THE ROLE OF SEMANTICS, ARGUMENT STRUCTURE, AND LEXICALIZATION IN COMPOUND STRESS ASSIGNMENT IN ENGLISH

INGO PLAG
University of Siegen

GERO KUNTER
University of Siegen

SABINE LAPPE
University of Siegen

MARIA BRAUN
University of Siegen

It is generally assumed that noun-noun compounds in English are stressed on the left-hand member (e.g. *cóurtroom*, *wátchmaker*). However, there is a large 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 a study in which three kinds of factors held to play a role in compound stress assignment are tested: argument structure, lexicalization, and semantics. The analysis of 4,353 noun-noun compounds extracted from the Boston University Radio Speech Corpus shows that there is indeed a considerable amount of variation in stress assignment. Overall, semantics turns out to have the strongest effect on compound stress assignment, whereas an approach relying on argument structure is much less successful in predicting compound stress. The article presents for the first time large-scale empirical evidence for the assumption that lexicalization has an effect on compound stress assignment. The article also makes a methodological contribution to the debate in showing that (and how) corpus-based studies using acoustic 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 or exemplar-based models 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 gradience, fuzziness, semiregularity, and irregularity of many phenomena on all levels of linguistic description (see e.g. Bod et al. 2003, Hay & Baayen 2005). The present article deals with one area where this semiregularity is pervasive: stress assignment in English noun-noun compounds.

In general, it has often been claimed that English 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

* The first three authors wish to dedicate this article to their academic teacher, mentor, and friend Rüdiger Zimmermann, Professor and Chair of English Linguistics at the University of Marburg, on the occasion of his retirement in September 2008. Rüdiger has been a great source of inspiration for all three of us, and our research team owes much to his generosity and support.

We also would like to thank two sets of *Language* referees and the editor Brian Joseph for their very helpful and detailed remarks on earlier versions of this article. We are also grateful for the comments we received at the following talks, colloquia, and conferences where we presented material from this article: Directions in English Language Studies, University of Manchester, April 2006; Sprachwissenschaftliches Kolloquium at Universität Siegen, May 2006; Quantitative Investigations in Theoretical Linguistics 2 at Universität Osnabrück, June 2006; Lecture series 'Products of the Symbolic Mind' at Humboldt-Universität Berlin, July 2006; Workshop Marginale Morphologie at Universität Stuttgart, July 2006; University of Toronto, October 2006; 5th Mental Lexicon Conference at McGill University, October 2006; University of Essex, February 2007; Mediterranean Morphology Meeting in Ithaca, Greece, September 2007; and Paderborn University, November 2007. We are also indebted to Heinz Giegerich for his support and critical comments, and to Harald Baayen for his methodological suggestions, and for sharing his expertise in statistics with us. The research presented here was made possible by two grants from the Deutsche Forschungsgemeinschaft (grants PL151/5-1 and PL 151/5-3), which we gratefully acknowledge.

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 bóard*) differ significantly not only 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 a large proportion of nominal compounds, it has been pointed out, for example, by Jespersen (1909 [1961]:153ff.), Kingdon (1958), Schmerling (1971), Fudge (1984), Liberman and Sproat (1992), Sproat (1994), Bauer (1998), Olsen (2000, 2001), and Giegerich (2004), that there are also numerous exceptions to the proposed rule.[1] In other words, there are structures that are stressed on the right-hand side in spite of the fact that these structures should be regarded as compounds by most analysts. Some of these forms are listed in 1.[2] 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	spring bréak	silk tíe

In view of this situation, the obvious question is how we can account for this variability in the stress assignment of noun-noun constructs. Systematic empirical or experimental work on the problem is scarce, but many studies on compounding contain pertinent remarks and data. Basically, one finds three kinds of hypotheses, which are spelled out in the literature to different degrees of explicitness. These hypotheses, which are discussed in more detail in §2, refer to either structural, semantic, or analogical factors that are held responsible for the stress of noun-noun constructs.[3] Structural and semantic factors rest primarily within a given compound, while analogical factors refer to the relationship a given compound has to other compounds.

The aim of the present article is to test the adequacy of the two hypotheses concerned with the compound-internal factors. We do so in a systematic fashion using a large amount of data that were gathered independently from the present investigation. Why

[1] The extent of this variation is unclear, and seems to depend on the kind of data one looks at. For example, Plag and colleagues (2007) find 90 percent left stress in CELEX, while we found 83 percent left stresses among the 2,599 noun-noun compounds in Teschner & Whitley 2004 (our analysis). Both figures are based on dictionary data. Sproat (1994) counts 70 percent left stresses in his newspaper corpus.

[2] Examples in §§1 and 2 either are taken from various sources or were collected by the authors in an anecdotal fashion. The stresses given may sometimes differ from the intuitions of the individual reader. See also n. 3 concerning potential sources of variability in use or intuitions.

[3] It has to be pointed out that all current approaches try to explain across-type variability. That means that all existing hypotheses about compound stress are based on the assumption that stress assignment to a given compound (i.e. a given type) is triggered by certain properties of this compound and therefore will always be the same for all (noncontrastive) realizations (i.e. tokens) of that compound. This assumption is to some extent questionable, as pointed out by Bauer (1983a:103), and as shown in more detail more recently in Kunter 2007. Since the present article tests existing hypotheses, we are primarily concerned with across-type variation. Within-type variation is also taken into account in so far as our investigation of across-type variability is based on tokens from a speech corpus, with many types being represented by more than one token. See also §7 for more discussion of this problem.

Yet another dimension of variation in compound stress assignment has been observed repeatedly in the literature (e.g. Giegerich 2004, Plag 2006), namely across-variety variation (e.g. British English vs. Scottish English vs. American English). In this article we investigate American English data only.

is this necessary? First, with few exceptions, the provenance of the data in earlier studies often remains obscure, and the selection of data does not seem to be in any way systematic but more designed to prove the point of the respective author. The second problem is that the amount of data is usually quite small, ranging from only a handful of pertinent examples to a few hundred forms. The third problem is that most of the studies do not discuss the details of their methodological decisions, such as the assignment of particular examples to a given analytical category.

In sum, there is still a need for large-scale empirical investigations of compound stress variability using independently gathered sets of data. The present article provides such a study, and it is the first study to use a large body of actual speech data. We present the results of the investigation of all noun-noun compounds extracted from the Boston University Radio Speech Corpus (Boston Corpus for short; Ostendorf et al. 1996). It is shown that some of the traditional claims about noun-noun stress are indeed supported by our data, while other claims are not. For example, in a separate analysis of argument structure (excluding semantic factors) we find that the distinction between argument-head compounds and modifier-head compounds plays a role in stress assignment, but only with synthetic compounds ending in the agentive suffix *-er*. We also find that most of the semantic categories and relations assumed in the literature to trigger rightward stress show the expected effects, while some of the categories and relations do not. Furthermore, we discovered effects for categories and relations that have not been reported before in the literature. In a regression analysis including all potential factors, only the semantic and lexicalization effects are robust, while argument structure disappears as a significant predictor of stress assignment. Another important finding is that none of the effects we find is of a categorical nature. In other words, the distribution of stress cannot be captured adequately in a deterministic rule-based model of grammar. Our article also makes a methodological contribution to the debate in showing that speech corpus-based studies using acoustic measurements can shed new light on the issue of variable compound stress.

Before we turn to the discussion of the hypotheses to be tested, a word is in order with regard to the notorious problem of whether noun-noun constructions should be analyzed as compounds or phrases. In general we remain agnostic in this article with regard to this issue, because, first, the a priori exclusion of certain types of data might have biased our results in an undesired fashion. Thus, in the scarce literature on the variability of compound stress, the notion of noun-noun compound is usually taken for granted, so that in a study that wants to test any claims in this domain a restrictive definition of noun-noun compound is inappropriate. Second, it has often been pointed out (e.g. more recently in Bauer 1998 or 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). Other criteria, such as separability, orthography, or semantic transparency, do not yield consistent results either (cf. Bauer 1998). Hence we sometimes, and conservatively, speak of 'noun-noun constructs' in this article, although the structures under investigation would probably be regarded as proper compounds by most analysts. We also use the term 'compound' for convenience's sake, but without theoretical commitment. The discussion of the structural hypothesis below elaborates on this point.

In what follows, we first review the hypotheses put forward in the literature and then describe the Boston Corpus and our data-coding procedure, with attention to the methodological problems involved. We then present the results for the structural hypothesis and for the semantic hypothesis, and follow with a probabilistic analysis in which

all factors are entered into a regression analysis. Finally, we provide a discussion of our findings.

2. HYPOTHESES ON STRESS ASSIGNMENT TO COMPOUNDS.

2.1. OVERVIEW. Four types of approach have been taken to account for the puzzling facts of variable noun-noun stress. Three of these are investigated in this article, namely the effects of argument structure, lexicalization, and semantics. Another approach, not dealt with here, draws on the idea of analogy and hypothesizes that stress assignment is generally based on analogy to existing noun-noun constructions in the mental lexicon. The analogical hypothesis has recently been tested formally in Plag et al. 2007 and Lappe & Plag 2007, and the reader is referred to these papers for details. In the next two subsections we introduce the hypotheses in focus in this study.

2.2. THE STRUCTURAL HYPOTHESIS: ARGUMENT STRUCTURE AND LEXICALIZATION. The first one is what Plag (2006) has called the 'structural hypothesis'. Proponents of this hypothesis (e.g. Bloomfield 1933, Lees 1963, Marchand 1969, 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. Such an approach, however, would face the problem of explaining why not all forms that have the same superficial structure, that is, noun-noun, 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 construct 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. This point is taken up by Giegerich (2004) and is 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 happen only 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 *kick the bucket* 'die').

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 *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.[4]

[4] Giegerich characterizes modifier-head structures in terms of their lack of argument-predicate semantics. We prefer the term 'argument-head' to 'argument-predicate' in the context of this article because of its parallelism with 'modifier-head'.

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.

The idea that there is a relationship between lexicalization and stress assignment is not new and has also been advocated by other authors. For example, Bauer (1983b: 51) mentions irregular stress assignment in English derivatives and Danish compounds as prototypical cases of (phonological) lexicalization. And Adams (1973:59) 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' (emphasis added). To the best of our knowledge, however, the relationship between lexicalization and compound stress assignment has never been tested empirically, probably due to a number of methodological problems. The first is that lexicalization is not a categorical notion, but rather a gradual one, and second, that it is not exactly clear how it can be decided whether a given item is lexicalized or not, or, under a gradient view, more lexicalized or less lexicalized than another item. For compounds, four criteria come to mind: frequency, spelling, semantic transparency, and phonological transparency.[5]

In this study we test the potential effects of lexicalization using frequency and spelling as indicators of lexicalization. Higher frequency indicates a higher degree of lexicalization,[6] and one-word spellings should be most prevalent with lexicalized compounds, while less lexicalized compounds should prefer two-word spellings.[7] Given frequency and spelling as correlates of lexicalization, we can make interesting and falsifiable predictions about stress assignment according to Giegerich's hypothesis. First, with regard to corpora data, we should expect that the amount of leftward-stressed modifier-head compounds should vary according to frequency. Modifier-head compounds with a higher token frequency should be more prone to leftward stress than modifier-head compounds at the lower end of the frequency range. In addition, we would expect a higher proportion of left-stressed modifier-head compounds among those spelled as one word than among those spelled as two words.

Second, the structural hypothesis predicts that we should never find rightward stress among those noun-noun constructs that exhibit complement-head order. The latter point

[5] Although it is very common in generative approaches to assume that a given item is either part of the lexicon (i.e. 'lexicalized') or not, such a view is not supported by psycholinguistic findings. Current psycholinguistic models of the mental lexicon assume the existence of lexical representations of different degrees of strength in memory (e.g. Butterworth 1983). In such models the question arises of how often one needs to be exposed to establish a lexical representation. Recent research indicates that even a single exposure is enough to leave detectable traces in memory, but that more frequent exposure is necessary to firmly establish a lexical representation in all its facets in long-term memory (e.g. Bloom 2000, de Vaan et al. 2007). Lexicalization must therefore be conceived of as a gradual phenomenon.

[6] Compare 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, emphasis added).

[7] Note that a connection between spelling and lexicalization does not mean that stress would be dependent on orthography (to the effect that only literate speakers would know how to stress correctly). If anything, it is the other way around. Given the options of English orthography, speakers would express their intuition that a given compound is felt to be more or less integrated by choosing a more or less integrated spelling. This has been shown to be the case in Plag et al. 2007. Obviously, for unwritten languages one would have to look for other sorts of external cues relevant to determining the degree of lexicalization.

is, however, not always true, as pointed out by Giegerich himself (2004:19), who cites *Tory léader* as a counterexample.[8]

Third, the structural hypothesis also entails 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*) may suggest that this prediction is probably wrong, but thorough empirical testing is needed for verification.

Before turning to the discussion of what we call the semantic hypothesis we would like to emphasize that what has been labeled the 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 our reason for calling the hypothesis 'structural', although the underlying distinction might be semantic.

2.3. 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, Sproat 1994, 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 categories and relationships. Pertinent examples are copulative compounds like *geologist-astrónomer* and *scholar-áctivist* (see Plag 2003:146), which are uncontroversially considered to be regularly right-stressed.[9] Other meaning relationships that are often, if not typically, accompanied by rightward stress are temporal or locative (e.g. *a summer níght*, *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 semantic 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 noun-noun constructs in which the first noun 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 of 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 (*Mádison Street* vs. *Madison Ávenue*). And again, we have to state that, apart from the copulative compounds (see Olsen 2001) and compounds expressing an authorship relation (see Plag 2006), detailed and systematic empirical studies are lacking for the classes postulated to trigger rightward stress.

Note that we use the label 'semantic hypothesis' in this article to refer to approaches that set up semantic categories or semantic relations 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, the structural hypothesis and the semantic hypothesis converge on the point that they expect rightward stress to be largely restricted to modifier-head com-

[8] Note that such aberrant behavior may result from different interpretations. Thus, *Tory leader* may also be interpreted as a copulative compound denoting someone who is at the same time a leader and a Tory.

[9] As pointed out by one referee, even this generalization has its (apparently very few) exceptions, for example, *mán-servant*, which is left-stressed. The pertinence of this example was contested by another referee, who claimed that *man-servant* is not a copulative compound.

pounds. Note also that some of the semantic categories proposed in the literature, for example 'N1 is a proper noun', could also be labeled 'structural'. Our use of 'structural' in 'structural hypothesis' and 'semantic' in 'semantic hypothesis' is to be taken as a convenient label for these approaches, and not as a specific claim about the theoretical status of each of the particular phenomena in question. Thus, subsuming, for instance, 'N1 is a proper noun' under the structural hypothesis would not make sense because the proponents of the approach we label 'structural hypothesis' have never referred to that category, although this category might be considered 'structural' (and 'semantic' at the same time).

2.4. TESTING THE HYPOTHESES: PREVIOUS RESEARCH. Few systematic empirical studies are available that investigate variable compound stress experimentally or with independently gathered data. Sproat (1994) discusses a variety of methods for stress assignment in English compounds for the purpose of text-to-speech synthesis. Sproat uses stress intuitions by native speakers, and as we do in this study he uses thousands of compounds from a text corpus for testing. However, Sproat does not use speech data, nor does he explicitly test the hypotheses that are in the focus of the present article. Some of his results are nevertheless pertinent. Sproat's algorithm makes use of, among other things, semantic rules 'derived from observations that at least some accentual patterns . . . have a semantic basis' (1994:82). The algorithm also employs semantic relationships coded as a cross-product of the pertinent semantic categories of the compound constituents as found in *Roget's Thesaurus* (Chapman 1977). The resulting category combinations are utterly strange from a theoretical point of view due to their lack of clear criteria of analysis and selection.[10] An additional problem is that, apart from one combination (N1: MATERIAL, N2: CLOTHING), the category combinations are nowhere near the categories mentioned in the theoretical literature as factors influencing compound stress. On the empirical side, the semantic information did not contribute much to successful compound stress classification in Sproat's study, neither in the form of semantic rules, nor in the form of cross-products of semantic categories instantiated in the two constituents.

Two recent papers have shed some new light on the issues at hand. Plag 2006 is an experimental study of compound stress, and Plag et al. 2007 is a corpus-based study using the CELEX lexical database (Baayen et al. 1995). In both studies an argument-structure effect was found. In the CELEX database, however, the vast majority of modifier-head compounds do not behave as expected and the argument-structure effect is restricted to compounds ending in the suffix *-er*. In neither study was there a measurable lexicalization effect when frequency was used as a correlate of lexicalization. Using spelling, Plag and colleagues (2007) found a lexicalization effect that, however, was not restricted to modifier-head compounds. Plag 2006 also tested whether the semantic hypothesis makes the right predictions for compounds with a causative relation (as in a *Kauffmann sonata*) against a relation that is not predicted by the literature to trigger right-hand stress (as in the *Twilight Sonata*). It turned out that the data show either no effect, or show an effect in the opposite direction of what the semantic hypothesis would have predicted. Plag et al. 2007 tested many more semantic relations and found many effects, but only some of the effects predicted by the literature. In general, large parts of the CELEX data were ill-behaved.

[10] Some examples of such category combinations are ANIMALS-INSECTS – ASSEMBLAGE, ASSOCIATION – PART, and DEPUTY-AGENT – INFORMATION.

Although these previous studies yielded interesting results, there are some general problems involved. Sproat's study did not systematically and explicitly investigate the structural and semantic factors as put forward in the literature, nor was it based on natural speech. The experimental study in Plag 2006 tested only a very limited amount of data and only one aspect of the semantic hypothesis (the causative relation). The CELEX study in Plag et al. 2007, in contrast, covered many more data and predictor categories, but was based on a corpus that has a relatively strong bias toward lexicalized compounds. Furthermore, it is unclear where the stress information in CELEX comes from. It is mostly taken from dictionaries, but this in turn raises the question of how the dictionaries arrived at their stress judgments. Given the general variability of compound stress and compound stress judgments (see below for detailed discussion), the reliability of such information is somewhat questionable.

A study is therefore called for that includes fewer lexicalized compounds and that is based on natural speech instead of stress judgments by dictionary makers. The present investigation is such a study.

3. METHODOLOGY.

3.1. GENERAL REMARKS. Our corpus, the Boston University Radio Speech Corpus, was collected primarily to support research in text-to-speech synthesis, particularly the generation of prosodic patterns. The corpus consists of professionally read radio news data and includes speech from seven (four male, three female) FM radio news announcers associated with WBUR, a public radio station. The main radio news portion of the corpus consists of over seven hours of news stories recorded in the WBUR radio studio during broadcasts over a two-year period. In addition, the announcers were also recorded in a laboratory at Boston University. For the latter recordings (the so-called 'lab news'), the announcers read a total of twenty-four stories from the radio news portion. The announcers were first asked to read the stories in their non-radio style and then, thirty minutes later, to read the same stories in their radio style. Each story read by an announcer was digitized in paragraph size units, which typically include several sentences. The files were digitized at a 16k Hz sample rate using a 16-bit A/D. The orthographic transcripts were generated by hand.

We chose this corpus for the following reasons. First, due to the topics covered in the news texts, we expected a sufficiently large number of compounds to be present in the corpus. Second, for the acoustic analysis we needed high-quality recordings. Third, given that the speakers were trained news announcers we expected them to produce relatively standard, error-free speech. All three expectations were met by the corpus. In all texts we manually annotated all sequences consisting of two (and only two) adjacent nouns, one of which, or which together, functioned as the head of a noun phrase. From this set we eliminated proper names such as *Barney Frank* and those with an appositive modifier, such as *Governor Dukakis*.

We finally arrived at an overall number of 4,353 tokens of noun-noun constructs, representing 2,450 word types.[11] This set of words was subjected to an acoustic and statistical analysis using the speech analysis software Praat (Boersma & Weenink 2005) and the statistical package R.

3.2. ACOUSTIC ANALYSIS AND DETERMINATION OF PROMINENCE. One major problem when dealing with compound stress in natural speech is to determine whether a given

[11] For illustration of our data, a random sample of five hundred forms from our corpus is given in Appendix B.

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, individual intuition should be substituted by more systematic evidence. We therefore devised a methodology that combines experimentally derived perception ratings with measurements of the acoustic correlates of compound stress.

In the literature on English, pitch, intensity, and duration are standardly taken to be acoustic correlates of stress (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 (see Lehiste 1970:120, Ladefoged 2005:92). This seems also true for compound stress, where we deal with the relative prominence of two stressed syllables, and not with the contrast between a stressed and an unstressed syllable. This difference in prominence is realized by differences in the distribution of pitch accents: in words with compound stress, only the first constituent is accented, while in words with phrasal stress, there is a pitch accent on each constituent (Gussenhoven 2004).

With regard to the acoustic correlates of compound stress, in their investigation of minimal pairs of compounds and phrases in English (e.g. *blackberry*, *black berry*), Farnetani and Cosi (1988) as well as Ingram and colleagues (2003) found that pitch and duration differences between the two constituents are the most reliable indicators of compound stress. Plag (2006) also uses pitch measurements as correlates of compound stress, measuring the F0 difference (transformed into semitones) between the main-stressed syllable of the left constituent and the main-stressed syllable of the right constituent.[12] In this article, we build on the methodologies of these previous studies but take a more sophisticated approach.

We started our investigation by taking a pseudo-random sample of 105 compounds (fifteen from each speaker) from the Boston Corpus and had thirty-one native speakers of American English listen to and rate these words according to which of the two constituents of the compound they regarded as more prominent.[13] Subjects had to make their prominence rating on a continuous scale (by moving a slider on a computer screen), with the ratings being later transformed into values from 0 through 999 (with higher numbers indicating more rightward stress).[14] If subjects felt, for example, that both constituents were equally prominent, they would choose to place the slider at a point somewhere in the middle (corresponding to a value of around 500). If they felt that the left constituent was very prominent they moved the slider far to the left, and far to the right for those compounds that they felt to be clearly right-prominent.

[12] See Plag 2006 for a detailed discussion of the methodological problems involved.

[13] See Kunter 2009 for the detailed description and discussion of that experiment and its results. For the purposes of this article we only summarize the most important points that are necessary for an understanding of our corpus analysis.

[14] Previous perception experiments on compound stress (e.g. Lutstorf 1960) have exclusively relied on forced choice between two or three categories of prominence. In doing so, the issue of how many stress levels exist (left, right, and level?) is prejudged. Furthermore, subjects often feel uncomfortable with forced choice because they simply have problems in deciding and classifying accordingly what they perceive. In general, listeners' judgments about stress are extremely variable (e.g. Fry 1958, Bauer 1983b, Gussenhoven 2004:3), so that a methodology is called for that allows a more fine-grained statistical analysis of the variability of the judgments in experiments on stress. Using a gradient rating scale is such a method.

The resulting (more than 3,000) ratings were then subjected to a linear regression analysis, in which the mean rating for each compound was modeled as a function of five acoustic parameters. The values for these parameters were derived from the syllable with primary stress in both the left and the right constituent. As prominence between constituents is held to be expressed by the relation of pitch, loudness, and duration between the two constituents, these parameters were included in the regression equation as difference calculations. The pitch difference, for example, was derived by measuring the mean F0 (as an indicator of pitch) of the left and right constituents and transforming the difference between the two into semitones. This difference δ_{pitch} is positive if the left value is larger than the right value. All other things being equal,[15] clearly left-stressed compounds should have a positive pitch difference. A corresponding relation is assumed for differences in loudness (measured as mean intensity) and duration.

Another acoustic correlate of stress that has been identified in recent years is the degree of spectral tilt. It has been shown (e.g. Sluijter et al. 1997) that prominent syllables show a more balanced distribution of energy across the spectrum, while a decrease of energy in the higher range of the spectrum indicates less prominence. In our data, only in the left constituent did spectral tilt turn out to be a significant predictor (T_{left}) of prominence ratings, but not in the right constituent. We also included the mean local pitch variation for both constituents, derived as the absolute pitch change between every two consecutive frames, averaged over the number of frames. Thus, an interval in which the fundamental frequency is fairly stable shows little differences between the points of the pitch curve, and the mean local pitch variation will be low. In an interval with a falling or rising pitch, the difference between points, and hence the pitch variation, is large.[16] The regression analysis showed that left-stressed compounds have a fairly steady pitch in the right constituent (S_{right}), while the pitch variability in the left constituent was insignificant for the prediction of prominence.

The final regression model, which can account for 70 percent of the variation (R^2), is given in 2.[17] It shows that the left constituent is perceived as more prominent if it has a higher pitch, a longer duration, and a higher intensity than the right constituent. In addition, little spectral tilt in the left constituent and a low pitch variability in the right constituent also contribute to the perception of left prominence in compounds.

$$(2)\quad y = 401.10 - 17.73 \times \delta_{pitch} - 679.39 \times \delta_{dur} - 12.54 \times \delta_{int} - 4.38 \times T_{left} - 0.73 \times S_{right}$$

The good fit of the regression model is illustrated in the following figure, which, in the form of a density plot, compares the predictions of the model with the experimental ratings.

[15] Here we abstract away from differences in vowel quality and length, which may also affect the measurements. A further abstraction is that we assume a direct association between prominence and a high pitch. Of course, low pitch accents are also possible in English, and are indeed highly frequent in questions (see Hedberg et al. 2004). In the radio news context of the present corpus, however, questions are only rarely found, and low accents in compounds play a negligible role. This is illustrated by an investigation of the prosodic labeling that is provided for 1,223 compounds from the Boston Corpus. A L* or L* + H accent is present in only forty (3.3 percent) left elements and thirty-seven (3.0 percent) right elements. Thus, the association of high prominence with high pitch holds for most of our data.

[16] In Praat, the mean local pitch variation is referred to as 'average absolute pitch slope'.

[17] δ_{pitch} was measured in semitones, δ_{dur} in seconds, δ_{int} in dB, T_{left} in dB, and S_{right} in semitones/second.

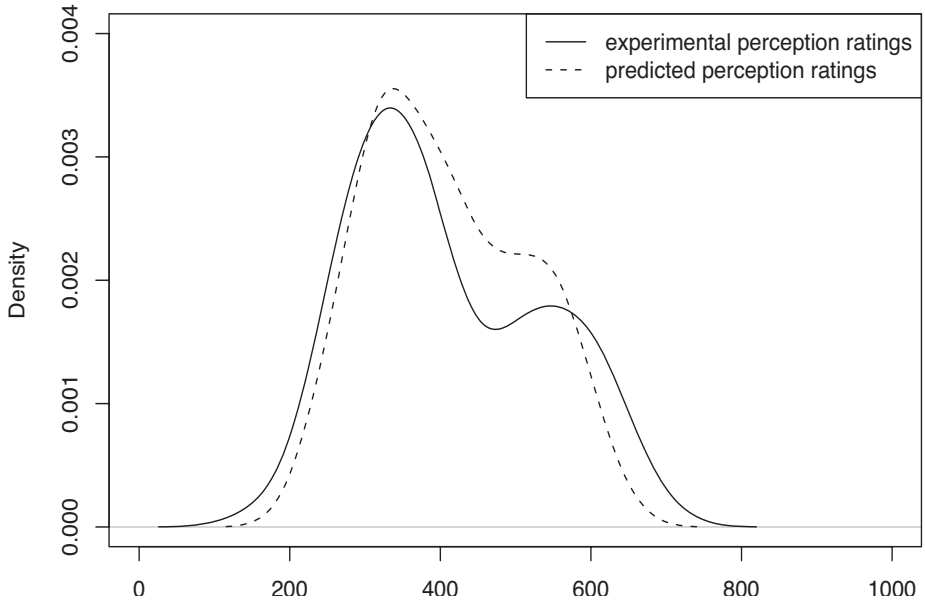

FIGURE 1. Experimental vs. predicted perception ratings.

Using the regression model in 2 with its five acoustic parameters, we can now estimate the degree of left and right prominence of every compound in the Boston Corpus. In what follows we refer to these estimated perception scores as PERCEPTION SCORES. The estimated perception score per se does not, however, solve the major problem, that is, the detection of rightward and leftward stresses in the data, since we are dealing with a gradient measurement and not with a binary one. Now, given a certain perception score, what is to be regarded as leftward and rightward stress?

This problem is analogous to that of taking simple pitch measurements as indicators of stress, as done in Ingram et al. 2003 or Plag 2006. Both studies have shown that in using such gradient measurements there is a considerable overlap in the measurements between different categories of stress. Along the lines of Ingram et al. 2003 and Plag 2006 we can assume, however, that statistically significant differences between two kinds of compounds in pitch (as in their studies) or perception scores (as in the present study) indicate different stress categories. Let us briefly illustrate this with an example from the Boston Corpus. In general it can be assumed that left-headed compounds such as *attorney general* would receive rightward stress.[18] The prediction would now be that such compounds differ significantly in their perception score from the right-headed compounds in the corpus, since the majority of the latter should be left-stressed, if we believe the literature (especially the compound stress rule). This is indeed the case: a Welch-modified *t*-test shows a very highly significant difference ($t(15.265) = -4.3301, p < 0.001$) in the predicted direction: the perception score of the left-headed compounds is higher than that of the right-headed ones (means: 515 vs. 428). A similar

[18] The analysis of such constructs is somewhat controversial, since they might also be analyzed as noun-adjective combinations with an unusual word order. There is no controversy, however, about whether these constructs are right-stressed. We have included these forms (coded as left-headed compounds) for the sake of broad empirical coverage.

result is obtained when only pitch difference is taken into account ($t(15.426) = 4.9711$, $p < 0.001$, means 0.1 vs. 3.4 semitones), which supports the assumption that pitch alone is already a good indicator of compound stress. What this example of left-headed and right-headed compounds shows is that two groups that, according to the literature, should differ in the preferred stress pattern have significantly different average prominence ratings.

This raises the question, however, of how the gradient measurement is to be interpreted phonologically. One might think that the gradient measurement could be (somewhat absurdly) interpreted in such a way that the degree of stress, that is, how strongly the prominence in a given token of a compound is phonetically encoded and perceived, is taken as a function of the category in question. Such an interpretation is to a large extent erroneous, however. Since we are interested in differences between categories, we compare the means of perception scores over sets of compounds, which means that the higher average perception score for one set of compounds (say, left-headed ones in the above example) generally reflects the higher frequency of right-stressed forms in this group, and not (primarily) a stronger articulation of rightward stress. It may of course be the case that, in addition, compounds of a particular category may indeed have a more strongly articulated prominence of one or the other kind, but this is not a necessary assumption for our analysis, nor a particularly likely one.

One other methodologically possible alternative would have been to take advantage of the gradient experimental setup (see again n. 14), and then transform the gradient scores into a binary stress distinction (i.e. left vs. right). This might have been attractive from a theoretical phonological point of view that wants to abstract away from surface variation. It is generally the case, however, that when gradient values are transformed into binary ones, a considerable amount of information is lost. Furthermore, statistical techniques that turn gradient into binary values (such as cluster analysis or linear discriminant analysis) necessarily introduce classification errors. Overall this may lead to less powerful statistics if one wants to test specific hypotheses, and comparative exploratory analyses with a transformed binary stress variable confirmed this for our investigation. The logistic regression models with a binary dependent stress variable showed fewer significant effects than the logistic regression models using the gradient perception score as dependent variable.[19] We therefore decided to use the gradient measurements.[20]

A potential alternative to the use of our gradient estimated perception scores would have been to employ the prosodic labels (ToBI, e.g. Silverman et al. 1992) as given in the annotations of the corpus, since one would expect only one pitch accent on left-stressed compounds (namely on the left constituent), and two pitch accents on right-stressed compounds (namely one on each of the two constituents). Apart from potential other problems (such as rater reliability; see e.g. Ostendorf et al. 1996), this method could not be employed since only a small fraction of the publicly available Boston Corpus is prosodically annotated.

[19] We tried out two different kinds of transformation. One was to use the weighted acoustic cues in an automatic classification (employing a hierarchical cluster analysis based on acoustic cues; see Kunter & Plag 2007 for details), and the other was to use the local minimum between the two peaks of the rating scale distribution as shown in Fig. 1 as the cut-off point for the classification. We carried out parallel statistical analyses for both kinds of transformed binary stress values. It turned out that the results were largely parallel.

[20] In Lappe & Plag 2007, we use binary data derived from a subset of the Boston Corpus compounds by automatic classification (see n. 19). In Lappe & Plag 2008 we employed forced choice (left or right) listener ratings for this data set. In both studies, deterministic stress-assignment rules yielded very bad results.

3.3. THE ROLE OF DISCOURSE FACTORS. Using natural speech (instead of citation forms as found in dictionaries) for investigations of stress raises the question of the influence of discourse factors.[21] It is well known that discourse factors may influence the prominence of particular forms in running speech. Such factors fall under the broad label of 'information status', with the pertinent categories contrastiveness, focus, and the given/new distinction. In general the influence of these factors on pitch-accent assignment is very hard to predict, and attempts to do so are only partially successful (e.g. Hirschberg 1993). In principle, however, these factors might have an influence on compound stress, too, and it appears to be advantageous to incorporate these into models of compound stress assignment in running speech. Given the enormous methodological problems involved (see Hirschberg 1993 for some discussion), the question is, however, whether this is really necessary in order to achieve substantial results. This is at least partly an empirical question. The arguably most influential of the pertinent categories in our context is contrastiveness, which may be responsible for violations of the canonical stress pattern of words, and also of compounds. The examples in 3 illustrate this point.[22]

(3) a. I said A dog, not THE dog.
b. John tried to be helpful, but only succeeded in being UNhelpful.
c. I said Park STREET, not Park AVENUE.

Example 3a shows a normally unstressed function word receiving nuclear, hence contrastive, stress, and 3b shows a normally unstressed (or secondarily stressed) prefix that receives primary, hence contrastive, stress. In 3c a normally less prominent compound constituent, *street*, unexpectedly receives main stress for contrastive purposes.

The obvious question now is how far such discourse factors may influence stress assignment to compounds in running speech, and, in particular, in our corpus. With regard to the general influence, very little is known. In Plag 2006 the effects of focus and the given/new distinction were experimentally tested via clausal position and clause type. Plag's data show a general downstepping effect caused by clausal position, that is, the respective pitch values of both left and right constituents steadily decreased from initial to final position. No clear picture emerges, however, as to how clausal position and clause type, in combination with the structural and semantic factors tested, affect stress assignment. Hirschberg (1993) finds that Sproat's *np* algorithm accounted for 59 percent of the compound stresses in her running text speech data, but the author states that half of the wrongly classified cases represented 'accent strategies which were in fact clearly acceptable to native speakers' (1993:322).[23] This would in fact raise the accuracy of stress prediction based on citation forms and rules to a potential 80 percent. Given that even citation forms may vary across speakers and dialects, and given that *np*'s rules themselves are far from fully successful in predicting compound stress, there is little left for an error margin that can be attributed directly to discourse factors. This impression is corroborated by Hirschberg's finding that for the whole corpus (i.e. not just for the compounds) the modeling of the given/new contrast and of contrastiveness adds only 5.4 percent to the overall score of 82.4 percent correct pitch-

[21] Apart from discourse factors, the variability of stress assignment in some languages may be subject to additional constraints that apply only in certain domains, such as poetry or music. For example, in Spanish songs, word stress can shift for metrical reasons (see e.g. Morgan & Janda 1989). Such factors are not at issue here.

[22] Example 3b is adapted from Hirschberg 1993, ex. 7.

[23] Hirschberg (1993) uses the term 'accent strategies' where our teminology would have 'stress patterns'.

accent predictions. In sum, the general influence of discourse factors in stress assignment to compounds in running speech seems quantitatively not very decisive. This may, however, vary from corpus to corpus, so that it seems advisable to have a closer look at the corpus at hand to estimate the influence of such factors in this particular kind of speech.

To test the potential effect of contrastive stresses as the presumably most pertinent discourse factor on our data, we checked the occurrence of contrastive stresses in our above pseudo-random sample of 105 items, and in a further, completely random sample of 200 compounds from our list. In this overall sample of 305 compounds we found only three environments that were interpretable as potentially contrastive, but in none of them was the stress actually shifted, since the constituent to be stressed would have received main stress anyway.[24] Thus, the most pertinent discourse factor, contrastive stress, can be ruled out as potential confound for the results presented below.

An independent argument for the nondecisiveness of discourse factors in our study arises from a comparison of our results with those of related studies. Plag and colleagues (2007) come up with very similar results concerning the structural and semantic factors on the basis of dictionary data. In other words, no matter whether one takes citation forms, as Plag and colleagues (2007) did, or forms from running speech, as we do here, the overall tendencies in the data are roughly the same. Hence, there is good reason to believe that discourse factors in our speech corpus study of compound stress did not have an undue influence on our results.

Another potential confound may be what is known as the 'rhythm rule' or 'iambic reversal', that is, the tendency to shift stresses in order to achieve alternating patterns of stressed and unstressed syllables (cf. *Dundèe mármalade* → *Dùndee mármalade*; see for example Spencer 1996:260). In our compound data, the rhythm rule is unlikely to be of great importance since shift from one syllable to another within one of the compound constituents would not affect the prominence relation that is of interest to us, namely that between the two constituents. Incidentally, this is exactly what Spencer's *Dundee marmalade* example shows. This structure is right-prominent, and this right-prominence is not affected by the application or nonapplication of the rhythm rule. That neighboring words would trigger stress shifts in the compound (from one compound constituent to the other) is unlikely since our compounds are normally heads of their NPs, and thus typically surrounded by less prominent material (e.g. an adjective or a determiner on the left) or by intonation breaks.

3.4. CODING OF PREDICTOR CATEGORIES. Apart from determining the acoustic parameters and estimated perception score, we coded every compound according to the categories held to be responsible for stress assignment in the literature (and some more, to be discussed below). For those variables where categorization proved to be problematic due to the ill-defined nature of the categories mentioned in the literature, each compound was coded independently by two raters, who did not interpret and categorize the compounds in isolation, but took into account the context in which the compounds occurred in the news texts. We analyzed only that subset of the data where the two raters came

[24] For illustration, consider the following example, in which *AIDS patient* could potentially have a constrastive stress on *AIDS* ('brth' indicates short breathing pauses): 'Doctors have long struggled to draw the line between appropriate medical intervention and overly aggressive treatment. brth If a terminally-ill *cancer patient* has a heart attack for instance, brth doctors often recommend that the patient be allowed to die. brth The rationale being that intervention would merely extend suffering. brth Doctors apply the same argument to *AIDS patients*' (M3B20P1).

up with the same categorization. Overall, three raters were engaged in the coding, all of them holding both an M.A. and a Ph.D. in English linguistics. To test the structural and semantic hypotheses, we used the standard multivariate statistical procedures. Further details of the methodologies employed are discussed as we go along.

4. TESTING THE STRUCTURAL HYPOTHESIS.

4.1. ARGUMENT-HEAD VS. MODIFIER-HEAD COMPOUNDS. Let us first take a look at the role of the argument-structure distinction, taking into account only those compounds that were rated by both raters as either argument-head or modifier-head. This reduces our data set to 4,035 compounds. If we test the difference in perception score between modifier-head and argument-head compounds, we should expect a significantly lower score, that is, more leftward stress, with argument-head compounds. This is indeed the case ($t(920.091) = 5.7701$, $p < 0.001$, Welch-modified), with only a small to medium effect size (Cohen's $d = 0.25$) due to large overlap of the two categories.

Although the effect goes in the direction expected under the structural hypothesis, a very clear difference between the two sets of compounds does not emerge. In order to take a closer look at what is going on here we coded the morphological make-up of the head noun and investigated whether we would find an interaction of argument structure and the outermost suffix of the head noun. Under the structural hypothesis we should expect that there would be significant differences between the argument-head compounds and the modifier-head compounds sharing the same head morpheme. Table 1 gives examples of the kinds of combinations we found in the data.

MORPHOLOGY OF HEAD	ARGUMENT-HEAD	MODIFIER-HEAD
-er	law makers	house speaker
	service providers	women voters
	screw driver	city workers
-ing	fundraising	spring training
	flag waving	defense spending
	check kiting	morning shooting
-ion	jury selection	health education
	safety evaluations	bar association
	tourist attraction	restaurant information
conversion	tax increase	litmus test
	drug abuse	housing advocates
	campaign reform	day care

TABLE 1. Morphology of compound heads.

In addition to those types of compounds listed here, we also found a few compounds whose heads ended in the deverbal suffixes *-age*, *-al*, and *-ance*, respectively, but these compounds were too rare to be included in the statistical analysis. Overall, the heads of 1,168 items contained one of the suffixes shown in Table 1 as the outermost suffix.

A type-III analysis of variance of this subset of compounds revealed a significant main effect of right-hand morpheme ($F(3,1160) = 12.5004$, $p < 0.05$), and—most importantly—an interaction of argument structure and right-hand morpheme ($F(3,1160) = 10.0590$, $p < 0.001$, partial $\eta^2 = 0.0057$).[25] A post-hoc test using Tukey contrasts that looked at the four groups of morphologically distinct compounds showed that the only significant difference between argument-head and modifier-head com-

[25] A full documentation of the ANOVA can be found in Appendix A, Table A1.

pounds can be found among those compounds that have *-er* as their right-hand morpheme ($p < 0.001$). The boxplots in Figure 2 illustrate this.[26]

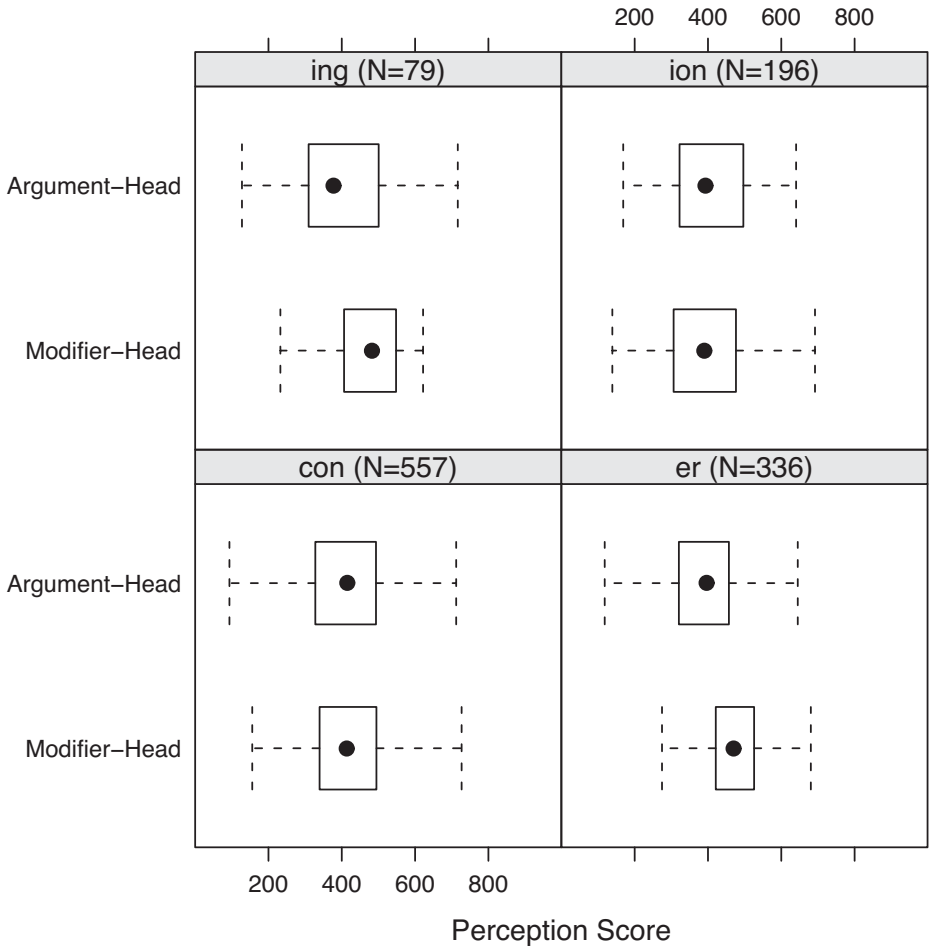

FIGURE 2. Perception score by right-hand morpheme and argument structure.

The fact that the assumed argument-structure effect is restricted to *-er* compounds may seem surprising. The same restriction, however, was found by Plag and colleagues (2007) in their CELEX study. These findings together must be interpreted in such a way that the argument-structure effect hypothesized in the literature is in fact an effect of only one particular subgroup of synthetic compounds, those ending in *-er*. Not surprisingly, this is the subgroup of argument-head compounds that is almost exclusively discussed in the literature, which the other subgroups are being largely ignored.

Apart from the restriction of argument-structure effects to *-er* compounds, an obvious problem of the structural hypothesis is the fact that a vast amount of modifier-head

[26] In these boxplots, the dots indicate the median, the boxes show the interquartile range, and the whiskers give 1.5 times the interquartile range in each direction. The suffix *-ing* also shows a considerable difference in the medians, but, due to a large overlap of the two categories, this difference does not reach significance (Tukey contrasts, adjusted $p = 0.28$).

compounds must be left-stressed, given the high proportion of rather low perception scores. If we take, for example, the mean perception score of *-er* argument-head compounds (403.3) as a baseline for clear leftward stress, still 39 percent of the modifier-head compounds (1,307 of 3,379 overall) have a perception score that is lower than that and should therefore be considered left-stressed, contra the hypothesis. In view of this situation, the obvious escape hatch for the structural hypothesis is lexicalization, which Giegerich (2004) takes to be the reason for the (in his view) noncanonical leftward stress on many modifier-head compounds.

4.2. LEXICALIZATION AND STRESS ASSIGNMENT. We first investigate lexicalization using frequency as a correlate. The problem with frequency is that compounds in general are comparatively rare. For example, Plag (2006) used a number of existing and quite familiar compounds in his experiment, but for some of them even the very large British National Corpus (100 million word tokens) had only very few attestations. Plag and colleagues (2007) had a similar problem with the COBUILD corpus (eighteen million word tokens), in which many of the CELEX compounds do not occur even once. These authors demonstrate, however, that the compounds' COBUILD frequencies nicely correlate with their Google frequencies. Google was therefore an obvious choice for us in spite of potential distortions of these frequency counts having to do with the specific indexing algorithms Google uses.[27] But even Google yielded no frequency for eighteen compounds, most of which refer to institutions that existed at the time the corpus recordings were made (around 1990), but that have since passed out of existence.[28] These compounds were excluded for the analyses involving frequency.

We first entertained an analysis of covariance (ANCOVA) for the 4,017 compounds for which the two raters agreed on their argument-head/modifier-head status and for which frequency information was available. In the type-III ANCOVA using the argument/modifier distinction and Google log frequency as predictors, and perception score as dependent variable, we find a main effect for frequency ($F(1,4013) = 7.3911$, $p < 0.01$), but no effect of argument structure and no interaction of the two predictors ($F(1,4013) = 0.1592$, $p = 0.69$).[29] In other words, we find a frequency effect in the sense that compounds with a higher frequency have a lower perception score, hence are perceived as more left-stressed, as expected by the hypothesis. However, and contra to Giegerich's hypothesis, this effect is not restricted to modifier-head compounds, but also holds for the argument-head compounds. In any case, the effect size of frequency in the ANCOVA is negligible (partial $\eta^2 = 0.001$).

As already mentioned above, spelling can be assumed to be a second correlate of lexicalization. The idea here is that modifier-head compounds spelled as two words should have a tendency to be more right-stressed than modifier-head compounds that are spelled as one word. This is corroborated by Sepp (2006), who finds a strong relation between compound spelling and compound prominence patterns. In her corpus, right-stressed compounds were only rarely written as one word, while left-stressed

[27] 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/.

[28] The procedure for obtaining Google frequencies was as follows. We made a list of all the compounds in our database using two different spellings, that is, one word and two words (hyphens are treated as spaces by Google), containing both the respective singular and plural forms. We then used the software made available by Hayes (2001) to search for these forms via Google on English-language webpages.

[29] A full documentation of the ANCOVA can be found in Appendix A, Table A2.

compounds occurred being spelled as one word or as two words. To substantiate our assumptions about the relationship between spelling, lexicalization, and stress assignment, we first tested the relationship between spelling and frequency, taking the spellings of the compounds as given in the transcripts of the Boston Corpus. The data set was the same as the one we used for the effect of frequency, but we excluded the thirty-one hyphenated words in the corpus in order to arrive at two clearly distinct classes. A Wilcoxon test for the modifier-head compounds showed a very highly significant difference in frequency for one-word spellings as against two-word spellings ($W = 667{,}318.5$, $p < 0.001$). Figure 3, left panel, illustrates the expected effect, that is, that the frequency of one-word modifier-head compounds is higher than that of two-word compounds.[30]

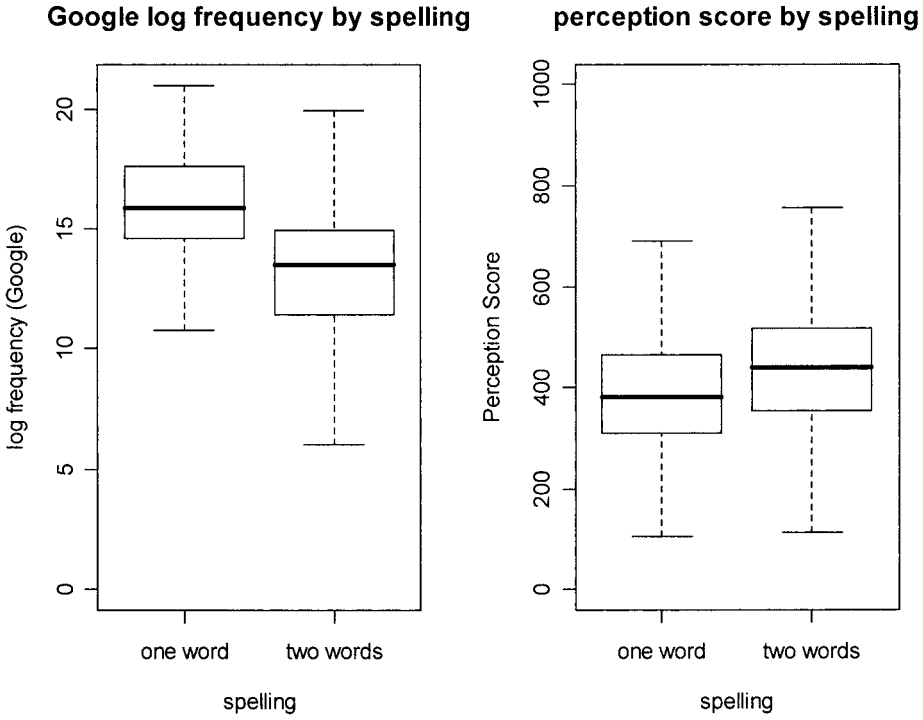

FIGURE 3. Spelling effects.

One-word compounds have a lower perception rating than two-word compounds ($t(419.294) = -7.7119$, $p < 0.001$, Cohen's $d = 0.43$, means 381.0 vs. 433.1). In a type-III ANOVA with argument structure and spelling as predictors and perception score as a dependent variable we find no main effect of argument structure ($F(1,3982) = 0.0001$, $p = 0.99$), a main effect of spelling ($F(1,3982) = 54.3771$, $p < 0.001$), and a significant interaction of spelling and argument structure ($F(1,3982) = 4.3382$, $p < 0.05$). The effect size of the interaction is negligible (partial $\eta^2 = 0.001$).[31]

[30] The same effect can be found if we take also the argument-head compounds into account ($W = 1{,}052{,}619$, $p < 0.001$).

[31] A full documentation of the ANOVA can be found in Appendix A, Table A3.

The interaction between spelling and structure is graphically shown in Figure 4, which indicates the differences between the four groups of compounds by connecting the mean perception scores of one-word spellings and two-word spellings for both argument-head compounds and modifier-head compounds, respectively. Roughly speaking, the graph shows that if words are lexicalized, they tend to be left-stressed across the board, but if they are not lexicalized, then argument structure makes a difference, but only a very small one.

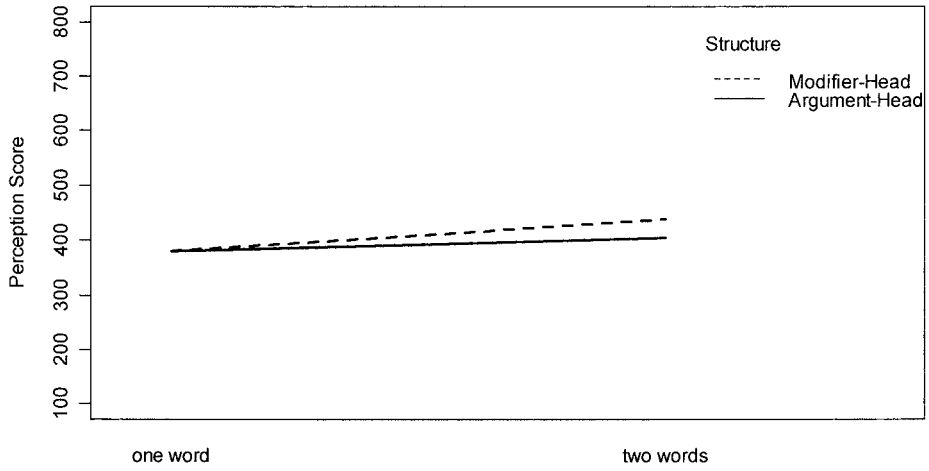

FIGURE 4. Interaction of spelling and structure.

To summarize our exploration of spelling and stress, we get a mixed picture. We find a significant lexicalization effect for compounds, in the sense that compounds written in one word (and thus, by extension, lexicalized) are more frequently left-stressed than compounds written in two words. The effect is only very slightly stronger for modifier-head compounds. These findings correspond quite well to those of Plag and colleagues (2007) who also found a small general lexicalization effect in the CELEX compounds.

A general assessment of the structural hypothesis looks therefore as follows. Overall the hypothesis is not very successful in predicting compound stress. It was shown that the argument-structure effect is restricted to compounds ending in *-er*. Using frequency and spelling as indicators, it was shown that there is a significant lexicalization effect. This effect is not restricted to modifier-head compounds, though a little bit stronger for this group than for argument-head compounds. Let us turn to the semantic hypothesis and see whether it fares any better.

5. TESTING THE SEMANTIC HYPOTHESIS.

5.1. SEMANTIC PROPERTIES OF CONSTITUENTS OR THE WHOLE COMPOUND. As mentioned in §2, we often find claims concerning rightward stress assignment that are based on semantic considerations. In general these considerations refer either to the semantic relationship between the two compound constituents, or to the properties of individual compound constituents or the compound as a whole. We present the analysis of the latter properties in this section, and the results of the analysis of the semantic relations in §5.2.

The literature (e.g. Gussenhoven & Broeders 1981, Fudge 1984:144ff., Zwicky 1986, Liberman & Sproat 1992) predicts rightward stress explicitly for the categories given

in 4a–e below.[32] To these five we added a sixth category, left-headedness; see 4f. N1 refers to the left constituent, N2 to the right constituent. Each compound was coded according to whether it displayed the pertinent category.[33]

(4) a. N1 refers to a period or point in time (as in *morning edition*, *holiday season*, *morning shooting*)
b. N2 is a geographical term (*Boston area*, *Jamaica plain*, *Beacon Hill*)
c. N2 is a type of thoroughfare (*Atlantic Avenue*, *Tiananmen Square*, *Sumner Tunnel*)
d. N1 and N2 form a proper noun (*Tufts University*, *Boston Celtics*, *Lynn Hospital*)
e. N1 is a proper noun (*Dukakis administration*, *King appointee*, *Hamlet machine*)
f. N1 and N2 form a left-headed compound (*attorney general*, *inspector general*)

The data were subjected to a type-III analysis of variance, in which five predictors remained significant. The ANOVA results are given in Table 2.[34]

SEMANTIC PROPERTIES	DIRECTION OF STRESS INFLUENCE	ANOVA STATISTICS	MEAN PERCEPTION SCORE
a. N1 refers to a period or point in time	rightward	$F(1,4076) = 5.5264$, $p < 0.05$	467.5999
b. N2 is a geographical term	rightward	$F(1,4076) = 8.0650$, $p < 0.01$	485.233
d. N1 and N2 form a proper noun	rightward	$F(1,4076) = 4.4110$, $p < 0.05$	475.4612
e. N1 is a proper noun	rightward	$F(1,4076) = 15.5082$, $p < 0.001$	454.4838
f. N1 and N2 form a left-headed compound	rightward	$F(1,4076) = 9.5337$, $p < 0.01$	515.3319

TABLE 2. Semantic properties: significant predictors and direction of influence.

Only one of the six categories does not show the predicted effect, that is, compounds where N2 refers to a type of thoroughfare are not significantly more right-stressed than other compounds. The five other categories show the expected effect toward more rightward stress. Compounds belonging to these five categories made up 15.5 percent of the 4,353 compounds in the Boston Corpus, or 16.5 percent of the 4,082 compounds in the above analysis.

5.2. SEMANTIC RELATIONS BETWEEN COMPOUND CONSTITUENTS. In addition to the semantic categories discussed in the previous paragraphs, the literature (e.g. Fudge 1984: 144ff., Zwicky 1986, Liberman & Sproat 1992) claims that rightward stress is triggered

[32] One can find in the literature some additional categories, such as compounds in which N2 refers to a dish (Gussenhoven & Broeders 1981). In this article we concentrated on those categories that are mentioned across sources and seemed large enough to merit investigation in a corpus study that uses uncontrolled data.

[33] One might think that one should exclude from the category mentioned in 4c ('N2 is a type of thoroughfare') those compounds in which the right constituent was *street*, since such compounds are unanimously considered to be left-stressed in the literature. To verify this claim we took all compounds of category 4c and compared its ten compounds with *street* as N2 to the remaining thirty-one compounds of this category. A Welch-modified *t*-test did not reach significance ($t(34.408) = 0.9878$, $p = 0.3302$), although the compounds with *street* as right constituent had the expectedly lower mean perception score than the rest of the compounds in this class (392.7143 vs. 429.8680). Since this difference turned out to be nonsignificant, we kept the *street* compounds in our data set and did not treat them as a special type of compound.

[34] We followed standard procedures of model simplification (e.g. Crawley 2005). We also tested for possible interactions with no interesting results. A full documentation of the ANOVA can be found in Appendix A, Table A4.

by the semantic relations given in 5. The relations are expressed by supposedly language-independent predicates that link the concepts denoted by the two constituents (see Levi 1978 for discussion).

(5) N2 DURING N1 (*summer vacations*)
N2 IS LOCATED AT N1 (*Newton residents*)
N2 IS MADE OF N1 (*canvas bags*)
N1 MAKES N2 (*Weld plan*)

There are a number of methodological problems with testing these claims. First of all, the semantic categories and semantic relations mentioned in the literature (such as 'N2 is a material', 'N2 is located at N1') seem generally ill-defined. Second, items are often ambiguous (i.e. they show more than one relation). Third, the number of potentially relevant semantic categories and relations is unclear, so that there may be many more than the ten categories and relations mentioned above that have an effect on stress assignment. On a theoretical level, it is also unclear how many and what kinds of relations and categories would be expected to play a role.

In order to deal with, if not solve, these problems with regard to semantic relations we used a set of eighteen semantic relations that are more or less established as useful in studies of compound interpretation. The bulk of these relations comes from Levi 1978, a seminal work on compound semantics, whose relations have since been employed in many linguistic (e.g. Liberman & Sproat 1992) and (more recently) psycholinguistic studies of compound structure and meaning (see for example Gagné & Shoben 1997, Gagné 2001). Levi's catalogue contains fewer than our eighteen relations, but we felt that some additions were necessary, especially to ensure the possibility of reciprocal relations. For example, Levi's list has a relation N2 USES N1 but no relation N1 USES N2. In such cases we added the missing relation to our set of relations to be coded. Furthermore, we added a few categories that we felt were missing from her set but were necessary to adequately categorize nonnegligible portions of the data, such as N2 IS NAMED AFTER N1. In Table 3 we present the final list of our relations, with three illustrative examples (unless the corpus did not provide at least three).

		SEMANTIC RELATION		EXAMPLE
1.	N2	CAUSES	N1	flu virus, AIDS virus
2.	N1	CAUSES	N2	hate crimes, computer problem, building boom
3.	N2	HAS	N1	wheelchair, stock market, department store
4.	N1	HAS	N2	state officials, army dentist, cabinet members
5.	N2	MAKES	N1	computer firm, survey researcher
6.	N1	MAKES	N2	footsteps, wheel ruts, computer image
7.	N2	IS MADE OF	N1	glass pellets, ivory towers, lead weights
8.	N2	USES	N1	litmus test, computer methods, phone services
9.	N1	USES	N2	(not attested in the corpus, only attestation in CELEX: handbrake)
10.	N1	IS	N2	exhibition games, home town, boyfriend
11.	N1/N2	IS LIKE	N2/N1	umbrella group, junk bonds, potholes
12.	N2	FOR	N1	court houses, payment system, charge cards
13.	N2	ABOUT	N1	law books, cancer committee, war story
14.	N2	IS LOCATED AT/IN/. . .	N1	Boston offices, Walpole residents, kitchen sink
15.	N1	IS LOCATED AT/IN/. . .	N2	classroom, combat zone, minority areas
16.	N2	DURING	N1	summer vacations, campaign appearances, morning shooting
17.	N2	IS NAMED AFTER	N1	Rockefeller Center, Abbot labs, Apollo computer
18.	OTHER			speed bumps, baby bills, generation gap

TABLE 3. List of semantic relations coded, illustrated with three examples each.

Some of the categories proved especially difficult to code consistently, so additional guidelines were developed. These concerned mainly the interpretation of the predicates CAUSE, MAKE, and IS. CAUSE was pertinent in cases where a cause (denoted by one constituent) triggers an effect (denoted by the other constituent), while MAKE was coded in cases of purposeful creation or of production. IS subsumes three cases, the first being that the left constituent denotes a subset of the denotation of the right constituent (*jail facilities*), the second being that left and right constituents are not in a subset-superset relation and IS works in both directions (*pet fish*, *girl-friend*), and the third being same-level copulative compounds (*actor-poet*). We treated the IS LIKE relation as nondirectional because we assumed that in those cases where the predicate holds in one direction it also holds in the other, since N1 and N2 share a (nondirectional) similarity on which the IS LIKE interpretation is essentially based.

Given that noun-noun compounds in English are in principle ambiguous (see for example Adams 2001:82–88, Plag 2003:148–51), a compound could be assigned multiple relationships. For example, *conference committee* was interpreted by our raters as exhibiting the two relations N1 HAS N2 and N2 FOR N1. Consequently, this compound entered the analysis with these two relations. Three of the relations (N1 CAUSES N2, N2 MAKES N1, N1 USES N2) had to be discarded because our corpus did not contain enough items for a statistical analysis. Furthermore, as mentioned above, we discarded all items for which the two raters did not agree. The rest of the data ($N = 2{,}041$) were subjected to a type-III analysis of variance, whose results are given in Table 4.[35]

SEMANTIC RELATION	DIRECTION OF STRESS INFLUENCE	ANOVA STATISTICS	MEAN PERCEPTION SCORE (overall mean: 423.4925)
4. N1 HAS N2	rightward	$F(1,2033) = 114.2071$, $p < 0.001$	472.0548
7. N2 IS MADE OF N1	rightward	$F(1,2033) = 7.0824$, $p < 0.01$	460.6647
8. N2 USES N1	leftward	$F(1,2033) = 5.9336$, $p < 0.05$	362.4201
10. N1 IS N2	rightward	$F(1,2033) = 22.2709$, $p < 0.001$	454.8547
14. N2 LOCATED AT N1	rightward	$F(1,2033) = 32.4254$, $p < 0.001$	454.3832
16. N2 DURING N1	rightward	$F(1,2033) = 25.0008$, $p < 0.001$	485.8757
17. N2 IS NAMED AFTER N1	rightward	$F(1,2033) = 10.8461$, $p < 0.01$	440.3148

TABLE 4. Semantic relations: significant predictors and direction of influence.

Of the four relations cited in 5 to trigger rightward stress, three behave as predicted: compounds expressing the relations N2 DURING N1, N2 LOCATED AT N1, and N2 IS MADE OF N1 have a higher perception score on average and are thus significantly more right-stressed. The relation N1 MAKES N2 (*Weld plan*) does not trigger more rightward stress. This finding is in line with the experimental results of Plag 2006, in which the predicted effect was also not found.

Compounds with the relation N1 IS N2 also tend toward rightward stress. This is expected, since these compounds correspond largely to the class of copulative compounds. From our analysis of the eighteen semantic relations, two new relations emerge that have an influence toward more rightward stress, N2 IS NAMED AFTER N1, and N1 HAS N2. Furthermore, we detect one new relation that has an influence toward more leftward stress, N2 USES N1. Of the 2,041 compounds, 42.4 percent ($N = 866$) have semantic relations that show a significant tendency toward rightward stress.

[35] A full documentation of the ANOVA can be found in Appendix A, Table A5.

To summarize our investigation of the semantic hypothesis, we can state that the categories and relationships claimed in the literature to trigger right-hand stress mostly behave in the expected way. There is one category ('N2 is a type of thoroughfare') and one relation (N1 MAKES N2), however, that do not behave in the predicted way. Furthermore, we have seen that there are also semantic categories and relations that have an effect on stress assignment and that have not been previously mentioned as such in the literature. The proportion of compounds that belong to the semantic categories and relations that tend toward rightward stress is quite high. If we take the subset of data used for the semantic relations, we find 43.8 percent of the compounds (893 of 2,041) showing at least one of the semantic properties or relations that have a tendency toward rightward stress.

This figure also shows that there is a considerable overlap between the semantic properties and the semantic relations. For this data set, we find 15.3 percent of the compounds ($N = 312$) showing one of the semantic properties, and 42.4 percent ($N = 866$) showing one of the semantic relations, with a tendency toward rightward stress, which would add up to more than 57 percent. However, only 43.8 percent of the compounds show one or more of the pertinent properties and relations, which means that 14 percent show more than one. An overlap is expected since, for example, compounds in which the first element denotes a period or point in time are likely to also encode a temporal relationship (N2 DURING N1, as in *morning edition*). Another case in point are compounds whose right consituent is a geographical term. These compounds are likely to also show a locative relation (N2 IS LOCATED AT N1, as in *Boston area*).

6. COMBINING STRUCTURAL AND SEMANTIC FACTORS. In order to see which of the factors show significant effects in a model that takes all kinds of factors into account, we carried out a regression analysis with the estimated perception score as the dependent variable and argument structure, morphology of the head, and all semantic categories and relations as predictors. The final trimmed model is documented in Table 5.[36]

	ESTIMATE	STD. ERROR	*t*-VALUE	Pr(>|t|)
(Intercept)	394.3156	16.8738	23.368	< 2e-16
compound is left-headed	123.9905	27.9411	4.438	9.62e-06
N2 is a geographical term	80.6706	21.8458	3.693	0.000228
compound is a proper noun	46.0590	11.1047	4.148	3.51e-05
4. N1 HAS N2	71.1338	6.6151	10.753	< 2e-16
7. N2 IS MADE OF N1	54.8301	21.8601	2.508	0.012217
10. N1 IS N2	63.9726	11.3205	5.651	1.84e-08
14. N2 LOCATED AT N1	36.8054	8.9265	4.123	3.90e-05
16. N2 DURING N1	80.0350	16.5692	4.830	1.47e-06
log frequency	−1.8619	0.8562	−2.175	0.029782
spelling (two words)	29.4599	10.1346	2.907	0.003693

TABLE 5. Linear regression with all factors as predictors (final model, $N = 1,898$, adjusted $R^2 = 0.1031$).

[36] The data set for this model was the largest one possible. Thus, it consisted of the data set for the semantic hypothesis, minus those items for which no frequency information was available, and minus the very few items that were hyphenated. There is a small overlap between the semantic categories and the structural relation of 'argument-head' ($N = 525$), which pertained to thirty-three tokens (e.g. *breath test*, which can be interpreted as a test that tests the breath, or as a test that uses breath to test the level of alcohol in the blood). In the process of model simplification we found some nonnormality in the distribution of the residuals, which points toward a nonlinearity between the predictors and the response variable. To address this nonlinearity we removed the thirty data points whose standardized residuals exceeded 2.5 standard deviations. The refitted model showed no harmful traces of nonlinearity. For this model we ran a bootstrap validation (see Baayen 2008:§6.2.4) in which all factors were kept, and which led to a corrected R^2 of 0.098.

Perhaps the most striking result is that argument structure and morphology of the head no longer play a role. Instead, the model has a rather low baseline of an estimated perception score of 394.3, which can be interpreted as indicating left stress (recall that the mean perception score of *-er* argument-head compounds was 403.3; see also Fig. 1). This baseline holds for compounds that are spelled as one word and do not belong to one of the semantic categories listed in the table. If a compound belongs to one of the semantic categories in the left column of the table, or is spelled as two words, the perception score increases by the estimate provided in the second column. For example, the estimate for compounds expressing a temporal relationship (N2 DURING N1) would be 474.4 (394.3 + 80.0) in this model. The only predictor that works in the direction of more leftward stress is frequency, where an increase of 1 in log frequency leads to an estimated decrease of 1.9 points on the perception-score scale.

It does not come as a surprise that argument structure and, among semantic relations, N2 USES N1 are no longer significant. These two factors were found to go in the direction of leftward stress. Leftward stress is, however, taken as a baseline by our model. Of the semantic category effects, three survive. Left-headed compounds, those that have a geographical term as their right constituent, and compounds that are proper nouns show significantly more rightward stress, while compounds where N1 is a proper noun or N2 denotes a type of thoroughfare do not show the expected effect. Apart from N2 IS NAMED AFTER N1, all semantic relations found to go together with more rightward stresses survive in the final model.

Both orthography and frequency emerge as significant predictors, but with no significant interaction between indicators of lexicalization and argument structure. This means that our data provide strong empirical evidence for a general lexicalization effect, with more lexicalized compounds having a tendency toward more leftward stress. In the final model the lexicalization effect is best seen with spelling, where two-word compounds have a perception score of 424 (394.3 + 29.5) as against 394 for one-word compounds. If lexicalization is measured in terms of frequency, the effect of frequency is rather marginal. The overall range in log frequency for our compounds is from 0.0 to 19.42, with half of the data falling into the interval between 11.76 and 15.38. The model therefore estimates a maximal effect of frequency of $19.42 \times -1.86 = -36.12$ perception-score units. In other words, frequency is capable of shifting the perception score toward leftward stress by maximally 36.12 when comparing the most frequent with the least frequent compounds. In comparison with the other effects, this is relatively little.

The final model provides very good evidence for a semantic approach to compound stress assignment and against the structural hypothesis. The most successful approach to variable stress assignment is to consider leftward stress assignment as the default, which can be overruled if the compound shows certain semantic properties or relations. These effects are, however, not deterministic, but rather are probabilistic in nature. Although compounds of the pertinent types show a strong tendency toward more rightward stress, not all compounds of a specific category show the expected behavior. For illustration, let us look at some bar plots for the semantic effects in Figure 5.

We can see that the perception scores for compounds of the pertinent categories have a strong tendency to be above the baseline (indicated by the dark bars in the graphs), but that in almost all of the categories we find compounds whose perception score is below the baseline, in the range that we would consider left-stressed. Similarly, the compounds not belonging to one of the semantic categories that tend toward rightward

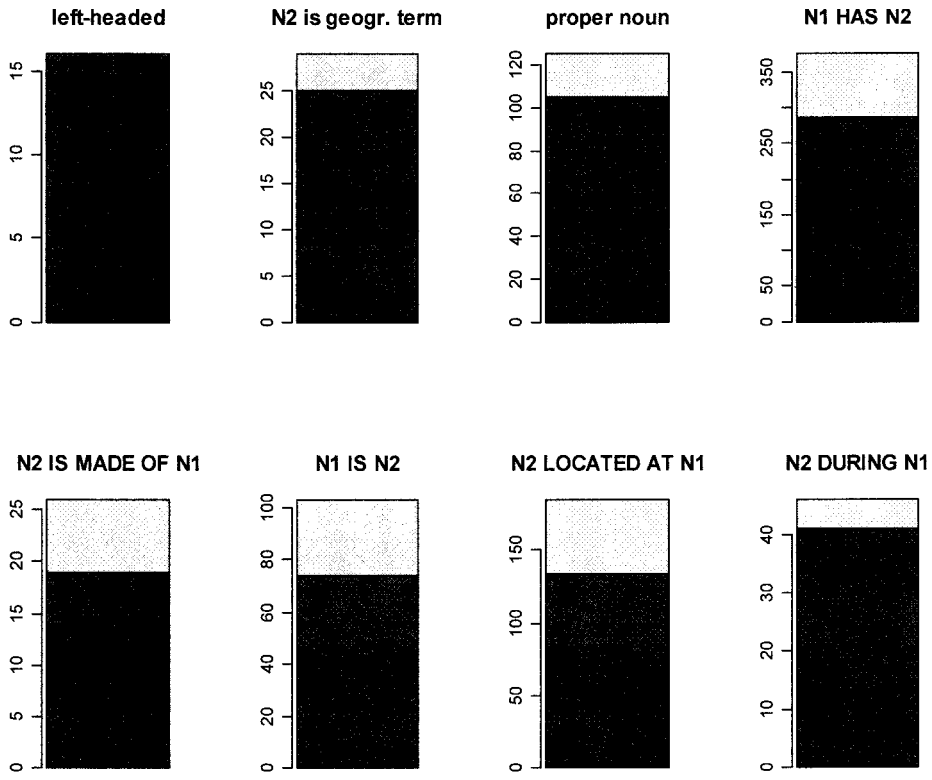

FIGURE 5. Distribution of items with perception scores above and below the baseline for different kinds of compounds. The dark portion of each bar indicates the proportion of items above the baseline; the light portion represents the proportion of items below the baseline. The height of the bars gives the number of pertinent compounds.

stress have a distribution in which only little more than half of the items are below the baseline (50.3 percent), that is, in the range that we would consider clearly left-stressed.

7. SUMMARY AND DISCUSSION. In this article we have looked at the variability of stress assignment to English noun-noun structures. In particular, we tested whether some existing hypotheses about compound stress can accurately predict the stresses as found in natural speech instantiated in the Boston University Radio Speech Corpus. Overall, it turned out that such speech corpus data can provide pertinent evidence.

The structural hypothesis and the semantic hypothesis were shown to be untenable in their existing forms. Although parts of the data show the expected behavior, other (and sometimes large) parts of the data do not behave as predicted. Thus, in the separate analysis of argument structure it was shown that an effect of argument structure is restricted to those compounds whose heads end in the suffix *-er*. This seemingly awkward effect not only is statistically robust, but has also been found in an independent study of compound stress (Plag et al. 2007). In that study the data came from the lexical database CELEX, which contains more than 4,000 compounds from two dictionaries. These parallel results raise two questions, the first of which is why this effect has never been noticed before. A careful look into the literature reveals a very simple reason: nobody seems to have ever looked in more detail at the stress behavior of synthetic

compounds (but see Giegerich 2006 for some pertinent remarks). While there is a host of studies on the intricacies of the internal and external syntax of synthetic compounds (see e.g. Spencer 1991 for an overview), stress assignment in these compounds has not been very systematically studied. The second and much more interesting question is of course why we should find this strange interaction between head morphology and stress assignment. One possibility is that it is not the morphology itself but rather the semantics that is responsible for the peculiar behavior of *-er* compounds. Thus, *-er* compounds usually refer to persons or instruments, while compounds ending in *-ing* (e.g. *fundraising*, *spring training*), *-ion* (e.g. *jury selection*, *health education*), or zero (*tax increase*, *litmus test*) are usually action or result nouns. If we take this seriously, Giegerich's (2004) distinction between argument-head and modifier-head compounds would have to be replaced by a distinction referring to the semantic category of the head (or the compound), that is, a distinction between person noun and action noun. This line of reasoning already points in a direction that seems generally more promising for an account of compound stress assignment, semantics.[37]

In the separate analysis of semantic factors, most of the semantic categories and relations mentioned in the literature show an effect in the predicted direction. Thus compounds that are left-headed (e.g. *attorney general*), compounds in which the first element denotes a period or point of time (e.g. *morning edition*), compounds that have a proper noun as their first element (e.g. *Dukakis administration*), and compounds that have a geographical term as their second element (e.g. *Boston area*) are all significantly more prone to rightward stress than other compounds. One other category ('N2 is a thoroughfare') does not show a significant effect. With regard to semantic relations we found that from a set of eighteen relations, six had a significant effect toward rightward stress, and one toward leftward stress.

This result goes far beyond what previous semantic approaches have claimed. First, a study of the pertinent literature revealed only four relations held to trigger righthand stress. Among these four, only three were found to act as significant predictors in the present study (N2 DURING N1 as in *summer vacations*, N2 IS LOCATED AT N1 as in *Newton residents*, and N2 IS MADE OF N1 as in *canvas bags*), while no effect is found for N1 MAKES N2 (*Weld plan*). A noneffect for the latter relation was also found in the experimental study in Plag 2006. This part of the semantic hypothesis is therefore likely to be wrong. Second, in addition to the three significant relations just mentioned, we found effects for four other relations: N1 HAS N2, N1 IS N2, N2 IS NAMED AFTER N1, and N2 USES N1. Surprisingly, the latter one works in the direction of leftward stress. Unsurprisingly, N1 IS N2 has been found to be significant in assigning more rightward stress. This category (taken from Levi 1978) largely overlaps with what are traditionally referred to as copulative compounds, and these are generally considered to be right-stressed (e.g. Olsen 2001).

[37] Interestingly, Giegerich (2005, 2006) has also recently proposed a more semantic approach to his original distinction. Whereas he acknowledges the general variability of stress in NN constructs, he predicts that rightward stress should not occur in compounds in which the semantics of the relation between N1 and N2 is nonattributive. The reason is that, in his view, leftward stress can only be assigned if the construction arises in the lexicon, and the lexicon is the place where nonattributive constructions arise. Attributive constructions, by contrast, arise in the syntax, where they are assigned rightward stress. In Giegerich's view, such constructions may, however, migrate to the lexicon, where they may or may not change their stress pattern from right to left. It is as yet unclear how the semantic relations and categories found to be significant predictors of compound stress in this article relate to Giegerich's (2005, 2006) proposal. Note, however, that the two relations that Giegerich mentions as central attributive relations, N2 IS N1 and N2 IS MADE OF N1, are among those relations that we found to favor rightward stress (cf. Table 4).

In a regression analysis including all factors, most of the semantic effects survived, while argument-structure and morphological effects disappeared. This means that we have collected very robust evidence for a view that the semantics is the most important factor in the prediction of the stress pattern of a given compound. The structural hypothesis has been shown to underdetermine stress assignment. Contra to the hypothesis, and irrespective of lexicalization, only certain subsets of the modifier-head compounds tend toward rightward stress. These subsets are semantically defined, as shown above.

The simultaneous analysis of all factors also revealed a significant, but only rather small, lexicalization effect that did not interact with argument structure. The prediction of the structural hypothesis concerning lexicalization is therefore also partially refuted, in that we find a general effect of spelling and frequency, and not an effect that is restricted to modifier-head structures.

Let us compare our speech-corpus results briefly with those based on dictionary data, as presented in Plag et al. 2007. Using the same sets of predictor variables, they found significant effects for two semantic categories and nine semantic relations, most of them overlapping with those found to be significant in the present study. Furthermore, Plag and colleagues also found a general lexicalization effect, not restricted to modifier-head compounds. These parallels in the main results are strong support for the significance of semantics and, though to a much lesser degree, lexicalization in compound stress assignment. In addition, the similarity of results between the two studies is independent support for the reliability of the kind of methodology we have developed and applied in the present article.

Another important parallel between the present article and Plag et al. 2007 is that the observed effects of certain factors on stress assignment are not categorical in nature. Plag et al. 2007 shows that the predictive accuracies of categorical rules are generally much worse than those of probabilistic or analogical models. The results of the present study point in the same direction. We find statistically significant effects, but these effects are of a probabilistic nature and often quite small. This means that there are tendencies in the expected directions, but there are often compounds of the pertinent category that do not behave in the predicted way (see again Fig. 5). This fact speaks strongly against a deterministic rule-based approach to compound stress and adds fuel to alternative approaches.

In recent studies employing such alternative modelings, such as Plag et al. 2007 and Lappe & Plag 2007, 2008, it is shown that probabilistic and exemplar-based models are generally more successful in predicting compound stress in English than rule-based models, and that analogical models in turn outperform probabilistic models. In analogical models, stress is assigned to new exemplars on the basis of similarity to existing compounds that are already stored in the mental lexicon. With such models, Lappe and Plag (2007, 2008) obtained the best results if the analogical algorithms worked exclusively on the basis of the left and right constituent families. A ‘constituent family’ is the set of compounds that share the first (or the second) constituent with a given compound. This constituent-family effect is reminiscent of the effect of certain constituents that has already been mentioned in the literature, for example, categorical leftward stress in all compounds that have *street* as their right constituent, with the decisive difference that analogical models provide systematic across-the-board evidence for such family effects. With such models, assigning stress solely only on the basis of the majority stress pattern of all stored compounds with the same left or right constituent is much more successful in assigning stress correctly to the new exemplar than assigning stress on the basis of similarities computed over all kinds of semantic or structural information.

Notably, this strong effect of the constituent families in compound stress assignment is in line with investigations of the morphological behavior of compounds in other languages (e.g. by Krott and collaborators, 2001, 2002, 2004a, 2004b, Kuperman et al. 2008), which have also shown that variable compound behavior is best accounted for by models that have constituent family among their most important predictors.

Analogical models may raise the question of why one should find robust semantic effects such as those found here in the first place. Interestingly, such effects could emerge even in an analogical approach that relies chiefly on constituent families (instead of directly on semantics). We know from psycholinguistic studies (e.g. Gagné & Shoben 1997, Gagné 2001) that compounds that share one constituent with each other tend to show the same semantic relation. For example, compounds with the right constituent *magazine* tend to show the relation N2 ABOUT N1, as in *mountain magazine*. The semantic effects shown to exist in our compounds may thus potentially emerge in an analogical model as a by-product of the similarities computed over large numbers of exemplars. More research is necessary to show whether this is indeed the case.

Apart from developing the right formal models, future investigations would also have to provide more detailed information concerning two other important issues. The first would be lexicalization. Given that both dictionary data and speech-corpus data show significant (though not very strong) lexicalization effects (based on frequency and spelling) it would be interesting to take other measures of lexicalization into account, such as semantic transparency.

Another issue is the variability of stress across different tokens of the same compound (i.e. type). The data from the Boston Corpus strongly suggest that stress variability may hold not only across types, but also within types, that is, across tokens of one type.[38] Thus, it seems that certain compounds are always stressed in a certain way, while other compounds may be stressed sometimes leftward and sometimes rightward.[39] One case in point is *state official*, which shows distinct within-speaker and cross-speaker variation in the Boston Corpus. It seems reasonable to assume that token frequency may play an important role in preventing within-type variability: high-frequency items have a higher representational strength, including their phonological properties. We would therefore predict that compounds with a relatively high frequency will show less variability than compounds with a relatively low frequency. This hypothesis, however, needs refinement. Given that the constituent families of a compound have an influence on the stress pattern of a compound, one could hypothesize that variability may emerge in those cases where the constituent families of the left and the right element of a given compound suggest different stress patterns, respectively. Taking both the token frequency of the compound and the size of the two constituent families into account, we can make the following predictions. If the token frequency is low, the item will have a low representational strength and the stress pattern is largely dictated by the two constituent families. If the constituent families suggest competing stress patterns, the outcome will be unclear and leave room for variability. If the token frequency is high, variability may occur if the sizes of the two stress-competing constituent families are both also large. In this scenario, there is competition between the stress bias induced by the two constituent families and competition between the constituent-based stress

[38] See Kunter 2007 for a pilot study of this type of variability using data from the Boston Corpus.

[39] In his experimental phonetic study, Lutstorf even comes to the conclusion that 'most compounds may shift their stress pattern' (1960:141, emphasis omitted). This may be an artefact of the methodology, which involved forced choice between three stress levels (left, right, and level).

bias and the stress representation stored in the memory. That this refined hypothesis seems to be on the right track is evidenced, for example, by the highly variable stress patterns of *state officials* or *budget cuts* found in Kunter 2007, 2009. The constituent family of *state* shows a strong bias for right stress, while there is a strong left-stress bias emerging from the constituent family of *officials*. The compound is highly frequent in the news context, but at the same time consists of two highly type-frequent elements (the two constituent families are the largest in the Boston Corpus). The case is similar with *budget cuts*, where competing stress biases of the constituent families and a high degree of variability cooccur, as predicted by the hypothesis. The hypothesis still allows low-frequency forms to show only little variability if there is a clear bias introduced by the constituent families. This hypothesis makes clear and testable predictions about which compounds are likely to show variability and which ones are not, and these predictions may be tested in future work.

Appendix A: Analyses of variance tables

	Sum Sq	Df	*F* value	Pr(>F)
(Intercept)	54,294,584	1	4,033.1864	< 2.2e-16
isAH	14,139	1	1.0503	0.3057
morphRight	504,842	3	12.5004	4.775e-08
isAH:morphRight	406,240	3	10.0590	1.518e-06
Residuals	15,615,871	1,160		

Table A1. Analysis of covariance (type III) showing the effect of right-hand morpheme ('morphRight') and argument structure ('isAH') on perception score.

	Sum Sq	Df	*F* value	Pr(>F)
(Intercept)	43,569,239	1	2,942.5784	< 2.2e-16
logFreqsum	109,437	1	7.3911	0.006583
isAH	17,317	1	1.1695	0.279564
logFreqsum:isAH	2,357	1	0.1592	0.689937
Residuals	59,418,419	4,013		

Table A2. Analysis of variance (type III) showing the effect of frequency ('logFreqSum') and argument structure ('isAH') on perception score.

	Sum Sq	Df	*F* value	Pr(>F)
(Intercept)	38,463,912	1	2,624.7436	< 2.2e-16
orth	796,861	1	54.3771	2.002e-13
isAH	2	1	0.0001	0.99084
orth:isAH	63,573	1	4.3382	0.03733
Residuals	58,353,624	3,982		

Table A3. Analysis of variance (type III) showing the effect of spelling ('orth') and argument structure ('isAH') on perception score.

	Sum Sq	Df	*F* value	Pr(>F)
(Intercept)	612,370,212	1	41,533.2839	< 2.2e-16
isLH	140,566	1	9.5337	0.002031
isTIM	81,482	1	5.5264	0.018779
isGEO	118,912	1	8.0650	0.004535
isN1PN	228,655	1	15.5082	8.352e-05
isPropN	65,036	1	4.4110	0.035769
Residuals	60,096,885	4,076		

Table A4. Analysis of variance (type III) showing the effect of semantic properties ('isLH', 'isTIM', 'isGEO', 'isN1PN', 'isPropN') on perception score.

	Sum Sq	Df	*F* VALUE	Pr(>F)
(Intercept)	174,539,225	1	12,014.1729	< 2.2e-16
semRel4	1,659,176	1	114.2071	< 2.2e-16
semRel7	102,891	1	7.0824	0.007846
semRel8	86,202	1	5.9336	0.014940
semRel10	323,546	1	22.2709	2.529e-06
semRel14	471,069	1	32.4254	1.419e-08
semRel16	363,207	1	25.0008	6.222e-07
semRel17	157,570	1	10.8461	0.001007
Residuals	29,534,971	2,033		

TABLE A5. Effect of semantic relations ('semRel4', 'semRel7', 'semRel8', 'semRel10', 'semRel14', 'semRel16', 'semRel17') on perception score.

APPENDIX B.

List of 500 randomly selected compounds from the Boston University Radio Speech Corpus. The spelling was changed to normal standards.

access road	biotechnology industry	career day	computer intelligence
adult prisons	bird species	car phones	computer makers
advance copy	blood screenings	case studies	computer sales
advertising agency	board meeting	cement patio	computer screen
advertising battle	bombshell	chain reaction	concrete spaghetti
advertising dollars	bond rating	chair manufacturers	condo boom
aerobics research	bookkeeper	chairwoman	condo market
African-American	boom decade	changing tables	conference committee
AIDS office	Boston agencies	check kiting	consumer advocate
AIDS patients	Boston area	checkpoint	consumer advocates
AIDS tests	Boston College's	Chelsea agency	consumer attitudes
air officials	Boston home	chewing coca	consumer awareness
air pollution	Boston mayor	Chicago experience	consumer reports
air time	Boston meeting	chief economist	contract offer
air-time	Boston offices	chief justice	Corico River
aluminum cans	Boston police	chief justices	cornerstones
amnesty program	Boston schools	Christmas ads	corps members
art copies	boycott organizers	Christmas carol	cost-containment
art market	breakfast meeting	Christmas season	cost control
arts grants	breath test	church protest	cost savings
art trustees	budget mess	cigarette tax	cost sharing
assault rifle	budget news	city hall	court houses
assault rifles	budget paranoia	city planners	court reform
attorney general	budget process	city streets	court ruling
attrition program	budget proposal	class meeting	credit card
auto insurance	budget savings	class participation	credit unions
auto mechanics	Bulger breakfast	Cleo award	crewmen
aviation officials	Burger King	Cleveland Cavaliers	cross street
baby bills	business analyst	clinic workers	dairy cow
Baghdad University	business man	clothing boutique	damage award
bank scam	businessmen	club house	damage awards
bargaining table	business owners	coffee maker	defense budget
batting practice	business people	Colinnet software	defense secretary
Bay State	business principles	college scholarships	development officials
beaver thing	cabinet posts	commodity auction	diets experts
bicycle stores	cable splicer	community activists	dinosaur magnets
bidding process	Cambridge corporation	community service	dinosaur mania
bill collectors	Cambridge resident	compromise bill	dinosaur merchandise
biotechnology companies	Cambridge residents	computer company	dinosaur products
biotechnology firms	camp grounds	computer design	dinosaur researcher

dinosaur society
Dorchester rink
drug activity
drug education
drug habit
drug profits
drug programs
drug trade
drug treatment
drug users
dune fields
DWI charges
economics professor
education efforts
emergency spending
energy audit
engineer entrepreneurs
EPA warnings
equipment trucks
expansion plans
family lands
fare increase
features editor
fee hikes
fighting skills
Finneran version
fire fighter
fire fighting
fish communication
fledgling organization
fleet representative
flu viruses
fossil fuels
freebie breakfast
fringe candidates
front-runner
funding increase
funding pump
fundraising
gambling clubs
gang neighborhoods
gas tax
generation gap
gravel replacement
ground surface
ground swell
gun charges
gun control
hair combs
hairstylist
hallways
hand guns
Hannover Street
Harrison Avenue
Harvard study
headaches
health advisory
health class
health insurers
health study
hearing room
Helms amendment
HIV carriers
HIV infection
HMO plan
HMO proposal
holiday spending
homebuilders
horror story
horse owners
hotel rooms
house efforts
housekeeper
housekeepers
housing advocates
housing subsidy
hunger strike
immigrant advocates
immigration policy
industry analysts
industry giant
information age
initiative process
insurance charges
insurance officials
IRS agent's
jail facilities
job candidates
job front
journal line
justice system
Kansas City
key deterrent
King candidacy
laboratory studies
land holdings
law books
law clerk
lawsuits
leadership opportunities
leadership team
lead standards
lemon survey
licensing fee
life-span
Lincoln parents
listening device
literacy test
loan company
lobbying campaign
loon calls
lottery participants
love affair
lunchboxes
Lynn Hospital
machine Democrats
mafia captain
Maine wilderness
Malibu beach
management skills
manufacturing plants
Marshall Plan
Mashentucket Pequots
Massachusetts administrator
Massachusetts ballot
Massachusetts chapter
Massachusetts elections
Massachusetts hospitals
Massachusetts lobbyist
Massachusetts school
Massachusetts senator
Massachusetts taxes
Massachusetts towns
Massachusetts voters
mass exodus
media campaigns
memory chips
metal plate
metal shell
Metco parents
minority donors
minority patients
minority voters
MIT engineer
money problems
mortgage scam
name calling
nanny work
Navy doctor
negotiating sessions
neighborhood groups
neighborhood schools
NHL play-offs
nose-dive
nude works
nutcracker
obscenity charges
office parks
oil facilities
oil spill
organ transplants
package tours
pancakes
Paris review
parking lot
Parkland
party affiliation
party infighting
Patriot missile
pension benefit
people skills
pest killers
phone calls
phone service
phone services
photo journalist
pine trees
Pittsburgh Pirates
planning perspective
plant operators
point guard
Polaroid company's
police academy
police departments
police detective
police force
police killer
police misconduct
police scanners
policy reversal
policy statements
pool room
position paper
practice balls
prison overcrowding
privacy laws
Providence Journal
rap group
rat traps
rate hike
rate hikes
rate payers
recidivism rates
record amounts
recording industry
record numbers
recycling capacity
referral shop
registry police
rental market
rescue effort
research funding
rest homes
restructuring program
retailers association
retirement age
river front
roadside
road surface
sail fin
salary arbitration
sales performance
sand beds
Saronno labs
Saugus restaurant
school budget
school complex
school systems
score board
Seabrook
search warrant
security captain

selling point
senate president
service people
service providers
sewage system
sex lives
shellfish
signature collection
sign language
skin protein
slugfest
Somerville mayor
Sox prospects
spending authorizations
spending cuts
spring training
stairway
state biologists
state budget
state coffers
state colleges
state decision
state dignitaries
state economy
state jobs
state parks
state police
state prison
state prisons
state representative
state resident
state roads
state subsidies
state treasurer
state workers
stock market
storm windows
strategy session
strike vote
student athletes
Suffolk Downs
summer day
sunlight
survey researchers
survey results
TA associates
target neighborhoods
tax advantage
tax amnesty
tax assessments
tax expert
tax hike
tax lawyers
tax payer
taxpayers
tax proposal
tax returns
tax vote
telephone company
telephone equivalent
television anchor
teller machines
temper tantrums
test authority
test-case
testing ground
theater critic
THM standards
tobacco companies
tourism industry
town halls
town officials
tracking polls
trade group
trade mission
trail blazers
training directors
training schedule
transit alternatives
transition process
tree trunks
trial lawyer
tuition hikes
TV commercials
union members
US House
US Senate
voting booth
Walpole site
Waltham company
war effort
warplanes
water bottles
watershed
water users
weekend's
Weld administration
Weston Observatory
whirlwind
wind erosion
Winter Haven
work week
world champion
Yale biologist
Yale students
youth counselors
youth services
zoo officials

REFERENCES

ADAMS, VALERIE. 1973. *An introduction to Modern English word-formation*. London: Longman.

ADAMS, VALERIE. 2001. *Complex words in English*. Harlow: Pearson Education/Longman.

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

BAAYEN, R. HARALD; RICHARD PIEPENBROCK; and LEON GULIKERS. 1995. *The CELEX lexical database* (CD-ROM). Philadelphia: Linguistic Data Consortium.

BAUER, LAURIE. 1983a. *English word-formation*. Cambridge: Cambridge University Press.

BAUER, LAURIE. 1983b. Stress in compounds: A rejoinder. *English Studies* 64.47–53.

BAUER, LAURIE. 1998. When is a sequence of two nouns a compound in English? *English Language and Linguistics* 2.65–86.

BLOOM, PAUL. 2000. *How children learn the meanings of words*. Cambridge, MA: MIT Press.

BLOOMFIELD, LEONARD. 1933. *Language*. New York: Holt, Rinehart, & Winston.

BOD, RENS; JENNIFER HAY; and STEFANIE JANNEDY (eds.) 2003. *Probabilistic linguistics*. Cambridge, MA: MIT Press.

BOERSMA, PAUL, and DAVID WEENINK. 2005. Praat: Doing phonetics by computer (Version 4.3.01). Amsterdam: University of Amsterdam. Online: http://www.praat.org/.

BUTTERWORTH, BRIAN. 1983. Lexical representation. *Language production, vol. 2: Development, writing and other language processes*, ed. by Brian Butterworth, 257–94. London: Academic Press.

CHAPMAN, ROBERT L. 1977. *Roget's International Thesaurus*. 4th edn. New York: Harper and Row.

CHOMSKY, NOAM, and MORRIS HALLE. 1968. *The sound pattern of English*. New York: Harper and Row.

CLAHSEN, HARALD. 1999. Lexical entries and rules of language: A multidisciplinary study of German inflection. *Behavioral and Brain Sciences* 22.6.991–1060.

CRAWLEY, MICHAEL J. 2005. *Statistics: An introduction using R*. Chichester: John Wiley.

DE VAAN, LAURA; ROB SCHREUDER; and R. HARALD BAAYEN. 2007. Regular morphologically complex neologisms leave detectable traces in the mental lexicon. *Mental Lexicon* 2.1.1–23.

FARNETANI, EDDA, and PIERO COSI. 1988. English compound versus non-compound noun phrases in discourse: An acoustic and perceptual study. *Language and Speech* 31.157–80.

FRY, DENNIS BUTLER. 1958. Experiments in the perception of stress. *Language and Speech* 1.126–52.

FUDGE, ERIK C. 1984. *English word-stress*. London: George Allen and Unwin.

GAGNÉ, CHRISTINA L. 2001. Relation and lexical priming during the interpretation of noun-noun combinations. *Journal of Experimental Psychology: Learning, Memory and Cognition* 27.236–54.

GAGNÉ, CHRISTINA L., and EDWARD J. 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, HEINZ J. 2004. Compound or phrase? English noun-plus-noun constructions and the stress criterion. *English Language and Linguistics* 8.1–24.

GIEGERICH, HEINZ J. 2005. Associative adjectives and the lexicon-syntax interface. *Journal of Linguistics* 41.571–91.

GIEGERICH, HEINZ J. 2006. Attribution in English and the distinction between phrases and compounds. *Englisch in Zeit und Raum – English in time and space: Forschungsbericht für Klaus Faiss*, ed. by Petr Rösel. Trier: Wissenschaftlicher Verlag Trier.

GUSSENHOVEN, CARLOS. 2004. *The phonology of tone and intonation*. Cambridge: Cambridge University Press.

GUSSENHOVEN, CARLOS, and A. BROEDERS. 1981. *English pronunciation for student teachers*. Groningen: Wolters-Noordhoff-Longman.

HAY, JENNIFER, and R. HARALD BAAYEN. 2005. Shifting paradigms: Gradient structure in morphology. *Trends in Cognitive Sciences* 9.7.342–48.

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

HAYES, BRUCE. 2001. Query Google. Los Angeles: University of California, Los Angeles. Online: http://www.linguistics.ucla.edu/people/hayes/QueryGoogle/.

HEDBERG, NANCY; JUAN M. SOSA; and LORNA FADDEN. 2004. Meanings and configurations of questions in English. *Proceedings of Speech Prosody 2004, Nara (Japan)*, 309–12.

HIRSCHBERG, JULIA. 1993. Pitch accent in context: Predicting intonational prominence from text. *Artificial Intelligence* 63.305–40.

INGRAM, JOHN; THI ANH THU NGUYEN; and ROB PENSALFINI. 2003. An acoustic analysis of compound and phrasal stress patterns in Australian English. St. Lucia: University of Queensland, MS.

JESPERSEN, OTTO. 1909 [1961]. *A modern English grammar on historical principles, part 1: Sounds and spelling*. London: Allen and Unwin.

KINGDON, ROGER. 1958. *The groundwork of English stress*. London: Longmans, Green and Co.

KROTT, ANDREA; R. HARALD BAAYEN; and ROBERT SCHREUDER. 2001. Analogy in morphology: Modeling the choice of linking morphemes in Dutch. *Linguistics* 39.51–93.

KROTT, ANDREA; PETER HAGOORT; and R. HARALD BAAYEN. 2004a. Sublexical units and supralexical combinatorics: The case of Dutch interfixes in visual processing. *Language and Cognitive Processes* 19.453–71.

KROTT, ANDREA; GARY LIBBEN; GONIA JAREMA; WOLFGANG DRESSLER; ROBERT SCHREUDER; and R. HARALD BAAYEN. 2004b. Probability in the grammar of German and Dutch: Interfixation in triconstituent compounds. *Language and Speech* 47.1.83–106.

KROTT, ANDREA; ROBERT SCHREUDER; and R. HARALD BAAYEN. 2002. Analogical hierarchy: Exemplar-based modeling of linkers in Dutch noun-noun compounds. In Skousen et al., 181–206.

KUNTER, GERO. 2007. Within-speaker and between-speaker variation in compound stress assignment. Paper presented at the second International Conference on the Linguistics of Contemporary English, Université de Toulouse II, July 4, 2007.

KUNTER, GERO. 2009. *The phonetics and phonology of compound stress in English*. Siegen: Universität Siegen dissertation, to appear.

KUNTER, GERO, and INGO PLAG. 2007. What is compound stress? *16th International Congress of Phonetic Sciences, University of Saarbrücken, 6–10 August 2007*, ed. by Jürgen Trouvain and William Barry, 1005–8. Saarbrücken: Universität Saarbrücken.

KUPERMAN, VICTOR; ROBERT SCHREUDER; RAYMOND BERTRAM; and R. HARALD BAAYEN. 2008. Reading polymorphemic Dutch compounds: Towards a multiple route model of lexical processing. Nijmegen: Radboud Universiteit Nijmegen, MS.

LADD, D. ROBERT. 1984. English compound stress. *Intonation, accent and rhythm*, ed. by Dafydd Gibbon and Helmut Richter, 253–66. Berlin: Mouton de Gruyter.

LADEFOGED, PETER. 2005. *Phonetic data analysis: An introduction to fieldwork and instrumental techniques*. Oxford: Blackwell.

LAPPE, SABINE, and INGO PLAG. 2007. The variability of compound stress in English: Towards an exemplar-based alternative of the compound stress rule. *Proceedings of the ESSLLI workshop on exemplar-based models of language acquisition and use. Dublin, Ireland, August 2007*.

LAPPE, SABINE, and INGO PLAG. 2008. The variability of compound stress in English: Rules or exemplars? Paper presented at the 13th International Morphology Meeting, Vienna, February 3–6, 2008.

LEES, ROBERT B. 1963. *The grammar of English nominalizations*. The Hague: Mouton.

LEHISTE, ILSE. 1970. *Suprasegmentals*. Cambridge, MA: MIT Press.

LEVI, JUDITH N. 1978. *The syntax and semantics of complex nominals*. New York: Academic Press.

LIBERMAN, MARK Y., and RICHARD SPROAT. 1992. The stress and structure of modified noun phrases in English. *Lexical matters* (Lecture notes 24), ed. by Ivan Sag and Anna Szabolcsi, 131–81. Stanford, CA: CSLI Publications.

LIPKA, LEONHARD. 1994. Lexicalization and institutionalization. *The encyclopedia of language and linguistics*, 1st edn., ed. by Roland E. Asher, 2164–67. Oxford: Pergamon.

LUTSTORF, HEINZ THEO. 1960. *The stressing of compounds in Modern English: A study in experimental phonetics*. Bern: Walter Fischer.

MARCHAND, HANS. 1969. *The categories and types of present-day English word-formation*. 2nd rev. edn. Munich: C. H. Beck.

MORGAN, TERRELL A., and RICHARD D. JANDA. 1989. Musically-conditioned stress shift in Spanish revisited: Empirical verification and nonlinear analysis. *Studies in Romance linguistics: Selected proceedings from the 17th Linguistic Symposium on Romance Languages*, ed. by Carl Kirschner and Janet Ann DeCesaris, 273–88. New Brunswick, NJ: Rutgers University Press.

OLSEN, SUSAN. 2000. Compounding and stress in English: A closer look at the boundary between morphology and syntax. *Linguistische Berichte* 181.55–69.

OLSEN, SUSAN. 2001. Copulative compounds: A closer look at the interface between syntax and morphology. *Yearbook of Morphology* 2000.279–320.

OSTENDORF, MARI; PATTI PRICE; and STEFANIE SHATTUCK-HUFNAGEL. 1996. *Boston University Radio Speech Corpus*. Philadelphia: Linguistic Data Consortium, University of Pennsylvania.

PAYNE, JOHN, and RODNEY HUDDLESTON. 2002. Nouns and noun phrases. *The Cambridge grammar of the English language*, ed. by Rodney Huddleston and Geoffrey K. Pullum, 323–524. Cambridge: Cambridge University Press.

PLAG, INGO. 2003. *Word-formation in English*. (Cambridge textbooks in linguistics.) Cambridge: Cambridge University Press.

PLAG, INGO. 2006. The variability of compound stress in English: Structural, semantic and analogical factors. *English Language and Linguistics* 10.1.143–72.

PLAG, INGO; GERO KUNTER; and SABINE LAPPE. 2007. Testing hypotheses about compound stress assignment in English: A corpus-based investigation. *Corpus Linguistics and Linguistic Theory* 3.2.199–233.

SAMPSON, GEOFFREY RICHARD. 1980. Stress in English N + N phrases: A further complicating factor. *English Studies* 61.264–70.

SCHMERLING, SUSAN F. 1971. A stress mess. *Studies in the Linguistic Sciences* 1.52–65.

SEPP, MARY. 2006. *Phonological constraints and free variation in compounding: A corpus study of English and Estonian noun compounds*. New York: City University of New York dissertation.

SILVERMAN, KIM E. A.; MARY BECKMAN; JOHN F. PITRELLI; M. STENDORF; COLIN WIGHTMAN; PATTI PRICE; JANET PIERREHUMBERT; and JULIA HIRSCHBERG. 1992. ToBI: A standard for labeling English prosody. *Proceedings of the 1992 International Conference on Spoken Language Processing*, 867–70. 12–16 October, Banff, Alberta, Canada.

SKOUSEN, ROYAL; DERYLE LONSDALE; and DILWORTH S. PARKINSON (eds.) 2002. *Analogical modeling: An exemplar-based approach to language*. Amsterdam: John Benjamins.

SLUIJTER, AGAATH M. C.; VINCENT J. U. VAN HEUVEN; and JOS J. A. PACILLY. 1997. Spectral balance as a cue in the perception of linguistic stress. *Journal of the Acoustic Society of America* 101.503–13.

SPENCER, ANDREW. 1991. *Morphological theory: An introduction to word structure in generative grammar*. Cambridge: Cambridge University Press.

SPENCER, ANDREW. 1996. *Phonology*. Oxford: Blackwell.

SPENCER, ANDREW. 2003. Does English have productive compounding? *Topics in morphology: Selected papers from the third Mediterranean Morphology Meeting (Barcelona, September 20–22, 2001)*, ed. by Geert E. Booij, Janet DeCesaris, Angela Ralli, and Sergio Scalise, 329–41. Barcelona: Institut Universitari de Lingüística Applicada, Universtitat Pompeu Fabra.

SPROAT, RICHARD. 1994. English noun-phrase accent prediction for text-to-speech. *Computer Speech and Language* 8.79–94.

TESCHNER, RICHARD, and M. STANLEY WHITLEY. 2004. *Pronouncing English: A stress-based approach with CD-ROM*. Washington, DC: Georgetown University Press.

ZWICKY, ARNOLD M. 1986. Forestress and afterstress. *Ohio State University Working Papers in Linguistics* 32.46–62.

English Linguistics
Fachbereich 3
Universität Siegen
Adolf-Reichwein-Str. 2
D-57068 Siegen
[plag@anglistik.uni-siegen.de]
[kunter@anglistik.uni-siegen.de]
[arndt-lappe@anglistik.uni-siegen.de]
[braun@anglistik.uni-siegen.de]

[Received 11 September 2007;
accepted 12 May 2008]