J. Linguistics **48** (2012), 485–520. © Cambridge University Press 2012
doi:10.1017/S0022226712000199 First published online 30 July 2012

Informativeness is a determinant of compound stress in English[I]

MELANIE J. BELL

Anglia Ruskin University

INGO PLAG

Universität Siegen

(Received 29 October 2010; revised 23 May 2012)

There have been claims in the literature that the variability of compound stress assignment in English can be explained with reference to the informativeness of the constituents (e.g. Bolinger 1972, Ladd 1984). Until now, however, large-scale empirical evidence for this idea has been lacking. This paper addresses this deficit by investigating a large number of compounds taken from the British National Corpus. It is the first study of compound stress variability in English to show that measures of informativeness (the morphological family sizes of the constituents and the constituents' degree of semantic specificity) are indeed highly predictive of prominence placement. Using these variables as predictors, in conjunction with other factors believed to be relevant (see Plag et al. 2008), we build a probabilistic model that can successfully assign prominence to a given construction. Our finding, that the more informative constituent of a compound tends to be most prominent, fits with the general propensity of speakers to accentuate important information, and can therefore be interpreted as evidence for an accentual theory of compound stress.

1. Introduction

An idiosyncratic aspect of Present-day English is that, while many noun–noun (henceforth NN) combinations are pronounced with the left-stress pattern characteristic of Germanic compounds, there are also many combinations that normally have right prominence. Some examples are given in

[1] The authors wish to thank Sabine Arndt-Lappe, Kristina Kösling, Gero Kunter and three anonymous *Journal of Linguistics* referees for their feedback on earlier versions. Special thanks also to Harald Baayen for discussion and support. This work was made possible by an AHRC postgraduate award (114 200) and a major studentship from Newnham College, Cambridge, to the first author as well as two grants from the Deutsche Forschungsgemeinschaft (PL151/5-1, PL 151/5-3) to the second author, all of which are gratefully acknowledged.

https://doi.org/10.1017/S0022226712000199 Published online by Cambridge University Press

(1a, b). Here, and in examples of compounds throughout the text, capital letters indicate the syllable that is usually most prominent:

(1)	(a)	TAble lamp	(b)	silk SHIRT
		CREdit card		christmas DAY
		TEA cup		kitchen SINK

Most native speakers of English would produce the types in (1a) with main stress on the first element but the types in (1b) with main stress on the second element. Of course, this can change in a contrastive context. In *They're coming on CHRISTmas day, not BOXing day*, the compound *christmas day*, normally stressed on the second element, receives contrastive stress on the first element. However, in this paper we are concerned not with contrastive contexts, but with the characteristic prominence patterns of NN combinations spoken in a neutral context or in their citation form.

A note on terminology is in order here. The terms used in studies of compound prominence patterns are somewhat confused. Some scholars prefer the term 'stress', others speak of '(prosodic) prominence'. The particular choice of terminology is often dependent on the authors' theoretical assumptions about the phenomenon. For example, the term 'prominence' seems to be used by people favouring an analysis of compound stress in terms of pitch accents, rather than lexical stresses. Since there is no theory-neutral term available, we use both terms more or less interchangeably in this paper.

Despite the issue of compound stress assignment having received considerable attention from scholars of English over more than a hundred years, there is still no fully satisfactory explanation of the facts and no completely successful way of predicting which prominence pattern will apply to any given combination of nouns. This paper uses data from the British National Corpus (BNC) and from a large-scale production experiment to investigate a particular hypothesis about compound stress assignment. The hypothesis rests on two assumptions. Our first assumption, based on acoustic studies of compound stress (Farnetani, Torsello & Cosi 1988, Plag 2006, Kunter & Plag 2007, Kunter 2011), is that different compound prominence patterns are realised by differences in the distribution of pitch accents: in left-prominent compounds, only the first constituent is accented, while in right-prominent compounds, both constituents are accented. Our second assumption is that, in general in language, uninformative elements tend to be unaccented, while more informative and unexpected information is accented. On the basis of these assumptions, we hypothesise that a compound's stress pattern is at least partly determined by the informativeness of its second constituent. The hypothesis predicts that an uninformative constituent in the right-hand position will not receive an accent, i.e. the compound will be left-stressed. On the other hand, a highly informative constituent in the right-hand position

https://doi.org/10.1017/S0022226712000199 Published online by Cambridge University Press

will receive an accent, i.e. the compound will be right-stressed (see discussion of phonetically-grounded studies in Section 2.1).

We investigate a number of measures of informativeness. By combining these with other established determinants of compound stress, we are able to construct a probabilistic model that achieves a higher rate of success in predicting NN prominence than any method previously proposed in the literature. Overall, our study provides strong empirical evidence for the important role of informativeness in compound stress assignment.

The paper is organised as follows: Section 2 outlines previous attempts to explain the variation in NN prominence, Section 3 describes the methodology used in the present study, Sections 4 and 5 describe and discuss the results of our analyses, and Section 6 summarises the findings and discusses their implications for theories of compound stress.

2. The issues

2.1 *General background*

Throughout the twentieth century, in keeping with the prevalent linguistic paradigms of the time, linguists sought rules by which to assign prominence to English NN combinations. One of the most influential proposals was that of Chomsky & Halle (1968: 15–18), who coined the 'Compound Rule' and the 'Nuclear Stress Rule'. The 'Nuclear Stress Rule' was said to account for the normal prominence pattern in English phrases, in which the last strong syllable is the most prominent, i.e. carries the nuclear stress. On the other hand, the 'Compound Rule' was taken to assign primary stress to the main-stressed vowel of the first element of a binomial compound. Taken at face value, this suggests that NN combinations with prominence on the first element are compounds, whereas those with prominence on the second element are phrases. However, Chomsky & Halle (*ibid.*) make no attempt to define by other criteria the strings to which each of these rules applies. This means that, if taken literally, the rules are circular: left stress is assigned to compounds and compounds are defined as those combinations that receive left stress. Obviously, such a rule is unworkable, and in fact very few authors have taken it literally; Chomsky & Halle themselves (*ibid.*: 156) point out that there are several kinds of exception to the Compound Rule and that there is a need for 'an investigation of the conditions, syntactic and other, under which the Compound Rule is applicable'. So, if some NN combinations are to be analysed as phrases, there are two questions to answer: (i) What are the criteria by which the two classes can be distinguished? and (ii) Why do some combinations that most scholars regard as compounds nevertheless have right prominence?

In fact, as argued by Bauer (1998), Olsen (2000) and Bell (2005, 2011), there is very little evidence for a class of phrasal NNs. In this paper we will

https://doi.org/10.1017/S0022226712000199 Published online by Cambridge University Press

therefore use the term 'compound' for all NN constructions. We will investigate sequences consisting of two, and only two, adjacent nouns, in which one modifies the meaning of the other, or where together they have a single meaning different from the meaning of either constituent individually. However, proper names, such as *Laurie Bauer*, and constructions with an appositive modifier, such as (*my*) *sister Lillian*, are excluded.

The question then becomes: Under what circumstances do Present-day English NN compounds receive left prominence and under what circumstances do they receive right prominence? The problem for any account that tries to answer this question in terms of a rule is that, in order to achieve coverage of the empirical facts, it becomes necessary to append a significant list of exceptions; at least, that has been the case for all rules so far postulated. For example, Fudge (1984: 144–146) states that the majority of English NN compounds are 'initially-stressed', i.e. left-prominent, and then lists seven classes of exception. The first six of these involve identifiable semantic categories, and might therefore be regarded as rule-governed, but the final class, labelled 'miscellaneous cases', is not susceptible to such an analysis. Furthermore, within his lists of exceptions, Fudge (*ibid.*: 144) asterisks exceptions to the exceptions, where 'initial stress is an alternative possibility'. Another example of a rule-based approach is that of Giegerich (2004), who suggests that the basic distinction between left- and right-prominent compounds is that in left-prominent types the first noun (N1) is a complement of the second noun (N2), as in *OPera singer*, whereas in right-prominent types N1 is an attribute of N2, as in *steel BRIDGE*. The problem for this hypothesis is that there are many left-prominent combinations where N1 clearly takes the role of attribute rather than complement (e.g. *OPera glass*), and Giegerich accounts for these exceptions by suggesting that they are the product of a diachronic process of lexicalisation. Yet this implies that, contrary to the facts, attribute-type compounds cannot be coined with left prominence: so Giegerich, like Fudge, has to invoke exceptions to the exceptions. His solution is to argue that, once certain attribute-type compounds become listed with left prominence through a diachronic process of lexicalisation, others can be directly formed by analogy.

One of the reasons why categorical approaches struggle to account for the facts is that compound prominence itself is not as categorical as a rule-based approach would suggest. It is well known that some compounds vary with dialect, e.g. *BOY scout* in American English but *boy SCOUT* in British English, and that others show free variation, e.g. *ICE cream* or *ice CREAM*. Jespersen (1909: 155), for example, writes that 'individual pronunciations vary not a little on this point, and – what is very important – … "*level stress*" [i.e. right prominence] *really means* "*unstable equilibrium*"' (original emphasis). Other authors, e.g. Bauer (1978, 1983), Levi (1978), Pennanen (1980) and Kunter (2010, 2011), have shown that prominence varies not only between but also within speakers, both in production and perception.

https://doi.org/10.1017/S0022226712000199 Published online by Cambridge University Press

In addition to this variation, there are cases where pairs of compounds with evidently similar structure and semantics consistently receive contrary prominence patterns, depending on the identities of their constituents, for example *APPle cake* and *LEMon cake* compared with *apple PIE* and *lemon PIE* (see Sampson 1980 for an overview of the problems).

Apart from the difficulties caused by the complexities of the language itself, there are also problems associated with the methodology used in most twentieth-century studies. In virtually all cases, authors used fairly small datasets, usually selected to illustrate specific points, and assigned prominence to them on the basis of their own intuition, thus effectively ruling out any opportunity to study the variation in prominence found in actual speech. Many accounts in fact recycle the same small number of examples from previous papers. Furthermore, as hinted in the preceding paragraphs, there is considerable variation in terminology. 'Uneven stress' (Sweet 1892: 288), 'unity stress' (Jespersen 1909: 152), 'one high stress' (Bloomfield 1935: 180), 'compound stress' (Chomsky & Halle 1968: 15–18), 'single stress' (Jones 1972: 257), 'fronted stress' (Levi 1978: 41), 'initial stress' (Fudge 1984: 144), 'forestress' (Zwicky 1986) and 'lefthand stress' (Liberman & Sproat 1992) are amongst the many terms used for what we are calling 'left prominence'. The respective terms for our 'right prominence' are 'even stress', 'level stress', 'two high stresses', 'nuclear stress', 'double stress', 'normal (nonfronted) stress', 'final stress', 'afterstress' and 'righthand stress'. It can be seen from the variety of terms that there is some uncertainty about whether the relevant distinction is between one and two stresses or between two possible positions of a single stress. In some analyses three possibilities are admitted: stress on the first constituent only, stress on the second constituent only, or stress on both.

Phonetically-grounded studies (Farnetani et al. 1988, Plag 2006, Kunter & Plag 2007, Kunter 2011) have shown, however, that at the phonological level, only two patterns should be assumed, which we call 'left prominence' and 'right prominence'. Right-prominent compounds are acoustically characterised by more or less level pitch and intensity across the two constituents, while left-prominent compounds exhibit lower pitch and intensity on the right element. These acoustic characteristics are in correspondence with theories that take different compound prominence patterns to be realised by differences in the distribution of pitch accents: in left-prominent compounds, only the first constituent is accented, while in those with right prominence, there is a pitch accent on each constituent (see Gussenhoven 2004: 19, 276f.; Kunter 2011: 66–67). When both constituents are accented, even if the second accent is no higher than the first, the right-hand constituent is perceived as more prominent because the expected declination in pitch does not occur. The question that we will investigate in this paper thus boils down to whether the second noun will receive an accent or not.

As we have seen, categorical rules are unsuccessful at predicting which of these patterns of accentuation will apply to a particular NN compound.

https://doi.org/10.1017/S0022226712000199 Published online by Cambridge University Press

Recognising this, Kingdon (1958: 149–152), for example, declines to give rules and states only that 'one or two indications can be given which will help in deciding on the correct stress' (*ibid.*: 150). The indications he gives, involving firstly semantics and secondly the possibility of contrast (an indication of the relative informativeness of the constituents), have been recognised by many authors before and since, and the ideas date back at least to Sweet (1892: 288–289). In fact, there has been significant consensus amongst Anglicists for over a century about the factors that influence compound prominence. The problem has been that most, if not all, of these factors do not apply in a categorical fashion, and it was not until early in the twenty-first century that alternative, non-categorical approaches were systematically applied to the question. For example, Plag, Kunter & Lappe (2007) explicitly compared rule-based models of compound prominence with two types of non-categorical model: logistic regression and analogical. They found that both types of non-categorical model far outperformed the rule-based approaches in correctly predicting prominence.

The factors held responsible for the distribution of the two stress patterns are numerous, and include semantics, argument structure, lexicalisation, constituent length, constituent identity and informativeness. The recent large-scale studies by Plag and colleagues (Plag 2006, 2010; Plag et al. 2007, 2008; Arndt-Lappe 2011; Kunter 2011) have found substantial evidence for the influence of semantic factors, lexicalisation and constituent identity, but empirical evidence for the role of informativeness is lacking. In this paper we will fill this gap and investigate the role of informativeness, alongside other factors that have already been shown to influence the assignment of compound prominence.

2.2 *Informativeness*

The idea that informativeness may play a role in compound stress assignment can be traced back at least to Sweet (1892), who uses the term 'logical prominence' for what we call 'informativeness'. He writes that the left prominence of compounds 'seems to be the result of the second element being less logically prominent than the first, through being a word of general meaning and frequent occurrence in compounds' (*ibid.*: 288). Obviously, Sweet's reasoning rests on the still widely held assumption that, in general in language, uninformative elements tend to be unaccented, while more informative and unexpected information is accented. He identifies two ways in which a constituent can be relatively uninformative: by 'being a word of general meaning', i.e. semantically non-specific, and by occurring frequently in compounds, i.e. being relatively likely to be found in that position.

Several authors have followed in Sweet's footsteps. Marchand (1969: 23) states that '[t]he frequent occurrence of a word as second constituent is apt to give compound character [left prominence] to combinations with such

https://doi.org/10.1017/S0022226712000199 Published online by Cambridge University Press

words'. Although he does not refer to the notion of informativeness, the effect of a high positional frequency of a compound constituent can be explained as an informativeness effect. The frequent occurrence of a word as second constituent makes that constituent more predictable, and hence less informative, vis-à-vis the other constituent, hence prone to be deaccented.

Jones (1972: 259) makes essentially the same point: 'When the second element of a compound is felt to be of special importance, double stress [right prominence] is used'. Bolinger (1972) and Ladd (1984) suggest that this effect of informativeness can be understood in terms of implicit contrast: so in *silk SHIRT*, for example, there is an implied contrast with *silk DRESS*, *silk TIE*, etc.

Despite the long existence of these perhaps intuitively plausible claims, the role of informativeness in compound prominence assignment has, until recently, never been empirically tested. However, there have been attempts to predict pitch accent placement in general on the basis of measures of informativeness. In information theory, a standard measure of 'information content' is the negative log likelihood of a word in a corpus (Shannon 1948): the more frequent a word, and therefore the more likely, the less information it is taken to convey. Pan & McKeown (1999) show that pitch accent placement in texts can be quite successfully predicted on the basis of this variable: the less frequent a word, and therefore the more informative, the more likely it is to be accented. They also investigated another measure of informativeness, which was based on semantic specificity and used the distribution of a word across different texts in a corpus. Again, the greater the information content of a word, the more likely it was to receive an accent. Pan & Hirschberg (2000) looked at the effect of local context on accent placement. They found that bigram predictability, the probability of a particular word occurring in a context immediately following the previous word, was a useful predictor of accentuation. The more likely a word, given the occurrence of the previous word, the less likely it was to be accented. These studies did not, however, address the particular issue of stress assignment in NN compounds.

Plag & Kunter (2010) is the only previously published study to have investigated the effects of informativeness on compound prominence. As a measure of informativeness, they used the positional family size of the constituents, i.e. the number of compound types in a corpus that share a constituent in a given position. Each compound has a left constituent family and a right constituent family. For example, the left constituent family of *country house* would include compounds such as *country club*, *country music* and *countryside*, while the right constituent family would feature compounds like *town house*, *jailhouse* and *summer house*. The informativeness hypothesis can therefore be formulated in terms of family sizes: the larger the positional family of N2, the more expected is N2 in this position and so the less informative it will be. However, as well as family size, Plag & Kunter (*ibid.*)

https://doi.org/10.1017/S0022226712000199 Published online by Cambridge University Press

also took the so-called 'constituent family bias' into account (see Plag 2010), which is analogical in nature and denotes the tendency for compounds with a particular word in N1 or N2 position to have the same prominence pattern as other compounds with that word in the same position. In all three of their corpora, they found only weak effects of family size, and these effects interacted with the constituent family bias in such a way that for larger families the family bias effect became stronger. The informativeness measure thus only acted as a modulator for the much more significant family bias effect. However, their family sizes were based on quite small corpora, and it is possible that more robust measures of informativeness would have produced stronger effects.

In this study we use all three types of informativeness measure discussed above: absolute predictability, relative predictability and semantic specificity. Given that N1 will be accented irrespective of its informativeness, we focus our attention on the informativeness of N2, either by itself, or in relation to the informativeness of N1.

Our first set of informativeness measures are based on the frequencies of the constituents and of the compound. We will use a slightly simpler variant of Shannon's (1948) 'information content', namely the frequency of N2. The more frequent the second constituent, the less informative it is, and so our hypothesis is that the less likely it is to receive stress. Our second measure expresses the relative informativeness of N2 vis-à-vis N1. Mathematically, this is the conditional probability of N2 given N1, which can be expressed as the compound's frequency divided by the frequency of N1. This is the proportion of times that, when N1 occurs, it is in the context of a compound with N2. The higher this conditional probability, the less informative the second constituent will be and, according to our hypothesis, the less likely it is to be accented.

We also include two conditional probability measures based on family sizes. The first of these is simply the family size of N2, which is a measure of the probability of N2 occurring as the second element in a NN compound. The second family size measure is the probability of the occurrence of N2 given the family size of N1. If N1 has a large family, a particular N2 is less probable, and therefore more informative. Mathematically this conditional probability based on family size can be expressed as 1 divided by the family size of N1. These two family size measures are type-based measures, i.e. measures that count the number of different compounds in which a constituent occurs. It would also be possible to use token frequency, i.e. the summed frequencies of all such compounds. However, evidence from psycholinguistics suggests that, in the case of compounds, the psychologically most salient measure is the type frequency (Schreuder & Baayen 1997). We would therefore expect that, if family sizes are predictive of compound prominence, the type frequency would be the best predictor. Preliminary tests using both type-based and token-based family size measures confirmed

https://doi.org/10.1017/S0022226712000199 Published online by Cambridge University Press

this: the type-based measures were indeed more highly predictive, and so we used only type-based family size measures in our subsequent models.

Our final parameter of informativeness is semantic specificity. Semantically highly-specific words can be considered more informative than words that have a less specific meaning. In psycholinguistics, specificity of meaning has been operationalised with the help of so-called 'synsets'. This measure is based on the Wordnet lexical database (Fellbaum 1998), where a synset is a group of words with similar meanings: the greater the number of synsets to which a word belongs, the more highly polysemous and the less specific in meaning it is. Words with more restricted meanings are assumed to convey more specific information, and therefore to be more informative, than words with a greater number of meanings. Hence, we would expect right prominence to be less likely in compounds where N2 has a high synset count, i.e. belongs to a large number of synsets, and is therefore relatively non-specific. If relative informativeness is important, then we might also expect the relationship between the synset count of N1 and the synset count of N2 to be significant: if N1 is highly specific, we would predict that N2 will have to be even more specific to receive an accent than if N1 were non-specific.

2.3 *Semantics*

The semantics of a compound has three elements: the meaning of the first noun (N1), the meaning of the second noun (N2) and the relationship between them (R). For example, in the case of *table lamp*, the relationship between TABLE and LAMP is that a table lamp is a lamp designed to stand on a table. In the case of *silk shirt*, the relationship between SILK and SHIRT is that a silk shirt is a shirt made of silk. Because R is not overtly expressed in the compound, there is considerable scope for ambiguity and semantic opacity. A much quoted example refers to compounds with *oil*, where *olive oil* denotes oil extracted from olives, but *baby oil* refers to oil for rubbing on babies; similarly *peach thing* could denote something peach coloured, made of peaches, or designed for holding peaches, amongst other possibilities.

There are many claims in the literature that right prominence in compounds is determined by the semantics of the constituents or the semantic relation between the constituents (see, for example, Plag et al. 2008 for a review of that literature). Large-scale empirical studies have found that some of the predicted effects are indeed significant, but they are probabilistic in nature, rather than being categorical rules. For example, Plag et al. (2007, 2008) found the effects shown in Table 1. There is some variation across the corpora (CELEX, Baayen, Piepenbrock & Gulikers 1995 and BURSC, Ostendorf, Price & Shattuck-Hufnagel 1996).

For the present study, we include three of these relations that occur frequently in our data, and which we define as N1 = TEMPORAL LOCATION DEFINING N2 ('temporal'), N1 = SPATIAL LOCATION DEFINING N2 ('locative'),

https://doi.org/10.1017/S0022226712000199 Published online by Cambridge University Press

Semantic property	Influence	CELEX	BURSC
N1 refers to a period or point in time	Rightward		✓
N2 is a geographical term	Rightward		✓
N1 and N2 form a proper noun	Rightward	✓	✓
N1 is a proper noun	Rightward	✓	✓
N1 and N2 form a left-headed compound	Rightward		✓
Semantic relation			
N1 has N2	Rightward		✓
N2 is made of N1	Rightward	✓	✓
N1 is N2	Rightward	✓	✓
N2 is located at N1	Rightward		✓
N2 occurs during N1	Rightward		✓
N2 is named after N1	Rightward	✓	✓
N2 is for N1	Leftward	✓	
N2 uses N1	Leftward		✓

Table 1
Semantic categories found to influence compound prominence (Plag et al. 2007, 2008).

and N1 = MATERIAL OR INGREDIENT OF N2 ('made of'). We also noticed that many of the compounds in this third group were names of food items, which is a category mentioned by Gussenhoven & Broeders (1981) as favouring right stress. We therefore decided to add this category to our classification: NN = NAME OF A FOOD ITEM ('name of food item').

The fact that all four classes occur relatively frequently in the BNC data means that, if we do not find a significant effect, we can be confident that this is not due to there being too few examples of the categories.

2.4 *Lexicalisation*

Another influential factor is lexicalisation. According to many scholars, perhaps beginning with Sweet (1892: 289), more highly lexicalised compounds seem to be more prone to left prominence (see also Giegerich 2004 for a more recent claim in this direction). However, lexicalisation is an intricate concept, and generally recognised to be not a categorical notion, but rather a gradual one. Furthermore, there is no agreed test by which to decide whether a given item is lexicalised, or, under a gradient view, more lexicalised or less lexicalised. In general, the following criteria can be used: frequency, orthography, semantic transparency and semantic institutionalisation. We discuss each briefly in turn.

https://doi.org/10.1017/S0022226712000199 Published online by Cambridge University Press

It is generally assumed that lexicalisation strongly correlates with frequency (e.g. Lipka 1994: 2165). The more frequent a word, the more likely it is to be listed in the mental lexicon of speakers. It has also been shown that frequency correlates with orthography (e.g. Plag et al. 2007, 2008). Compounds usually written as one word tend to have higher frequencies than compounds usually written as two separate words, which is a strong indication that orthographically concatenated compounds are more lexicalised on average than non-concatenated, i.e. spaced, compounds. And concatenated compounds have indeed been shown to have a very strong tendency to be left-stressed (e.g. Plag 2006; Plag et al. 2007, 2008).

The orthography also relates to semantics, since we assume that writers tend to concatenate those compounds that are felt to be a single unit instead of the sum of the two individual parts. This is most clearly the case with semantically non-transparent compounds such as *butterfly*. What exactly do we mean by non-transparency? In the case of compounds, a compound is fully transparent if the meanings of both constituents are part of the meaning of the compound. For example, a *software company* is a company that produces or sells software, a *honey bee* is a bee that produces honey. These compounds are fully transparent, since the meanings of the two constituents are part of the meaning of the compound as a whole. This is different from *butterfly* 'an insect with a long thin body and four usually brightly coloured wings' (*Oxford Advanced Learner's Dictionary* 1995), where neither butter nor flies are involved in the interpretation. One has to concede, however, that semantic transparency is a gradient notion for at least two reasons. Firstly, there are at least three degrees of transparency, depending on how many constituents' meanings are part of the meaning of the compound (none, one, or two). Secondly, the degree of similarity between the constituent and its free form correspondent is also gradient. Thus one could argue that *fly* is only partially opaque in *butterfly*, since a fly is also a flying insect. What all this boils down to with regard to lexicalisation is the fact that a compound with an even partially opaque meaning needs to be memorised with that meaning, and hence is lexicalised.

The final aspect of lexicalisation is what has been called 'institutionalisation'. We mean by this term that a potentially ambiguous compound takes on a particular meaning in a given context or in the speech community at large. For example, a *city hall* is not just any hall in a city but the building where the administration of the municipality is situated. In such cases, particular meanings of the two constituents may still be part of the meaning of the whole, but the compound is nevertheless memorised with a particular sense, and hence is lexicalised.

To operationalise semantic opacity and institutionalisation one can use dictionaries. In general, it can be assumed that dictionaries, for economic and practical reasons, tend to list those complex words that are in some sense idiosyncratic, for example, have a meaning that is not inferable from the

https://doi.org/10.1017/S0022226712000199 Published online by Cambridge University Press

constituent parts, or a particular meaning amongst several theoretically possible ones. Hence, 'listedness', that is to say, having an entry in a dictionary, can be taken as an indication that a compound is likely to be institutionalised or semantically opaque. Of course, dictionaries also list some fully transparent complex words, but one can assume that amongst those compounds listed in a dictionary there is a large proportion of non-transparent ones.

In the light of these various considerations, the present study includes three variables that are taken to be correlates of lexicalisation. These are listedness, compound frequency and 'spelling ratio', which is the number of non-spaced tokens of a compound found in a corpus divided by the number of spaced tokens, i.e. the ratio of non-spaced frequency to spaced frequency. The prediction is that compounds that are listed or have a high frequency, or a high spelling ratio will be more lexicalised and hence more prone to left stress.

2.5 *Length of constituents*

Another factor which may influence compound prominence is the length of the constituents. Jespersen (1909: 153) remarks that right stress is often found with longer right constituents. This factor has not been empirically tested in previous studies, and we include various measures of length in the present analysis with a view to testing this prediction. These include the number of syllables in N2, the total number of syllables in the compound, and the number of syllables in the compound following the main-stressed syllable of N1. This latter variable is a measure of the length of the unaccented 'tail' that would result from the compound being left-stressed: it was included to follow up a suggestion by Ladd (2008: 244) that there may be a cross-linguistic tendency for phonological constituents with left accentuation to receive an additional accent on their final element when they are long (see further discussion in Section 4.2).

2.6 *Identity of constituents*

It has often been mentioned that certain compound constituents give rise to particular stress patterns. A textbook example compares compounds with either *street* or *avenue* as the right-hand constituent: these are predictably left-prominent in the case of *street*, e.g. *PARK street*, and right-prominent in the case of *avenue*, e.g. *park Avenue* (where capitalisation indicates the most prominent syllable). In empirical studies of this effect, constituent identity has turned out to be the most successful single predictor amongst all known factors (Plag et al. 2007, Plag 2010, Arndt-Lappe 2011), no matter whether regression models or computational analogical models have been used.

Plag (2010) used logistic regression modelling to investigate compound stress. He operationalised the constituent identity effect as 'constituent

https://doi.org/10.1017/S0022226712000199 Published online by Cambridge University Press

family bias': for a given corpus, this is the proportion of compounds with a shared constituent, in the same position, that have a particular stress pattern. In other words, it is a measure of the probability that a compound with a particular constituent in a given position will receive left or right prominence. Because our dataset contains few compounds that share both left and right constituents with other compounds in the data, we were unable to calculate meaningful biases in this way. An alternative would have been to calculate biases on the basis of stress patterns given in dictionaries. However, the compounds found in dictionaries tend to have a great preponderance of left stress, corresponding to the high incidence of lexicalisation. In the dictionary-based CELEX lexical database (Baayen et al. 1995), for example, over 90% of the NN compounds have left stress. Dictionary data would therefore not be representative of the kinds of compounds found in our data, where the incidence of left and right prominence is much more balanced. In any case, it is interesting that, by incorporating informativeness and constituent length, we were able to build a successful model without using constituent identity. Further research will be needed to tease apart the relationship between these various effects.

3. Methodology

3.1 *Data*

Previous empirical studies of compound prominence variability have either used experimental data (e.g. Plag 2006), dictionary data (e.g. Plag et al. 2007), or speech data from more specialised genres, such as the news text data from the Boston University Radio Speech Corpus (BURSC) used by Plag et al. (2008). This paper extends the empirical scope to another kind of data, namely compounds used in everyday conversation, whose stress patterns were elicited in a controlled experimental procedure. This procedure is described below, and further details, as well as the full dataset, can be found in Bell (2012).

The corpus from which the compound types were sampled is the British National Corpus (BNC), a balanced corpus of 100 million words of British English dating from between 1960 and 1994. This corpus consists of about 90 million words from written texts and 10 million from spoken texts, the latter being further subdivided into a contextual section (speeches, radio broadcasts, lectures, etc.) and a demographic section. The demographic section is so called because it consists of 4.23 million words of spontaneous conversation, recorded by informants selected to represent a demographically representative sample of the population. The compounds in our study were sampled from this demographic section, since it represents a source of compounds that are actually used by speakers in everyday conversations.

The compound types selected as experimental items all occurred in the corpus with spaced orthography, i.e. transcribed as two orthographic words.

https://doi.org/10.1017/S0022226712000199 Published online by Cambridge University Press

Some of them also occurred, elsewhere in the corpus, with hyphenated or concatenated orthography. By selecting compounds that had at least one spaced token, it was hoped to get a balanced distribution of left and right prominence: hyphenated and concatenated compounds are more likely to be lexicalised and left-prominent, whereas spaced orthography occurs with both kinds of prominence. Mark Davies kindly supplied a list of randomly ordered NN collocates occurring in the demographic part of the BNC, representing a total of about 20,000 types. Starting from the top of that random list, each collocate type was checked to establish whether it was used as a compound in the corpus or not. Strings that were not compounds in the sense defined in Section 2.1, such as *tea mother* occurring in the context of *Would you like some tea mother?* were excluded. Sampling ended after 1000 compounds had been found. At a later point, we excluded those compounds from this list where either constituent was itself a compound, since these did not conform to our definition of NN.

The prominence pattern of these compounds was established using a production experiment, which is described in the next subsection.

3.2 *Experimental setup*

The compounds were presented visually on a computer screen to participants in the context of a standard carrier sentence to be read out: *She told me about the* (*compound*). This sentence was selected because it is sufficiently general for any compound to be inserted without the sentence becoming meaningless, and also places the compound in final position, thus eliminating the possibility of the compound prosody being influenced by a following accent which might make it more difficult to perceive. Using the same carrier sentence for all items was intended to minimise the possibility that differences in prosody between compounds were due to differences in the contexts in which they occurred. Presenting the items in the context of a sentence was also intended to reduce the risk that they would be read with a list intonation which might affect prosody.

The presentations were produced as follows. To the 1000 sentences containing NN compounds, were added 1000 similar sentences containing simplex nouns and 1000 containing adjective–noun (AdjN) combinations, also selected randomly from the demographic BNC, to be used as fillers. This list of 3000 sentences was reproduced four times to give a total of 12,000 tokens. These were divided semi-randomly into lists of 300 sentences, so that each list contained about 100 experimental (NN) items and no item appeared twice in the same list. Lists were then assigned to participants, again semi-randomly, in such a way that no speaker would repeat an item. In this way, four tokens of each item were elicited from four different speakers, but each token appeared in a different list and position to guard against effects of list position or reader fatigue. The sentences containing items or fillers were

https://doi.org/10.1017/S0022226712000199 Published online by Cambridge University Press

converted to slides with one sentence per slide, and the slides presenting the items and fillers were alternated with blank slides showing only a small cross. These blank slides were intended to make the readers pause between reading successive items, again to guard against developing list intonation.

The participants in the experiment were seventeen adult native speakers of educated British English: eight with northern English accents, and nine with southern accents. Nine of them were volunteers and eight were teachers of English as a Foreign Language who were given remission from teaching in return for participation. The readers were told that they were participating in a study of variation in pronunciation, and were asked to read the sentences naturally, at their own speed, pausing between each one. The slides were presented on a laptop computer, and readers used the keyboard to move on to successive slides. No more than one list of 300 sentences was read at any session, and these were presented in groups of 100 at a time, with readers taking a five minute break between groups. Sessions were separated by at least one day. In this way, it was hoped to enable readers to stay as natural as possible and avoid the reading becoming mechanical. Most participants read two or three lists in total, but one person read only a single list and three people read five lists each.

The readings were recorded onto magnetic tape using Sony Walkman professional recorders and these recordings were subsequently digitised using Audacity sound editing software. Using the same software, the compound tokens were then edited from the carrier sentences.

After they had read their whole quota, participants were asked whether they were aware of any rule about the assignment of prominence to NN combinations. One of the teachers said she had been aware of such a rule, and so her data were excluded because of the possibility that this belief had influenced her pronunciation. The items lost were redistributed amongst other participants who had not already read them. Similarly, any items where the reader stumbled, hesitated, repeated themselves or developed a 'sing-song', list-like intonation were excluded and redistributed in subsequent sessions. Finally, because of the random way items were assigned to lists, some ended up in contrastive contexts, for example *She told me about the oak tree* immediately following *She told me about the old tree*. Items in such contexts were excluded and reassigned. In this way, 4000 acceptable tokens were elicited: four tokens, each from a different reader, for each of the 1000 types. The number of NN compound tokens spoken by each participant ranged between 97 and 465 tokens, reflecting the different numbers of lists they had read.

3.3 *Prominence rating*

Not only is there variation in the prominence assigned by speakers to English compounds, but there is also considerable variation in the way it is perceived: that is to say that one listener might perceive a token as having

https://doi.org/10.1017/S0022226712000199 Published online by Cambridge University Press

right prominence, while a different listener, or even the same listener on a different occasion, might perceive the same token as having left prominence. Kunter (2010, 2011) showed that some listeners show greater inter-rater agreement than others in this respect, and he was able to identify from his listeners a sub-group of relatively reliable raters, in the sense that they agreed with each other to a statistically significant extent. For the purposes of the present data collection, two native speakers of English participated in Kunter's experiment and were found to be amongst the reliable group. These raters subsequently produced three ratings for each of the compounds in the experiment. The first rating was made online by the first rater while the participants were reading the compounds, assigning either left or right prominence to each item that was read out. To address potential concerns about the reliability of these online ratings, all tokens were again rated by the first rater, as well as the second rater, using a special computer programme developed by Kunter for his own experiment on the perception of compound prominence (Kunter 2011).

In that experiment, Kunter (*ibid.*) presented compound tokens to his raters using a graphical interface that included a slider bar with which they could indicate the relative perceived prominence of left and right constituents, treating it as a continuous rather than categorical variable. Moving the slider to the extreme left or right of the bar indicated completely left or right prominence, while placing the slider at intermediate positions indicated less contrast in the perceived prominence of the two constituents, and placing it centrally indicated that they were perceived as equally prominent. His listeners could hear items as many times as required before making a decision, and the programme then converted the position of the slider into a number between 0 and 999, where extreme left = 0, extreme right = 999 and centre point = 499. Kunter generously made his programme available for the present data collection. Both raters used the programme to listen to all the tokens in isolation, with the option of listening repeatedly rather than having to make an instantaneous judgement. This meant that for each token there were three prominence ratings, one categorical and two numerical. The numerical prominence ratings were converted to categorical ones, with values of 0–499 indicating left prominence and values of 500–999 indicating right prominence, and the three ratings for each token were then compared. In 97% of cases, the first rater's two judgements were the same; of these, the second rater's judgement also concurred in 97% of cases, giving a total of 3764 tokens for which all three judgements agreed. Those tokens for which the three ratings were not unanimous were excluded from further analysis.

3.4 *Categories coded*

In this section we describe the variables that we used as predictors in our models. We were primarily interested in the effect of informativeness, but

https://doi.org/10.1017/S0022226712000199 Published online by Cambridge University Press

Informativeness	Semantic relations	Lexicalisation	Length
Frequency of N2	Temporal	Listed in dictionary	N2 syllables
Conditional probability of N2 based on N1 frequency	Locative	Compound frequency	Total syllables
Family size of N2	Made of	Spelling ratio	Syllables after N1 main stress
Conditional probability of N2 based on N1 family size	Name of food item		
Synset count N1			
Synset count N2			

Table 2
Predictors initially present in the analysis.

also included as covariates measures related to effects found previously in the literature, i.e. semantic relations and lexicalisation, and, additionally, constituent length. Table 2 summarises our predictors.

3.4.1 *Informativeness*

The frequency of N2 was taken from the whole 100 million words of the BNC using the BYU-BNC interface (Davies 2004–). The frequencies used were lemmatised, which is to say they included all inflectional variants of the word in question. To calculate the conditional probability of N2 based on frequency, we also extracted lemmatised frequencies for N1 and for the compound as a whole. Separate frequencies were obtained for the compound written as two words (spaced) and one word (non-spaced), with hyphenated tokens included in the non-spaced count. The compound frequency was then taken as the sum of these two values, and conditional probability calculated as compound frequency divided by frequency of N1.

For the predictors based on family size, we again used the BNC and the BYU-BNC interface (*ibid.*). However, obtaining family size measures from a corpus presents special challenges for the researcher. We first estimated family sizes from the whole BNC, by searching for spaced noun–noun collocates in which the relevant values of N1 or N2 occurred in first or second position respectively. Since, for the reasons mentioned in Section 3.1, not all collocates are compounds, we then had to take some measure to alleviate the danger of the raw data giving us unrepresentative family sizes. The raw data would definitely over-estimate family sizes: the important question was whether this overestimation was consistent across compounds. Considerable care was taken to address this potential problem.

https://doi.org/10.1017/S0022226712000199 Published online by Cambridge University Press

Initially, a semi-random sample of 187 families was selected, representing a balance of N1 and N2 families from both left- and right-prominent types. This sample included more than 10% of all families in the data. For each of these families, all the constituent NN collocates were checked to ascertain whether they actually occurred as compounds in the corpus. Since this involved the checking of each and every token of each of the collocates, assessment was restricted to the spoken corpus of the BNC, still 10 million words in size. The accurate family sizes thus obtained were then regressed against the estimated family sizes. There was a highly significant correlation ($r = .98$, $p < .0001$), but a number of outliers could be identified, all of which showed estimated family sizes that were disproportionately large, as against the (manually checked) actual family sizes. A closer look at these outliers revealed that they belonged to the following categories: the constituents were vocatives in N2 position (as in the example *tea mother* above); had homonymic forms (e.g. *lead*); were part of high-frequency formulae such as *morning* meaning *good morning*; had a high probability of being mistagged, e.g. the post-nominal adjective *present*, as in *highlights any special feature present*; or had very small family sizes.

In order to address the issue of outliers revealed by the sample, the complete set of families was then searched for compounds potentially belonging to these five categories. In total, 331 pertinent families were found. For all these additional families the actual family sizes were also manually determined by checking all tokens in the spoken part of the BNC. In this way, a sample of 518 families was established with both estimated raw counts for family sizes and the actual family sizes (all based on the spoken part of the BNC). As before, the actual and estimated family sizes for this sample were regressed onto each other, and again a very high correlation ($r = .97$, $p < .0001$) was found. As expected, an inspection of the errors in the linear model revealed a number of outliers: those with residuals larger than 1.5 standard deviations were removed from the data to prevent them exerting undue influence on our statistical models. After exclusion of these family size outliers we were left with 3252 tokens for our analysis. For this set of data, it is safe to assume that the estimated family sizes are highly correlated with the actual family sizes.

Synset counts were manually extracted from the Wordnet index file for nouns, for all values of N1 and N2 in the data. In cases where a word did not appear in Wordnet (usually because of dialectal differences between British and American English), the number of senses was taken from *OED Online*, the online version of the *Oxford English Dictionary*. For three proper nouns, there was no entry in Wordnet or *OED Online*: these cases were assumed to have one sense.

All these measures (constituent frequency, family size, and synset counts) are known to influence the response times for visual lexical decision and

https://doi.org/10.1017/S0022226712000199 Published online by Cambridge University Press

word naming (Baayen 2005), and it can be safely assumed that they have some psychological reality. For the present study we use these measures to assess the effect of informativeness on compound prominence.

3.4.2 *Semantic relations*

Three annotators classified the compounds according to whether they belonged to any of the four semantic categories identified in Section 2.2, namely N1 = TEMPORAL LOCATION DEFINING N2 ('temporal'), N1 = SPATIAL LOCATION DEFINING N2 ('locative'), N1 = MATERIAL OR INGREDIENT OF N2 ('made of'), and NN = NAME OF A FOOD ITEM ('name of food item').

Two of the annotators were native speakers of English, one also a trained linguist and the other with an interest in linguistics; the third rater, although not a native speaker of English, had native-like competence and a Ph.D. in English linguistics. For each compound in the dataset, classifications were given by two of these three annotators; in cases where the two judgements did not agree, the annotators discussed the classification to see if a consensus could be reached. In the small number of cases where no consensus was possible, the items were excluded from further analysis.

3.4.3 *Lexicalisation*

To check whether the compounds were listed, *OED Online* was manually searched for each one. There is considerable variation in how compounds are listed in the dictionary, sometimes as full entries and sometimes under one of their constituents, usually the modifier. Because of this inconsistency, any hit from the main electronic search page (i.e. not including the full text) was counted as an entry.

As described above, lemmatised compound frequencies were taken from the whole 100 million words of the BNC using the BYU-BNC interface (Davies 2004–). Separate frequencies were obtained for the compound written as two words (spaced) and one word (non-spaced), with hyphenated tokens included in the non-spaced count. Compound frequency was then defined as the sum of the two different spelling frequencies, while spelling ratio was the non-spaced frequency divided by the spaced frequency.

In accordance with the previous literature, we take listedness, compound frequency and spelling ratio to be primarily indications of semantic opacity. The assumptions are, firstly, that combinations which are not semantically transparent are more likely to be listed in dictionaries; secondly, that opaque compounds are on average more frequent than transparent ones; and thirdly, that concatenated or hyphenated orthography is used by speakers when they perceive the two components of a compound as constituting a single conceptual unit.

https://doi.org/10.1017/S0022226712000199 Published online by Cambridge University Press

3.4.4 *Length*

Two native speakers of English determined the number of syllables in N1 and N2 for each compound: any discrepancies were resolved by reference to the *Longman Dictionary of Contemporary English* (1978). In cases with an optional schwa, this was included in the count: for example, *secretaries* was recorded as consisting of four syllables rather than three. The main-stressed syllable of N1 was established in a similar way. The total number of syllables in the compound was taken as the sum of N1 syllables and N2 syllables. The number of syllables after N1 main stress was obtained by subtracting, from this total, the number of syllables in N1 up to and including the syllable with main stress.

3.5 *Statistical analysis*

To alleviate the potentially harmful effects of extreme values on our statistical models, the spelling ratio and all measures of informativeness were first logarithmatised.

For the statistical analysis, we used different kinds of multiple regression. In regression analysis, the outcome of a dependent variable can be predicted on the basis of independent variables. Multiple regression has the advantage of showing the effect of one variable while holding all other variables constant, which is especially welcome for an investigation like this one, where many different predictors seem to play a role at the same time.

In addition to standard generalised linear models and mixed effects models (see Baayen 2008), we also fitted generalised additive models (Hastie & Tibshirani 1990, Wood 2006), which provide a more precise way of modelling interactions involving two (or more) numerical predictors, compared with standard linear models. The results of the additive models were very similar to those of the standard linear models. Given that most readers will be unfamiliar with the technicalities of generalised additive models, we decided to report the results of the more straightforward standard models in this paper.[2] The statistical procedures will be explained in more detail as we go along.

In Section 4, we will present the analysis of the full dataset, which includes all tokens for which the listeners agreed on the stress pattern. Since many compounds were accented differently by different speakers, this dataset includes many compounds with both right- and left-stressed tokens, in varying proportions. Section 5 presents an analysis based only on the compounds

[2] The smooth terms for the measures of informativeness in the generalised additive models sometimes showed some wriggly, non-linear behaviour. A more detailed investigation of potential non-linearities in the effects of informativeness measures remains a topic for future research.

https://doi.org/10.1017/S0022226712000199 Published online by Cambridge University Press

Number of tokens	3252
Proportion of tokens with left stress	.60
Number of compound types	864
Number of N1 lemma types	590
Number of N2 lemma types	574
Median family size of N1	221
Median family size of N2	276

Table 3
Dataset for token analysis.

that were stressed in the same way by all four speakers, which are taken to be those with the most nearly categorical prominence.

4. Token-based analysis

4.1 *Results*

The full dataset contains the 3252 tokens that remained after the computation of the family sizes, and includes tokens of 864 of the original 1000 compounds. Table 3 gives an overview of some properties of this dataset. Of the 864 compound types in the data, 323 (37%) have at least one token with each of the two stress patterns; in other words, 37% of the compounds show inter-speaker variation in prominence. To address this variation, we used mixed effects regression modelling, which allowed us to include speaker as a random effect (see Baayen 2008: 241ff.).

All the variables from Table 2 were initially included as potential predictors. A closer look at the numerical variables, however, revealed that a number of them were highly correlated with one another. This collinearity is potentially harmful in regression analysis since it means that different predictors compete in accounting for the same part of the observed variation in the dependent variable. In order to reduce any negative impact on the statistical models, the level of collinearity needs to be kept below a certain threshold, and so, where predictors are highly correlated, it may be best to use only one of them.

In our data, the frequency of N2 and the family size of N2 were highly correlated. Exploratory analyses showed that the predictive power of the family size measure was greater, so we kept that measure and excluded the frequency of N2. Similarly, the conditional probability of N2 based on N1 frequency (frequency of NN/frequency of N1) was highly correlated with compound frequency. Here, our exploratory analyses revealed that neither variable had a significant influence on prominence, and so we excluded them both from further analysis. In the models that follow, 'conditional probability of N2' therefore refers to conditional probability based on N1 family

https://doi.org/10.1017/S0022226712000199 Published online by Cambridge University Press

size. Finally, all three measures of length were correlated with one another. We found that, overall, the one that produced the best and most interpretable models was the number of syllables after N1 main stress, and we therefore kept just this one. Removing these correlated variables from the data resulted in a satisfactory fall in collinearity, as indicated by a decrease in the condition number κ from 58.3 to 18.0.[3]

We fitted a logistic generalised mixed effects regression model with speaker as a random effect and with the other predictors as fixed effects. In addition to the individual predictors, we also included an interaction term for the synsets, to see the combined effect of N1 and N2 synsets: recall that if relative informativeness is important, then we might expect this interaction to have a significant effect on prominence. We report the results of the final model, from which all non-significant predictors were removed step-wise, following standard procedures of model simplification.

We document the final model in Table 4. The significant main effects for predictors unrelated to informativeness are plotted in Figure 1, the significant main effects for informativeness predictors based on the probability of N2 are shown in Figure 2, and the significant interaction for informativeness predictors based on semantic specificity appears in Figure 3. For the categorical variables, the dots indicate the mean estimated probability of N2 receiving an accent, in compounds that have the pertinent value of that variable. For the continuous variables, the graphs show regression lines. To show the effect of each predictor in turn, the other predictors are adjusted to their reference level (for categorical variables) or to their means (for continuous predictors). The reference level for the categorical variables is 'no': in other words, the model shows the effect of independently varying each predictor in a situation where none of the (other) semantic categories applies and the compound is not listed (except where this is the predictor in question). In the table, positive coefficients indicate a tendency towards right prominence, negative coefficients towards left prominence. The model is quite successful in its predictions ($C = .846$), in spite of the fact that so many compound types are variably stressed in this dataset.[4]

We will first discuss the measures that are not related to informativeness. Their effects are illustrated in Figure 1. Three of the four semantic relations show a significant main effect in the expected direction. The positive coefficients for these predictors mean that compounds expressing a 'locative', 'temporal' or 'made-of' relation have a significantly higher chance than

[3] The condition number κ is used to assess the danger of collinearity in the data. According to Baayen (2008: 182), condition numbers of 30 and more may indicate potentially harmful collinearity.

[4] C is the concordance index, and it measures the concordance between the outcome predicted by the model and the value actually observed. A C-value of .5 indicates total randomness, 1 indicates perfect predictions. A C-value of below .8 indicates that the model is not very useful in making predictions because of the high error rate (see Kutner et al. 2005).

https://doi.org/10.1017/S0022226712000199 Published online by Cambridge University Press

Random effect:	Variance	Std.Dev.			
Speaker	0.23177	0.48143			
Fixed effects:	**Estimate**	**Std.Error**	***z*-value**	**Pr(>|z|)**	**Sig.**
(Intercept)	−1.54510	0.34227	−4.514	$6.35e^{-06}$	***
Temporal	2.59219	0.28451	9.111	$<2e^{-16}$	***
Locative	1.81599	0.15704	11.564	$<2e^{-16}$	***
Made of	2.71829	0.14471	18.785	$<2e^{-16}$	***
Listed in dictionary	−0.74772	0.09992	−7.483	$7.25e^{-14}$	***
Log spelling ratio	−0.28695	0.03047	−9.419	$<2e^{-16}$	***
Syllables after N1 main stress	0.34126	0.04692	7.272	$3.53e^{-13}$	***
Log family size of N2	−0.27984	0.04293	−6.518	$7.11e^{-11}$	***
Log conditional probability of N2	−0.09968	0.04033	−2.472	0.0134	*
Log synset count N1	0.59942	0.14180	4.227	$2.37e^{-05}$	***
Log synset count N2	0.01310	0.11872	0.110	0.9122	
Log synset count N1: log synset count N2	0.20115	0.08493	−2.368	0.0179	*

Significance codes: *** $p<.001$; ** $p<.01$; * $p<.05$

Table 4
Mixed effects regression model, final token-based analysis (N = 3252, $C = .846$).

other compounds of having an accent on N2. The fourth category ('name of food item') was non-significant in the full model ($p = .642$). This may have been because the overwhelming majority of 'food item' compounds also fell into the larger 'made of' class, which could therefore have subsumed any independent effect of the smaller category. We also see that both predictors based on lexicalisation work in the expected direction. Firstly, if a compound is listed it has a greater chance of being left-stressed. Secondly, the higher the spelling ratio, i.e. the greater the proportion of a compound's tokens that are written non-spaced, the lower is the chance that N2 will be accented. Finally, there is a main effect of length, such that the more syllables there are in the compound after the main-stressed syllable of N1, the greater is the chance that N2 will be accented.

Let us now turn to those predictors that measure informativeness. As we can see from Table 4, both probability-based measures turned out to be significant. The plot for the family size of N2, given in the left-hand panel of Figure 2, shows an effect in the expected direction. The larger the family size,

https://doi.org/10.1017/S0022226712000199 Published online by Cambridge University Press

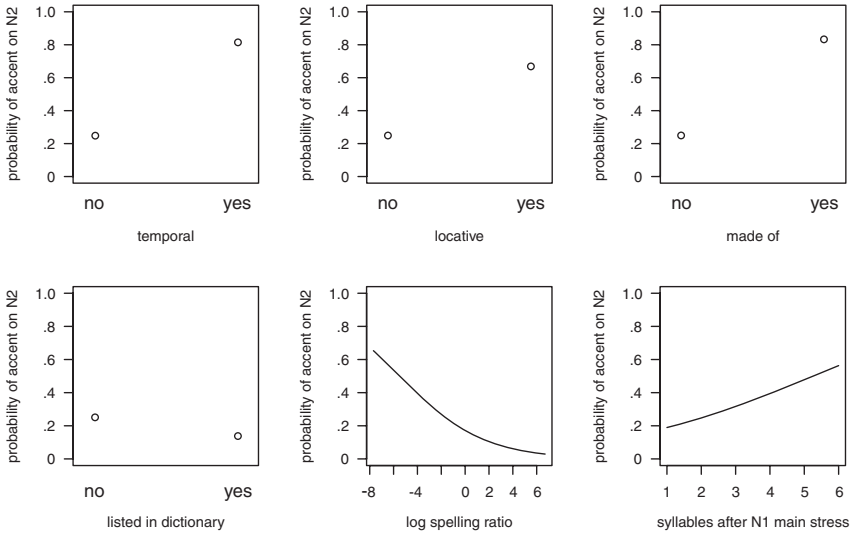


Figure 1
Partial effects in the final mixed model, predictors unrelated to informativeness.

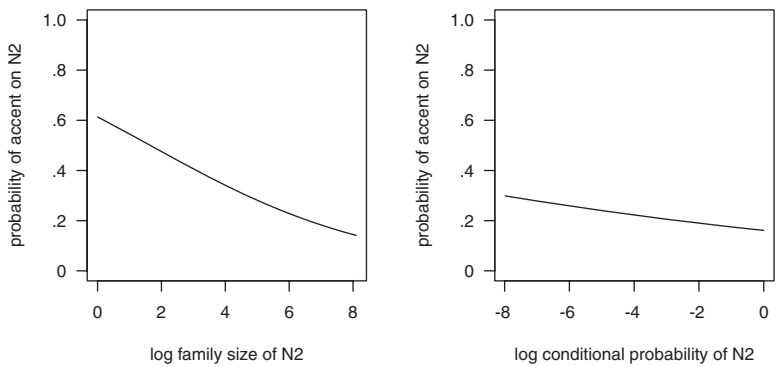


Figure 2
Partial effects in the final mixed model, informativeness predictors based on probability of N2.

i.e. the less informative N2 is, the less likely it is to be accented. The right-hand panel of Figure 2 shows the expected effect for conditional probability. The greater the likelihood of N2, given N1, the lower is the probability of an accent on N2.

There is also a significant interaction term for the synset counts. In order to understand this interaction it is useful to inspect the contour plot, shown in Figure 3. The contour plot is three-dimensional in nature and works in

https://doi.org/10.1017/S0022226712000199 Published online by Cambridge University Press

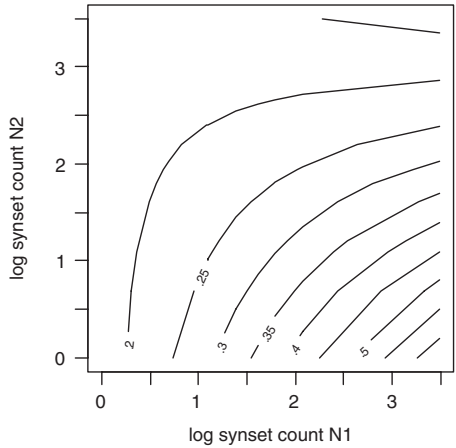


Figure 3
Partial effects in the final mixed model, informativeness predictors based on semantic specificity (contour lines indicate probability of accent on N2).

analogy to topographic maps, with valleys and mountains indicating opposing values of the dependent variable. The x and y axes show the values of the two predictors. The contour lines indicate regions with the same probability of right stress, with the probability of accent on N2 given as a figure on the line. The contour plot shows a growing probability of stress on N2 for compounds towards the lower right corner of the plot. This means that with non-specific N1s and highly specific N2s, we find the highest probability of N2 being accented. This is fully in accordance with the hypothesis: accentuation of N2, and therefore right prominence, is most likely when N2 is highly informative relative to N1. We also see that the effect of N2 specificity only kicks in beyond a certain threshold of N1 non-specificity. Thus, for the most specific N1s the specificity of N2 does not have a strong influence on stress placement. Similarly, for the least specific N2s, the specificity of N1 has little effect.

To summarise the results, we find clear effects of informativeness that are in line with the predictions, with less informative second constituents being less likely to be accented. Beside the informativeness of N2 itself, as indicated by its family size, the informativeness of N2 relative to N1 also has a say in stress placement, as shown by the effects of conditional probability and the synset interaction.

4.2 *Discussion*

With regard to established determinants of compound stress assignment, we found the same effects as previous studies. Our findings confirm the role of semantic relations and lexicalisation as determinants of prominence.

https://doi.org/10.1017/S0022226712000199 Published online by Cambridge University Press

We also found a new effect, that of length. Why should longer 'tails', i.e. longer strings of syllables following the first main stress, attract prominence to N2? Ladd (2008: 244) suggests that there may be a cross-linguistic tendency for phonological constituents with left accentuation to receive an additional accent on their final element when they are long. He explains this in terms of a metrical theory of sentence stress, whereby longer sentences are composed of more than one intermediate phonological phrase, each of which carries an accent (Ladd 2008: 275–276). Given that phonetic studies of compound prominence (e.g. Kunter 2011) indicate that 'right' prominence is more accurately described as double accentuation, Ladd's (*ibid.*) analysis can easily be extended to NN compounds: compounds of sufficient length constitute two intermediate phonological units. However, the motivation for such a metrical structure is not entirely clear. Ladd (*ibid.*: 275) states that 'when … sentences are longer, it is difficult to treat the whole sentence as a single intermediate phrase', but does not specify where this difficulty lies. Nevertheless, we might speculate that the difficulty, if it exists, is related to the mechanics of speech production.

In addition to the effects just discussed, we found robust effects of informativeness on compound prominence assignment. Recall that left-prominent compounds carry one accent, while right-prominent compounds carry two, one on each of the constituents. Hence, placement of an accent on the first constituent is a given, and the issue is whether the second constituent also receives an accent. According to the hypothesis, the decision about placing this accent would naturally involve assessing primarily the informativeness of the entity that is to receive that accent (or not), i.e. the informativeness of N2. The regression models show effects in the predicted direction. More informative second constituents attract stress, whether informativeness is measured in terms of probability, relative probability or semantic specificity.

The robust effects of informativeness on accent placement in NN compounds are in line with the findings by Pan & McKeown (1999) and Pan & Hirschberg (2000). These authors quite successfully employed word frequency measures and the conditional probability of the second word in a two-word sequence to predict accentuation in medical texts. As in this study, more informative constituents were more likely to receive an accent.

Our results differ from those of Plag & Kunter (2010). These authors found that, in their corpora, family size only added to the much more significant effect of family bias. One probable reason for the discrepancies between their findings and ours could be that their family sizes were based only on dataset-internal families, hence very small, while our family sizes are derived from a very large corpus, hence very large, and thus potentially more highly correlated with actual family sizes. However, a more important difference between the two studies is that we included different, and more, informativeness measures in this study, whereas Plag & Kunter (2010) used only one, namely

https://doi.org/10.1017/S0022226712000199 Published online by Cambridge University Press

family size. The fact that a variety of measures of informativeness reach significance in the present models lends further support to the idea that informativeness is a powerful determinant of compound prominence.

5. Type-based analysis: Compounds without inter-speaker variation in stress

5.1 *Results*

Of the 864 compound types in our data, 541 types were accented in the same way by all four speakers. Of these, 341 were consistently left-prominent, and 200 were consistently right-prominent. The analysis presented in this section uses only these non-variable compounds. Because there is no inter-speaker variation in this dataset, we did not need to include speaker as a random effect, and were therefore able to use standard logistic regression rather than mixed effects modelling. In the statistical programme R, logistic regression modelling has the advantage that there are readily available procedures for model criticism, which help to increase confidence in the resulting models.

The results of a logistic regression analysis of this dataset were very similar to the results described in the previous section. In the full model, we initially included the same predictors as in the token-based analysis, but without the random effect for speaker. The full model, using all predictors, was simplified in the usual fashion by stepwise elimination of non-significant predictors. The resulting model contained the same predictors as the final token-based mixed model described above, except for the random effect of speaker. To check that the model did not over-fit the data, we used bootstrap validation (with 200 bootstrap runs in the simulation), which is a way of detecting and eliminating any superfluous predictors in a model: none of the predictors were eliminated during this process. To ameliorate the effects on the model of any extreme values in our dataset, and thereby increase the model's predictive power for new data, we then applied penalised maximum likelihood estimation (see Baayen 2008: 225). This involves introducing a penalty factor into the model to guard against over-large coefficients, i.e. to avoid exaggerating the effect of any of the predictors.

The final penalised model is shown in Table 5. Again, positive coefficients indicate a tendency towards right prominence, negative coefficients towards left prominence. Notably, we now find a higher *C*-value of .918, which means that this model is even better in its predictions than the token-based analysis. This is not unexpected, since the token-based model includes within-type variation whereas the present analysis focuses on those types that show no inter-speaker variation in our data. After penalisation, the effect of the synset count interaction is now only marginally significant, but it does survive subsequent validation of the penalised model. The graphs for the type-based

https://doi.org/10.1017/S0022226712000199 Published online by Cambridge University Press

https://doi.org/10.1017/S0022226712000199 Published online by Cambridge University Press

	Coefficient	Std.Error	Wald Z	p	Penalty scale	Sig.
Intercept	−3.0992	0.98508	−3.15	.0017	0.0000	**
Temporal	3.7436	0.90451	4.14	<.0001	0.3162	***
Locative	2.5597	0.46334	5.52	<.0001	0.3162	***
Made of	3.7208	0.42326	8.79	<.0001	0.3162	***
Listed in dictionary	−0.9263	0.29258	−3.17	.0015	0.3162	**
Log spelling ratio	−0.4145	0.08636	−4.80	<.0001	0.8409	***
Syllables after N1 main stress	0.4490	0.13622	3.30	.0010	0.4440	**
Log family size of N2	−0.3919	0.12943	−3.03	.0025	0.6246	**
Log conditional probability of N2	−0.3181	0.12521	−2.54	.0111	0.5641	*
Log synset count N1	1.0600	0.40764	2.60	.0093	0.3376	**
Log synset count N2	0.1198	0.35253	0.34	.7339	0.3518	
Log synset count N1 : log synset count N2	−0.4812	0.24706	−1.95	.0515	0.6905	•

Significance codes: *** $p<.001$; ** $p<.01$; * $p<.05$; • marginal

Table 5
Logistic regression model, final type-based analysis (N = 541, $C = .918$).

model are extremely similar to the ones shown in Figures 1–3 for the token-based model and are therefore not reproduced here. To summarise, the analysis of the non-variable compounds further substantiates the contribution of informativeness to compound stress assignment as already found in the token analysis.

5.2 *Discussion*

It is interesting to compare the success of our models with those previously published in the literature. Plag (2010) used regression analysis to model compound prominence on the basis of semantic variables, lexicalisation and constituent bias, as defined above in Section 2.6. Table 6 shows a comparison between his results and those presented here. The statistic *C* gives an indication of the models' success in predicting stress: the higher the *C* value, within the range .5–1.0, the greater a model's success. Overall, it can be seen that models that include information about length and informativeness perform at least as well as those that have information about constituent bias but not about informativeness or length.

To look more closely at the predictive accuracy of some of these different models, and also to compare them with published analyses using analogical modelling, we calculated the probability of right stress, as predicted by our type-based model, for each of the 541 compounds in the non-variable dataset. These probabilities were then converted into categorical predictions, left or right, with all probabilities below .5 counting as left, all others as right. This allows us to directly compare the model's predictions with the observed stresses, as shown in Table 7.

From the information in Table 7, it is possible to calculate the proportion of compounds in the data for which the model predicts the attested stress pattern. It is also possible to analyse the proportion of correct predictions for right-stressed and left-stressed compounds separately. Table 8 compares the predictive accuracy of different type-based models: the logistic regression models presented here and in Plag (2010) as well as analogical models described in Arndt-Lappe (2011), which are based on the same databases as Plag (2010) and use the computational algorithm AM::Parallel (Skousen & Stanford 2004). Looking first at the two models based on the CELEX lexical database (Baayen et al. 1995), we see that they are extremely successful overall, with 95% success for both regression analysis and analogical modelling. However, a look at the left-stressed and right-stressed compounds separately shows that, although left stress can be predicted with 99% accuracy, the predictive accuracy for right stress is far below chance. In other words, both the regression analysis and the analogical model far over-predict left stress for this dataset: the high predictive accuracy overall is explained by the fact that 94% of compounds in the data are left-stressed.

https://doi.org/10.1017/S0022226712000199 Published online by Cambridge University Press

https://doi.org/10.1017/S0022226712000199 Published online by Cambridge University Press

			Predictors in analysis				
Corpus	Data	*C*	Informativeness	Semantics	Lexicalisation	Length	Constituent bias
BNC	Types	.918	✓	✓	✓	✓	
BNC	Tokens	.846	✓	✓	✓	✓	
BURSC	Types	.794		✓	✓		✓
BURSC	Tokens	.828		✓	✓		✓
CELEX	Types	.899		✓	✓		✓

Table 6
Comparison of model fits across different regression analyses (BURSC and CELEX models from Plag 2010).

	Predicted left	Predicted right
Observed left	316	25
Observed right	44	156

Table 7
Predicted vs. observed stress, final type-based model.

Approach	Corpus	Proportion left stress	Predictive accuracy for left stress	Predictive accuracy for right stress	Predictive accuracy overall
Regression analysis	BNC	.63	.93	.78	.87
	BURSC	.67	.92	.50	.78
	CELEX	.94	.99	.32	.95
Analogical modelling	BURSC	.67	.90	.61	.80
	CELEX	.94	1.00	.20	.95

Table 8
Comparison of predictive accuracy across different type-based models (BURSC and CELEX regression models from Plag 2010; analogical models from Arndt-Lappe 2011).

It can be seen from Table 8 that, in terms of the proportion of left stresses, the BNC-based dataset used here is much more similar to the Boston Radio Speech Corpus (BURSC, Ostendorf et al. 1996) than it is to CELEX. Comparing the BURSC models with our BNC model we see that, overall, the BNC model using informativeness and length is somewhat more successful than either of the models using constituent bias. However, when we look at the figures for left stress and right stress separately, we see that all three models are actually very similar in terms of their ability to predict left stress, but the present model is much more successful at predicting right stress. It therefore seems that including information about the informativeness of N2, and the number of syllables after the main-stressed syllable of N1, significantly increases the power of probabilistic models to predict right stress. This can be understood in terms of accentuation: if N2 is more highly informative, then it is more likely to be accented, i.e. the compound is more likely to be right-stressed. Similarly, if left stress would produce a long, unaccented string of syllables, then a second accent, i.e. right stress, is more likely.

6. Conclusion

In this paper, we have investigated the question of whether informativeness is a determinant of compound prominence in English. Analyses both of

https://doi.org/10.1017/S0022226712000199 Published online by Cambridge University Press

compounds showing within-type variation in prominence and of compounds showing no within-type variation in our data have provided very robust evidence for an effect of informativeness on stress assignment. In accordance with the predictions, it has been shown that all measures of informativeness can help to predict the probability of a compound having a particular stress pattern. In general, the more informative N2 is, the more likely it is to receive an accent; in other words, the more informative N2 is, the more likely is the compound to be right-stressed. Predictably, this effect can be modulated by the informativeness of N1, as shown by the significant interaction effect of the synset counts and by the effect of the conditional probability of N2. Given the significance of the family-based measures of informativeness, our results also substantiate the role of constituent families in compound structure and processing. Constituent families have been shown to be influential not only in stress assignment (e.g. Plag 2010, Arndt-Lappe 2011), but also in compound semantics (e.g. Gagné & Shoben 1997, Gagné 2001) and compound morphology (e.g. Krott, Schreuder & Baayen 2002, Krott et al. 2007).

The finding that informativeness is a strong predictor of stress pattern has repercussions for the phonological analysis of compound prominence. As mentioned in the introduction, we follow an analysis of compound prominence in terms of pitch accent placement, instead of lexical stress. In this view, the role of informativeness is to be expected and is naturally accounted for: uninformative elements do not receive a pitch accent. In an approach where compound stress assignment is a lexical-phonological process, the role of informativeness is unexpected and seems inexplicable. Our results therefore provide independent evidence for an accent-based theory of compound prominence.

We also found, for the first time, empirical evidence for a length effect. The greater the number of syllables after the main-stressed syllable of N1, the higher the chances of the compound bearing two accents and hence of N2 being prominent. Finally, in addition to these new effects, we found the same effects of semantic relations and lexicalisation as reported from previous studies. The reader should note that all effects were measured while controlling for the effects of other predictors, which means that we have good evidence that they are all independent determinants of compound prominence.

The semantic effects and the lexicalisation effect largely replicate those found by Plag and colleagues for other corpora (e.g. Plag et al. 2007, 2008). The informativeness effect is unprecedented in compound research, but in line with the studies by Pan and colleagues mentioned above (Pan & McKeown 1999, Pan & Hirschberg 2000), who found general effects of informativeness on accent placement in texts. Our results differ, however, from the results of Plag & Kunter (2010). They found that the most significant predictor of prominence is the family bias, that is to say the tendency for compounds that share the same first or second constituent to receive the same kind of

https://doi.org/10.1017/S0022226712000199 Published online by Cambridge University Press

prominence. Although we did not include this variable in our models, it is to be expected that family bias effects would be found in our data, since any predictor that is specific to N1 or N2 will inevitably give rise to such an effect. For example, all compounds with the same second constituent will also have the same number of syllables, frequency, synset count and positional family size for N2. If long constituents in N2 position increase the chance that N2 receives an accent, then longer words in N2 position will tend to have greater bias for right stress than shorter words. Similarly, if informative constituents in N2 position are more likely to be accented than less informative constituents, then more informative words in N2 position will tend to have greater bias for right stress than less informative words. Thus, it is even conceivable that length and informativeness measures could underlie the family bias effect. This in turn would explain why the results of models that include family bias, but not length or informativeness, are similar to the results presented here. The main difference between our results and previously published models of compound prominence is that our models are more successful at predicting right prominence. However, because we used a different corpus, we cannot be sure that this difference is not an artefact of the data. Clearly, what is called for is a study that includes both informativeness measurements and family bias in order to tease these effects apart.

A further question for future research concerns the variation in prominence between different tokens of the same compound. Very little is known about what influences the extent of this variation, although some compounds appear to be more variable than others (see Kunter 2011: Chapter 8). However, the involvement of constituent informativeness in prominence assignment suggests that this might also contribute to variability. The family sizes of constituents will vary across the mental lexicons of different speakers, and unconscious perceptions of informativeness might also therefore differ. In general, however, one might expect more variability where N2 is neither very informative nor very uninformative, either in absolute terms or relative to N1. Furthermore, variability might arise when informativeness conflicts with some other determinant of prominence: for example, when a compound has a semantic relation that predisposes it to right stress, but a very uninformative right-hand constituent. Initial results (Bell & Plag 2012) suggest that such mechanisms might indeed be involved.

REFERENCES

Arndt-Lappe, Sabine. 2011. Towards an exemplar-based model of stress in English noun–noun compounds. *Journal of Linguistics* 47.3, 549–585.

Baayen, R. Harald. 2005. Data mining at the intersection of psychology and linguistics. In Anne Cutler (ed.), *Twenty-first century psycholinguistics: Four cornerstones*, 69–83. Mahway, NJ: Lawrence Erlbaum.

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

https://doi.org/10.1017/S0022226712000199 Published online by Cambridge University Press

Baayen, R. Harald, Richard Piepenbrock & Leon Gulikers. 1995. *The CELEX lexical database* (CD-ROM). Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.
Bauer, Laurie. 1978. *The grammar of nominal compounding with special reference to Danish, English and French* (Odense University Studies in Linguistics 4). Odense: Odense University Press.
Bauer, Laurie. 1983. Stress in compounds: A rejoinder. *English Studies* 64.1, 47–53.
Bauer, Laurie. 1998. When is a sequence of two nouns a compound in English? *English Language and Linguistics* 2.1, 65–86.
Bell, Melanie J. 2005. Against nouns as syntactic premodifiers in English noun phrases. *Working Papers in English and Applied Linguistics* 11, 1–48. University of Cambridge, Research Centre for English and Applied Linguistics.
Bell, Melanie J. 2011. At the boundary of morphology and syntax: Noun noun constructions in English. In Alexandra Galani, Glynn Hicks & George Tsoulas (eds.), *Morphology and its interfaces*, 137–167. Amsterdam & Philadelphia, PA: John Benjamins.
Bell, Melanie J. 2012. *The English noun noun construct: Its prosody and structure*. Ph.D. thesis, University of Cambridge.
Bell, Melanie J. & Ingo Plag. 2012. Within-type variation in English compound stress: Towards an explanation. Ms., Anglia Ruskin University & Universität Siegen.
Bloomfield, Leonard. 1935. *Language*. London: George Allen & Unwin.
Bolinger, Dwight. 1972. Accent is predictable (if you're a mind-reader). *Language* 48.3, 633–644.
The British National Corpus, version 3 (BNC XML Edition). 2007. Distributed by Oxford University Computing Services on behalf of the BNC Consortium. http://www.natcorp.ox.ac.uk/ (accessed 20 May 2012).
Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper and Row.
Davies, Mark. 2004–. BYU-BNC (based on the British National Corpus). http://corpus.byu.edu/bnc/ (accessed 20 May 2012).
Farnetani, Edda, Carol Taylor Torsello & Piero Cosi. 1988. English compound versus non-compound noun phrases in discourse: An acoustic and perceptual study. *Language and Speech* 31.2. 157–180.
Fellbaum, Christiane (ed.). 1998. *WordNet: An electronic lexical database*. Cambridge, MA: MIT Press.
Fudge, Erik C. 1984. *English word-stress*. London: George Allen & Unwin.
Gagné, Christina L. 2001. Relation and lexical priming during the interpretation of noun–noun combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27.1, 236–254.
Gagné, Christina L. & Edward J. Shoben. 1997. Influence of thematic relations on the comprehension of modifier–noun combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 23.1, 71–87.
Giegerich, Heinz J. 2004. Compound or phrase? English noun-plus-noun constructions and the stress criterion. *English Language and Linguistics* 8.1, 1–24.
Gussenhoven, Carlos. 2004. *The phonology of tone and intonation*. Cambridge: Cambridge University Press.
Gussenhoven, Carlos & Ton Broeders. 1981. *English pronunciation for student teachers*. Groningen: Wolters-Noordhoff-Longman
Hastie, Trevor J. & Robert J. Tibshirani. 1990. *Generalized additive models* (Monographs on Statistics and Applied Probability 43). London: Chapman & Hall/CRC.
Jespersen, Otto. 1909. *A modern English grammar on historical principles. Part 1, sounds and spellings*. Heidelberg: Carl Winter's Universitätsbuchhandlung.
Jones, Daniel. 1972. *An outline of English phonetics*, 9th edn. Cambridge: W. Heffer & Sons.
Kingdon, Roger. 1958. *The groundwork of English stress*. London: Longmans, Green & Co.
Krott, Andrea, Robert Schreuder & R. Harald Baayen. 2002. Linking elements in Dutch noun–noun compounds: Constituent families as analogical predictors for response latencies. *Brain and Language* 81.3, 708–722.
Krott, Andrea, Robert Schreuder, R. Harald Baayen & Wolfgang U. Dressler. 2007. Analogical effects on linking elements in German compound words. *Language and Cognitive Processes* 22.1, 25–57.

https://doi.org/10.1017/S0022226712000199 Published online by Cambridge University Press

Kunter, Gero. 2010. Perception of prominence patterns in English nominal compounds. *Speech Prosody* 2010 (102007), 1–4. http://speechprosody2010.illinois.edu/papers/102007.pdf (accessed 20 May 2012).
Kunter, Gero. 2011. *Compound stress in English: The phonetics and phonology of prosodic prominence* (Linguistische Arbeiten 539). Berlin: Walter de Gruyter & Co.
Kunter, Gero & Ingo Plag. 2007. What is compound stress? In Jürgen Trouvain & William J. Barry (eds.), *Proceedings of the 16th International Congress of Phonetic Sciences, University of Saarbrücken, 6–10 August 2007*. Saarbrücken: Universität Saarbrücken.
Kutner, Michael H., Christopher J. Nachtsheim, John Neter & William Li. 2005. *Applied linear statistical models*, 5th edn. Boston, MA: McGraw-Hill Irwin.
Ladd, D. Robert. 1984. English compound stress. In Dafydd Gibbon & Helmut Richter (eds.), *Intonation, accent and rhythm: Studies in discourse phonology*, 253–266. Berlin: Walter de Gruyter.
Ladd, D. Robert. 2008. *Intonational phonology*, 2nd edn. Cambridge: Cambridge University Press.
Levi, Judith N. 1978. *The syntax and semantics of complex nominals*. New York: Academic Press.
Liberman, Mark & Richard Sproat. 1992. The stress and structure of modified noun phrases in English. In Ivan A. Sag & Anna Szabolcsi (eds.), *Lexical matters* (CSLI Lecture Notes 24), 131–181. Stanford, CA: CSLI Publications.
Lipka, Leonhard. 1994. Lexicalization and institutionalization. In Ronald E. Asher (ed.), *The encyclopedia of language and linguistics*, vol. 4, 2164–2167. Oxford: Pergamon.
Longman Dictionary of Contemporary English. 1978. London: Longman.
Marchand, Hans. 1969. *The categories and types of Present-day English word-formation: A synchronic–diachronic approach*, 2nd edn. Munich: C. H. Beck'sche Verlagsbuchhandlung.
Olsen, Susan. 2000. Compounding and stress in English: A closer look at the boundary between morphology and syntax. *Linguistische Berichte* 181, 55–70.
Ostendorf, Mari, Patti Price & Stefanie Shattuck-Hufnagel. 1996. *Boston University Radio Speech Corpus*. Philadelphia, PA: Linguistic Data Consortium.
Oxford Advanced Learner's Dictionary of Current English, 5th edn. 1995. Oxford: Oxford University Press.
The Oxford English Dictionary, 2nd edn. 1989. *OED Online*. Oxford University Press. http://www.oed.com (accessed 20 May 2012).
Pan, Shimei & Julia Hirschberg. 2000. Modeling local context for speech accent prediction. *The 38th Annual Meeting of the Association for Computational Linguistics*, 233–240.
Pan, Shimei & Kathleen R. McKeown. 1999. Word informativeness and automatic pitch accent modeling. *Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora* (EMNLP/VLC'99), 148–157.
Pennanen, Esko V. 1980. On the function and behaviour of stress in English noun compounds. *English Studies* 61.3, 252–263.
Plag, Ingo. 2006. The variability of compound stress in English: Structural, semantic and analogical factors. *English Language and Linguistics* 10.1, 143–172.
Plag, Ingo. 2010. Compound stress assignment by analogy: The constituent family bias. *Zeitschrift für Sprachwissenschaft* 29.2, 243–282.
Plag, Ingo & Gero Kunter. 2010. Constituent family size and compound stress assignment in English. In Susan Olsen (ed.), *New impulses in word-formation* (Linguistische Berichte Sonderheft 17), 349–382. Hamburg: Buske.
Plag, Ingo, Gero Kunter & Sabine Lappe. 2007. Testing hypotheses about compound stress assignment in English: A corpus-based investigation. *Corpus Linguistics and Linguistic Theory* 3.2, 199–233.
Plag, Ingo, Gero Kunter, Sabine Lappe & Maria Braun. 2008. The role of semantics, argument structure, and lexicalization in compound stress assignment in English. *Language* 84.4, 760–794
Sampson, Rodney. 1980. Stress in English N + N phrases: A further complicating factor. *English Studies* 61.3, 264–270.
Schreuder, Robert & R. Harald Baayen. 1997. How complex simplex words can be. *Journal of Memory and Language* 37.1, 118–139.
Shannon, Claude E. 1948. A mathematical theory of communication. *The Bell System Technical Journal* 27.3, 379–423 & 27.4, 623–656.

https://doi.org/10.1017/S0022226712000199 Published online by Cambridge University Press

Skousen, Royal & Theron Stanford. 2004. AM::Parallel (am2.3). Brigham Young University. http://humanities.byu.edu/am (accessed 20 May 2012).
Sweet, Henry. 1892. *A new English grammar logical and historical. Pt. 1: Introduction, phonology and accidence*. Oxford: Clarendon Press.
Wood, Simon N. 2006. *Generalized additive models: An introduction with R*. London: Chapman and Hall/CRC.
Zwicky, Arnold M. 1986. Forestress and afterstress. *Ohio State University Working Papers in Linguistics* 32, 46–62.

Authors' addresses: (*Bell*)
Department of English, Communication, Film and Media,
Anglia Ruskin University, Cambridge CB11PT, UK
melanie.bell@anglia.ac.uk

(*Plag*)
Anglistik/Sprachwissenschaft, Universität Siegen,
D-57068 Siegen, Germany
plag@anglistik.uni-siegen.de

https://doi.org/10.1017/S0022226712000199 Published online by Cambridge University Press