitivity and extrametricality revisited

Ulrike Domahs*[a], Ingo Plag[b], and Rebecca Carroll[c]

[a] *Institute of Germanic Linguistics, University of Marburg, Germany*

[b] *English Language and Linguistics, University of Duesseldorf, Germany*

[c] *Institute of Physics, Medical Physics Group, University of Oldenburg*

*Corresponding author:

Ulrike Domahs

Institut für Germanistische Sprachwissenschaft

Philipps-Universität Marburg

Wilhelm-Röpke Str. 6a

35032 Marburg

Telephone: ++49-(0)6421-2824536

Email: ulrike.domahs@staff.uni-marburg.de

Abstract

English, German, and Dutch show very similar word stress patterns, in that word stress is not fixed to a certain position within a word, but realized within the final three syllables. There is, however, no consensus on the actual stress-assigning algorithms and the role of quantity (e.g. Kiparsky 1982; Wiese 2000; Hayes 1995; Giegerich 1985, 1992; Trommelen and Zonneveld 1999a, 1999b). Existing studies are methodologically problematic since they largely depend on convenience samples of existing words and do not test their claims with new words. Using mixed effects regression and classification trees as analytical tools, this paper presents the results of a production experiment with pseudowords and an analysis of large random samples as found in the CELEX lexical database. It is shown that stress assignment is sensitive to syllabic weight in all three languages, though in slightly different ways. The implications of these results for the metrical structure of the three languagues are discussed.

Keywords: *Germanic word stress, quantity-sensitivity, pseudoword production task, corpus analysis, metrical prosody*

1 Introduction

In German, English, and Dutch monomorphemic words, primary word stress is assigned to one of the last three syllables. Three possible stress positions can be observed: stress on the final syllable (e.g. German *Argumént* / Dutch *argumént* / English *kangaróo*), on the penultimate syllable (e.g. *Agénda* / *agénda* / *agénda*), and on the antepenultimate syllable (e.g. *Léxikon* / *léxicon* / *léxicon*). This raises the question how the speaker knows which position within a word is to be stressed.

Four different approaches are possible. First, one could assume that words simply have to be learned together with a particular stress pattern. Another possibility is that stress is assigned on the basis of stress rules. Such rules would govern a particular type of foot structure (trochaic vs. iambic), the direction of foot structure formation (from right to left or vice versa), the alignment of the head foot in a prosodic word (left or right), and quantity-sensitivity (i.e. syllable weight influences the selection of the stress position or not). A third alternative combines computed and lexically retrieved stress patterns in which a default stress pattern is selected in words with a lack of lexically determined metrical structure. Finally, there is the possibility of an analogical mechanism, which assigns stress on the basis of similarity to existing words in the lexicon.

Linguistic theories assume for none of the three languages that stress is purely lexically determined, although psycholinguistic or computational models have provided some evidence for this assumption (e.g. Cutler & Norris 1988; Cutler & van Donselaar 2001; Daelemans et al. 1994; van Donselaar, Koster, & Cutler 2005). Instead, it has been uncontroversially proposed that German, Dutch and English are trochaic languages that are regularly stressed on the penult, if the final syllable is reduced (Giegerich, 1995; Vennemann, 1990; Kager, 1989; Trommelen & Zonneveld, 1999a, 1999b). Native words in West Germanic languages are frequently bisyllabic with a reduced final syllable, automatically leading to penultimate stress. In trisyllabic and longer words that have a final full vowel, matters are less straightforward since the three-syllable window allows three different possible stress patterns. It is controversial whether the type of stress pattern depends on the syllable weight of the final and prefinal syllables (Féry 1998; Giegerich 1985; Hayes 1982; Kager 1989; Trommelen and Zonneveld 1999a, 1999b). Although word stress patterns in German, English, and Dutch appear to be rather similar, metrical analyses of their stress systems lead to different results. Part of the problem could be the methodology of these studies itself since they are primarily based on convenience samples and do not test their hypotheses systematically against larger, independent samples, or against neologisms or pseudowords.

The aim of the present paper is to investigate the controversies raised in the literature on the quantity-sensitivity of these languages on a broader empirical basis and to explore more systematically the similarities and differences of the three West Germanic stress systems. The empirical findings will be used to evaluate theoretical proposals about the kinds of mechanism that regulate stress assignment in the three languages. We present the results of a production experiment with trisyllabic pseudowords and of an analysis of a large set of words gleaned from the CELEX lexical database (Baayen, Piepenbrock and Gulikers 1995). Both types of data provide clear evidence for quantity effects in all three languages and for the role of foot structure in stress assignment. Our results thus call into question a number of claims in the literature concerning the weight-sensitivity of the languages under discussion, and concerning some aspects of their metrical organization.

The paper is structured as follows. In the next section we give an overview of existing approaches to word-stress in the three languages and develop our research questions. Section 3 presents the results of the production experiment, section 4 the results of the CELEX analysis. Section 5 presents the comparison of the CELEX data and the experimental data. Section 6, finally, summarizes our results and discusses the theoretical implications.

2 Word stress in German, Dutch and English: an overview

The approaches presented in this section are mostly in the tradition of Metrical Phonology in which word stress is basically built upon feet of a certain type. Hayes (1995) proposed that languages differ basically with respect to the foot type they choose. It is assumed that feet are strictly binary consisting of either two syllables or two moras, where either the left (trochaic) or the right part (iambic) is strong.

The stress systems of the West-Germanic languages German, Dutch and English share some basic properties, but seem to be different with respect to others. Generally, it is assumed that German, English, and Dutch are trochaic languages. However, analyses differ according to the domain on which feet are constructed, namely either syllables or moras (cf. Hyman 1985; Roca 1992). Furthermore, there are some accounts arguing in favor of a trochaic-dactylic system (for German: Eisenberg 1991; Vennemann 1995; for English: Burzio 1987, 1991, 1994).

In monomorphemic words of all three languages, one of the three final syllables is stressed and stress is assigned starting from the right edge of the word to the left. Furthermore, for Dutch and German a generalization can be found that words containing a final schwa-syllable are regularly stressed on the penultimate syllable (e.g. Bruce & Árnason 1999; Giegerich 1985; Kager 1989; Trommelen & Zonneveld, 1999b; Jessen, 1999). In the following subsections we will discuss each language in turn.

2.1 German

For this language, quantity-sensitivity is still under debate. Many researchers postulate that syllable weight is decisive for German stress assignment (e.g. Domahs, Wiese, Bornkessel-Schlesewsky, and Schlesewsky 2008; Féry 1986, 1998; Giegerich 1985; Ramers 1992; Wurzel 1970, 1980). In particular, it is suggested that the final syllable is stressed in words with a heavy final syllable (e.g *Argumént* 'argument'), but is unstressed in words with a light final syllable, in which case the penultimate syllable receives primary stress (e.g. *Agénda*). According to these approaches, final and prefinal stress is predictable by the weight of the final syllable whereas antepenultimate stress seems to be prespecified in the lexicon. Giegerich (1985), however, claims that the antepenult is computed as a stressed syllable if both the final and prefinal syllables are light (e.g. *Rísiko* 'risk').

The situation becomes more complicated if we look at the notion of quantity-sensitivity as such because some approaches for German develop their own notion of syllable weight. As Hyman (1985) points out, languages with a quantity-sensitive stress system are defined in terms of moras, i.e. units of syllable weight. A syllable is normally counted as monomoraic, or light, if its rhyme consists of a short vowel, whereas a bimoraic, or heavy, syllable comprises a rhyme with either a long vowel or a short one followed by a consonant. According to Féry (1998) only superheavy syllables (i.e. syllables with three filled rhyme positions as in VVC or VCC) are taken to be heavy while Vennemann (1990, 1991, 1995) postulates that any closed syllable is heavy in contrast to open syllables, which he throughout classifies as light, irrespective of vowel length. Thus a VV rhyme is heavy in the traditional approach, but light in Féry's and Vennemann's approach, and a VC rhyme is heavy in the traditional and Vennemann's approach, but light in Féry's. According to Giegerich (1985), final consonants are extrametrical, therefore final syllables are heavy if consisting of a long vowel or of a short vowel followed by two consonants.

Such inconsistencies regarding the role of vowel length, consonant extrametricality as well as the amount of counter-examples to stress rules based on syllable weight led some phonologists (e.g. Eisenberg 1991; Kaltenbacher 1994; Wiese 1996) to consider the German stress system to be insensitive to syllable weight rather than sensitive. Since most of the native words are bisyllabic and end in a reduced syllable that cannot bear main stress, the statistically predominant stress position is the penultimate syllable. Thus, it is suggested that only penultimate stress is regular, whereas for words with final and antepenultimate stress the stress position has to be lexically determined.

2.2 Dutch

The Dutch stress system is unanimously classified as quantity-sensitive in the literature. Van der Hulst (1984), Kager (1989), Trommelen and Zonneveld (1989, 1999b), Booij (1995), and Zonneveld and Nouveau (2004) propose a metrical theory of the Dutch stress system in which closed syllables are heavy and open syllables are light irrespective of the vowel length. In their accounts metrical feet consist of either one heavy (i.e. closed) or two light syllables. For words with an open final syllable, the unmarked stress is realized on the penultimate syllable (e.g. *sombréro*). In words with a closed final syllable, the stress pattern is constrained by the structure of the penultimate syllable. If the penult is light, the antepenultimate syllable receives the stress (e.g. *álcohol* 'alcohol'). If the penult is heavy it attracts primary stress (e.g. *Gibráltar*). In these words, the heavy final syllable itself cannot receive main stress because it is considered extrametrical at the word level. A systematic exception to this pattern concerns words with a super-heavy final syllable that is not extrametrical. Hence such words have final stress (*abrikóos* 'apricot'). All other exceptions to these stress regularities have to be marked lexically. In such cases a light stressed final syllable is for instance specified as a monosyllabic foot bearing main stress, or an unstressed super-heavy final syllable is marked as extrametrical.

Although there is consensus that Dutch is a quantity-sensitive language, it is debated what has to be considered as a heavy syllable. Most accounts favour the option that closed syllables build monosyllabic feet while open syllables with long vowels do not. This is justified by diverse theoretical considerations. Lahiri & Koreman (1988), for instance, suggest that long vowels in Dutch are associated with only one mora, Kager (1989) claims that weight is defined by the number of segment root nodes following the first mora of the rhyme, and van Oostendorp (1995) proposes that long vowels are not represented as long. Van der Hulst (2003), in contrast, assumes that there is no duration contrast at all but only a tenseness contrast, where lax vowels must be followed by a consonant and tense vowels occur in open syllables. Contrary to the postulation that only closed syllables are heavy, recent systematic phonetic analyses of Dutch vowels by Rietveld, Kerkhoff, and Gussenhoven (2004) revealed that durational differences between long vowels in open syllables and short vowels in closed syllables occurred only in stressed syllables. This finding is interpreted as evidence that not only closed syllables but also open syllables with long vowels are parsed as heads of feet and that vowel length therefore contributes to syllabic weight (Gussenhoven, 2009).

The present study was not designed to address the controversy about the interpretation of syllabic weight in Dutch, but to systematically compare the three languages under discussion. Dutch is the only one of the three languages in which vowel length is systematically encoded by the orthography. Using only consonant-final syllable as heavy across all languages allowed us to implement an uncontroversial coding of heaviness.

2.3 English

In the literature on English we find quantity-sensitive and quantity-insensitive models. Kiparsky (1982, 1985) and Booij and Rubach (1992) assume that regular word stress assignment is not regulated by syllable weight properties. Rather, default stress in monomorphemic nouns is suggested to fall on the penult. The most notable assumption they make is that only the default stress pattern is derived by a stress rule. This so-called "English Stress Rule" as described by Hayes (1982) builds a trochaic foot over the last two syllables, leading to penultimate stress. All other stress patterns are considered to be lexically specified.

Accounts that are designed to explain a larger range of data claim that the English stress system resembles the Latin Stress Rule and is sensitive to syllable weight (e.g. Chomsky & Halle, 1968; Liberman & Prince 1977; Giegerich 1985, 1992; Hayes 1982; Kager 1989; Roca 1992; Trommelen & Zonneveld, 1999a). Leaving the final syllable aside as extrametrical, the stress position depends on the structure of the prefinal syllable. If the penult is heavy (with a rhyme consisting of either VV or VC), the penult is stressed; otherwise the antepenult receives main stress. However, such an algorithm is not capable of explaining all cases of English stress patterns. For example, there are cases where the final syllable does receive primary stress (*Hallowéen, violín, lemonáde*). These must then be considered exceptions to the rule of extrametricality. Accordingly, Hayes (1982: 239) proposed that final syllables containing a long vowel are not extrametrical, but form monosyllabic feet and receive either primary (*Hallowéen*) or secondary stress (*'misan,thrope*). In contrast, final syllables containing short vowels are analyzed as being extrametrical.

2.4 Summary and research questions

Summarizing the parametric accounts[1] introduced above, similarities and differences between the languages are illustrated in Table 1. The summary is based on analyses by Janßen (2003: 191) for German, Kager (1989) and Trommelen & Zonneveld (1999b) for Dutch, and Hayes (1982), Giegerich (1985) for English. Controversial parameter settings are given in bold.

Table 1:

Metrical parameters of stress assignment in Germanic languages

Parameters	**German**	**English**	**Dutch**
Foot type	trochee	trochee	trochee
Direction	right-to-left	right-to-left	right-to-left
Quantity-sensitive	**yes / no**	**yes / no**	yes
Heavy syllable	**closed rhyme**	bimoraic syllable	closed rhyme
Extrametricality	no	**yes**	yes
		foot-level	word-level
Word level labeling	head right	head right	head right

Of these parameters, foot type and direction are not controversial, as shown in the previous subsections. The others require further examination not only from a language specific but also from a comparative point of view. The comparative view is relevant because evidence for

[1] In more recent analyses, the above mentioned stress systems have been modeled in the framework of Optimality Theory. In OT analyses proposed for German, English, and Dutch, RHYTHMTYPE/TROCHEE and FOOTBINARITY are either undominated or highly ranked (German: Alber 1997; Féry 1998; Knaus & Domahs 2009; English: Pater 2000; Dutch: Zonneveld & Nouveau 2004) which is compatible with the undisputed relevance of these metrical properties/constraints. Quantity sensitivity is expressed by various constraints and most importantly by Weight-to-Stress Principle (WSP) which demands heavy syllables to be parsed as head syllable of a foot. In most accounts, WSP is of intermediate importance suggesting that it is violable (Pater 2000 for English, Nouveau & Zonneveld for Dutch, Alber 1997, 2005 for German). Extrametricality is expressed in OT terms by the constraint NONFINALITY (Prince & Smolensky, 1993/2004) which militates against the existence of a head of a prosodic word in word final position. For instance, in the analysis of English (Pater 2000) and Dutch (Zonneveld & Nouveau 2004) such a constraint is ranked relatively high and thus rarely violated or, in the case of Pater's analysis, lexically indexed and therefore valid for some English words. In German, the issues concerning the stressability of the final syllable are less clear. Most accounts on German metrical analysis of words either render NONFINALITY as a low ranked constraint or do not consider it at all (Alber 1997; Féry 1998; Knaus & Domahs 2009). The directionality of parsing and the word rule in OT terms results from the ranking of the constraints ALLFEETLEFT (McCarthy & Prince 1993) and RIGHTMOST (Prince & Smolensky 1993/2004). According to Knaus and Domahs (2009), ALLFEETLEFT is ranked relatively low in German in comparison to RIGHTMOST, resulting in the preference of output forms with main stress on the word final foot. Since terminology between theoretical frameworks differ, we refer to metrical properties by using the terminology of parametric accounts in the remainder of this paper.

similar principles across related languages strengthens the likelihood of their existence.

A parameter controversially discussed is the parameter quantity, which distinguishes between quantity-sensitive and -insensitive languages. Many authors consider English, as well as German and Dutch, to be sensitive to syllabic weight (Trommelen and Zonneveld 1999a, b; Kager 1989; Giegerich 1985, 1992). For Dutch, there seems to be a strong consensus that the quantity of the final and penultimate syllable is crucial for the parsing of syllables into feet. By contrast, some have tried to establish a quantity-insensitive account of stress assignment for English (cf. Kiparsky 1982; Booij and Rubach 1992) and German (Wiese 1996) with default stress on the penult and lexicalized stress on the other positions. At least for German and English, the notion of syllabic weight is disputed, in particular the question of what renders a syllable heavy.

A parameter that may differentiate most clearly between the languages in question is extrametricality. While extrametricality does not seem to play a role in German (Giegerich, 1985; Vennemann, 1990; Féry, 1998; Janßen, 2003), it is a widely accepted property of the Dutch (Kager 1989; Trommelen and Zonneveld, 1999b) and English metrical system (Hayes 1982; Giegerich 1985, 1992; Trommelen & Zonneveld, 1999a). Two problems emerge here. First, the assumption of final syllable extrametricality runs into serious empirical problems since many words are stressed on the final syllable, and there are generalizations possible across these cases. Second, and probably as a response to the first problem, the details of what exactly should be considered extrametrical are contested. For Dutch, Trommelen and Zonneveld (1999b) devise a special account for heavy and superheavy final syllables in which the former are seen to be extrametrical and the latter not. In a comparative view, the notion of extrametricality as proposed for English is distinct from the one proposed for the Dutch stress system. In Dutch the final syllable is extrametrical only at the word level but not at the foot level, which means that a final heavy syllable may build a monosyllabic foot but cannot carry word accent, i.e. primary stress. In English final syllables are seen to be extrametrical at the foot level and therefore not involved in foot structure formation at all.

Finally, the parameter responsible for the localization of main stress in prosodic words is called end rule, or word rule. At the word level, stress is assigned to the right edge of the prosodic word, meaning that the rightmost foot is the strong foot bearing main stress (cf. Hayes 1982; Trommelen and Zonneveld 1999a,b; Giegerich 1985, 1992; Kager 1989; Alber 1997; Féry 1998). This is similar in all three languages. However, the interaction of the word level rule in combination with the extrametricality rule seems to lead to different results in the three languages. Assuming extrametricality in the sense proposed by Hayes (1982), English final syllables are not involved in structure building, which results in a rightmost, but not word final, strong foot. In Dutch, a final heavy syllable builds a non-branching foot but is extrametrical at the word level, also leading to non-final stress in words. In accounts on German stress assignment, extrametricality does not play a crucial role. However, many words with a heavy final syllable are stressed on the antepenult, suggesting that extrametricality may be active in some German words as well.

As already mentioned in the introduction, the discussion and the controversies suffer from a scarcity of systematic empirical evidence. Most studies have used convenience samples, so that the empirical coverage of proposed generalizations is often unclear. Furthermore, it would be important to know how the proposed analyses would transfer to words that are unknown to the speakers.

For German and Dutch, Janßen (2003) analyzed data from a production experiment with pseudo-words and also the corresponding CELEX data, but the statistical analysis remained purely descriptive. We will reanalyze her data sets in the present study in a comparative perspective, and using state-of-the-art statistical tools. Féry (1998) also used CELEX and observed correlations of syllable structures and stress positions speaking in favor of a quantity-sensitive system in German. Janßen (2003) obtained a similar result with a different suggestion on what is to be counted as heavy. Recently, Röttger, Domahs, Grande, and Domahs (2012) observed in a pseudoword production task that the weight of the final syllable is the strongest predictor for German stress assignment, but that the weight of all syllables and also the orthographic coding of weight in written language plays a role for pseudoword reading.

For English, Guion, Clark, Harada, and Wayland (2003) used pseudowords to determine factors affecting stress placement in English, but they only looked at bisyllabic words. More recently, Ernestus and Neijt (2008) investigated word stress in Dutch, English, and German by eliciting stress judgements on polysyllabic pseudowords in an elicitation task, and compared it to words in CELEX. Their study focused, however, on the very specific question of the location of stress as affected by the syllables preceding the three final syllables.

In sum, we still need systematic studies of large random samples of existing words and we still need studies that investigate how speakers stress words that are unknown to them. The present study will provide such data.

The problems discussed in the previous paragraphs lead to the following research questions:

1. What is the role of syllabic weight in stress assignment in the three languages?
    a. Does syllabic weight influence the position of word stress?
    b. If so, which syllables contribute to that decision, and which syllables can attract stress based on their weight?
2. Do we find evidence for extrametricality? If so, at which level, the foot or the word?
3. How can we explain the distribution of the stresses as found with the unknown words that speakers pronounce? Can the distribution be explained by the metrical rules proposed in the literature?

3. Stress assignment in pseudowords

3.1 Methodology

We carried out a production experiment with each of the three languages in which participants had to pronounce trisyllabic pseudowords.

3.1.1 Stimuli and participants

Given the three-syllable window, trisyllabic words are a good way to systematically investigate the stress systems of the languages in question. In all three production experiments, we therefore presented trisyllabic pseudowords and varied them in their syllable structure. We ensured that all three syllables could potentially carry primary stress. Open and closed syllables were combined in different positions within the word in order to investigate the interplay between syllable structure and syllable position during stress assignment. In addition to open and closed syllables in all three syllables, complex syllables of the structure CVCC were included in final position. In the following, open syllables are referred to as light, closed as heavy and complex syllables as super-heavy. Though this classification does not necessarily obey weight classifications under those specific accounts that assume final consonant extrametricality (e.g. Hayes 1982, Giegerich 1985), we will use this weight differentiation following Kager (1989) and Trommelen & Zonneveld (1999a, b) for English and Dutch, van der Hulst (1984), Lahiri & Koreman (1988) and van Oostendorp (1995) for Dutch and Vennemann (1990) and Janßen (2003) for German. An additional advantage of this decision is that it allows us to treat syllables equally, irrespective of their position within words, which is important for the investigation of interactions in the statistical analysis.

However, the possibility of final consonant extrametricality will be discussed in the results section 3.2.1. The eight structural conditions are exemplified in Table 2 below.

Table 2: Examples of pseudowords in each condition and language

Condition	**German**	**English**	**Dutch**
1 v.v.v	*Pa.go.ta*	*ca.bo.ra*	*pa.go.ta*
2 v.vc.v	*Bu.mol.ta*	*bo.mol.ta*	*bo.mol.ta*
3 vc.vc.v	*Las.fon.ta*	*lis.fon.ta*	*las.fon.ta*
4 v.v.vc	*Kä.ga.fur*	*ca.ga.foth*	*ke.ga.for*
5 v.vc.vc	*Bo.kam.was*	*bo.cam.vas*	*bo.kam.was*
6 vc.v.vc	*Bin.sa.kaf*	*bin.sa.cub*	*bin.sa.kaf*
7 v.vc.vcc	*Ru.kol.menk*	*ru.col.mest*	*ru.kol.menk*
8 vc.v.vcc	*Rul.ko.menk*	*rul.co.mest*	*rul.ko.menk*

The conditions were mainly constructed on the basis of structure combinations found in existing words In particular, those conditions were selected that enabled us to examine specific questions on the influence of quantity on stress assignment: Words with a light final syllable were examined in three conditions, containing three light syllables (condition 1), a heavy penult (condition 2) or a heavy penult and antepenult (condition 3). The first condition enables to study stress assignment without quantity distinctions between syllables. According to the approaches introduced in section 2, penultimate stress is expected to occur predominantly in German (e.g., Féry, 1998; Vennemann, 1990, 1991; Wiese, 1996; Janßen, 2003) and Dutch (e.g., Kager, 1989; Trommelen & Zonneveld, 1999b), and antepenultimate stress in English (e.g Trommelen & Zonneveld, 1999a) and also in German according to Giegerich (1985). Conditions 2 and 3 allow us to test whether the heavy penult attracts main stress and whether the heavy antepenult competes with the heavy penult (condition 3).

For words containing heavy final syllables, three conditions were constructed: Light penults (condition 4 and condition 6) are used to test whether the antepenult is stressed in Dutch and English, if the penult is light, and even more, if the antepenult is heavy (condition 6). A heavy penult in condition 5 should show whether the heavy penult attracts main stress. For German, some theories predict the heavy final syllable to receive main stress in all three conditions (e.g., Vennemann, 1990, 1991; Giegerich, 1985) and some others that the penult is stressed (e.g., Eisenberg, 1991; Féry, 1998; Wiese, 1996).

In words with super-heavy final syllables we tested whether super-heavy final syllables receive main stress or whether a heavy penult (condition 7) or heavy antepenult (condition 8) competes for main stress. For German and Dutch, the super-heavy final syllable should be stressed, for English, the predictions are less clear. Although English super-heavies may receive main stress, this pattern is seen as exception to the extrametricality of the final syllable (e.g., Trommelen & Zonneveld, 1999a). Thus, words with a heavy penult should be stressed on the penult and with a light penult on the antepenult.

These eight conditions allow us to examine the role of the quantity of the final syllable on stress assignment in the first place, as it is seen to be the most influential one, but also the quantity of the penultimate and antepenultimate syllable. However, the pseudoword studies did not include all logically possible combinations of syllable structures. Conditions with three heavy syllables (CVC.CVC.CVC and CVC.CVC.CVCC) were excluded because such words are not attested in the three languages. Furthermore, words with super-heavy syllables and light penult and antepenultimate (CV.CV.CVCC) as well as with light final and penult and heavy antepenult (CVC.CV.CV) were not tested because such conditions would not necessarily add further insights into the role of quantity on stress assignment.

In the item construction, resyllabifications of coda consonants as onset consonants of the following syllable were avoided by filling each onset position. In addition, in syllable contacts the sonority of segments avoids the parsing of segments into complex onsets (e.g. a word like *bat.ram* could be syllabified as *ba.tram*, while *las.fon.ta* cannot be syllabified as **la.sfon.ta*).[2]

Potential similarities to existing words were avoided as far as possible by including only items whose final two syllables did not rhyme with existing words (based on CELEX)[3]. In particular, the orthographic form should not be similar to or rhyme with existing words. Given that English orthography is opaque and allows for certain pronunciation variants, it is almost impossible to predict the actual pronunciation. Our criterion of controlled orthography not only differs from the design by Guion and colleagues (2003) but specifically avoids the often cited correlation of stress assignment and association with other words (e.g. Guion et al. 2003; Hammond 2004; Smith and Baker 1976). Here, we want to specifically avoid the potential orthographic association with stress of existing words in order to isolate syllabic weight as a factor. Additionally, any obviously marked graphemic combinations were avoided, and the

---

[2] Accidentally, three items (e.g., *Lüt.ra.palf ~ Lü.tra.palf*) presented in the German and Dutch experiments violated this prerequisite and were excluded from the data-analysis. For all other items with potentially ambiguous syllabification, only instances with the intended syllabification were considered.

[3] In the analysis it turned out that some Dutch pseudowords with a closed final syllable (conditions 4 – 6) were indeed similar in their endings to existing words. See below for discussion.

graphemes <e, i, y> were completely left out in word final position, since they are highly likely to be realized as reduced vowels (German: <Tomate>: [toˈma:tə], Dutch: <ekstase>: [ɛksˈta:sə]) or since they tend to be associated with a certain stress pattern. The latter is, for example, the case with words ending in <y>, which are predominantely stressed on the antepenultimate syllable in English (e.g. <ecstasy> [ˈeks.tə.si]).

Concerning the segmental make-up of pseudowords some additional properties were controlled for. The experimental items were constructed following the phonotactic rules of each individual language (Booij, 1999; Hammond, 2004; Hall, 1992; Wiese, 1996/2000). Since the phonotactic constraints vary, there are minor segmental differences between the three sets of items, while the combinations of syllable structures were identical in all languages. Despite our attempt at following the pertinent phonotactic rules, some unidiomatic consonantal combinations (e.g., <mk> in German *Ga.dom.kust*) occurred at syllable boundaries. While such combinations are not necessarily found in (German) monomorphemic words, they are still structurally legitimate by adhering to the sonority hierarchy. However, unidiomatic syllable contacts may have promoted a compound reading, with stress falling on one of the syllables before the problematic syllable contact. A second cue for potential compound readings may have been that some Dutch final syllables did not fully adhere to the prerequisite of non-resemblance to existing words. If true, such pseudowords would not be stressed like monomorphemic words but on the first syllable, like compounds. Crucially, compound stress (here antepenult stress) should occur independent of the structure of the last two syllables. The data show, however, that this is not the case, a compound interpretation of these items on behalf of the participants is therefore unlikely. We return to this issue in the discussion.

Finally, there was no graphemic indication of vowel length. Therefore, subjects had to solely rely on syllable structure information for vowel length. German and Dutch participants generally realized vowel letters in open syllables as long vowels (or tense vowels, depending on theory (van der Hulst, 1984, Kager, 1989)), and vowel letters in closed syllables as short vowels. In German and Dutch, open syllables do not show contrastive vowel length and it is therefore generally assumed that vowel length (or tenseness), unlike in English, does not contribute to syllable weight (e.g., Wiese, 1996 for German; Kager, 1989 for Dutch). Open syllables in our Dutch and German data were therefore coded with only one vowel slot (as proposed by Oostendorp, 1995). English participants varied in their pronunciation between neutralized schwa, full long vowel (or diphthongal) pronunciations and full short vowel pronunciations. Due to the underdetermination of vowel quality and length in the English spelling system vowels were coded as perceived by the phonologically aware transcribers.

According to these criteria, pseudowords were constructed in eight different conditions, ten items each (see appendices A to C). Since stress in English is distinctive with respect to lexical category, all pseudowords were disambiguated as nouns by presenting the stimuli in a sentence context, and by instructing the participants to regard them as nouns. In contrast to the cross-linguistic pseudoword study by Ernestus and Neijt (2008), our items consist of comparable segments across all languages and were varied only for phonotactic reasons. Furthermore, our presentation modes were identical for all three languages.

In addition to the experimental items, mono-, bi-, and quadrisyllabic filler items (15 each) were used in order to force the participants to produce prosodic words differing in syllable number. This procedure should reduce potential automatic repetition of identical prosodic structures. The segmental constraints as described for the test items did not hold for the filler items, therefore <e, i, and y> were included to reach more variability in segmental combinations within the test corpus. The items were randomized and presented in a carrier context sentence as in (1) – (3) in order to ensure a natural intonation and to avoid the realization of a boundary tone which could occur when presented as a list of isolated words.

(1)	*Ich habe gehört, dass Peter*	***Binsakaf***	*gesagt hat.*
(2)	*I heard that Peter said*	***binsacub***	*yesterday.*
(3)	*Ik heb gehoord dat Flora*	***binsakaf***	*heeft gezegd.*

The items' pronounceability and the status as possible words were pre-tested by native speakers of the respective language. Four different randomizations were used to avoid order effects to influence the overall results.

In each experiment, participants were asked to first read the sentences including the critical words silently to acquaint themselves with the unknown word, and then to read out the sentences aloud. German and Dutch participants were recorded using a SONY digital recorder and a Sennheiser "Electret" microphone and American participants using a PC laptop computer and a HAMA headset microphone. All responses were transcribed according to their stress patterns and each transcription was controlled twice by phonetically trained raters (the interrater reliability was 97% for the German and English data and 98% for Dutch). In most cases, the identification of primary stress positions was unambiguous, especially for English, where the judgment of stress patterns was facilitated by the reduction of unstressed syllables to schwa syllables. Nevertheless, not in all responses could a stress position be identified in which case the items were discarded. Furthermore, responses where the subjects altered the syllabic structure had to be excluded as well. This included, but was not limited to, resyllabification into unintended structures (e.g., *Ta.klu.tarp* instead of *Tak.lu.tarp*), as well as leaving out or adding consonants. Interestingly, American participants seemed to have more difficulties in reading the unknown words, with an error rate of 16% in comparison to 13% in German and only 6% in Dutch participants.

For the German experiment, 25 native speakers were recruited, 14 females, 12 males, ranging in age between 20 and 34 years. All participants were students of the University of Duesseldorf (Germany). The English experiment was carried out with 23 monolingual native speakers of American English (12 females, 11 males) between 18 and 57 years of age, recruited at the universities of Marburg and Gießen in Germany (all of them exchange students) and in Eastern Massachusetts. The Dutch experiment had 16 native speaker participants (12 females, 4 males) between the age of 19 and 34 (all students at Radboud Universiteit Nijmegen).

3.1.2 Statistical analysis

For the analysis of the production data we used two different methods, generalized mixed effects regression and classification trees. We devised mixed effects regression models (e.g. Baayen 2008, Baayen et al. 2008) to test whether the structure of the three syllables has an influence on stress assignment to a particular syllable. Furthermore, the models tested whether these effects differ from language to language. Mixed effects regression has the advantage of bringing subject and item variation under statistical control and of being able to deal with unbalanced data sets. This is most welcome in our case since not all combinations of syllable structures are represented in the stimuli with equal frequency. However, mixed effects regression has the disadvantage that it cannot handle complex three-way or four-way interactions in an easily interpretable way. We therefore complement the regression analysis with an analysis using classification trees of the CHAID type (CHi-squared Automatic Interaction Detection, e.g. Kass 1980), which are more suitable for investigating the potential influence of particular constellations of the values of a large number of predictor variables. For this analysis we used the statistical package R (R Development Core Team, 2011) together with the partykit and CHAID packages (Hothorn & Zeileis 2012)

For the mixed effects analysis we used R and the lme4 package (Bates, Sarkar, Bates & Matrix, 2007). We first fitted generalized mixed effects models with the weight-related predictors STRUCFIN (i.e. 'structure of the final syllable', with the values light (L), heavy (H), superheavy (sH)), STRUCPENULT (i.e. 'structure of the penultimate syllable', with the values L, H), STRUCANTEPEN (i.e. 'structure of the antepenultimate syllable', with the values L, H). Additionally, LANGUAGE was entered as a predictor interacting with the other predictors to assess differences between the three languages concerning their sensitivity to syllable structure effects.

We ran three different analyses, one for each type of stress (final, penultimate and antepenultimate) as dependent variable. First we fitted a model with the above-mentioned predictors and stressFinal as the dependent variable. If the stress for a given item ended up on the final syllable, this was coded as `yes` for this variable, if the stress did not end on the final syllable this was coded as `no`. For the other two analyses we defined stressPenult and stressAntepen as dependent variables, respectively, with `yes` and `no` as values, depending on the presence or absence of stress on the respective syllable.

In order to keep subject and item variation under statistical control, subject and item were included as random effects. We tested the necessity of these random effects with log-likelihood tests, which always showed that the inclusion of these random effects was justified. We also tested more complex random effect structures, for example with random contrasts for subjects and some of the other predictors. In some of the models the inclusion of random contrasts further improved the predictive power but did not change the nature of the effects. We therefore report the simpler models that contain only random intercepts for subject and item. The regression models were simplified following standard procedures of stepwise removal of non-significant predictors and non-significant interactions (e.g. Baayen 2008).

CHAID constructs decision trees with binary and non-binary branching. CHAID trees are especially well suited for large data sets where predictors interact in complex ways. The algorithm works through all predictors and partitions the data into subsets that differ significantly in their distribution of the response variable from other subsets, with the subsets being characterized by particular constellations of the values of the predictor variables. As suggested by its name, CHAID uses chi-square tests for determining the best split at each step of the partitioning process. In our analyses we set the (Bonferroni-adjusted) alpha-levels for the merging and splitting of categories to $p < 0.001$. For our analyses we used the statistical package 'CHAID' in R (Hothorn 2009).

From the top to the bottom of the tree the subsets become increasingly structurally homogeneous. One other important advantage of this statistical method is that it deals with multinomial response data in an easily interpretable way, which is most welcome in our case, since in our data stress can fall on one of three syllables, giving us a dependent variable with three possible outcomes, instead of two outcomes, as in the regression models described in the previous subsection. Furthermore, and as already mentioned, using decision trees, we can model more complex interactions than those we could implement in our mixed effects regression models.

Why do we use both methods alongside each other? Mixed effects regression models have the disadvantage that more complex interactions like those at issue in this study are not so easily interpretable. Classification trees, on the other hand, do not bring subject and item variation under proper statistical control, as including subject or item into the model increases the number of nodes to such an extent that the model is no longer interpretable. We therefore present the results of both types of analysis in order not to miss out on important sources of variation and still arrive at interpretable results. As we will see, regression models and classification trees converge on the same basic results.

3.2 Results 1: Regression analysis

In section 2 we reviewed the claims about the effects of weight on stress assignment in the three languages. The design and analysis of our data with LANGUAGE as a co-variate necessitates, however, also a different perspective, namely one that focuses on the position of stress on a particular syllable (e.g. the final syllable) across languages. In other words, in order to make sense of some of our results, the language-based hypotheses need to be restated as position-based hypotheses. This will be done in the pertinent subsections.

3.2.1. Overview

Only responses exhibiting an unambiguous stress pattern and without reading errors were considered for further analyses. Overall, the German participants produced 1724 analyzable responses, the English 1660 and the Dutch participants 1173.

The barplot in Figure 1 gives an overview of the distribution of stresses. We can see that in all languages, penultimate stress is by far the preferred stress position. In German, antepenult and final stress are much less preferred, but equally frequent, while in Dutch and

English, final stress is clearly in the minority (only 16.1% for Dutch, and 11.4% for English)

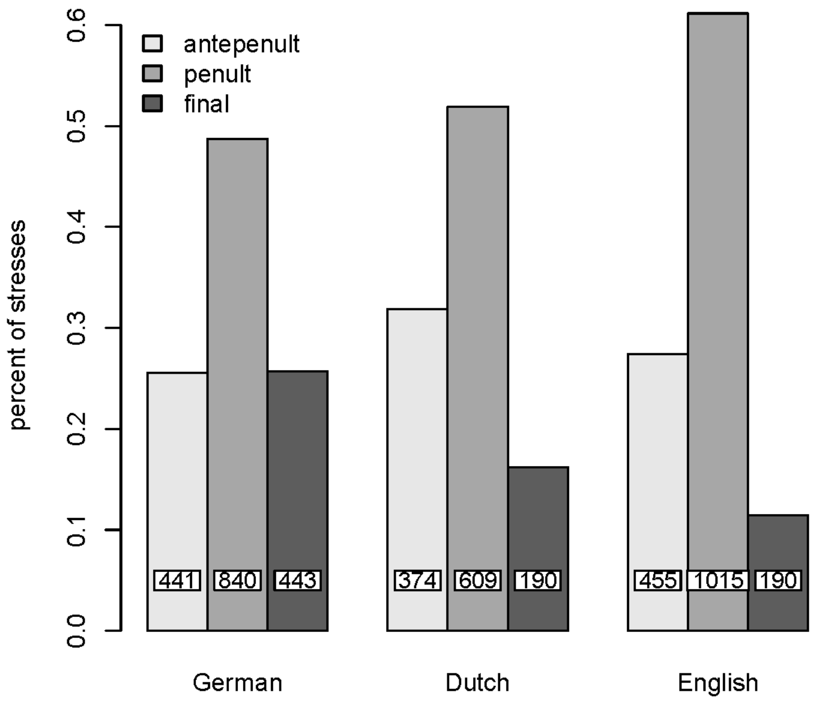

Figure 1: Distribution of stresses by language. Figures in the boxes give the number of pertinent observations (N_{German}=1724, N_{Dutch}=1173, and $N_{English}$=1660)

3.2.1 Final stress

For the final syllable, we can hypothesize that in German this syllable receives stress if it is heavy (e.g. Vennemann, 1990, 1991) or superheavy (Giegerich, 1985). Dutch final syllables would be stressed if superheavy (Trommelen & Zonneveld, 1999b), while English finals are extrametrical and their structure should not be relevant for main stress assignment (Trommelen & Zonneveld, 1999a).

Table 3 presents the final model. In this table and the tables to follow variables are presented in small capitals and values in `typewriter script`. The baseline is a German pseudoword with light syllables in each position (i.e. LANGUAGE `German`, STRUCFIN `L`, STRUCPENULT `L`). Positive coefficients indicate an increase in the likelihood of the final syllable being stressed; negative coefficients indicate a decrease in this likelihood.

Table 3

Mixed effects regression model for final stress

Random effects				
Groups	Name	Variance	Std.Dev.	
item	(Intercept)	0.52099	0.7218	
subject	(Intercept)	4.22577	2.0557	
Number of obs: 4057, groups: item, 237; subject, 64				
Fixed effects				
	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-3.3881	0.4856	-6.978	< 0.000 ***
STRUCFIN H	2.2559	0.2792	8.080	< 0.000 ***
STRUCFIN sH	3.6548	0.3188	11.464	< 0.000 ***
LANGUAGE Dutch	-2.3383	0.8373	-2.799	0.00513
LANGUAGE English	-0.2610	0.7728	-0.338	0.73551
STRUCPENULT H	-0.5859	0.1499	-3.908	< 0.000 ***
STRUCANTEPEN H	-0.3286	0.1539	-2.135	0.03280 *
STRUCFIN H : LANGUAGE Dutch	1.3613	0.5585	2.437	0.01479 *
STRUCFIN sH : LANGUAGE Dutch	0.8102	0.6078	1.333	0.18250
STRUCFIN H : LANGUAGE English	-1.2100	0.5331	-2.270	0.02322 *
STRUCFIN sH : LANGUAGE English	-1.5759	0.5452	-2.890	0.00385 **

C	**AIC**	**BIC**	**logLik**	**deviance**
0.9295973	2588	2670	-1281	2562

The model shows main effects for all predictor variables and a significant interaction of language and the structure of the final syllable. The overall predictive accuracy is very high, with a concordance index of 0.93. Let us look at the individual effects in more detail.

For all languages, the placement of final stress is dependent on the structure of the final syllable, such that an increase in weight leads to a higher probability of final stress. This effect varies significantly in strength depending on the language we look at, as shown by the significance of the interaction of STRUCFIN and LANGUAGE. The interaction is plotted in Figure 2.

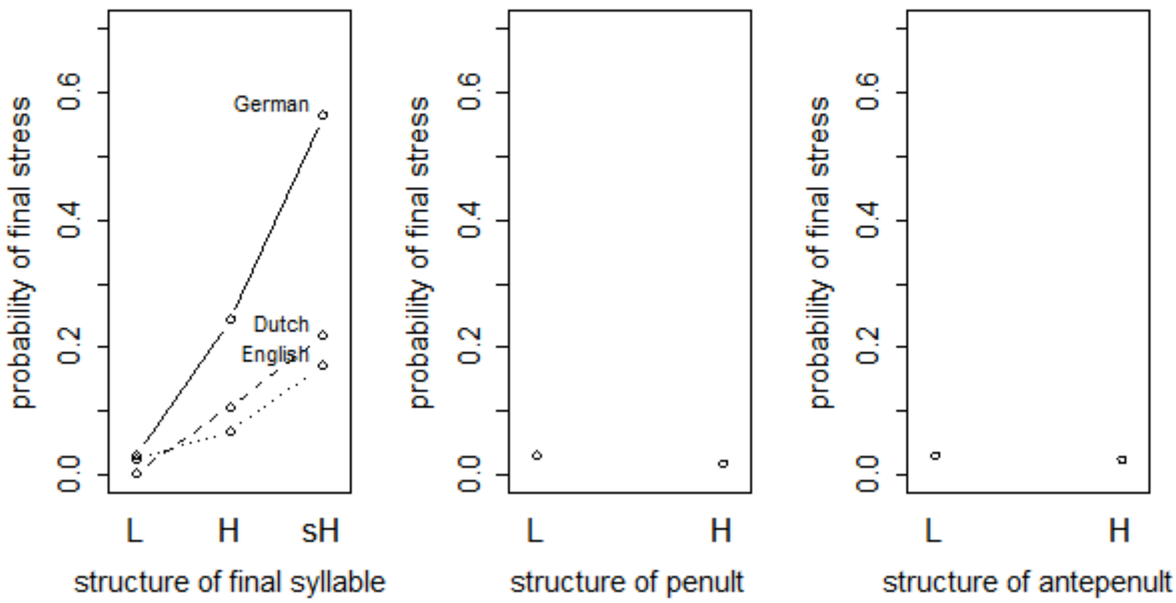

Figure 2: Probability of final stress by language and syllable structure

We can see that a light final syllable is generally not stressed in any of the languages. As the weight of the final syllable increases, the chances of attracting stress increase in all three languages, but the languages differ in effect strength. As can be seen from the model in Table 3 as well as from Figure 2, there is a significant difference between German on the one hand and Dutch and English on the other. German shows the strongest effect, followed by Dutch and English. In German, a superheavy final syllable is more likely to be stressed than to be unstressed. In contrast, an increase in weight of the final syllable in Dutch or English does not increase final stress to such a great extent, and even superheavy syllables are much more likely not to be stressed. The effects of the penultimate and antepenultimate structures are significant but very weak as shown by the very small coefficients in Table 3 and in the two right panels of Figure 2.

These results support the idea that German is quantity-sensitive and stresses superheavy final syllables. The results for Dutch are not so clear. We find, quite expectedly, a significant increase in the likelihood of stress for superheavy syllables, but the effect size, i.e. the increase in likelihood is by far not as strong as the phonological literature would have predicted. The significant effect of the final syllable for English is surprising and not in accordance with the literature. The effect is not very strong, however.

3.2.2 *Penultimate stress*

Weight-sensitive approaches to German stress predict that the penult is stressed either if the final syllable is light, or if the penult itself is heavy. Approaches that assume weight-insensitivity predict default stress on the penult irrespective of its structure. In Dutch the penult should be stressed in words that have a final light syllable, or a heavy penult. If the final syllable is heavy, only a heavy penult can be stressed. English stresses the penult (only) if it is heavy.

We fitted a mixed effects model analogous to the one for final stress. The final model is documented in Table 4. The baseline is again a German pseudo-word with light syllables in each position.

Table 4

Mixed effects regression model for penultimate stress

Random effects				
Groups	Name	Variance	Std.Dev.	Correlations
item	(Intercept)	0.68740	0.82910	
subject	(Intercept)	0.96704	0.98338	
Number of obs: 4057, groups: item, 237; subject, 64				

Fixed effects				
	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.27657	0.31495	4.053	< 0.000 ***
STRUCFIN H	-2.84644	0.27951	-10.184	< 0.000 ***
STRUCFIN sH	-4.01475	0.32629	-12.304	< 0.000 ***
LANGUAGE Dutch	0.62649	0.48583	1.290	0.197216
LANGUAGE English	-1.62166	0.50703	-3.198	0.001382 **
STRUCPENULT H	1.28722	0.24393	5.277	< 0.000 ***
STRUCFIN H : LANGUAGE Dutch	-0.47069	0.41807	-1.126	0.260227
STRUCFIN sH : LANGUAGE Dutch	0.05712	0.47652	0.120	0.904582
STRUCFIN H : LANGUAGE English	1.56206	0.44074	3.544	0.000394 ***
STRUCFIN sH : LANGUAGE English	1.90808	0.47477	4.019	< 0.000 ***
LANGUAGE DUTCH:STRUCPENULT H	-0.21228	0.35709	-0.594	0.552194
LANGUAGE ENGLISH:STRUCPENULT H	1.91190	0.32351	5.910	< 0.000 ***

C	**AIC**	**BIC**	**logLik**	**deviance**

0.9167233	3529	3618	-1751	3501

The model shows main effects for the structure of the final syllable and for the structure of the penultimate syllable. In addition we find a significant interaction of STRUCFIN and LANGUAGE and of STRUCPENULT and LANGUAGE. The overall predictive accuracy is again very good, with a concordance index of 0.92. Figure 3 illustrates the results.

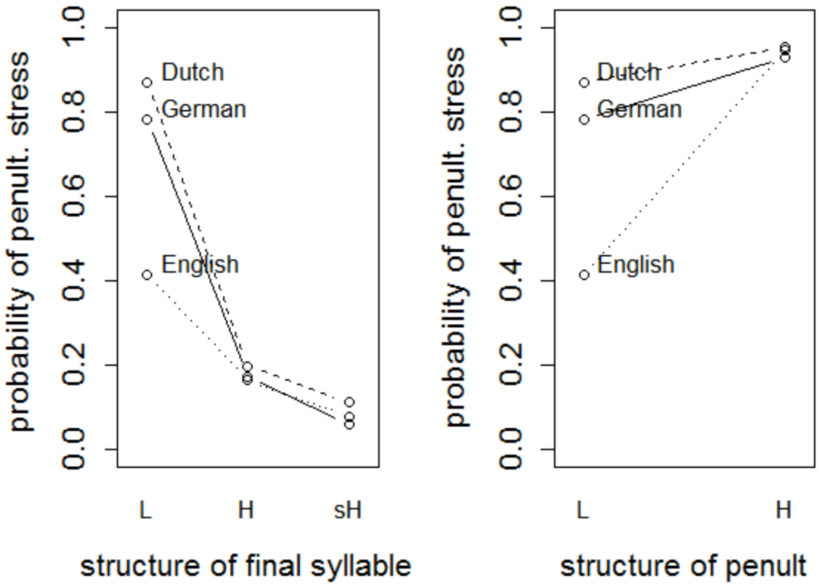

Figure 3: Partial effect of penultima structure (left panel) and interaction of final structure and language in mixed effects regression model for penultimate stress.

With regard to the final syllable we can state that for all three languages a heavy or superheavy final syllable goes together with a low probability of penultimate stress. If the final syllable is light, however, Dutch and German show a very strong preference for penultimate stress, while English shows only a moderate increase in the probability of penultimate stress.

With regard to the role of the penultimate syllable itself, we can see that its structure is highly influential in English, but not in Dutch or German. All languages stress heavy penults almost categorically (given a light final syllable in German and Dutch).

Our data support quantity-sensitive accounts of stress assignment in German and Dutch as outlined above. The effect of heavy penults in English stress assignment is also in accordance with existing models. However, we find an unexpected effect of the structure of the final syllable, which runs counter to the expectation that this syllable is extrametrical. The relatively high proportion of stressed light penults (about 40%) is also surprising.

3.2.3 *Antepenultimate stress*

There are three hypotheses for German antepenultimate stress. One approach says it is irregular (e.g. Féry, 1998, Vennemann, 1990), and should therefore be strongly dispreferred in a pseudo-word experiment. A second approach (e.g. Giegerich 1985) predicts antepenult stress if the last two syllables are both light, and a third approach (Janßen, 2003; Domahs et al. 2008; Janßen und Domahs, 2008) claims that antepenult stress occurs with words that have a heavy final syllable.

Dutch words should be stressed regularly on the antepenult if the final syllable is heavy and the penultimate light, and one should not observe a weight effect for the antepenult itself (Trommelen and Zonneveld, 1999). English should stress the antepenult if the penultimate syllable is light.

The baseline is again a German pseudoword with light syllables in each position. Again, positive coefficients indicate an increase in the likelihood of the antepenultimate syllable being stressed.

Table 5

Mixed effects regression model for antepenultimate stress

Random effects				
Groups	Name	Variance	Std.Dev.	
item	(Intercept)	0.46075	0.67878	
subject	(Intercept)	2.26229	1.50409	
Number of obs: 4057, groups: item, 237; subject, 64				

Fixed effects				
	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-2.6118	0.3917	-6.668	< 0.000 ***
STRUCFIN H	2.1159	0.2657	7.964	< 0.000***
STRUCFIN sH	2.0497	0.2920	7.020	< 0.000***
LANGUAGE Dutch	-0.1172	0.6021	-0.195	0.84571
LANGUAGE English	1.7775	0.6134	2.898	0.00376 **
STRUCPENULT H	-1.1084	0.1326	8.361	< 0.000 ***

STRUCANTEPEN H	0.2854	0.1354	2.108	0.03502 *
STRUCFIN H:LANGUAGE Dutch	0.4855	0.3935	1.234	0.21734
STRUCFIN sH:LANGUAGE Dutch	0.6859	0.4282	1.602	0.10918
STRUCFIN H:LANGUAGE English	-0.9128	0.4515	-2.022	0.04321 *
STRUCFIN sH:LANGUAGE English	-0.7487	0.4671	-1.603	0.10895
LANGUAGE Dutch:STRUCPENULT H	0.3775	0.3241	1.165	0.24411
LANGUAGE English:STRUCPENULT H	-1.6659	0.3102	-5.370	< 0.000 ***

C	**AIC**	**BIC**	**logLik**	**deviance**
0.9050374	3359	3454	-1665	3329

We find a main effect for all variables and two significant interactions: one for LANGUAGE and STRUCFIN and one for LANGUAGE with STRUCPENULT. The C value of the model is very satisfactory (concordance value of 0.91).

As illustrated in the left panel of Figure 4, and also shown by the positive coefficients for STRUCFIN, an increase in heaviness of the final syllable increases the chances of antepenultimate stress in all three languages, though to different degrees. This effect for the structure of the final syllable is in accordance with the literature on Dutch and supports the approach by Janßen and colleagues for German. However, contrary to expectation, the effect is not strong enough to actually lead to antepenultimate stress in the majority of cases, since the probability remains below 50%. For English, we would not have expected this effect at all. With approximately 60% probalitiy of antepenultimate stress in pseudowords with heavy final syllables, the effect is nevertheless the strongest for all three languages.

The effect of the structure of the penultimate syllable is that a heavy penult decreases the chances of antepenultimate stress in English, while in German and Dutch the structure of the penult is largely irrelevant for antepenultimate stress. These findings are in accordance with the hypothesis for English, but only partly for German and Dutch. Our data are compatible with Janßen (2003) and with an approach that considers antepenultimate stress as irregular and dispreferred (Féry 1998). The data falsify Giegerich's prediction (1985) that antepenultimate stress occurs if the final two syllables are light. The predictions of Trommelen & Zonneveld (1999b) for Dutch words go in the right direction but the effect is by far not as strong as expected.

Finally, there is also a main effect for the structure of the antepenultimate syllable but this effect is extremely small as shown in the rightmost panel of Figure 4 and the coefficient in Table 4.

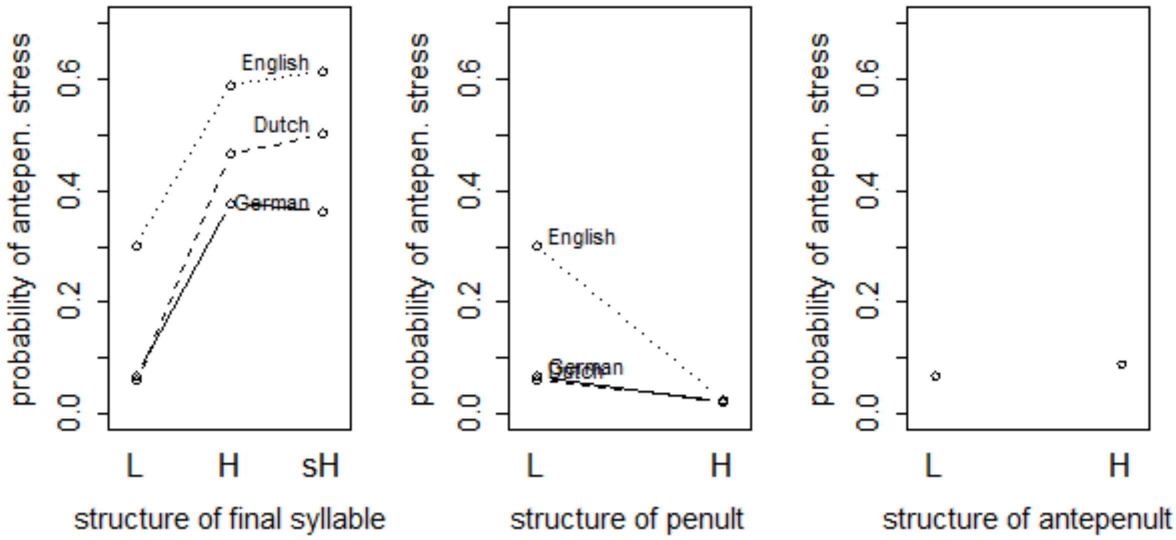

Figure 4: Partial effects of mixed effects regression model for antepenultimate stress.

We will now turn to the analysis using classification trees in order to investigate in more detail the potential interaction of the different predictors, in particular the effects of specific constellations of the three syllabic structures on stress assignment. In addition this analysis allows us to look more systematically at the crosslinguistic differences that the three languages present.

3.3 Results 2: Classification trees

We fitted a classification tree using the CHAID algorithm (Kass 1980) of the CHAID package in R (version 2.15.1). Alpha levels for the merging of predictor categories and for the splitting of a node in the most significant predictor were set to $p < .001$. The tree is plotted in Figure 5 with strucFin, strucPenult, strucAntepen and language as independent variables and stress (with the values final, penult, antepen) as dependent variable.

The tree is to be read as follows. Each node contains the name of the variable according to which the data show a significant split. Note that not all of the splits are theoretically interesting because the sensitivity of the algorithm sometimes finds splits that differ only slightly in their (otherwise clear) majority choice. Thus, classification trees have a tendency to overfit the data (Baayen 2008, Chapter 5). We will concentrate on those splits that show significant differences in their majority choice. The nodes are numbered for easy reference. The terminal nodes give the distribution of stresses for the respective constellation of features in terms of a bar chart for this subset and the total number of observation in this set.

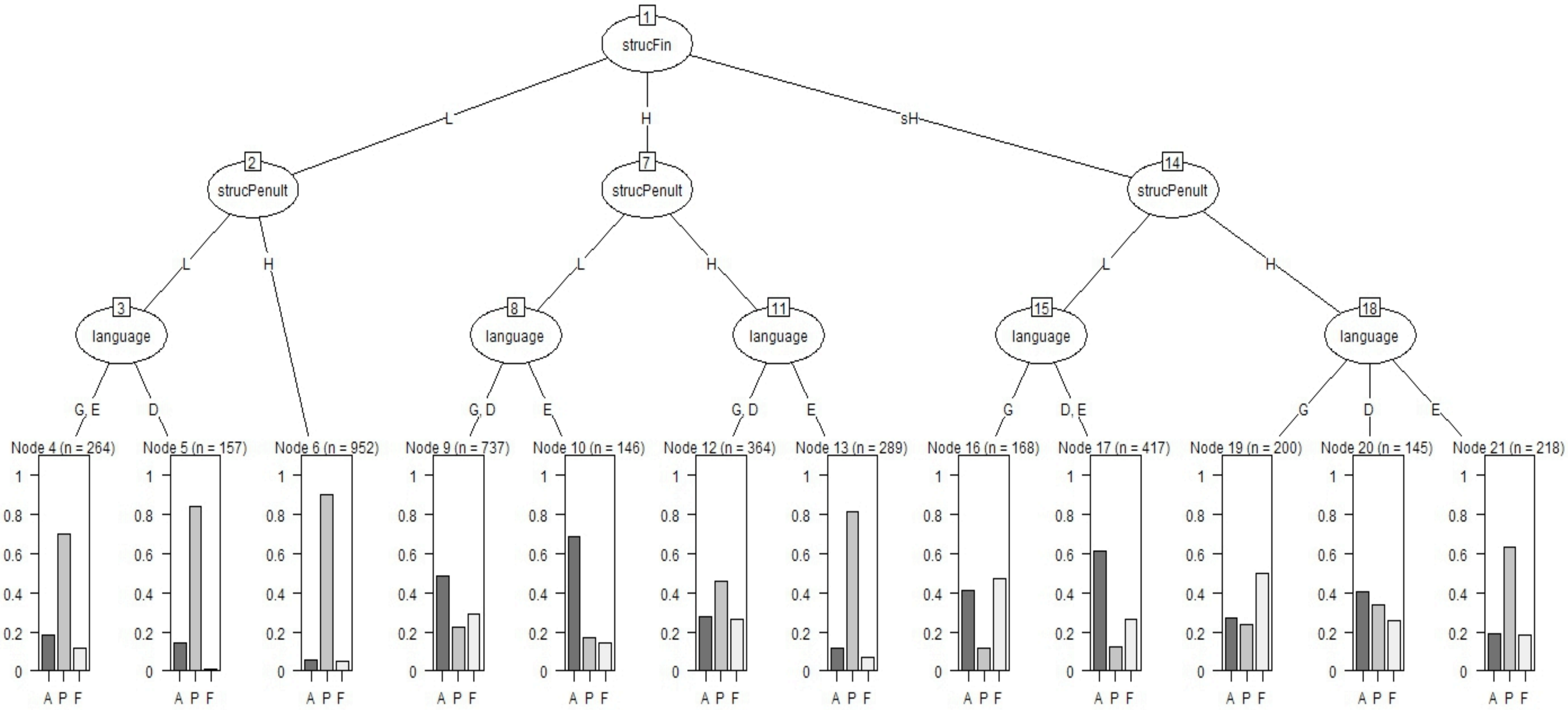

Figure 5: Classification tree, experimental data set of all languages

As Figure 5 shows, the structure of the antepenultimate syllable does not play any role. The most important effect concerns the structure of the final syllable (Node 1, the root node). In all languages, words with light final syllables behave significantly different from words with heavy or superheavy final syllables. Words with light final syllables are predominantly stressed on the penult in all three languages (Nodes 4, 5, and 6 in the graph).

For words with heavy final syllables there is an interaction with the structure of the penultimate syllable (Node 7): if that structure is light, all languages prefer antepenultimate stress (Nodes 9 and 10) with English in the lead (Node 10). If the penultimate syllable is heavy, English shows overwhelmingly stress on the penult (Node 13), whereas German and Dutch only have a moderate relative majority for penultimate stress and still sizable proportions of final and antepenultimate stress (Node 12).

Let us finally turn to words with superheavy final syllables. Here we also find an interaction with the structure of the penultimate syllable. This is the only subset of the data in which the differences between the languages become really striking. If the penult is light, German is divided between antepenultimate and final stress (Node 16) while Dutch and English have a strong tendency toward antepenultimate stress (Node 17). This suggests for German that words that end in the sequence LsH build a monosyllabic foot at the right edge, which either functions as the prosodic head of the word giving us the structure $(XL)_{Fw}(^{\text{ˈ}}sH)_{Fs}$, or not. If not, the antepenult is stressed (see left bar of Node 16), in accordance with the structure $(^{\text{ˈ}}XL)_{Fs}(sH)_{Fw}$ A similar metrical structure can be assumed for the almost 30% Dutch and English words of Node 17 that have final stress. However, in these languages, there is still a clear tendency of preferring antepenultimate stress over final stress. For Dutch this preference is rather unexpected since Kager (1989) and Trommelen & Zonneveld (1999b) predict stress on final superheavies. For English, antepenultimate is predicted in words with light penults due to the assumed extrametrical status of the final syllable (e.g. Giegerich 1992, Trommelen & Zonneveld, 1999a).

With words ending in HsH (Node 18), all three languages differ significantly from each other: German prefers final stress, English penultimate stress, and Dutch antepenultimate stress (with a much less pronounced majority choice). The German pattern is predicted by quantity sensitive accounts, the English pattern emerges naturally under the assumption of final extrametricality, but the Dutch pattern is unaccounted for by any existing approach.

3.4 Summary and discussion

Both, the regression analyses and the classification analysis have provided clear evidence that the structure of the final and penultimate syllables is influential in stress assignment to new words. Hence, German, Dutch and English must be considered quantity-sensitive languages, with the three languages showing very similar patterns overall. This is in line with theories suggesting that the quantity of the final two syllables restricts stress assignment (e.g. Giegerich, 1985, Vennemann, 1990, Féry, 1998 for German; Kager, 1989, Trommelen & Zonneveld, 1999b for Dutch; Hayes, 1982; Trommelen & Zonneveld, 1999a for English). And it is also in line with findings on German stress assignment reported in a more recent paper by Röttger et al. (2012).

Furthermore, the data provide strong evidence against final syllable extrametricality at the foot level in any of the languages, as the structure of the final syllable turned out to be a robust significant predictor of stress assignment in all models. Therefore, extrametricality at foot level proposed for instance by Chomsky & Halle (1968), Hayes (1982), Kager (1989), or Trommelen & Zonneveld (1999a) is not supported by the actual patterning of the data.

Regarding final consonant extrametricality in German, the data speak for an analysis in which a final coda consonant also contributes to the syllabic weight of the final syllable because very few of these words were stressed on the penult. Penultimate stress was mainly observed with words/pseudowords containing open (=light) final syllables.

For English and Dutch, we observe that words with final heavy syllables are less likely to be stressed on the penultimate syllable suggesting that the final heavy syllable is parsed as monosyllabic foot. However, this raises the question of why final syllable stress never shows up as a majority choice. Under final syllable extrametricality at the word level this fact is predicted. In other words, we find a situation in which for the selection of antepenultimate vs. penultimate stress the quantity of the final syllable is decisive, but it can nevertheless not bear main stress. For Dutch this fact is accounted for by extrametricality at the word level, but not at the foot level (e.g. Trommelen & Zonneveld 1999b). Our data suggest the same analysis for English.

An alternative explanation, which has been mentioned in section 3.1.1, is that, in the experiment, words with heavy or super-heavy final syllable were interpreted as compounds. Under this assumption we would expect to find an increased chance of antepenultimate stress assignment, as most Germanic compounds are stressed on the initial syllable. The comparison of pseudowords with super-heavy (node 14), heavy (node 7), and light (node 2) final syllables in Figure 5 could support such an interpretation. However, if we compare node 15 with node 18 and node 8 with node 11, it becomes apparent that not only the structure of the final syllable but also of the penult plays a role. For instance, in German pseudowords with final super-heavy syllables, pseudowords with a heavy penult are stressed predominantly on the final syllable while those with an open penult on either antepenultimate or final syllable. This is not a pattern expected under compound readings. Furthermore, pseudowords with a heavy final syllable are stressed predominantly on the heavy penult, and on the antepenult if the penult was light. In those cases, the heavy penult blocks the antepenult as landing site for stress, speaking in favor of right to left parsing in monomorphemic words, and against compound readings.

In order to further investigate the possibility of compound misinterpretation and its potential consequences for stress assignment, we conducted another test. Some of the Dutch pseudowords ended in strings that might have been interpreted as existing words (e.g. *was* 'wash'), which could also have triggered compound readings. We therefore coded an additional factor for each Dutch word (LAST SYLLABLE IS A WORD, with the values `yes` and `no`) and included it into our CHAID model. However, this additional factor did not turn out to be a significant predictor of antepenultimate stress. Thus, both the distribution of majority choices and the lack of influence of the factor LAST SYLLABLE IS WORD dismiss the possibility of compound readings of those pseudowords.

To summarize, we find that the quantity of the final syllable and the penult are strong predictors for stress assignment in all three languages. However, the pseudoword studies also reveal a certain amount of unclear stress preferences, which is a challenge to existing theories. In particular, words with heavy and super-heavy final syllable allow for equally strong majority choises of two positions. This amount of stress variation data suggests that certain aspects of existing accounts need to be revised in order to be able to understand the treatment of nonce words by the speakers of the respective languages. We will return to this issue below.

The - sometimes perhaps unexpected - distribution of stresses in the pseudowords may raise the question of whether the experimental data reliably reflect the speakers' intuitions (and ultimately their metrical system), or should be considered insignificant artefacts arrived at by improper methods. In order to address this concern, the experimental study was complemented by a study of the distribution of stresses in the lexicon, i.e. in the established vocabulary of the three languages. If the distribution of stresses in the lexicon is very similar to the one we found in the experiment, this would counter any attempt to dismiss the experimental findings as artefactual. In the next section we will therefore present a systematic comparison of lexical and experimental data.

4 Stress assignment in the lexicon

4.1 Method

CELEX is a lexical database that contains lexical data from German, Dutch and English, with different kinds of lexical information (e.g. orthographic, phonological, morphological) that can be accessed for very specific research questions. It has been used in very many investigations of the lexical structures of the three languages, and it is generally taken as a model of the established vocabulary of the three languages.

We first extracted all monomorphemic trisyllabic words and their stress specification and syllable structures. In order to be able to compare the CELEX data with the experimental data, these words were then recoded for syllable structure in the same way as the experimental data, using the values L, H, sH.

For the experimental data we used mixed effects regression in order to be able to get the subject and item variation under statistical control. Neither subject nor item variation applies to CELEX, which means that mixed regression is not applicable. We therefore used again CHAID trees. If the distribution of stresses as found in the experiment emerges from the lexicon, TYPE OF DATA should not come out as an influential variable for the partitioning of the different data sets.[4]

4.2 Results

An overall model including CELEX and experimental data for all three languages revealed effects for the variable TYPE OF DATA only at the terminal node level. The resulting tree is very large and has a root node split for STRUCFIN. In order to present readable trees, we present the three subtrees branching from the root node, each of which having one value of STRUCFIN. The details of the three subtrees will be discussed shortly. The overall predictive

[4] One reviewer raises the problem that the exclusion of words with three heavy syllables from the stimuli set of the experiment would make the two data sets more similar, because we do not consider the full range of potential differences. However, as already mentioned in section 2, none of the three languages has words of this structure. And if words with three heavy syllables do not occur in any of the three languages, one cannot compare them to

accuracy of the tree is 71%.

For words with a light final syllable (see Figure 6), the structure of the penult is the best predictor for the majority choices in each language (see Node 1). The main result of the classification tree is that the majority choices in both data sets are quite comparable (see Node 4 vs. Node 7, and Nodes 14 vs. Node 17). Only for English do differences in stress distributions between the lexicon versus the experimental data occur (see Node 11 vs. 12). English words ending in two light syllables (xLL, Node 11) are predominantly stressed on the antepenultimate syllable whereas pseudowords (Node 12) are stressed with equal frequency on either antepenult or penult.

Furthermore, there is an interesting effect observable in English for words with a heavy or superheavy penult (Node 17). While in the experiment such words are almost categorically stressed on the penult (Node 21), the CELEX data show more variation and an effect of the structure of the antepenultimate syllable (Node 18).

For words with a heavy final syllable (see Figure 7), we only find data sets effects if the penult is light. But even in those cases the majority choices remain unaffected (cf. Nodes 5, 6, 7, 9, and 10).

Finally, words with a superheavy final syllable show clear language-specific effects of TYPE OF DATA. German and English display the same majority choices but more variation in the experiment (Nodes 3 vs. Node 4, Node 9 vs. Node 10). With Dutch, CELEX data and experimental data show opposite trends. While Dutch words with superheavy final syllable are almost categorically stressed on the final syllable (Node 6), the experimental data show a preponderance of antepenultimate stress and still sizable proportions of words with penultimate and final stress (Node 7).

any experimental data set. Note also that we tested eight different conditions and this provided considerable opportunity for the two datasets to differ from each other.

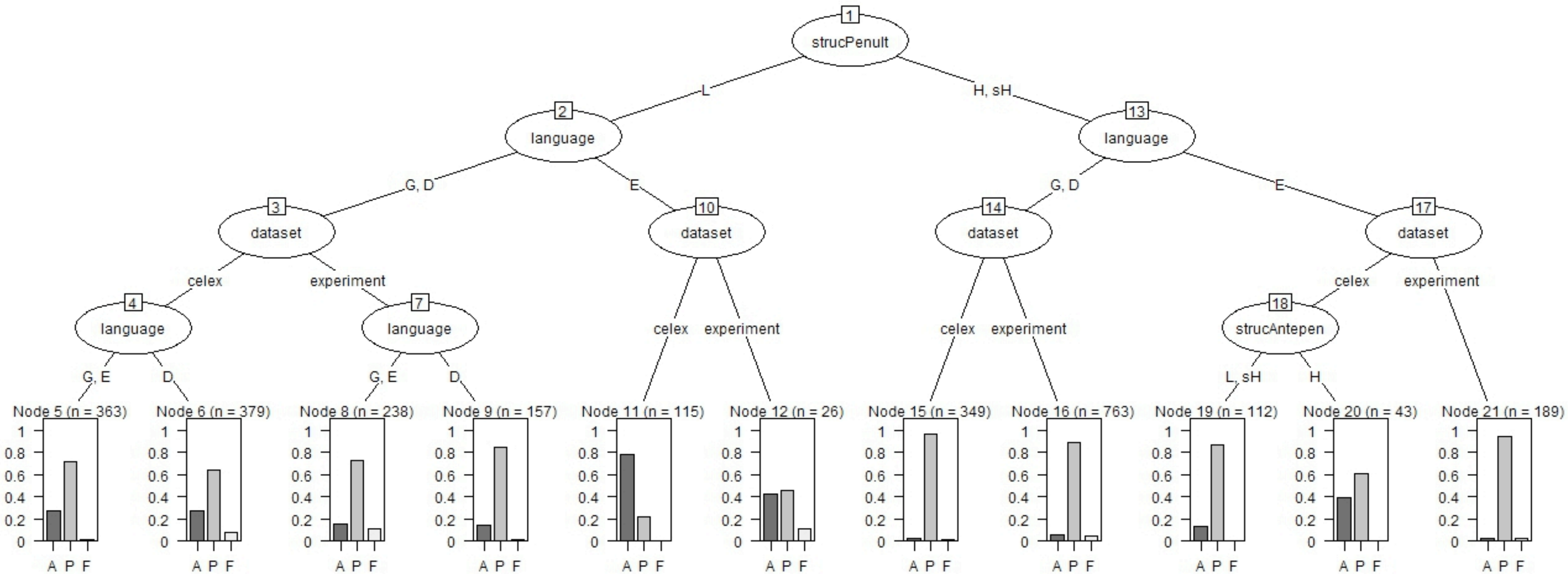

Figure 6: Subtree of the classification tree for words with a light final syllable, combined data sets, bar coding: dark grey = antepenult (A), grey = penult (P), light grey = final (F), German (G), Dutch (D), English (E)

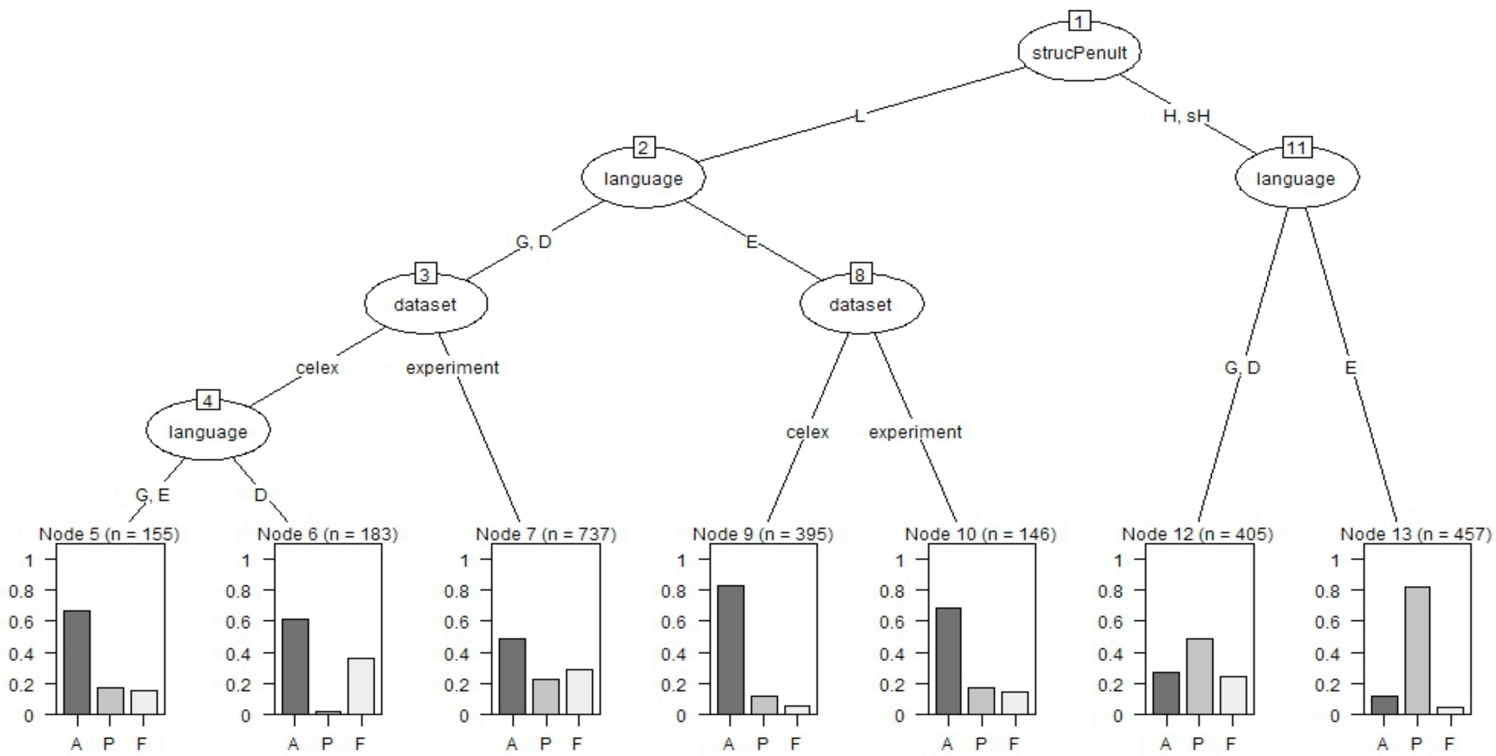

Figure 7: Subtree of the classification tree for words with a heavy final syllable, combined data sets, bar coding: dark grey = antepenult (A), grey = penult (P), light grey = final (F), German (G), Dutch (D), English (E)

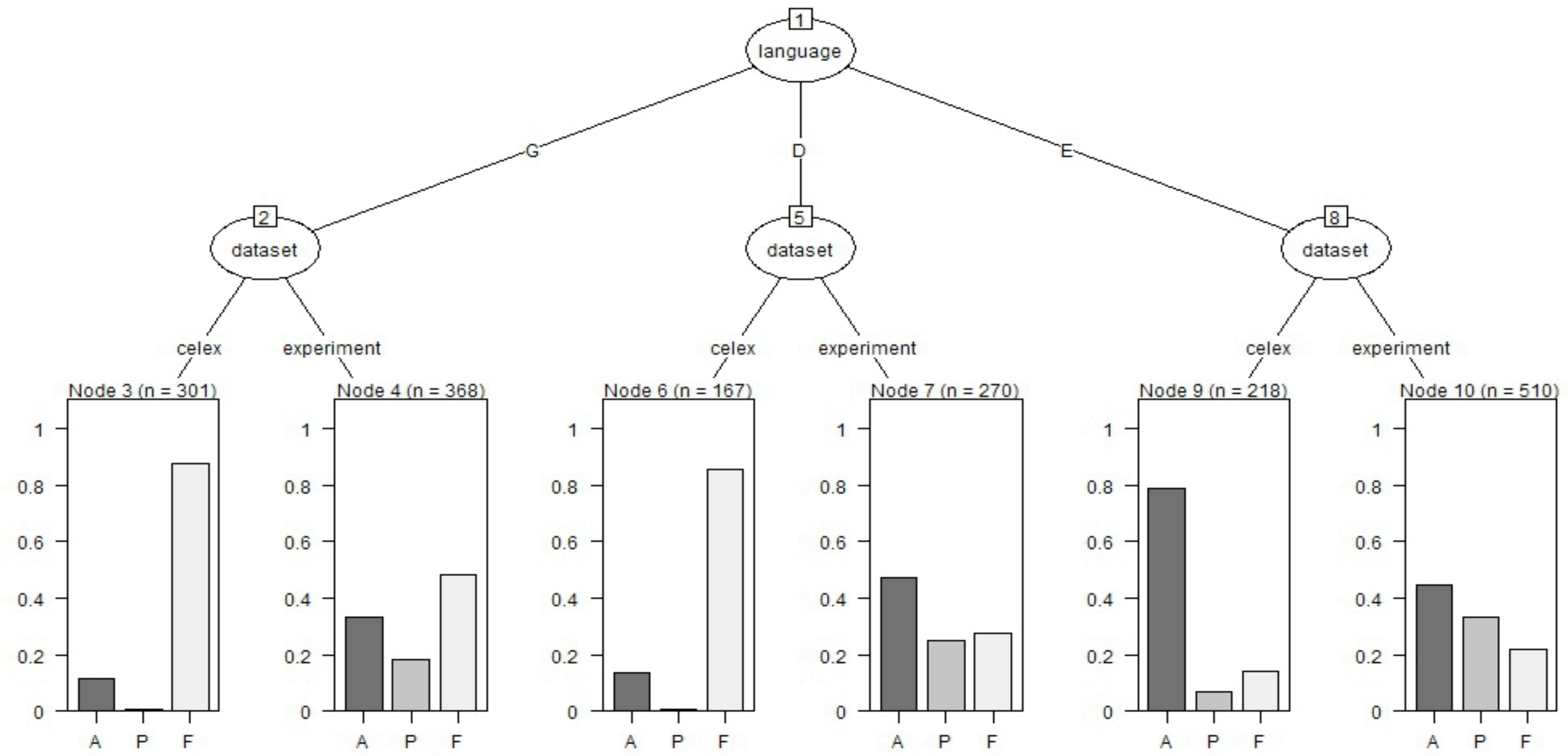

Figure 8: Subtree of the classification tree for words with superheavy final syllables, combined data sets, bar coding: dark grey = antepenult (A), grey = penult (P), light grey = final (F), German (G), Dutch (D), English (E)

4.3 Summary and discussion

To summarize, we can say that in all three languages, the lexical data and the experimental data largely show the same types of effect, with minor differences as to the degree of variability. The comparison of the two data sets demonstrates that the structure of the final syllable serves as a strong predictor for stress position in attested words as well. An interesting discrepancy between attested and unattested words can be found in the Dutch data, however, where superheavy final syllables are stressed categorically in attested words but not in pseudowords.

Overall, the similarities between the distribution of the two kinds of data sets and the occurrence of stress variation in forms with heavy final syllable strongly suggest that the experimental data are dependable and that probably the same metrical principles govern the distribution of stress in the two data sets. The nature of these principles will be discussed in more detail in the discussion in section 5.

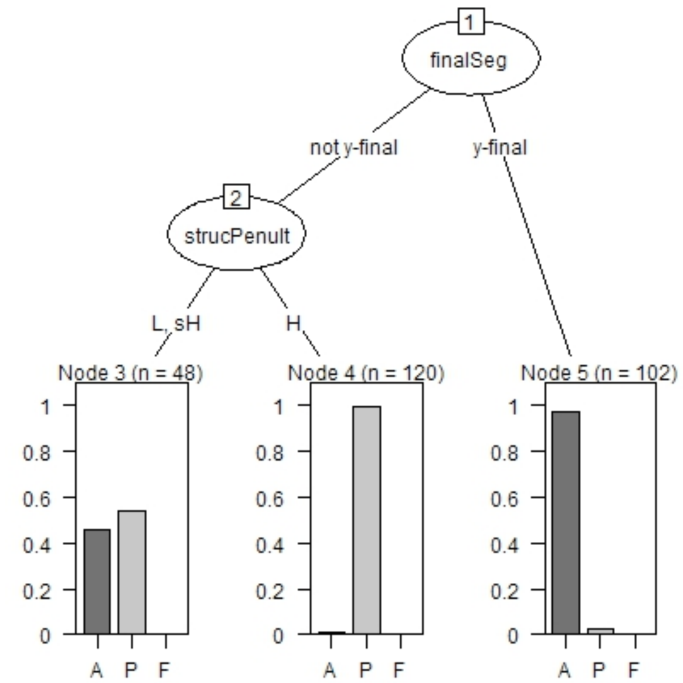

Figure 9: English CELEX data, effect of final <y>: dark grey = antepenult (A), grey = penult (P), light grey = final (F)

The close similarity in stress assignment between the experimental pseudowords and the established words may also have further implications for theories of stress assignment.

Given that in our experiment we sometimes find the effects predicted by the theoretical literature and sometimes we do not, and given that this behavior mirrors the distribution of stress in the lexicon, the question arises what may be responsible for the subjects' responses. One possibility is the existence of certain rules that assign stress in a deterministic fashion, given a certain input. A classic case of such a rule would be the 'Latin Stress Rule' for English, which says that the penultimate syllable is stressed if it is heavy, and that the antepenult is stressed if the penult is light. Although this rule would successfully predict a large proportion of the English speaker data in our experiment, there is considerable leakage. For example, the structure of the final syllable should not play any role. It does, however, with the speakers in our experiment. In a categorical, rule-based account such leakage is generally considered to be caused by 'exceptions', but it is unclear how large the number of exceptions should become before one starts doubting the rule. Overall, the amount of variability is so large in our experimental data that any categorical approach runs into very serious empirical problems (as outlined in the previouos section).

In recent years, much work has addressed the problem of variability in phonology in a non-categorical form. For example, Albright (2009) looked at word phonotactics in English, Plag and colleagues (e.g. Plag, Kunter, Lappe & Braun 2008, Plag 2010, Arndt-Lappe 2012) studied variable compound stress assignment in English, and Ernestus & Baayen (2005) investigated voice neutralization with word-final obstruents in Dutch. With all three phenomena, it is shown that the variable patterning of the data is best predicted on the basis of similarity to words in the lexicon. The emergence of phonological patterns from the lexicon can be formally modeled in various ways. Ernestus & Baayen, for example, compare five different kinds of model, Stochastic Optimality Theory, Generalized Linear Models, Classification and Regression Trees, and the two analogical models Spreading Activation and Analogical Modeling of Language (AML). All of these models are capabale of making rather successful and testable predictions based on the distribution of forms in the lexicon. Similarly, Plag and colleagues have implemented regression analyses and two kinds of analogical algorithms to predict stress assigment to compounds.

That word stress may also emerge from the lexicon is argued for in computational, analogy-based accounts, such as Arciuli, Monaghan, and Seva (2010), Daelemans, Gillis, and Durieux (1994), Gupta and Touretzky (1994), Zevin and Joanisse (2000). These studies use computational algorithms like TiMBL or AML that assign stress to a given form on the basis of that form's similarity with existing forms in the lexicon. Under the assumptions of such analogical or probabilistic models the distribution of forms in the lexicon would be predictive of the distribution of forms as produced by speakers in an experiment. In other words, the differences between the forms in the lexicon and the forms in an experimental data set should be minimal.

In contrast, under a categorical, rule-based approach, the lexicon would possibly contain a lot of forms that do not conform to the rules of the present-day speaker (e.g. long-established lexicalized forms), and the speakers in a production experiment would be expected to produce new forms in accordance with the assumed rules, irrespective of the distribution of forms in the lexicon.

Based on these considerations it is theoretically interesting to compare established words with experimental data. If the amount and kind of variation found in the lexicon mirrors that found in the experimental data this may be an indication that stress assignment to new forms works on the basis of similarity with existing lexical items. Of course, the distribution of forms in the lexicon may also be the result of categorical rules having produced forms that were then listed in the mental lexicon of the speakers. Categorical approaches would face, however, a serious problem if neither the distribution of the lexicalized forms nor of the experimental forms conforms to the output patterns expected under rule application.

It is beyond the scope of this paper to test in detail the hypothesis that stress assignment emerges from the lexicon. There are nevertheless some interesting subsets of data that are suggestive that such a study would be fruitful. We now turn and to these cases.

We noted above that the behaviour of Dutch final superheavies constitutes a peculiar exception to this general trend and merits closer inspection. An analysis of stressed superheavies in the Dutch CELEX data set reveals that 80% of them consist of VVC-rhymes (e.g. *mineraal* 'mineral', *kameraad* 'fellow') and that half of the VCC-rhymes end in the cluster <*nt*> (e.g. *argument*, *testament*). In contrast, the corresponding experimental stimuli do not share these properties. Instead they have the structure VCC and contain minority clusters such as <*mp*>, <*lk*>, or <*ms*>. This means that the lexicon does not contain many items over which speakers could generalize to these new forms. It does therefore not come as a surprise, and does in fact speak in favor of a lexicon-based account on stress assignment that we find the observed discrepancies of attested words vs. pseudowords with Dutch final superheavies. At a more theoretical level these facts may be interpreted against approaches that state a rule for superheavy final syllables to carry word stress. The would-be weight effect found in the

CELEX data could be reinterpreted as a segmental-skeletal effect on stress assignment, i.e. in words that end in a certain consonant cluster the final syllable is likely to attract main stress (in line with Daelemans, Gillis, and Durieux 1994).

A similar problem with a similar solution presents the discrepancy of English words with a light final syllable and a light penult. Let us take a closer look at the data to understand this discrepancy between the CELEX preference for antepenultimate stress and the even distribution of antepenultimate and penultimate stress in the experiment. Given that a majority of English trisyllables ending in <y> are stressed on the antepenultimate syllable (75% of the monomorphemes in CELEX and a vast majority of morphologically complex nouns with the suffix *–y*, Bauer, Lieber, & Plag, 2013: Chapter 12.2.5) it might be the case that the preference for antepenultimate stress in CELEX is due to the fact that many CELEX words end in <y>. An empirical analysis of the English CELEX data set including a variable encoding the final segment (`y` or `no y`) indeed brings out an effect of final segment as significant. As shown in Figure 9, words not ending in <y> prefer penultimate stress whereas words ending in <y> show antepenultimate stress. Taking the two kinds of words together, we end up with an almost even distribution of antepenultimate and penultimate stress as shown in Node 12 of Figure 6. None of the experimental stimuli ended in <y>, and the majority decision for these words is the same in CELEX and the experiment, i.e. penultimate stress.

Overall, the discrepancies between CELEX and the experiment turn out to result from particular constellations of certain structures, in which stress patterns as listed in CELEX are associated with specific segmental-skeletal properties of subsets of words. These specific properties, however, were not instantiated by the pseudowords in the experimental data sets. Under the assumption that speakers generalize to new forms on the basis of distributions in the lexicon, both the similarities and the discrepancies between the two data sets can be explained. We find similarities between pseudowords and existing words where speakers were able to find similar structures in the lexicon, and we find discrepancies between the two types of words where it is impossible for the speakers to detect similarities for the pertinent subsets.

It remains to be seen whether in future studies formal models of grammar such as Analogical Modeling of Language or Stochastic Optimality Theory are able to capture the patterning of the empirical data as gathered in this study.

5. Discussion and conclusion

In this paper we have systematically explored the similarities and differences of three West Germanic word stress systems as well as the controversies surrounding the quantity-sensitivity of these languages and related issues. In particular we posed a number of research questions, which we are now in a position to answer. We will discuss each in turn.

5.1 What is the role of syllabic weight in stress assignment in the three languages?

This question has three aspects, the first and most important of which is whether syllabic weight influences the position of word stress. We saw that for all three languages, this is indeed the case, independent of the type of data we look at. For German and English these results are unaccounted for under approaches that postulate that these languages are not quantity-sensitive (e.g. Eisenberg, 1991 and Wiese, 1996 for German; Kiparsky, 1982, 1985, and Booij & Rubach, 1992 for English).

The second, more specific, aspect is which syllables contribute to the stress decision. We found that in all three languages, the weight of the final and penultimate syllable strongly influence stress placement, while the structure of the antepenultimate syllable exerts a negligable influence (contrary to the results by Röttger et al., 2012 for German). The results for English are unaccounted for under (common) approaches that restrict the triggering of quantity effects to the penult (e.g. Chomsky & Halle 1968; Liberman & Prince 1977; Giegerich 1985, 1992; Hayes 1982; Kager 1989; Roca 1992; Trommelen & Zonneveld 1999a).

Thirdly, we wanted to know whether the weight of a given syllable determines whether this syllable itself receives stress. For German, this is indeed very clearly the case. If the final syllable is superheavy, chances are high that it is stressed, and if the penult is heavy it is also significantly more likely to be stressed than if it is light. For Dutch, similar effects hold, though not across both data sets. In the attested words the final superheavy syllable is almost categorically stressed, whereas in the nonce words stress on the final syllable is never dominant. If the penult is heavy it is also significantly more likely to be stressed than if it is light. For English, we find clear evidence for stress attraction to heavy or superheavy penults. In sum, all three languages show robust effects of heavy syllables attracting stress, lending further support to the claim that they are all quantity-sensitive.

5.2 Do we find evidence for extrametricality?

Given the dependence of stress assignment on the weight of the final and penultimate syllable in all three languages, there is no evidence for extrametricality at the level of foot structure formation. In other words, no syllable is completely extrametrical in either language. However, although the structure of the final syllable plays a role for the metrical analysis of English and Dutch words (in both CELEX and the experiment), the final syllable is least likely to receive main stress in these languages. This suggests that the final syllable is extrametrical at the word level in both languages, supporting earlier claims for Dutch (e.g. Kager, 1989, Trommelen & Zonneveld, 1999b), and contradicting widely-held assumptions for English that maintain that the final syllable be extrametrical at the foot level (Chomsky & Halle, 1968; Liberman & Prince 1977; Giegerich 1985, 1992; Hayes 1982; Kager 1989; Roca 1992; Trommelen & Zonneveld, 1999a). The data suggest that heavy final syllables build monosyllabic feet but do not receive main stress.

With respect to syllable weight and the function of final syllables in metrical structure, we can summarize that those accouts can best explain our data that assume quantity-sensitivity and the contribution of final syllables to foot structure formation. However, words with heavy and super-heavy final syllables show various cases of stress variation. This is problematic for any deterministic account on stress assignment. We therefore discuss alternative explanations for stress distributions in our studies in the following section.

5.3 Stress and foot structure

Our results also have implications for the parsing of syllables into feet. The overall picture taken from our pseudoword studies and corpus analyses is that the formation of foot structure seems to be mainly guided by the parameters shown in Table 1: heavy and superheavy final syllables build non-branching feet that allow trisyllabic words to consist of two feet, while words ending in a light syllable build only one single foot. Thus, the variability of stress assignment arises mainly from the fact that in words with a heavy final syllable there are two feet that are in principle stressable. The variation obtained in the pseudoword experiments for words with two stressable syllables shows that main stress assignment is not deterministic and is perhaps better accounted for by a probabilistic approach. This is also supported by the tree analyses, which show that the distribution of main stress in the pseudoword data is quite comparable to main stress placement in existing words.

5.4 How can we explain the distribution of the stresses in unattested words?

It is an interesting theoretical question whether the stress distributions we find in the experimental data can be better explained by the metrical rules proposed in the literature, or by making reference to the structural similarity of the new words to existing lexical items.

The comparison of the stress distributions by syllable structure of pseudowords with those already in the lexicon is suggestive in this repsect. In all three languages we found large overlaps in the majority choices of words with a particular structure, which speaks for a theory that takes stress to emerge from the lexicon, perhaps through analogical mechanisms. In Dutch and, to a lesser extent, in English, however, we also found interesting differences between pseudowords and attested words. The experiment could not replicate the generalization emerging from the lexicon that final superheavy syllables almost categorically attract stress in Dutch, or that words ending in two light syllables are stressed on the antepenultimate syllable in English. A closer analysis revealed for both languages, however, that these seeming discrepancies between CELEX and experiment can be explained as segmental-skeletal effects emerging from the lexicon.

It is still an open question by which particular mechanism the potentially lexically-driven stress patterns, and the variability in both experimental and lexical data can be modelled. Future research will have to show which of the analogy-based, connectionist, stochastic constraint-based or rule-based approaches is best suited to account for the intricate stress assignment patterns found in the empirical data.

5.5 Conclusion

To summarize, the results of a production experiment with trisyllabic pseudowords and of an analysis of large numbers of existing words from the CELEX lexical database provide clear evidence for quantity effects in all three languages and for the role of foot structure in stress assignment. In particular, our empirical results challenge some widely held theoretical assumptions about metrical properties related to quantity and extrametricality in the tree languages. German, English, and Dutch rely on the same basic principles of structure formation, but they differ with respect to the role of extrametricality and the way quantity influences stress assignment.

At a more abstract level the comparison of these effects in the lexical data and the experimental data revealed a great deal of similarity, which opens up promising research perpectives for the question of which mechanisms underlie the assignment of stress to new words.

Acknowledgments

Parts of the research presented here were supported by German Science Foundation Grant SPP 1234 Wi 853//-2 to Richard Wiese and Ulrike Domahs. Special thanks go to the Max-Planck Institute for Psycholinguistics (especially to Kerstin Mauth) in Nijmegen where the Dutch data were collected and to four anonymous reviewers.

References

Alber, Birgit. 1997. Quantity sensitivity as the result of constraint interaction. In *Phonology in progress: Progress in phonology*, ed. Gert Booij and J. van de Weijer, 1-45. The Hague: Holland Academic Graphics.

Albright, Adam. 2009. Feature-based generalisation as a source of gradient acceptability. *Phonology* 26, 9-41.

Arciuli, Joanne, Monaghan, Padraic, Seva, Nada. 2010. Learning to assign lexical stress during reading aloud: Corpus, behavioural and computational investigations. *Journal of Memory and Language* 63: 180-196.

Arndt-Lappe, Sabine 2011. Towards an exemplar-based model of English compound stress. *Journal of Linguistics* 47(3): 549-585.

Baayen, Harald R. 2008. *Analyzing linguistic data: A practical introduction to statistics*. Cambridge: Cambridge University Press.

Baayen, R. H., D. J. Davidson, D. M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59, 390-412.

Baayen, Harald R., Piepenbrock, Richard and Gulikers, L. 1995. *The CELEX lexical database. Release 2 [CD-ROM]*. Philadelphia: Linguistic Data Consortium, University of Pennsylvania.

Bates, D., Sarkar, D., Bates, M., and Matrix, L. (2007). The lme4 package. *R Package Version 2.4.1*.

Bauer, Laurie, Lieber, Rochelle, and Plag, Ingo (2013). *The Oxford reference guide to English morphology.* Oxford: Oxford University Press.

Booij, Geert and Rubach, Jerzy. 1992. Lexical Phonology. In *International Encyclopedia of Linguistics*, ed. William Bright, 293-296. Oxford: OUP.

Booij, Geert. 1999. *The Phonology of Dutch*. Oxford: Oxford University Press.

Burzio, Luigi. 1987. English Stress. In *Certamen Phonologicum: Papers from the 1987 Cortona Phonology Meeting*, ed. Pier M. Bertinetto and Michele Loporcaro, 153-175. Torino; Rosenberg and Sellier.

Burzio, Luigi. 1991. English vowel length and foot structure. In *Certamen Phonologicum II. Papers from the 1990 Cortona Phonology Meeting*, ed. Pier M. Bertinetto and Michael Kenstrowicz, 121-145. Torino: Rosenberg and Sellier.

Burzio, Luigi. 1994. *Principles of English Stress*. Cambridge Studies in Linguistics 72. Cambridge: CUP.

Chomsky, Noam and Halle, Morris 1968. *The sound pattern of English*. New York: Harper and Row.

Cutler, Anne and Norris, Dennis. 1988. The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance* 14: 113-121.

Cutler, Anne and van Donselaar, Wilma. 2001. Voornaam is not (really) a homophone: Lexical prosody and lexical access in Dutch. *Language and Speech* 44: 171-195.

Daelemans, Walter, Gillis, Steven and Durieux, Gert 1994. The acquisition of stress: a data-oriented approach. *Computational Linguistics* 20: 421-451.

Domahs, Ulrike, Wiese, Richard, Bornkessel-Schlesewsky, Ina and Schlesewsky, Matthias. 2008. The processing of German word stress: Evidence for the prosodic hierarchy. *Phonology* 25(1): 1-36.

Donselaar, van Wilma, Koster, Mariëtte, and Cutler, Anne. 2005. Exploring the role of lexical stress in lexical recognition. *Quarterly Journal of Experimental Psychology A* 58**:** 251-273.

Eisenberg, Peter. 1991. Syllabische Struktur und Wortakzent: Prinzipien der Prosodik deutscher Wörter. *Zeitschrift für Sprachwissenschaft* 10: 37-64.

Ernestus, Mirjam and R. Harald Baayen. 2003. Predicting the unpredictable: Interpreting neutralized segments in Dutch. *Language,* 2003*, 79*, 5-38

Ernestus, Miriam and Neijt, Anneke. 2008. Word length and the location of primary word stress in Dutch, German, and English. *Linguistics* 46 (3): 507-540.

Féry, Caroline. 1998. German word stress in Optimality Theory. *Journal of Comparative Germanic Linguistics* 2: 101-142.

Giegerich, Heinz. 1985. *Metrical phonology and phonological structure: German and English.* Cambridge: Cambridge University Press.

Giegerich, Heinz. 1992. *English phonology: An introduction*. Cambridge: CUP.

Guion Susan G., Clark, J. J., Harada, Tetsu and Wayland, Ratree P. 2003. Factors affecting stress placement for English nonwords include syllabic structure, lexical class, and stress patterns of phonologically similar words. *Language and Speech* 46(4): 403-427.

Gupta, Prahlad and Touretzky. Dave. 1994. Connectionist models and linguistic theory: Investigations of stress systems in language. *Cognitive Science* 18: 1-50.

Gussenhoven, Carlos. 2009. Vowel duration, syllable quantity, and stress in Dutch. In Kristin Hanson and Sharon Inkelas (eds). *The Nature of the Word. Essays in Honor of Paul Kiparsky.* Cambridge, MA/London, UK: MIT Press. 181-198.

Hammond, Michael. 2004. Gradience, phonotactics, and the lexicon in English phonology. *International Journal of English Studies* 4: 1–24.

Hall, Tracy A. 1992. *Syllable structure and syllable-related processes in German*. Tübingen: Max Niemeyer Verlag.

Hayes, Bruce. 1982. Extrametricality and English stress. *Linguistic Inquiry* 13:227-276.

Hayes, Bruce. 1995. *Metrical stress theory: Principles and case studies.* Chicago: University of Chicago Press.

Hothorn, Torsten 2009. CHAID package for R. http://r-forge.r-project.org/projects/chaid/ http://r-forge.r-project.org/projects/chaid/

Hothorn, Torsten and Achim Zeileis (2012) partykit: A Toolkit for Recursive Partytioning. Version 0.1-4.

Hulst, Harry G. van der 2003. Dutch syllable structure meets Government Phonology. In: Takeru Honma, Masao Okazaki, Toshiyuki Tabata and Shin-ichi Tanaka (eds.). *A new century of phonology and phonological theory: A festschrift for professor Shosuke Haraguchi on the occasion of his sixtieth birthday*. Tokyo, Japan: Kaitakusha, 313-343.

Janßen [Domahs], Ulrike. 2003. *Untersuchungen zum Wortakzent im Deutschen und Niederländischen*. Doctoral dissertation. University of Düsseldorf. Available (September 2012) http://deposit.ddb.de/cgi-bin/dokserv?idn=972217770&dok_var=d1&dok_ext=pdf&filena me=972217770.pdf

Janßen [Domahs], Ulrike and Domahs, Frank. 2008. Going on with optimised feet: Evidence for the interaction between segmental and metrical structure in phonological encoding from a case of primary progressive aphasia. *Aphasiology* 22(11): 1157-1175.

Jessen, Michael 1999. German. In *Word prosodic systems in the languages of Europe*, ed. Harry van der Hulst, 515-545. Berlin: Mouton de Gruyter.

Kager, René 1989. *A metrical theory of stress and distressing in English and Dutch.* Doctoral dissertation. (= Linguistic Models 14). Dordrecht: Foris Publications.

Kaltenbacher, Erika. 1994. Typologische Aspekte des Wortakzents: Zum Zusammenhang von Akzentposition und Silbengewicht im Arabischen und Deutschen. *Zeitschrift für Sprachwissenschaft* 13: 20-55.

Kass, G. V. 1980. An exploratory technique for investigating large quantities of categorical data. *Applied Statistics* 29(2): 119–127.

Kiparsky, Paul. 1982. From cyclic phonology to lexical phonology. In *The structure of phonological representations*. Vol.1, ed. Harry van der Hulst and Norval Smith, 131-175. Doordrecht: Foris.

Kiparsky, Paul. 1985. Some consequences of lexical phonology. *Phonology Yearbook* 2: 83-138.

Knaus, Johannes and Domahs, Ulrike. 2009. Experimental evidence for optimal and minimal metrical structure of German word prosody. *Lingua* 119: 1396-1413.

Lahiri, Aditi & Koreman, Jacques. 1988. Syllable weight and quantity in Dutch, *Proceedings of the 7th West-Coast Conference on Formal Linguistics,* 217-228.

Liberman, Marc Y. and Prince, Alan. 1977. On Stress and Linguistic Rhythm. *Lingustic Inquiry, 8*, 249-336.

Oostendorp, Marc van. 1995. *Vowel quality and syllable projection*. PhD Dissertation, Tilburg University.

Pater, Joe. 2000. Non-uniformity in English secondary stress: The role of ranked and lexically specific constraints. *Phonology* 17: 236-274.

Plag, Ingo, Kunter, Gero, Lappe, Sabine, and Braun, Maria. 2008. The role of semantics, argument structure, and lexicalization in compound stress assignment in English. *Language* 84.4, 760-794.

R Development Core Team. 2011. *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing.*Vienna: Austria. ISBN 3-900051-07-0.

Rietveld, Toni, Kerkhoff, Joop, & Gussenhoven, Carlos. 2004. Word prosodic structure and vowel duration in Dutch. *Journal of Phonetics*, 32, 349-371.

Roca, Iggy. 1992. Constraining extrametricality. In *Phonologica 1988,* ed. Wolfgang Dressler, Hans Lüschutzky, Oskar Pfeiffer and John Rennison, 239-248. Cambridge: CUP.

Röttger, Timo, Domahs, Ulrike, Grande, Marion, and Domahs, Frank. 2012. Structural factors affecting the assignment of word stress in German. *Journal of Germanic Linguistics* 24 (1): 53-94.

Smith, Philip T. and Baker, Robert G. 1976. The influence of English spelling on pronunciation. *Journal of Verbal Learning and Verbal Behavior* 15: 267-285.

Trommelen, Mike and Zonneveld, Wim. 1999a. English. In *Word prosodic systems in the languages of Europe*, ed. Harry van der Hulst, 478-491. Berlin: Mouton de Gruyter.

Trommelen, Mike and Zonneveld, Wim. 1999b. Dutch. In *Word prosodic systems in the languages of Europe*, ed. Harry van der Hulst, 492-515. Berlin: Mouton de Gruyter.

Vennemann, Theo. 1990. Syllable structure and simplex accent in Modern Standard German. *Chicago Linguistic Society,* 26(2): 399–412.

Vennemann, Theo. 1991. Syllable structure and syllable cut prosodies in Modern Standard German. In *Certamen Phonologicum II: Papers from the Cortona Phonology Meeting 1990*, ed. Pier M. Bertinetto, Michael Kenstowicz and Michele Loporcaro, 211-245. Torino: Rosenberg and Sellier.

Vennemann, Theo. 1995. Der Zusammenbruch der Metrik im Spätmittelalter und sein Einfluss auf die Metrik. In *Quantitätsproblematik und Metrik,* ed. Hans Fix. Greifswalder Symposion zur germanistischen Grammatik, 185-223. Amsterdam: Radopi.

Wiese, Richard. 1996/2000. *The phonology of German.* Oxford: Oxford University Press.

Zevin, Jason and Joanisse, Marc. 2000. Stress assignment in nonword reading. *Journal of Cognitive Neurosciences* 41B: S5.

Zonneveld, Wim, Trommelen, Mike, Jessen, Michael, Bruce, Gösta and Árnason, K. 1999. Word stress in West-Germanic and North-Germanic languages. In *Word prosodic Systems in the Languages of Europe*, ed. Harry van der Hulst, 477-603. Berlin: Mouton de Gruyter.

Zonneveld, Wim and Nouveau, Dominique. 2004. Child word stress competence: an experimental approach. In *Constraints in phonological acquisition*, ed. Kager, René, Pater, Joe and Zonneveld, Wim, 369-408. Cambridge: Cambridge University Press.

Appendix Pseudowords

A German Pseudowords

vc.v.vcc	**v.vc.vcc**	**vc.v.vc**	**v.v.vc**
Gam.do.kust	*Ga.dom.kust*	*Bin.sa.kaf*	*Bä.lu.ful*
Hul.sa.domp	*Hu.sal.domp*	*Dim.pu.sat*	*Dä.mu.pok*
Kon.fa.sorp	*Ko.fan.sorp*	*Ful.go.but*	*Fe.ko.mot*
	Lü.ras.palf	*Gos.ta.ful*	*Go.nü.bak*
	Mä.fal.rolk	*Hol.ta.pok*	*Hü.ba.guf*
Nom.pu.lams	*No.pur.lams*	*Kis.nu.mot*	*Ke.ga.fur*
Pas.ro.garf	*Pa.ros.garf*	*Lan.do.guf*	*La.fö.was*
Rul.ko.menk	*Ru.kol.menk*	*Mur.la.was*	*Va.so.haf*
Ser.da.nulz	*Se.daf.nulz*	*Pel.no.fur*	*Wo.ta.sat*
Tak.lu.tarp	*Ta.luf.tarp*		*Zö.da.but*

v.vc.vc.	**v.v.v**	**v.vc.v.**	**vc.vc.v**
Bo.kam.was	*De.to.pu*	*Bu.mol.ta*	*Dir.san.ra*
Da.pön.bak	*Fu.sa.fo*	*Me.fal.bo*	*Kat.lön.bo*
Fä.lus.fur	*Ga.rö.so*	*Na.dur.so*	*Kum.sak.pu*
Go.rum.ful	*Ho.bu.lo*	*Pa.lön.fo*	*Las.fon.ta*
Ra.bosch.kaf	*Kä.na.ra*	*Ro.taf.ku*	*Mok.nas.fo*
Si.fas.mot	*Lö.mu.da*	*Sä.lot.ga*	*Nel.kum.lo*
Tu.lor.sat	*Mü.la.ku*	*Tö.pum.lo*	*Rän.gul.da*
Vu.pal.but	*Ne.kü.ga*	*We.läs.ra*	*Sap.wur.ku*
Wo.sol.pok	*Pa.go.ta*	*Vü.sol.da*	*Täm.buf.ga*
Zü.bal.guf	*Ru.ga.bo*	*Zi.lat.pu*	*Wis.top.so*

B English Pseudowords

vc.v.vcc	**v.vc.vcc**	**vc.v.vc**	**v.v.vc**
rulcomest	*sadufnalf*	*binsacub*	*baloguth*
masforuld	*parosgarf*	*gostafaz*	*lamopog*

nompolans	*gadomcust*	*holtanof*	*lafowoth*
serdanuls	*huzaldomp*	*folmadoth*	*fekomof*
lusrapalf	*bofangond*	*lundogof*	*gonusab*
tilcopalt	*mafalrolt*	*molravos*	*hanogaf*
confagond	*nopumlans*	*fulgobog*	*cagafoth*
holsadomp	*rucolmest*	*pelnofut*	*sudabod*
posragols	*niraspalf*	*zalfolup*	*votasat*
gamdocoft	*tolufpalt*	*thimravas*	*masocath*

v.vc.vc.	**v.v.v**	**v.vc.v.**	**vc.vc.v**
bocamvas	*hobalu*	*tholatpo*	*tambufga*
falosfuth	*dotopu*	*visalda*	*molnasfo*
zefasmof	*gerosu*	*rocafta*	*nelcumlo*
sudalgaf	*cabora*	*tosumlo*	*cumzacto*
goromfod	*mulako*	*nadalco*	*lisfonta*
dapunbod	*losuda*	*velasra*	*domsanro*
raboshgat	*pulota*	*palonfo*	*cathlonbo*
godolpog	*fotafo*	*silatpa*	*wistocso*
tulasrup	*nipago*	*bomolta*	*rongalda*
nopalbol	*rogaba*	*mefalbo*	*lupvulco*

C *Dutch Pseudowords*

vc.v.vcc	**v.vc.vcc**	**vc.v.vc**	**v.v.vc**
gandokost	*gadomkost*	*binsakaf*	*belufol*
holsadomp	*hosaldomp*	*dimposat*	*demopok*
konfasorp	*kofansorp*	*falgobot*	*fekomot*
	luraspalf	*gostafol*	*gonubak*
metfarolk	*mefalrolk*	*holtapok*	*hubagof*
nompolams	*noporlams*	*kisnomot*	*kegafor*
pasrogarf	*parosgarf*	*landogof*	*lafuwas*

rulkomenk	*rokolmenk*	*morlawas*	*vasohaf*
serdanols	*sedafnuls*	*pelnofor*	*wotasat*
taklotarp	*taloftarp*		*zudabot*

v.vc.vc.	**v.v.v**	**v.vc.v.**	**vc.vc.v**
bokamwas	*detono*	*bomolta*	*dirsanra*
daponbak	*fosafo*	*mefalbo*	*katlondo*
felosfar	*garoso*	*nadorso*	*komsakpo*
goromfol	*hobolo*	*palonfo*	*lasfonta*
raboskaf	*kenara*	*rotafko*	*moknasfo*
sifasmot	*lumoda*	*selotga*	*nelkomlo*
tolorsat	*mulako*	*silatpu*	*rengolda*
wosolpok	*nekuga*	*tupomlo*	*sapworko*
vopalbot	*pagota*	*vusolda*	*tembofga*
zubalgof	*rogabo*	*welesra*	*wistopso*