English Language and Linguistics 10.1: 143–172. © Cambridge University Press 2006
doi:10.1017/S1360674306001821 Printed in the United Kingdom

The variability of compound stress in English: structural, semantic, and analogical factors[1]

INGO PLAG
Universität Siegen
(Received 4 May 2005; revised 4 October 2005)

It is generally assumed that noun–noun (NN) compounds in English are stressed on the left-hand member (e.g. *cóurtroom*, *wátchmaker*). However, there is a considerable amount of variation in stress assignment (e.g. *silk tíe*, *Madison Ávenue*, *singer-sóngwriter*), whose significance and sources are largely unaccounted for in the literature. This article presents an experimental study in which three competing hypotheses concerning NN stress assignment are tested. The stress patterns of novel and existing compounds, as obtained in a reading experiment with native speakers of American English, were acoustically measured and analyzed. The results show that there is indeed a considerable amount of variation in stress assignment, and that all three hypothesized factors, i.e. structure, semantics, and analogy, are relevant, though to different degrees. On a theoretical level, the findings strongly suggest that a categorical approach cannot be upheld and that probability and analogy need to be incorporated into an adequate account of stress assignment in noun–noun constructions. The article also makes a methodological contribution to the debate in showing that experimental studies using pitch measurements can shed new light on the issue of variable compound stress.

1 Introduction

The last decade has seen a growing interest in alternative ways of describing what has traditionally been called a linguistic 'rule'. In both psycholinguistic and theoretical-linguistic circles there is a debate about the nature and role of symbolic rules, associative networks, and analogical mechanisms in the organization of language (see, for example, Clahsen, 1999, or Skousen et al., 2002). This interest has been fed by an increasing awareness even in generative linguistics of the fuzziness, semi-regularity and irregularity of many phenomena on all levels of linguistic description. The present article deals with one area where this semi-regularity is pervasive but

[1] This article is dedicated to Günter Rohdenburg on the occasion of his 65th birthday. I would like to thank the two *ELL* reviewers for their remarks on the first version of this article. I am also grateful to the audiences at the DUtKöMarSie-Workshop 2005, the Sprachwissenschaftliches Kolloquium at Universität Siegen 2005, and the Jahrestagung der Deutschen Gesellschaft für Sprachwissenschaft 2005 in Cologne for providing comments and suggestions. I am also very grateful to Heinz Giegerich, Laurie Bauer, Maria Braun, Miriam Ernestus, Sabine Lappe, Hiromi Noda, Gero Kunter, and Mareile Schramm for commenting on earlier versions. Thanks are also due to Gero Kunter for his help with some of the acoustic analyses and Holger Mitterer for his help with the Praat scripts. A very special thanks goes to Harald Baayen for his critical support, detailed suggestions, and for sharing his expertise in statistics with me. I also thank my student Julia Albrecht for starting all this. Needless to say, the usual disclaimers apply. This work was supported by a research grant from the Deutsche Forschungsgemeinschaft.

https://doi.org/10.1017/S1360674306001821 Published online by Cambridge University Press

has not yet received an explanation: stress assignment in English noun–noun (NN) compounds.[2]

While, for the majority of compounds, stress has traditionally been considered to be the output of a regular stress assignment rule, scholars have long acknowledged that there are a substantial number of forms which this rule cannot account for (e.g. Schmerling, 1971). Furthermore, there is cross-varietal variation (e.g. British English vs. American English), which makes it even harder to systematically investigate this type of semi-regularity.

In general, it has often been claimed that compounds tend to have a stress pattern that is different from that of phrases. This is especially true for nominal compounds, which is the class of compounds that is most productive. While phrases tend to be stressed phrase-finally, compounds tend to be stressed on the first element. This systematic difference is captured in the so-called nuclear stress rule and compound stress rule (Chomsky & Halle, 1968: 17). Phonetic studies (e.g. Farnetani & Cosi, 1988; Ingram et al., 2003) have shown in addition that segmentally identical phrases and compounds (such as *blackboard* vs. *black board*) differ not only significantly in their stress pattern, but also in length, with phrases being generally longer than the corresponding compounds. While the compound stress rule apparently makes correct predictions for the vast majority of nominal compounds, it has been pointed out, e.g. by Kingdon (1958), Fudge (1984), Liberman & Sproat (1992), Bauer (1998), Olsen (2000, 2001), and Giegerich (2004), that there are also numerous exceptions to the rule. Some of these forms are listed in (1). The most prominent syllable is marked by an acute accent on the vowel.

(1)

geologist-astrónomer	apple píe	scholar-áctivist
apricot crúmble	Michigan hóspital	Madison Ávenue
Boston márathon	Penny Láne	summer níght
aluminum fóil	May flówers	silk tíe

In view of this situation, the obvious question is how we can account for the variability in stress assignment of noun–noun constructs. Systematic empirical or experimental work on this variability is scarce, but many studies on compounding contain pertinent remarks and data. Basically, one finds three kinds of hypotheses that are spelled out in the literature, to different degrees of explicitness. These hypotheses, which will be discussed in more detail shortly, refer to either structural, semantic, or analogical factors that are held responsible for the stress of NN constructs.

[2] In general this article remains agnostic with regard to the issue of whether NN constructions should be analyzed as compounds or phrases. Empirically, the a priori exclusion of certain types of data might have had a bearing on the results. Theoretically, it has often been pointed out (e.g. more recently by Bauer, 1998; Spencer, 2003) that the stress criterion is inadequate to distinguish between the two types of construction (if one believes in this dichotomy in the first place). Hence I mostly speak of 'NN constructs' in this article, although the structures under investigation would be regarded as proper compounds by most analysts. I also use the term 'compound' for convenience's sake, but without theoretical commitment. See also the discussion of the structural hypothesis below on this point.

https://doi.org/10.1017/S1360674306001821 Published online by Cambridge University Press

The aim of the present article is to test the adequacy of these hypotheses in an experimental pilot study with native speakers of American English, in which the stress patterns of novel and existing NN constructs were acoustically measured. The results show that there is indeed a surprising amount of variation in stress assignment even within one variety of English. A statistical analysis of the acoustic data further reveals that all three of the above-mentioned factors are influential in assigning stress to NN constructs, and that probability and analogy must be incorporated into an adequate model of these phenomena.

The structure of the article is as follows. Section 2 reviews the three hypotheses mentioned above, which sets the scene for the experimental investigation. In section 3, I describe the experiment and discuss the methodological problems involved. Section 4 presents and discusses the results, which is followed by section 5, the final discussion and conclusion.

2 Three hypotheses on stress assignment to NN constructs

Three types of approaches have been taken to account for the puzzling facts of variable NN stress. The first is what I call the 'structural hypothesis'. Proponents of this hypothesis (e.g. Bloomfield, 1933; Lees, 1963; Marchand, 1969; or Payne & Huddleston, 2002) maintain that compounds are regularly left-stressed, and that word combinations with rightward stress cannot be compounds, which raises the question of what else such structures could be. One natural possibility is to consider such forms to be phrases. However, such an approach would face the problem of explaining why not all forms that have the same superficial structure, i.e. NN, are phrases. Second, one would like to have independent criteria coinciding with stress in order to say whether something is a lexical entity (i.e. a compound) or a syntactic entity (i.e. a phrase). This is, however, often impossible: apart from stress itself, there seems to be no independent argument for claiming that *Mádison Street* should be a compound, whereas *Madison Ávenue* (or *Madison Róad*, for that matter) should be a phrase. Both kinds of words seem to have the same internal structure, both show the same meaning relationship between their respective constituents, both are right-headed, and it is only in their stress patterns that they differ. Spencer (2003) also argues that we find compounds with phrasal stress and phrases with compound stress, and hence that stress is more related to lexicalization patterns than to structural differences, a point taken up by Giegerich (2004, to be discussed in more detail shortly). A final problem for the phrasal analysis is the fact that the rightward stress pattern seems often triggered by analogy to other combinations with the same rightward element. This can only happen if the forms on which the analogy is based are stored in the mental lexicon. And storage in the mental lexicon is something we would typically expect from words (i.e. compounds), and only exceptionally from phrases (as in the case of *jack-in-the-box*).

Most recently, Giegerich (2004) has proposed a new variant of the structural hypothesis. On the basis of the fact that in English syntax complements follow the head, he argues that, due to the order of elements, complement–head structures like

https://doi.org/10.1017/S1360674306001821 Published online by Cambridge University Press

trúck driver cannot be syntactic phrases, hence must be compounds, hence are left-stressed. Modifier–head structures such as *steel brídge* display the same word order as corresponding modifier–head phrases (cf. *wooden brídge*), hence are syntactic structures and regularly right-stressed.[3]

This means, however, that many existing modifier–head structures are in fact not stressed in the predicted way, since they are left-stressed (e.g. *ópera glasses, táble cloth*). Such aberrant behavior, is, according to Giegerich, the result of lexicalization. Lexicalization as an explanation for leftward stress makes an interesting prediction: we should expect that novel modifier–head compounds should generally receive rightward stress. Furthermore, the amount of leftward-stressed compounds should vary according to frequency in corpora, since we know that lexicalization strongly correlates with frequency.[4] Thus, we should find more modifier–head structures with leftward stress among the more frequent items. And we should never find rightward stress among those NN constructs that exhibit complement–head order. The latter point is, however, not always true, as pointed out by Giegerich himself, who cites *Tory léader* as a counterexample. Furthermore, the structural hypothesis predicts that compounds with the same rightward member exhibit different stress patterns, depending on whether the leftward member is an argument or a modifier. Pairs such as *yárd sale* vs. *bóok sale* (or *trúck driver* vs. *Súnday driver*) suggest that this prediction is probably wrong. In any case, none of these predictions has ever been systematically tested against larger amounts of data.

Before turning to the discussion of what I call the 'semantic hypothesis', I would like to point out that what I have labeled 'structural hypothesis' is the hypothesis that rests largely on the argument–modifier distinction. Although this distinction clearly has strong semantic implications, there are, as pointed out above, crucial structural facts that correlate with this distinction. This is my reason for calling the hypothesis structural, although the underlying distinction might be semantic.

The second approach to variable compound stress is what can be called the semantic hypothesis. A number of scholars have argued that words with rightward stress such as those in (1) above are systematic exceptions to the compound stress rule (e.g. Sampson, 1980; Fudge, 1984; Ladd, 1984; Liberman & Sproat, 1992; Olsen, 2000, 2001; Spencer, 2003). Although these authors differ slightly in details of their respective approaches, they all argue that rightward prominence is restricted to only a limited number of more or less well-defined types of meaning relationships. For example, compounds like *geologist-astrónomer* and *scholar-áctivist* are copulative compounds, and these

[3] Giegerich characterizes modifier–head structures in terms of their lack of argument–predicate semantics. I prefer the term 'argument–head' instead of 'argument–predicate' in the context of this article because of its parallelism with 'modifier–head'.

[4] Cf. Lipka's definition, according to which lexicalization 'is defined as the process by which complex lexemes tend to become a single unit, with a specific content, *through frequent use*' (1994: 2165, my emphasis). Bauer (1983a: 51) mentions irregular stress assignment in English derivatives and Danish compounds as prototypical cases of (phonological) lexicalization. See also Adams (1973: 59), who writes that 'in established NPs *which are used frequently* and over a period of time the nucleus tends to shift from the second to the first element although this does not always happen' (my emphasis).

https://doi.org/10.1017/S1360674306001821 Published online by Cambridge University Press

are uncontroversially and regularly right-stressed.[5] Other meaning relationships that are often, if not typically, accompanied by rightward stress are temporal or locative (e.g. *a summer night*, *the Boston márathon*), or causative, usually paraphrased as 'made of' (as in *aluminum fóil, silk tíe*), or 'created by' (as in *a Shakespeare sónnet, a Mahler sýmphony*). It is, however, unclear how accurately the membership in a given class can really predict the kind of stress. The leftward stress on *súmmer school, súmmer camp* or *dáy job*, for example, violates Fudge's (1984: 144ff.) generalization that NNs in which N_1 refers to a period or point of time are right-stressed. Furthermore, it is unclear how many, and which, semantic classes should be set up to account for all the putative exceptions to the compound stress rule (see also Bauer, 1998: 71 on this point). Finally, semantically very similar compounds can behave differently in terms of stress assignment (*Fífth Street* vs. *Fifth Ávenue*). And again, we have to state that detailed and systematic empirical studies are lacking for most of the postulated classes.

Note that I use the label 'semantic hypothesis' in this article to refer to approaches that set up semantic categories and correlate these with stress patterns. Although these approaches actually never refer explicitly to the modifier–argument distinction, the semantic categories that are alleged to produce rightward stress would all involve modifier–head compounds, but never argument–head compounds. Thus, structural and semantic hypothesis converge on the point that they expect rightward stress to be largely restricted to modifier–head compounds.

Finally, a third approach can be taken which draws on the idea of analogy and hypothesizes that stress assignment is generally based on analogy to existing NN constructions in the mental lexicon. Plag (2003: 139) mentions the textbook examples of *street* vs. *avenue* compounds as a clear case of analogy. All street names involving *street* as their right-hand member pattern alike in having leftward stress (e.g. *Óxford Street*, *Máin Street, Fóurth Street*), while all combinations with, for example, *avenue* as right-hand member pattern alike in having rightward stress (e.g. *Fífth Ávenue, Madison Ávenue*). Schmerling (1971: 56) provides more examples of this kind, arguing that many compounds choose their stress pattern in analogy to combinations that have the same head, i.e. rightward member. It is, however, unclear how far such an analogical approach can reach. Along similar lines, Spencer (2003: 331) proposes that 'stress patterns are in many cases determined by (admittedly vague) semantic "constructions" defined over collections of similar lexical entries'. In a similar vein, Ladd (1984) proposes a destressing account of compound stress which would explain the analogical effects triggered by the same rightward members as basically semantico-pragmatic effects.

What is considered the effect of lexicalization in some approaches would emerge naturally in an analogical system, in which existing (i.e. lexicalized) compounds influence new (i.e. nonlexicalized) compounds to behave similarly. This raises the question on which basis similarity could be computed (cf. also Liberman & Sproat,

[5] As pointed out by one reviewer, even this nice generalization has its (apparently very few) exceptions, for example *mán-servant*, which is left-stressed.

https://doi.org/10.1017/S1360674306001821 Published online by Cambridge University Press

1992: 176 on this point). In principle, any property could serve that purpose, for example, the number of syllables of the right member, the semantic properties of the left member, or, perhaps absurdly, the third segment of the left member, or a combination of these. One rather simple assumption to start out with is that it is the right member that is responsible for the choice of the stress pattern. Given, for example, a set of compounds with the same right member, we would first expect that the vast majority of items in that set are stressed in a certain way, e.g. leftward, and that any novel form with that right-hand member will also receive leftward stress. A more sophisticated analogical model would incorporate of course more, and different, types of linguistic information (phonological, semantic, structural, frequential).

At this point a note is in order on the notion of analogy as used in different traditions. The traditional notion of analogy has been rightly criticized by many because it is difficult to see how any falsifiable prediction might be obtained with it. For instance, in Goldberg & Jackendoff (2005) we still find the statement that 'analogy is notoriously difficult to constrain' (2005: 475). Recent work in computational morphology has shown, however, that a formal, constrained, and computationally tractable notion is available that offers new ways of understanding the ways in which linguistic rules actually work. Such formal analogical models have been quite successful in predicting both regular and irregular morphology in general, and variable compound behavior in particular. For example, Krott and her collaborators (Krott et al., 2001, 2002, 2004) analyzed the semi-regular behavior of the linking morphemes in Dutch compounds in terms of analogy, using a memory-based analogical learning algorithm (TiMBL, Daelemans et al., 2000).[6] They compared the algorithm's performance with that of native speakers in an experiment with novel compounds and found that the variable occurrence of the three linking morphemes in Dutch compounds is much better accounted for by a dynamic analogical mechanism than by traditional symbolic rules. Although the analogical hypothesis has been evoked here and there (and quite informally) in treatments of English compounds, it has never been tested empirically or formally modeled (cf. Spencer's above-cited remark on the vagueness of possible analogical sets).

To summarize, there are three reasonable hypotheses available to account for the variability of NN stress, all of which are in some sense problematic and all of which are still in need of serious empirical testing. One way to do this is to carry out experimental studies, in which the data can be carefully controlled for the different potential factors involved. The present investigation is the first such study.

3 Methodology

3.1 Experimental setup

In order to test the three hypotheses I devised a reading experiment with novel and existing compounds, which was carried out with nine native speakers of American

[6] Analogical effects in compound interpretation have been shown to exist by Gagné and her collaborators (e.g. Gagné & Shoben, 1997; Gagné, 2001).

https://doi.org/10.1017/S1360674306001821 Published online by Cambridge University Press

English, three of them male, six of them female.[7] The experiment contained two kinds of stimuli. In order to test the productivity of stress assignment, stimuli had to be novel compounds. In order to test the lexicalization issue, stimuli had to be existing compounds. With regard to the semantic hypothesis, the experiment was designed to test the above-mentioned authorship relation (e.g. *Shakespeare sónnet*) as a case study. One subset of the stimuli contained 24 test items that were novel structures exhibiting either an authorship relation (e.g. *Kauffman symphony*) or no such relation (e.g. *Christmas symphony*). For ease of reference and lack of a better term, the latter relation will be referred to as the 'title relation' in the rest of the article. Note that both authorship relation and title relation are structurally modifier–head, and not argument–head, structures. The other subset of the stimuli contained 25 existing compounds, which were either argument–head structures or modifier–head structures (e.g. *opera singer* vs. *opera glasses*).

The experimental subjects had to read out sentences containing the randomized stimuli and some filler sentences. Potential sequencing effects were controlled for by presenting the sentences in two different orders (one the reverse of the other).[8] Most sentences expressed propositions having to do with classical music, such that the stimuli occurred in a more or less natural context. None of our subjects was an expert in (classical) music.

The acoustic and statistical analysis was done using the speech analysis software Praat (Boersma & Weenink, 2005) and the statistical package R. This analysis involved three major methodological problems, to which we now turn.

3.2 Methodological problem 1: Recognition of rightward and leftward stress

One major problem when dealing with compound stress is to determine whether a given form has leftward or rightward stress. In the linguistic literature forms are usually cited as having either leftward, rightward, or level stress, or as having variable stress. These classifications are normally based on the individual intuition of the researcher and are not the result of systematic investigations, let alone of acoustic or articulatory analyses. For some purposes, the reliance on intuition may be sufficient, but in an area where variation is prevalent and in the focus of the investigation, using individual intuition should be substituted by a more sophisticated methodology.

In our case, I first had listeners rate the individual items according to their perception. It turned out, however, that the assignment of items to the two stress categories was often very difficult for raters. One reason for this difficulty was that items were not tested in isolation, but embedded in a more or less natural context, which may make

[7] Six of the speakers have spent most of their lives in New Jersey, the other three come from other parts of the northeast of the US.

[8] A complete list of the stimuli and the sentences to be read out by the experimental subjects is given in the appendix. See also (4) below for a presentation of the 24 novel compounds according to their right and left members.

https://doi.org/10.1017/S1360674306001821 Published online by Cambridge University Press

it even harder for listeners to come up with clear judgments.[9] The rating methodology has the serious disadvantage that one has to deal with variation in the rating (within and across raters) and a significant portion of unclear cases where raters fail to make judgments, which reduces the number of observations that can enter the actual analysis. This method was therefore discarded after a trial.

A more objective method to investigate stress is of course the analysis of the acoustic correlates of stress, i.e. pitch, intensity, and duration (e.g. Hayes, 1995). Of the three factors, pitch is generally regarded as the strongest indicator of stress, with intensity and duration having ancillary function (cf. Lehiste, 1970: 120; Ladefoged, 2005: 92). This seems also true for compound stress. For example, in their investigation of minimal pairs of compounds and phrases in English (e.g. *blackberry, black berry*), Farnetani & Cosi (1988) as well as Ingram et al. (2003) found that pitch and duration are the most reliable indicators of compound stress. However, taking duration as an indication of stress usually involves minimal pairs, and hence was not applicable in the present study. This leaves us with pitch as the most important acoustically measurable correlate of stress. In order to capture the pitch patterns, I measured the fundamental frequency F0 in the middle of the main stressed vowels of the two compound members, respectively, and calculated the difference by subtracting the value for F0 of the right member from the value for F0 of the left member. The resulting difference will be referred to as 'pitch difference' in the rest of the article. We also measured the intensity at the same measure points, dividing the intensity value of the left member by the intensity value of the second member, in order to be able to also have a look at the relationship of pitch and intensity.

Measuring pitch and using the pitch measures as indicators of stress poses some additional methodological problems. First, there are coarticulation effects in that vowels in voiceless environments are higher in fundamental frequency than vowels in voiced environments. This effect is most pronounced at the transition points of consonants and vowels. Thus, in order to eliminate such coarticulation effects as far as possible we chose the middle of the stressed vowel as the measure point (this measure was also employed by Ingram et al., 2003).

Second, fundamental frequency drops over the course of an utterance, typically 20–40 Hz (e.g. Sternberg et al., 1980: 518ff). To my knowledge it has never been investigated how such downstepping may operate within compounds. Available studies (e.g. Pierrehumbert, 1979; Liberman & Pierrehumbert, 1984) have shown that there is a constant downstep ratio across the F0 peaks of an utterance. It is unclear, however, whether the two peaks of a compound should count as two peaks in such analyses. If so, there could potentially be a general bias in our data against rightward stress, such that all left-stressed items are more clearly left-stressed due to downstepping, and that the right-stressed items are less clearly right-stressed due to downstepping (which, incidentally, could be one reason for the impression of level stress often

[9] See Fry (1958), Bauer (1983b) or, more recently, Gussenhoven (2004: 3) on the variability of speaker/listener judgments in experiments on stress.

https://doi.org/10.1017/S1360674306001821 Published online by Cambridge University Press

mentioned in the literature). While this may indeed be the case, we still find significant differences in stress assignment in our data between different types of compounds, i.e. left-stressed and right-stressed ones. In terms of perception, speakers seem to normalize the expected F0 declination, thus making up for the less pronounced prominence of the right constituent in rightward-stressed compounds (see below for more discussion).

Third, we know that vowels differ with regard to their intrinsic pitch. Other things being equal, the higher a vowel, the higher is its pitch. This effect is quite small in general, and it seems negligible in our experiment, because higher and lower vowels occur in all positions across all types of compounds across most conditions (see below for details).

Fourth, it has been shown that stress is sometimes indicated by a downstep in pitch, instead of the expectable upstep (i.e. higher pitch; see e.g. Lehiste, 1970: 128; Ladefoged, 2005: 93). However, in the exceptional cases of downstep pitch discussed by these authors, intensity and length still show high values, i.e. intensity and length still behave in the usual way and thus somehow counterbalance the downstep effect of pitch. On the basis of these findings, one could hypothesize that any low pitch on the right-hand element in our data could just as well be an indication of rightward stress, instead of being an indication of leftward stress. Two objections can be raised against this hypothesis. One is that it is unclear whether the exceptional downstep marking of stress is observable in word stress at all, or whether it is restricted to sentence stress, since all examples of stress-marking downstep I found discussed in the literature are instances of sentence stress. The other objection is that the hypothesis is simply empirically not supported. In our data, there is a highly significant correlation between pitch difference and intensity ratio (Pearson, $r = .325$, $p < .000$), which means that we hardly find cases in which low pitch on the second member goes together with high intensity on that constituent. In other words, in the compounds in the experiment, vowels with low pitch usually have low intensity and vice versa, which means that in compounds downstepping is obviously not used to mark stress.

Measuring pitch and intensity in the described way does not, however, solve the major problem yet, namely that of detecting rightward and leftward stresses in the data. Given a pitch difference between the two members, what is to be regarded as leftward and rightward stress? Two approaches are possible, one absolute, the other relative. Under the categorical approach, one can simply assume that if the pitch difference is positive, the item is left-stressed, and if the difference is negative, the item is right-stressed. In (2) I have listed two examples from the sample of subject 1, where this approach yields the desired result:

(2)

	F0 difference	**Stress**
morning páper	−40.23 Hz	rightward
ópera glasses	+73.46 Hz	leftward

According to these figures this subject clearly has rightward stress on *morning páper* and leftward stress on *ópera glasses*, as indicated by the acute accents. This corresponds

https://doi.org/10.1017/S1360674306001821 Published online by Cambridge University Press

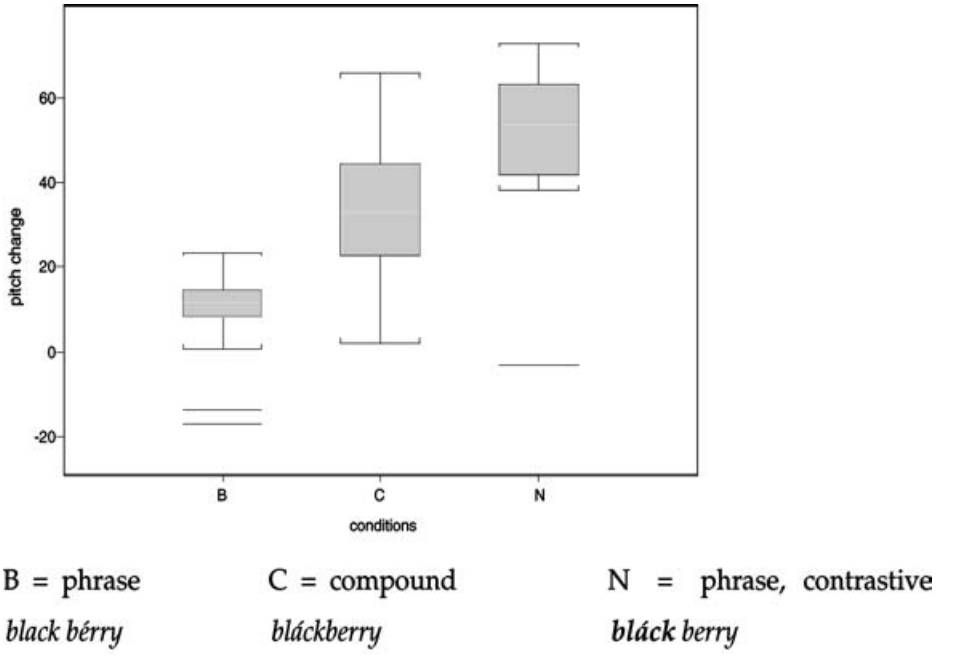


Figure 1. Pitch difference of three different stress patterns (Ingram et al., 2003)

consistently with the perceptual impressions of these two items with our raters. There is, however, a serious problem involved with using positive vs. negative pitch difference as indicators of leftward and rightward stress, respectively.

The problem is that what speakers perceive and produce as rightward stress does not necessarily imply a negative pitch difference. Ingram et al. (2003) show, for example, that there are clear differences between clearly leftward-stressed compounds (*bláckberry*), their segmentally identical corresponding syntactic phrases (*black bérry*), and the corresponding contrastively stressed syntactic phrases (as in *it was a bláck berry, not a blue one*). The interesting thing now is that the pitch difference for the phrases (such as *black bérry*) is generally positive, not negative, but still significantly smaller than the pitch differences observable with the other two categories. Compare figure 1, taken from Ingram et al. (2003). What I call 'pitch difference' in this article is called 'pitch change' in Ingram et al.'s paper. The boxes in figure 1 show the interquartile range with the line indicating the medians of the pitch differences of the three categories. The whiskers give 1.5 times the interquartile range. The differences between the three categories B, C, and N are all statistically highly significant, which means that we are dealing with three different stress patterns, i.e. a ternary contrast. Note, however, that the clearly rightward-stressed set of data in their vast majority do not show a negative pitch difference. Hence, what makes these items rightwardly stressed is not an absolute negative pitch difference but the relative difference between the pitch

https://doi.org/10.1017/S1360674306001821 Published online by Cambridge University Press

differences of this category and the pitch differences of the other categories (in spite of some overlap of categories). Similar results had already been obtained by Farnetani & Cosi (1988).

In sum, the absolute approach is theoretically inadequate, because it assumes an arbitrary and incorrect boundary between stress categories at 0 Hz pitch difference. From these results, we must draw the conclusion that the stress pattern of a set of data should be determined not in absolute terms of negative or positive pitch differences, but relative to the pitch differences of other sets of data.

Although the relative approach seems theoretically clearly preferable to the absolute one, I first used both kinds of analysis for our data to check the actual effect of this methodological difference. In accordance with the categorical approach I looked for positive and negative pitch differences. However, I also looked for statistically significant differences in F0 between different kinds of compound, along the lines of Ingram et al. (2003). Under the latter approach, statistically significant differences indicate different stress categories. In general, the larger one of two mean differences indicates what we know as leftward stress, the smaller mean difference indicates rightward stress. The relativity of the different stress categories may in fact explain the otherwise strange category of 'level' or 'double' stress sometimes found in the literature (e.g. Faiss, 1981: 132; Marchand, 1969: 2.1.2ff).[10] The actual number of different stress levels is, however, of no concern in this article and remains a topic of future research.

The overall tendencies with regard to the three hypotheses investigated in this article were very similar under both approaches, which is illustrated in an exemplary fashion for the effect of structural relation (argument vs. modifier) in figure 2.[11] The upper panel of figure 2, which plots the pitch differences of argument–head and modifier–head compounds in Hertz, shows that there are clear differences between the two categories both in terms of means and in terms of variance. The difference between argument–head and modifier–head compounds is statistically highly significant (see below for more detailed discussion). The lower panel of figure 2, which is created using the absolute approach, gives us less information than the upper panel. Only within the modifier–head compounds do we find rightwardly stressed items as defined in the absolute approach (19%), i.e. items with a negative pitch difference. The difference between the two categories is again statistically significant ($p = .01$, Fischer's Exact Test). What the lower panel does not show, however, is the full range of variability within each category, which is crucial for a more detailed and a more sophisticated analysis with statistical tools more powerful than chi-square or Fischer's Exact tests for nominal data.

Thus, apart from theoretical considerations on the nature of stress categories and their acoustic correlates, the relative approach is also advantageous from a practical point of view, because it allows for more fine-grained analyses and better statistical

[10] Giegerich (2004: fn 2) is also highly skeptical of the alleged category of level stress.
[11] Dots in this and the following boxplots represent individual outliers.

https://doi.org/10.1017/S1360674306001821 Published online by Cambridge University Press

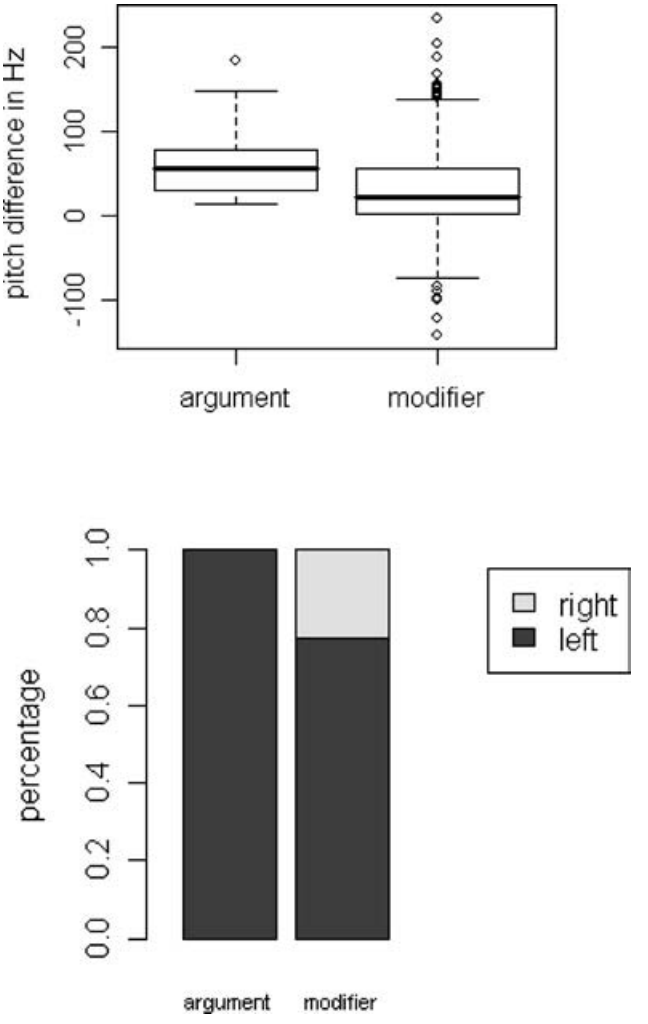


Figure 2. Measuring compound stress: relative vs. absolute approach

procedures. In this article I therefore only report the results obtained under the relative approach.

3.3 Methodological problem 2: The effect of clause type and sentential position

The second major problem concerning the investigation of compound stress is the effect of sentence intonation. It may seem, for example, that a compound in the final position of a question would show different F0 contours than the same compound in initial position of a declarative clause. How can this be dealt with in an empirical study? To control and at the same time investigate the potential effect of clause type and clause

https://doi.org/10.1017/S1360674306001821 Published online by Cambridge University Press

position, stimuli were placed in three kinds of clausal position in three kinds of clauses. Consider (3) for illustration of the three types of clauses and position, respectively. The items to be tested are given in bold:

(3) (a) declarative clause (n = 40, N = 333, m = 25)[12]
  i. initial position (n = 12, N = 102, m = 5)
  *This **concertmaster** is brilliant.*
  ii. medial position (n = 9, N = 78, m = 3)
  *The orchestra played the **Dream sonata** after the break.*
  iii. final position (n = 19, N = 153, m = 17)
  *At the moment we are rehearsing the **Twilight sonata**.*
 (b) interrogative clause, final position (n = 4, N = 34, m = 2)
  *What did you think about this new **piano arrangement**?*
 (c) preposed sub- or superordinate clause, final position (n = 5, N = 43, m = 1)
  *I always recognize a **Rossi sonata** when I hear one.*

NN constructs occurred in clause-initial, clause-medial, and clause-final NPs, and in declarative main clauses, in preposed sub- or superordinate clauses, and in interrogative clauses. The preposed sub- or superordinate clauses with the pertinent NP in final position were included to test and control the effect of level intonation as a continuance signal. In interrogative clauses the pertinent NP was also always placed in final position to test and control the effect of question intonation. In the declarative clauses 'clause-initial' refers to the first position in the sentence, and 'clause-medial' to the object NP preceding a sentence-final adjunct. 'Clause-final' generally means the last phrase in the clause. Not all positions and clause types were equally frequent in the stimuli in order to keep the number of items in the experiment manageable for the experimental subjects.

Looking at the different sentential positions, we can see that, quite expectedly due to downstepping, the pitch values steadily decrease from initial to final position. This is shown in figure 3, which plots the pitch values of the left and right members across sentential positions. The potential effects of position and clause type on stress assignment will be discussed as we go along.

3.4 Methodological problem 3: Pitch differences between men and women

The third and last major methodological problem concerns the problem of pitch variability according to sex. There seem to be important differences in pitch height and pitch range between men and women, both in general and in our experimental data.[13] For example, the females in our experiment have a mean pitch of 237 Hz

[12] For various reasons (most of them of a technical nature, having to do with the algorithms of the speech analysis software), not all items could be subjected to the acoustic analyses. The first figure in parenthesis gives the number of pertinent test sentences for each subject ('n'), the second figure in parenthesis gives the number of analyzable data points ('N'), and the third figure gives the number of data points that could not be analyzed ('m').

[13] Note, however, that there is no real consensus in the literature about the question of pitch range differences between male and female speakers. Different studies report different kinds of results (see, for example, Haan & van Heuven, 1999 for discussion).

https://doi.org/10.1017/S1360674306001821 Published online by Cambridge University Press

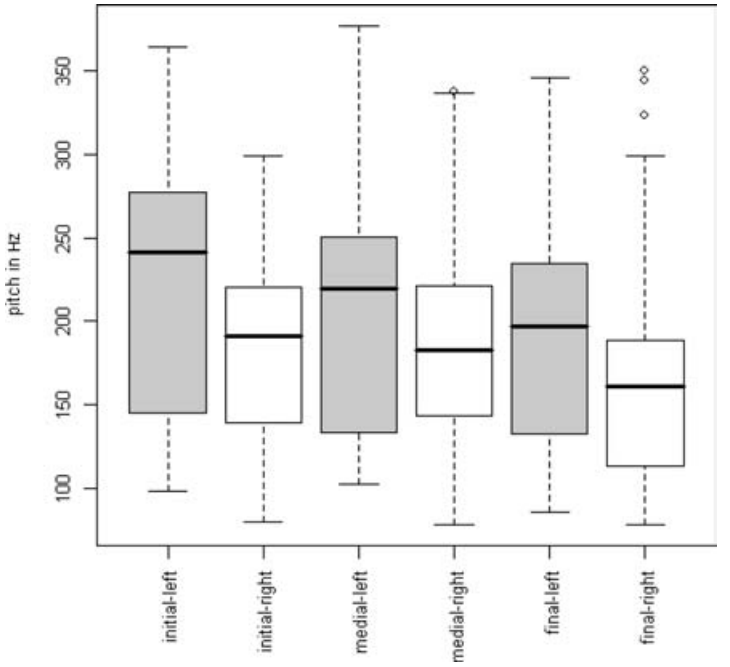


Figure 3. Downstep effects in compounds: pitch of left and right member by position (left pitches are shaded)

for the first member of the stimuli compounds, while the male data show only 125 Hz. Published data on the average fundamental frequency in speech show that the range of variation is approximately the same for men and women if expressed in perceptually more appropriate semitones (ST; see e.g. Henton, 1989; Traunmüller & Eriksson, 1995 for overviews; and Nolan, 2003 for logarithmic vs. other models of establishing equivalence between different pitch spans). We therefore transformed the pitch measurements logarithmically into semitones in order to statistically minimize the differences between men and women.[14] It should be noted, however, that even in terms of semitones, there is still a highly significant difference between men and women in our data with regard to their marking of compound stress (ST-difference: $F(1, 408) = 23.15$, $p < 0.001$). This is illustrated in figure 4 for pitch differences. Overall we found a robust main effect of sex across most conditions, but no significant interaction of sex with other predictors (structural relation, semantic relation, clause type or sentential position). I will document this in more detail as we go along. In general this means that, in our sample, women mark stress with more pronounced pitch differences and higher intensity ratios than men do.

[14] The pertinent model (cf. e.g. Henton, 1989: 302) is:

$$\text{difference in semitones} = 12^{*} \log (\text{left pitch}/\text{right pitch})/\log (2)$$

https://doi.org/10.1017/S1360674306001821 Published online by Cambridge University Press

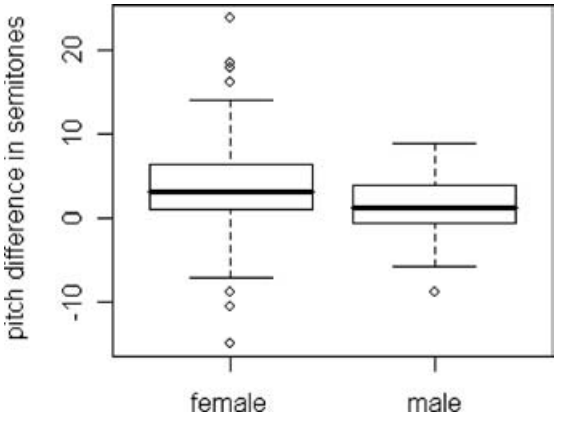


Figure 4. Pitch difference by sex

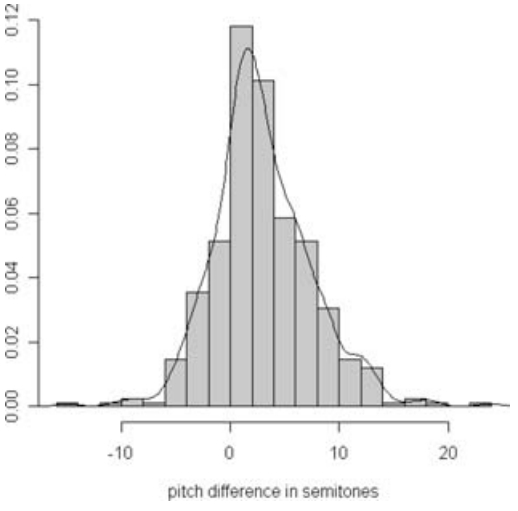


Figure 5. Histogram and density function for pitch difference

4 Predictions and results

In this section I will present the experimental results, each time first introducing the predictions of the different hypotheses for our experimental data and then giving the pertinent results.

Overall, the data revealed a considerable amount of variation. This is illustrated in figure 5, which shows the distribution of pitch differences in the data. The pitch differences show a range of roughly 30 semitones. Negative values indicate that the right constituent has a higher pitch than the left constituent. In Hertz, the pitch differences range from well below −100 Hz through more than 200 Hz. This variability calls for

https://doi.org/10.1017/S1360674306001821 Published online by Cambridge University Press

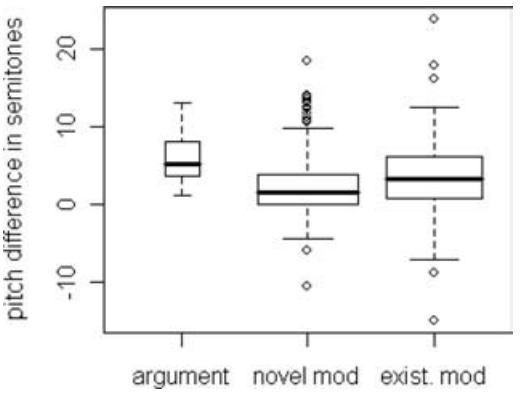


Figure 6. Pitch difference according to structural relation

an explanation, which I will attempt to provide in the following sections along the lines of the three competing hypotheses. I will start out with the structural hypothesis.

4.1 The structural hypothesis

This hypothesis makes three predictions for our experimental data. First, it predicts that all our novel compounds have rightward stress, because they do not exhibit argument–head relations, but modifier–head relations (either 'author' or 'title'). Second, existing compounds in our data should have rightward stress on modifier–head compounds and leftward stress on argument–head compounds. Third, it predicts that the leftward-stressed modifier–head compounds we might find in our data are lexicalized compounds that bear lexically marked exceptional leftward stress. How can this third prediction be tested? With regard to lexicalization, I follow the standard assumption that frequency is the surface correlate of lexicalization. Under this assumption the structural hypothesis makes the prediction that there should be a correlation between the frequency of a compound and its stress pattern. The higher the frequency of an existing compound, the more left-stressed tokens we find, because this is the lexicalized stress pattern of modifier–head structures.

Let us look at the results. We already saw in figure 2, upper panel, that there is a difference between argument–head and modifier–head compounds. Let us take a closer look at this difference by comparing the pitch differences for compounds with an argument–head relation (all of them existing compounds) with those of novel and existing modifier–head compounds. Figure 6 gives the pertinent boxplots.

A one-way ANOVA revealed the presence of significant differences between the three group means ($F(2, 407) = 8.47$, $p < 0.001$). A Welch-modified t-test showed that the mean pitch difference for argument–head compounds is significantly higher than that of modifier–head compounds ($t(29.46) = 4.6371$, $p < 0.001$; see also figure 3 above for a visualization in Hz). One further Welch-modified t-test revealed that the mean pitch difference for novel compounds is significantly lower than that for existing compounds ($t(402.737) = 3.1262$, $p < 0.002$). This is probably an effect of

https://doi.org/10.1017/S1360674306001821 Published online by Cambridge University Press

the high pitch difference of the argument–head compounds, which were also existing compounds. The two t-tests remain significant after a Bonferroni correction (both corrected p-values < 0.01).

An analysis using a linear mixed-effects model fit by Restricted Maximum Likelihood (REML) (see Bates & Sarkar, 2005), with subject and item as crossed random effects and structural relation and status as novel or existing form as fixed effects, yielded significant results ($F(2, 407) = 3.0697$, $p < 0.05$). After checking the residuals, which were non-normal, and the subsequent removal of nine outliers, the trimmed model showed an improved significance ($F(2, 398) = 3.7598$, $p < 0.03$). Including sex as a fixed effect and subject and item as crossed random effects in the multilevel analysis showed again a main effect of structural relation ($F(2, 397) = 3.8015$, $p < 0.03$) and of sex ($F(2, 397) = 5.1220$, $p < 0.03$), with no interaction. The experimental design did not allow me to add position as an explanatory variable to the model, as this would have led to a great many empty cells.

In sum, argument–head compounds have a higher pitch difference than modifier–head compounds, as predicted by the structural hypothesis. This effect is robust even if we take other sources of variability into account (subject, item, novelty, and sex).

We also see in figure 6 that argument–head compounds, apart from having a higher and consistently positive pitch difference, show a rather narrow range of variability. In contrast, a high proportion of modifier–head compounds have a negative pitch difference, and the overall variability is considerable. Furthermore, although there is a significant difference between the two categories, a fairly large number of modifier–head compounds clearly exhibit leftward stress, with pitch differences way above the mean pitch difference of argument–head compounds. This set of aberrant forms should, according to the hypothesis, be instances of lexicalization. For novel modifier–head compounds this is, however, impossible. The leftward stress of novel modifier–head compounds is inexplicable under the structural hypothesis and must have its source outside lexicalization.

In order to test a potential lexicalization effect with the existing modifier–head compounds, it is first necessary to establish the frequencies of our existing compounds. This was done in two ways. First, the frequencies were collected using Google. This is, however, not unproblematic due to distortions of frequency counts having to do with the specific indexing algorithms Google uses.[15] The Google frequencies were therefore complemented by frequency data from a large and carefully controlled corpus, the British National Corpus (BNC). In spite of its being a very large corpus of 100 million words, not all of our existing compounds were attested in the BNC, so that the use of the internet and Google, even if not totally adequate in its sampling procedure, was a welcome complementary procedure. An obvious drawback of the BNC is the fact that this corpus represents British English, but no corpus of American English of

[15] These problems with Google counts have recently been the subject of discussions in various internet forums, for example on corpora-list. Interested readers may consult the following websites for details: http://aixtal.blogspot.com/2005/02/web-googles-missing-pages-mystery.html, http://torvald.aksis.uib.no/corpora/

https://doi.org/10.1017/S1360674306001821 Published online by Cambridge University Press

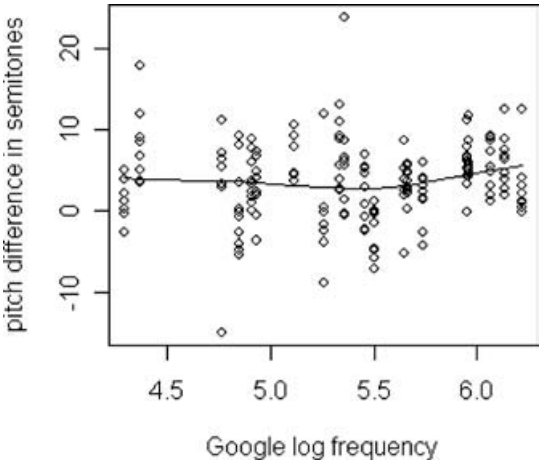


Figure 7. Pitch difference according to Google log frequency

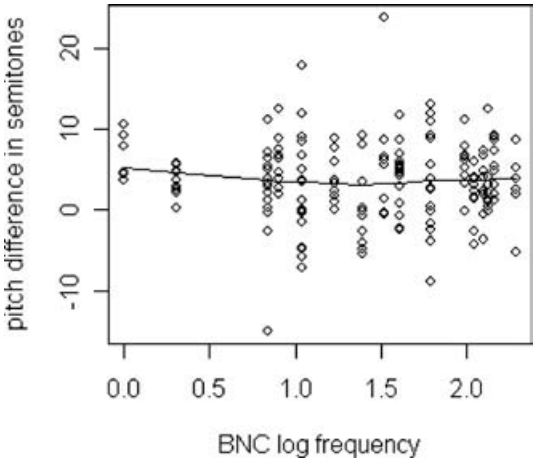


Figure 8. Pitch difference according to BNC log frequency

comparable size was available at the time of this study. There is no reason to assume, however, that there will be significant differences in the frequencies of these compounds between British and American English.

Recall that according to our interpretation of the structural hypothesis, one should find a correlation between the frequency of leftward stress among the observed items and the frequency of occurrence. The more frequent the word, the more likely it is that it has lexicalized, i.e. leftward, stress. Consider figures 7 and 8, which plot the Google log frequencies and the BNC log frequencies against the pitch difference measured in semitones. Each point in the plot represents one observation. In both graphs the general tendency in the data has been made visible with the help of a scatterplot smoother (Venables & Ripley, 2000: 228–32). Under the lexicalization hypothesis we would expect the dots to scatter around a line with a positive slope. This is, however,

https://doi.org/10.1017/S1360674306001821 Published online by Cambridge University Press

not the case. In neither of the two graphs do we find the predicted correlation between pitch difference and lexical frequency (Google: $p > 0.3$, BNC: $p > 0.8$). In other words, the frequency of an existing modifier–head compound does not help to predict its stress pattern. There is no lexicalization effect traceable in our data.

To summarize this section, we can say that the structural hypothesis is strongly supported with regard to the role of the argument–modifier distinction. Argument–head compounds are categorically left-stressed, while modifier–head compounds are not. There is a robust difference between these two categories of compound. However, there are also a substantial number of modifier–head compounds that stress-wise behave like argument–head compounds. According to the hypothesis, these forms should be explainable in terms of lexicalization. Higher-frequency items should have more leftward-stressed items among them than lower-frequency items. This is definitely not the case. We find instead that novel compounds, which can, by definition, not be lexicalized, show basically the same type of variability as existing modifier–head compounds. And even the variability of existing modifier–head compounds cannot be explained as a lexicalization effect, given their frequency distributions. Thus, neither the novel compounds nor the BNC and Google frequency data support the lexicalization hypothesis.

In essence, the structural hypothesis is correct in predicting leftward stress for argument–head compounds, but has nothing to say about all other compounds and thus underdetermines large parts of the data as they occur in the language. We will see in the following whether the other two hypotheses can shed some light on this data set.

4.2 The semantic hypothesis

The semantic hypothesis predicts for our stimuli that novel compounds expressing an authorship relation trigger rightward stress, while other modifier relations such as our title relation should trigger leftward stress. For existing compounds, the semantic hypothesis predicts leftward stress, unless pertinent semantic categories such as copulative, authorship, causative, etc., are involved. The latter prediction was not systematically tested since this would have required a large-scale study of its own. Figure 9 shows the results across all observations. As can be seen from this figure, there is hardly any difference observable between the two semantic categories. A one-way ANOVA confirms this impression; the difference is not significant ($F(1, 201) = .00082$, $p > 0.9$). This goes against the prediction of the semantic hypothesis, which made us expect that the author relationship triggers significantly smaller pitch differences than the title relationship.

Let us look, however, at the interaction of semantic relation with other variables, beginning with clausal position. Due to the design of the experiment (see section 3), it was possible to investigate the possible interaction of semantic relation and clausal position only in the largest group of items, those occurring in declarative clauses. To investigate the effect of subject and item in a linear mixed-effect model with item and subject crossed is difficult to apply because items vary with combinations of relation and position. Thus, combinations of position and relation are confounded with the particular items used, and further research is required to ascertain whether an effect of position and relation will be robust in a by-item design with repeated measures.

https://doi.org/10.1017/S1360674306001821 Published online by Cambridge University Press

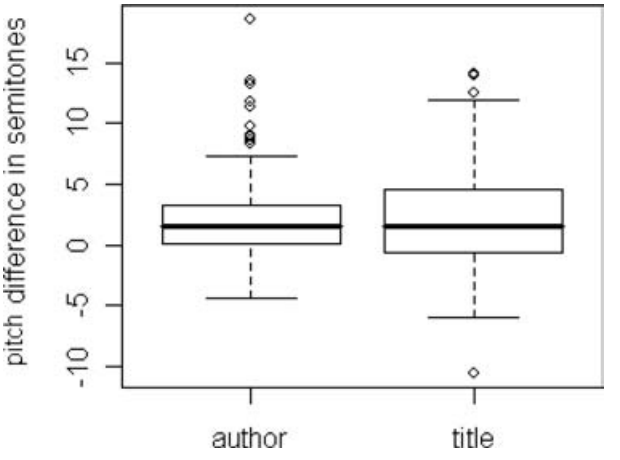


Figure 9. Pitch difference according to semantic relation, all clausal positions

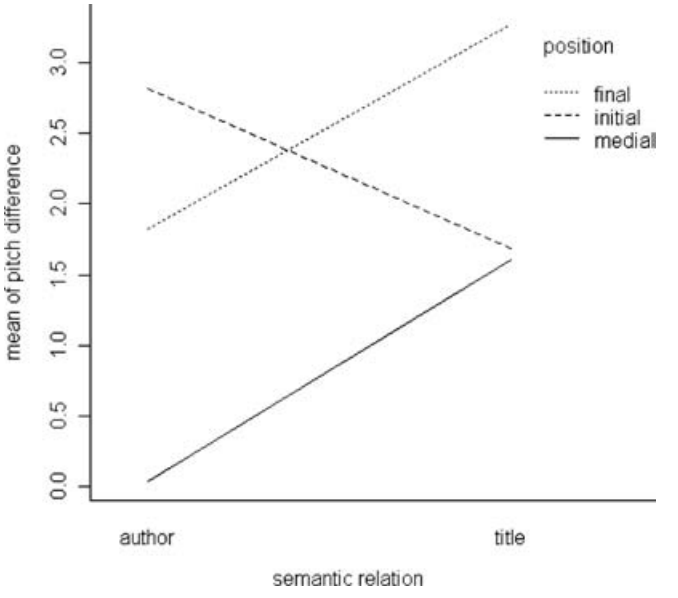


Figure 10. Interaction of clausal position and semantic relation in declarative clauses

A by-subject-only multilevel analysis was carried out, with position and relation as fixed effects. The inspection of the residuals revealed non-normality with a number of data points with undue leverage (according to Cooke's distance). After removing twelve outliers, we found a significant interaction of position and semantic relation ($F(2, 134) = 3.451$, $p < 0.05$), and no main effects (position: $F(2, 134) = 2.6948$, $p > 0.07$; semantic relation: $F(1, 134) = 0.044$, $p > 0.8$). Figure 10 gives the interaction plot for the three variables. We now see that the pitch difference in final position is generally higher than in medial position. What is more interesting for us, however, is

https://doi.org/10.1017/S1360674306001821 Published online by Cambridge University Press

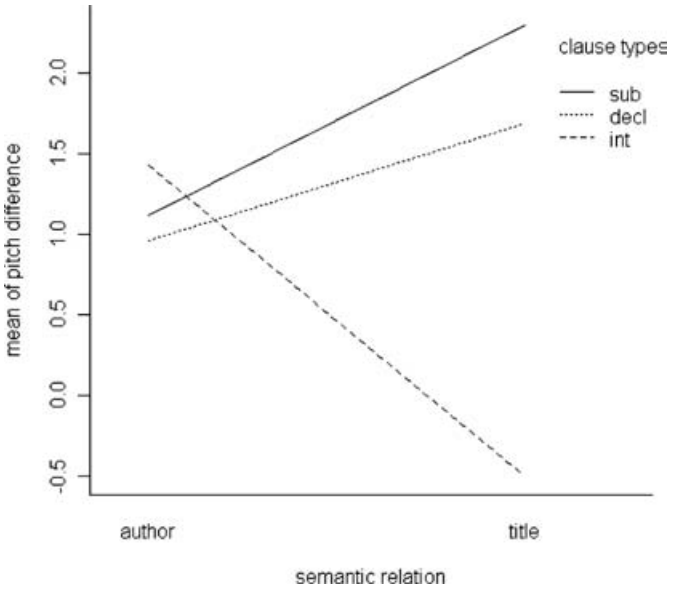


Figure 11. Interaction of clause type and semantic relation in final position ('decl' = declarative clauses, 'sub' = preposed sub- or superordinate clauses, 'int' = interrogative clauses)

that the pitch difference increases when going from author to title relation. Thus in this subset of the data we find that compounds with an authorship relation are more right-stressed than those expressing a title relation. This is in accordance with the semantic hypothesis. In initial position, however, this effect goes in the opposite direction. With compounds in initial position, the pitch difference decreases when going from author to title relation, which is the opposite of what the semantic hypothesis would predict.

Let us turn to the effect of clause type, which it was possible to test for all items in final position. To investigate the effect of subject and item, a linear mixed-effect model with item and subject crossed is again difficult to apply because items vary with combinations of semantic relation and clause type. Thus, combinations of clause type and semantic relation are confounded with the particular items used, and a different design would be required to ascertain whether an effect of clause type and semantic relation will be robust in a by-item analysis with repeated measures. A by-subject-only multilevel analysis, with clause type and semantic relation as fixed effects, yielded a significant effect of clause type and a highly significant interaction of clause type and semantic relation. However, after removing twelve outliers due to non-normal residuals, one is left with only the interaction effect ($F(2, 71) = 4.72$, $p < 0.02$). Figure 11 shows this interaction. We can see that only in final position of interrogative clauses do we find a significant difference between author and title compounds. This effect goes again in a direction that is unexpected under the semantic hypothesis. Including sex in the multilevel analysis shows no main effect of sex and no significant interaction of sex and semantic relation.

https://doi.org/10.1017/S1360674306001821 Published online by Cambridge University Press

To summarize, we have seen that the predicted effect of semantic relation only occurs in a subset of the data and sometimes interacts with other factors. This indicates that the semantic hypothesis may be on the right track, but that the influence of the pertinent semantic feature may not be strong enough to be observed in all clausal environments. It should be noted, however, that in some positions and clause types we find the opposite of what the semantic hypothesis leads us to expect. Thus in initial position of declarative clauses and in final position of interrogative clauses compounds expressing an authorship relation are more left-stressed. The semantic hypothesis is thus at best partially supported by our data and can explain only very little of the variance within the class of modifier–head compounds that the structural hypothesis leaves unaccounted for.

4.3 The analogical hypothesis

Starting out from the observation that in the case of *avenue*, *street*, *lane*, etc. all compounds with one of these forms as right constituents behave in the same way, we can venture the hypothesis that the right constituent may in general have an effect on stress. Under this assumption, we expect to find stress differences between constructs with different right elements, and to find no differences between constructs with the same right constituent. This potential effect may of course occur only in subclasses of constructs, *modulo* other, overriding, factors, such as the argument–modifier distinction.

In order to test the effect of the right constituent, I varied both the left and the right constituents of novel compounds. For example, I combined the right constituent *sonata* with three different names (*Kauffman*, *Rossi*, *Sydlosky*), for the author relation, and with three different nouns (*dream*, *twilight*, *winter*), for the title relation. A similar thing was done with compounds involving *symphony*, *anthem*, and *opera* as right constituents. In (4) the pertinent stimuli are given for illustration:

(4)

Left member	**Right member**	**Left member**	**Right member**
Easter	*anthem*	*Dream*	*sonata*
Emperor	*anthem*	*Twilight*	*sonata*
Farewell	*anthem*	*Winter*	*sonata*
Hoffman	*anthem*	*Kauffman*	*sonata*
Klebanoff	*anthem*	*Rossi*	*sonata*
Olderman	*anthem*	*Sydlosky*	*sonata*
Resurrection	*opera*	*Christmas*	*symphony*
Surprise	*opera*	*Moonlight*	*symphony*
Toy	*opera*	*Spring*	*symphony*
Groskinsky	*opera*	*Groskinsky*	*symphony*
Lieberman	*opera*	*Hoffman*	*symphony*
Rossi	*opera*	*Lieberman*	*symphony*

Under the assumption that the right constituent has an effect on the stress pattern, we would expect to find differences between the different compounds. For example, just as all *road* compounds are right-stressed, while all *street* compounds are left-stressed,

https://doi.org/10.1017/S1360674306001821 Published online by Cambridge University Press

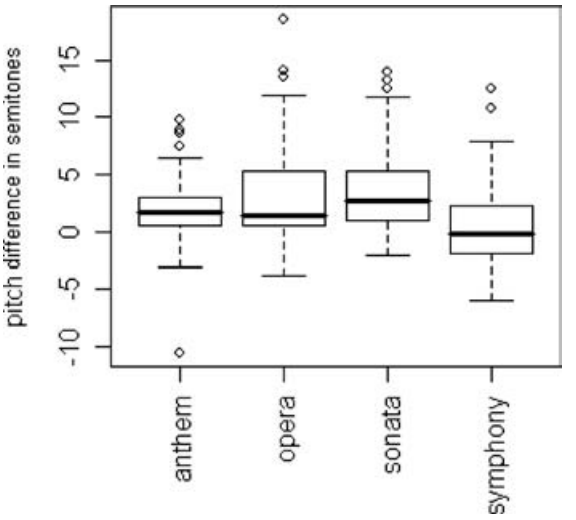


Figure 12. Pitch difference according to right member: novel compounds

we would expect to find that the four sets of compounds given in (4) show differences amongst each other. And this is indeed what can be observed.

A multilevel analysis of the pitch differences as dependent variable and right member and sex as fixed-effect predictors, and with subject and item as crossed random effects, revealed significant main effects of right member ($F(3, 196) = 3.48$, $p < 0.02$) and of sex ($F(1, 196) = 4.74$, $p < 0.05$) and no interaction between these two factors ($F < 1$). The standard deviation for the item random effect was 1.43, the standard deviation for subject was 1.17, and that of the residual error was 3.48. Thus, any variance that can be attributed in a principled way to subjects or items is accounted for in the model. When position is added to the model, it has no predictive value ($F(2, 196) = 1.32$, $p > 0.2$). Adding semantic relation to the model shows the same negative result ($F(1, 186) = 0.077$, $p > 0.7$). In other words, the effect of the right constituent is very robust and overrides a potential effect of semantic relation (author vs. title).

Let us inspect the main effect of right member in more detail by considering figure 12. Post-hoc pairwise Welch-modified t-tests testing all six possible contrasts revealed that the only significant difference is that between compounds headed by *symphony* and *sonata*, with *symphony* compounds being very highly significantly more right-stressed than *sonata* compounds ($t(99.603) = 4.0768$, $p < 0.001$ after Bonferroni adjustment). A simultaneous confidence intervals analysis using Tukey contrasts revealed again a contrast between *symphony* and *sonata* compounds (adjusted $p < 0.001$), and in addition a significant difference between *symphony* and *opera* compounds (adjusted $p < 0.05$).

To summarize, we can say that within the group of novel modifier–head compounds, we find a very robust effect of the right member on the stress pattern of a given compound. In particular, compounds with *symphony* as right member behave consistently

https://doi.org/10.1017/S1360674306001821 Published online by Cambridge University Press

differently from compounds with *sonata* or *opera* as right members, irrespective of the semantic relation expressed by the compound. This adds fuel to the idea that compound stress is at least partially driven by analogy, with right member as an important factor.

5 Discussion and conclusion

In general, the acoustic analysis of our data showed that there is indeed considerable variability in compound stress. This is a substantial finding by itself, which counterbalances the prevailing view in the literature that compound stress is a more or less categorical phenomenon.

The experimental study presented in this article tried to explain this variability with the help of three specific hypotheses. The results show a clear effect of the structural relation exhibited by a given compound. If the compound instantiates an argument–head relation, there is categorical leftward stress. This is in accordance with the general opinion expressed in the literature and in accordance with the structural hypothesis. However, we also found that modifier–head compounds show variably rightward and leftward stress, and that this variability, contra Giegerich (2004), cannot be explained as a lexicalization effect. We are thus left with a large number of compounds, i.e. the modifier–head compounds, whose variable stress patterns are unaccounted for under the structural hypothesis.

A closer analysis of this problematic set of compounds revealed that both semantic relation and analogy may play a role in determining the variability of stress assignment. This finding supports the two respective hypotheses and identifies these mechanisms as subsidiary to the basic structural distinction. While the analogical effect of the right member proved to be very robust, the effect of semantic relation did not, however, surface in all syntactic environments; moreover, in two environments even the opposite of the expected effect was observed. These facts certainly merit further investigation.

With regard to the more leftward stress of *sonata* compounds, which supports the idea of analogical effects based on the right member, our findings correspond with the pattern in Spencer (2003), who lists a number of *sonata* compounds, all of them left-stressed. Notably, this consistent leftward stress of compounds headed by *sonata* goes against the general trend of modifier–head compounds towards more rightward stress and overrides the potential semantic effect of author vs. title relation.

Spencer also claims that *symphony* compounds generally take leftward stress, but that 'there are semantic islands that behave differently. For instance, unique sets or types of work modified by a composer's name get right-stress: *a Mahler sýmphony*' (Spencer, 2003: 332). Our data show a somewhat different pattern. We consistently find rightward stress with *symphony* compounds, irrespective of semantic relation.

How do these discrepancies arise? I think two things may be pertinent here. The first is the problem of variety. Although Spencer says nothing about the variety he investigates, it is likely that this variety is Standard (Southern) British English. The data in this study, in contrast, are elicited from American speakers of English, which may be the reason for the said discrepancy between Spencer's and my own findings.

https://doi.org/10.1017/S1360674306001821 Published online by Cambridge University Press

This possibility indicates that any study of English compound stress needs to be clear about its reference variety. The second problem is methodological in nature and may concern the systematicity of data collection. Spencer does not give any information about how his data were sampled, which makes it hard to speculate about any additional reasons for the observed discrepancies between his data and mine.

To return to our central research question, we can summarize our findings by saying that all three types of factor play a role in stress assignment in English compounds, and that semantics and analogy interact in complex ways. This is of course not the end of the story. Many questions are still open, in particular questions concerning the nature of the semantic effects, the basis of analogical relationships, and the nature of the interaction of semantic and analogical effects.

For example, the semantic hypothesis must be further tested with regard to other potentially relevant semantic categories. I did not investigate temporal, locative, or other relationships that are held responsible for rightward stress, nor did I try out Spencer's (2003) idea that certain semantic fields trigger certain patterns (e.g. chess terminology or musical terminology). Furthermore, the strength of the semantic effect is an issue for further research, as well as the question of whether the seemingly semantic effect may be an epiphenomenon of underlying analogical mechanisms. Concerning the computation of analogical relationships, we need to test the potential influence of other properties apart from right-hand member and we need to model potential analogical effects computationally in order to substantiate the still tentative conclusions put forward in this article. In addition, the effects of clause type and clausal position merit further study.

Importantly, our article has also shown that acoustic analyses of experimental data can be very useful in determining the amount, the significance, and the potential sources of the variability we find in the stress assignment of NN constructs. This opens up rich perspectives for future studies to systematically test pertinent factors in order to overcome the shortcomings of the mostly impressionistic data sets that have prevailed in the literature until recently. In principle, such analyses could also be extended to speech corpus data, which may offer interesting insights also in this area of grammar.

Finally, on the theoretical plain, the present study adds to the growing body of evidence that morphological structure is intrinsically graded (cf. Hay & Baayen, 2005, for an overview). By allowing analogy and probability into the grammar, we can make progress toward solving some long-standing problems in compound research, and in morphological theory in general.

Author's address:
English Linguistics
Fachbereich 3
Universität Siegen
Adolf-Reichwein-Str. 2
D-57068 Siegen
Germany
plag@anglistik.uni-siegen.de

https://doi.org/10.1017/S1360674306001821 Published online by Cambridge University Press

References

Adams, V. (1973). *An introduction to Modern English word-formation*. London: Longman.
Bates, D. M. & D. Sarkar (2005). The lme4 library [Online]. Available at http://lib.stat.cmu.edu/R/CRAN/.
Bauer, L. (1983a). *English word-formation*. Cambridge: Cambridge University Press.
Bauer, L. (1983b). Stress in compounds: a rejoinder. *English Studies* **64**: 47–53.
Bauer, L. (1998). When is a sequence of two nouns a compound in English? *English Language and Linguistics* **2**: 65–86.
Bloomfield, L. (1933). *Language*. Chicago: Holt.
Boersma, P. & D. Weenink (2005). Praat: doing phonetics by computer (Version 4.3.01) [Computer program]. Retrieved from http://www.praat.org/.
Chomsky, N. & M. Halle (1968). *The sound pattern of English*. New York: Harper & Row.
Clahsen, H. (1999). Lexical entries and rules of language: a multidisciplinary study of German inflection. *Behavioral and Brain Sciences* **22**: 991–1060.
Daelemans, W., J. Zavrel, K. van der Sloot & A. van den Bosch (2000). TiMBL: Tilburg Memory Based Learner, version 3.0, Reference Guide. ILK Technical Report 00-01. Available from http://ilk.kub.nl/~ilk/papers/ilk0001.ps.gz.
Faiss, K. (1981). Compound, pseudo-compound and syntactic group especially in English. In Kunsmann, P. & O. Kuhn (eds.), *Weltsprache Englisch in Forschung und Lehre: Festschrift für Kurt Wächtler*. Berlin: Erich Schmidt Verlag. 132–50.
Farnetani, E. & P. Cosi (1988). English compound versus non-compound noun phrases in discourse: an acoustic and perceptual study. *Language and Speech* **31**: 157–80.
Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech* **1**: 126–52.
Fudge, E. C. (1984). *English word-stress*. London: George Allen & Unwin.
Gagné, Ch. (2001). Relation and lexical priming during the interpretation of noun–noun combinations. *Journal of Experimental Psychology: Learning, Memory and Cognition* **27**: 236–54.
Gagné, Ch. & E. Shoben (1997). Influence of thematic relations on the comprehension of modifier–noun combinations. *Journal of Experimental Psychology: Learning, Memory and Cognition* **23**: 71–87.
Giegerich, H. (2004). Compound or phrase? English noun-plus-noun constructions and the stress criterion. *English Language and Linguistics* **8**: 1–24.
Goldberg, A. & R. Jackendoff (2005). The end result(ative). *Language* **81**: 474–7.
Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge: Cambridge University Press.
Haan, J. & V. J. van Heuven (1999). Male vs. female pitch range in Dutch questions. In Ohala, J. (ed.), *Proceedings of the 14th International Congress of Phonetic Sciences*. Berkeley: University of California. 1581–4.
Hay, J. & H. Baayen (2005). Shifting paradigms: gradient structure in morphology. *Trends in Cognitive Sciences* **9**: 342–8.
Hayes, B. (1995). *Metrical stress theory: principles and case studies*. Chicago: Chicago University Press.
Henton, C. G. (1989). Fact and fiction in the description of female and male pitch. *Language and Communication* **9**: 299–311.
House, A. S. & G. Fairbanks (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *Journal of the Acoustical Society of America* **25**: 105–13.
Ingram, J., T. A. T. Nguyen & R. Pensalfini (2003). An acoustic analysis of compound and phrasal stress patterns in Australian English. Submitted for publication.

https://doi.org/10.1017/S1360674306001821 Published online by Cambridge University Press

Kingdon, R. (1958). *The groundwork of English stress*. London, etc.: Longmans, Green & Co.
Krott, A., H. Baayen & R. Schreuder (2001). Analogy in morphology: modeling the choice of linking morphemes in Dutch. *Linguistics* **39**: 51–93.
Krott, A., P. Hagoort & H. Baayen (2004). Sublexical units and supralexical combinatorics: the case of Dutch interfixes in visual processing. *Language and Cognitive Processes* **19**: 453–71.
Krott, A., R. Schreuder & H. Baayen (2002). Analogical hierarchy: exemplar-based modeling of linkers in Dutch noun–noun compounds. In Skousen, R., D. Lonsdale & D. S. Parkinson (eds.), *Analogical modeling: an exemplar-based approach to language*. Amsterdam: Benjamins. 181–206.
Ladd, D. R. (1984). English compound stress. In Gibbon, D. & H. Richter (eds.), *Intonation, accent and rhythm*. Berlin: Mouton de Gruyter. 253–66.
Ladefoged, P. (2005). *Phonetic data analysis: an introduction to fieldwork and instrumental techniques*. Oxford: Blackwell.
Lees, R. B. (1963). *The grammar of English nominalizations*. The Hague: Mouton.
Lehiste, I. (1970). *Suprasegmentals*. Cambridge, MA: MIT Press.
Liberman, M. & J. Pierrehumbert (1984). Intonational invariance under changes in pitch range and length. In Aronoff, M. & R. Oehrle (eds.), *Language sound structure*. Cambridge, MA: MIT Press. 157–233.
Liberman, M. & R. Sproat (1992). The stress and structure of modified noun phrases in English. In Sag, I. A. & A. Szabolcsi (eds.), *Lexical matters*. Stanford: Center for the Study of Language and Information. 131–81.
Lipka, L. (1994). Lexicalization and institutionalization. In Adger, R. E. (ed.), *The encyclopedia of language and linguistics*. Oxford: Pergamon Press. 2164–7.
Marchand, H. (1969). The categories and types of present-day English word-formation. 2nd edn. Munich: Beck.
Nolan, F. (2003). Intonational equivalence: an experimental evaluation of pitch scales. In Solé, I. M. J., D. Recasen & J. Romero (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona 3–9 August 2003*. Barcelona: Universitat Autònoma de Barcelona. 771–4.
Olsen, S. (2000). Compounding and stress in English: a closer look at the boundary between morphology and syntax. *Linguistische Berichte* **181**: 55–69.
Olsen, S. (2001). Copulative compounds: a closer look at the interface between syntax and morphology. In Booij, G. E. & J. van Marle (eds.), *Yearbook of morphology 2000*. Dordrecht, Boston, and London: Kluwer. 279–320.
Payne, J. & R. Huddleston (2002). Nouns and noun phrases. In Huddleston, R. & G. K. Pullum, *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press. 323–524.
Pierrehumbert, J. (1979). The perception of fundamental frequency declination. *Journal of the Acoustical Society of America* **66**: 362–70.
Plag, I. (2003). *Word-formation in English*. Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press.
Sampson, R. (1980). Stress in English N+N phrases: a further complicating factor. *English Studies* **61**: 264–70.
Schmerling, S. F. (1971). A stress mess. *Studies in the Linguistic Sciences* **1**: 52–65.
Skousen, R., D. Lonsdale & D. S. Parkinson (eds.) (2002). *Analogical modeling: an exemplar-based approach to language*. Amsterdam: Benjamins.
Spencer, A. (2003). Does English have productive compounding? In Booij, G. E., J. DeCesaris, A. Ralli & S. Scalise (eds.), *Topics in morphology. Selected papers from the*

https://doi.org/10.1017/S1360674306001821 Published online by Cambridge University Press

third Mediterranean morphology meeting (Barcelona, September 20–22, 2001). Barcelona: Institut Universitari de Lingüística Applicada, Universitat Pompeu Fabra. 329–41.
Sternberg, S. *et al.* (1980). Motor programs in rapid speech: additional evidence. In Cole, R. A. (ed.), *Perception and production of fluent speech*. Hillsdale, NJ: Lawrence Erlbaum Associates. 507–34.
Traunmüller, H. & A. Eriksson (1995). The frequency range of the voice fundamental in the speech of male and female adults. Ms.
Venables, W. N. & B. D. Ripley (2000). *Modern applied statistics with S-Plus*. New York: Springer.

Appendix

Existing compounds

Item	**Relation**	**Item**	**Relation**
music-lover	argument	bass voice	modifier
opera-singer	argument	soprano voice	modifier
concertmaster	argument	school choir	modifier
Bach choir	modifier	London orchestra	modifier
chamber ensemble	modifier	summertime	modifier
opera glasses	modifier	piano concerto	modifier
wind instruments	modifier	music lessons	modifier
violin lessons	modifier	music program	modifier
chamber-music	modifier	string quartet	modifier
Broadway musical	modifier	piano arrangement	modifier
symphony orchestra	modifier	concert hall	modifier
morning paper	modifier	concert pianist	modifier
Oxford Street	modifier		

Novel compounds

Item	**Relation**	**Item**	**Relation**
Klebanoff anthem	author	Groskinsky opera	author
Rossi opera	author	Kauffman sonata	author
Groskinsky symphony	author	Lieberman symphony	author
Hoffman anthem	author	Lieberman opera	author
Sydlosky sonata	author	Hoffman symphony	author
Olderman anthem	author	Moonlight symphony	title
Rossi sonata	author	Toy opera	title
Twilight sonata	title	Dream sonata	title
Christmas symphony	title	Spring symphony	title
Farewell anthem	title	Emperor anthem	title
Resurrection opera	title	Surprise opera	title
Winter sonata	title	Easter anthem	title

https://doi.org/10.1017/S1360674306001821 Published online by Cambridge University Press

List of sentences to be read out (sequence 1)

01. I forgot to bring my opera glasses.
02. I have always loved the Easter anthem since I first heard it.
03. I am a fan of chamber-music.
04. The Winter sonata was their best performance ever.
05. Sydlosky and Kauffman are my favorite composers.
06. I have always wanted to visit a Rossi opera.
07. As a child I had violin lessons.
08. I have always preferred wind instruments.
09. The London orchestra played a Groskinsky symphony.
10. Who are your favorite musicians?
11. She has a beautiful soprano voice.
12. The school choir was rehearsing an Olderman anthem when we came in.
13. We like to go to the little theater on Oxford Street.
14. The highschool orchestra has a new conductor.
15. Have you ever heard of the Surprise opera?
16. We attended a concert by the Bach choir.
17. I saw a piano concerto advertised in the morning paper.
18. He has a wonderful bass voice.
19. I always recognize a Rossi sonata when I hear one.
20. What did you think about the orchestration of the piece?
21. In the summertime I like to visit concerts in the park.
22. The Farewell anthem was sung at the end of the concert.
23. I prefer a string quartet to a full orchestra.
24. Did you have music lessons as a child?
25. We went to see a Lieberman opera last night.
26. They always have a musical evening on Sundays.
27. At the moment we are rehearsing the Twilight sonata.
28. I like to watch a music program on TV from time to time.
29. A Lieberman symphony will be played at the end of the evening.
30. She is a music-lover.
31. As a child he wanted to be an opera-singer.
32. A Hoffman anthem was sung right before the president's speech.
33. Have you ever dreamt of being the first violin?
34. This concertmaster is brilliant.
35. A Kauffman sonata was played as encore.
36. What did you think about this new piano arrangement?
37. Do you like the works of Klebanoff?
38. I am looking for a CD of a Klebanoff anthem.
39. You might even become a concert pianist when you grow up.
40. The Moonlight symphony is one of my favorite pieces of music.
41. The last time I went to see an opera was in December.
42. My brother used to play the English horn as a boy.
43. The Resurrection opera will have its premiere tomorrow.
44. I played in the highschool orchestra and I still attend their concerts.
45. I love classical music better than any other kind of music.
46. The orchestra played the Dream sonata after the break.
47. When we arrived at the concert hall the orchestra was still in rehearsal.
48. I love operas but I rarely ever go to see operettas.
49. We own a CD of the Christmas symphony.

https://doi.org/10.1017/S1360674306001821 Published online by Cambridge University Press

50. I have always dreamt of playing in a symphony orchestra.
51. Sometimes we go out to see a Broadway musical.
52. A Groskinsky opera will be played in Vienna next spring.
53. The Italian musician Rossi composed some very famous pieces of music.
54. Do you know the Emperor anthem?
55. The London Symphony Orchestra is in town right now.
56. This morning I listened to a Hoffman symphony on the radio.
57. I wish there were some decent concerts around here.
58. Do you remember that hymn in A minor?
59. I won two tickets for the Toy opera last week.
60. I really like the music of that chamber ensemble.
61. Have you ever listened to a Sydlosky sonata?
62. Olderman was a British nineteenth-century composer.
63. Would you like to go to a classical concert with me?
64. Everybody liked the Spring symphony last night.

https://doi.org/10.1017/S1360674306001821 Published online by Cambridge University Press