Prominence in triconstituent compounds:

Pitch contours and linguistic theory

Kristina Kösling[1], Gero Kunter[1], Harald Baayen[2] & Ingo Plag[1]

[1] Seminar für Anglistik, Universität Siegen
[2] Department of Linguistics, University of Alberta, Edmonton

Corresponding author

Gero Kunter
Universität Siegen
Fakultät I: Seminar für Anglistik
Adolf-Reichwein-Str. 2
D-57072 Siegen

Tel. +49 (0)271-740-4569
Fax +49 (0)271-740-3246
e-mail: kunter@anglistik.uni-siegen.de

ABSTRACT

According to the widely accepted Lexical Category Prominence Rule (LCPR, Liberman and Prince, 1977, see also Sproat 1994 for a more recent variant of the LCPR), prominence assignment to triconstituent compounds depends on the branching direction. Left-branching compounds, i.e. compounds with a left-hand complex constituent, are held to have highest prominence on the leftmost constituent, whereas right-branching compounds have highest prominence on the second of the three constituents. The LCPR is, however, only poorly empirically supported (e.g. Kvam, 1990, Sproat, 1994, Berg, 2009, Giegerich, 2009, Kösling and Plag, 2009). The present paper tests a new hypothesis concerning the prominence of triconstituent compounds and suggests a new methodology for the empirical investigation of compound prominence. According to this hypothesis the prominence pattern of the embedded compound has a decisive influence on the prominence of the whole compound. Using a mixed effects generalized additive model for the analysis of the pitch movements, it is shown that all triconstituent compounds have an accent on the first constituent irrespective of branching, and that the placement of a second, or even a third, accent is dependent on the prominence pattern of the embedded compound. The LCPR is wrong.

INTRODUCTION

Until recently, it was widely believed that phonological prominence assignment to triconstituent compounds such as *child care center* or *university textbook* depends on the branching direction. Left-branching compounds, i.e. compounds with a left-hand complex constituent (e.g. *child care*), are held to have highest prominence on the leftmost constituent, whereas right-branching compounds such as *university textbook* have highest prominence on the second of the three constituents. This idea was captured by Liberman and Prince (1977) in their 'Lexical Category Prominence Rule' (LCPR). Studies such as Berg (2009), Giegerich (2009) or Kösling and Plag (2009) have shown, however, that the branching direction is not a reliable predictor for NNN prominence placement, and that there exist compounds with all conceivable combinations of branching directions and prominence distributions. At present, it is unclear what exactly determines the prominence pattern of a given NNN compound.

The present paper has two major aims. First, it develops and tests a new hypothesis concerning the prominence of triconstituent compounds and, second, it suggests a new methodology for the empirical investigation of compound prominence. According to the hypothesis we develop in the next section, the main problem of the LCPR lies in the assumption that NN compounds in English are left-prominent. According to recent studies, about one third of the compound tokens in natural speech are right-prominent (e.g. Plag, 2010, Kunter, 2011, Bell and Plag, 2012), which has important repercussions for the distribution of prominence in larger compounds. We hypothesize that the prominence pattern of the embedded compound has a decisive influence on the prominence of the compound: that element which is most prominent in the embedded compound will also be the most-prominent element in the triconstituent compound.

Second, empirical investigations of compound prominence have been riddled with methodological problems. Pitch has been found to be a highly important cue to compound

prominence (e.g. Farnetani et al., 1988, Kunter and Plag, 2007, Kunter, 2011), and, consequently, the pertinent studies investigating the assignment of different prominence patterns often used pitch for their investigations.[1] However, the researchers used measures that all abstracted away from the actual pitch contours (such as mean pitch or minimum and maximum pitch), although the systematic variation of the contour is potentially highly informative for prominence perception. For instance, results in Kunter (2011) show that the slope of the pitch contour differs between the constituents of left- and right-prominent NN constructions. Yet, with the methodology employed in that analysis, it remained unclear in how far these findings reflect differences in the overall pitch contour of these constructions. We therefore propose a new method for modeling prominence in compounds that overcomes these problems by producing an abstraction of the pitch contour for the different compound types and makes use of the contour information: mixed effects generalized additive models (e.g. Wood, 2006, 2011).

Using this new statistical technique for the analysis of the pitch movements, it is shown that all triconstituent compounds have an accent on the first constituent irrespective of branching, and that the placement of a second accent is determined by the prominence pattern of the embedded compound together with branching. The findings are compared to those reported in Kösling (2011) using a different methodology for the same data set. The present

---

1   Intensity is another important cue to prominence in compounds (cf. Kochanski et al., 2005, Kunter and Plag, 2007, Kunter, 2011, Morrill, 2012). The relation between duration and prominence in compounds is fairly complex, and findings are somewhat contradictory (for a discussion of this cue, see Kunter, 2011 for noun-noun compounds, and Morrill, 2012 for adjective-noun compounds).

analysis yields compatible, yet more detailed results, which suggests that generalized additive models are a valid method of describing pitch contours in compounds.

The paper is structured as follows. In the next section we describe existing hypotheses on prominence assignment to NNN compounds and suggest an alternative. The third section will introduce the problems with modeling prominence in compounds on the basis of acoustic measurements. This is followed by a discussion of our methodology, and a presentation of the results. The last section concludes with a discussion of the results and an outlook of their implications.

Prominence in triconstituent compounds: Hypotheses

It is generally assumed that prominence assignment to triconstituent compounds depends on the branching direction of the compound. Left-branching compounds have highest prominence on the leftmost constituent, whereas right-branching compounds, i.e. compounds with a right-hand complex constituent, have highest prominence on the second constituent of the whole compound. The Lexical Category Prominence Rule (LCPR) proposed by Liberman and Prince (1977) has been evoked to account for this generalization. The LCPR labels metrical trees on the basis of strong-weak relations between two sister constituents. Hence, one constituent is always strong (S), i.e. more prominent, in relation to its weaker sister constituent (W). In particular, the LCPR makes a prediction as to which constituent of a binary construction will be prosodically strong:

(1) **LCPR (Lexical Category Prominence Rule)**

   'In a configuration [A B]: If C is a lexical category, B is strong iff it branches' (Liberman and Prince, 1977, p. 257)

The predictions of the LCPR are illustrated in Figures 1 and 2. In addition to the labels 'S' and 'W' (strong and weak, respectively), the 'N-level' refers to the level of the embedded NN compound and the 'IC-level' labels the level between the complex and the single constituent,

with 'IC' meaning 'immediate constituent'. The tree in Figure 1 shows a left-branching compound (labeled 'L'), while Figure 2 shows a right-branching compound (labeled 'R'). Sproat (1994) revises the LCPR in such a way that prominence assignment is seen as the result of deaccentuation: according to his *np* system, the second constituent of a compound retains its accent only if it is complex:

(2) ***np* system**

'For each node C dominating [A B], if B is not complex, and if C is N^0, then B is

deaccented; else both A and B retain their accents' (Sproat, 1994: 84)

Yet, even Sproat's revised version of the LCPR is only poorly empirically supported with respect to prominence assignment in complex compounds. In fact, available studies dealing with prominence assignment to NNN constructions (e.g. Kvam, 1990, Berg, 2009, Giegerich, 2009, Kösling and Plag, 2009) have shown that the LCPR fails to predict the correct prominence pattern for a considerable number of compounds.
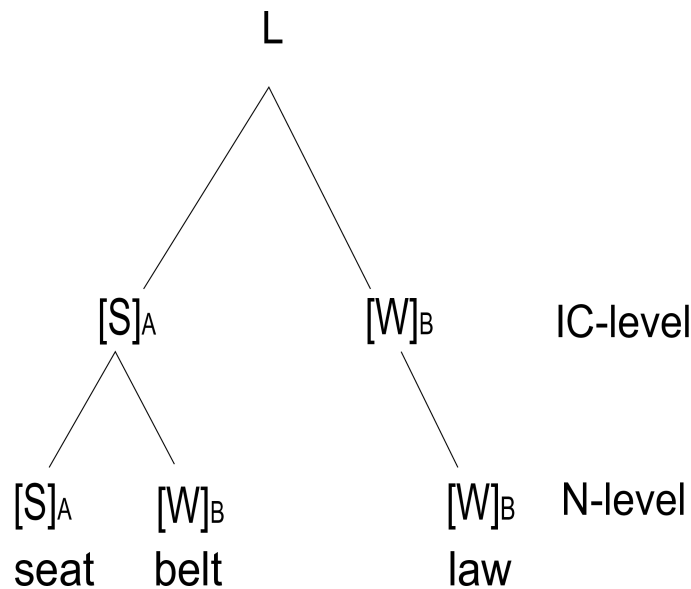

Figure 1: Left-branching compound

R

[W]$_A$ [S]$_B$ IC-level

[W]$_A$ [S]$_A$ [W]$_B$ N-level
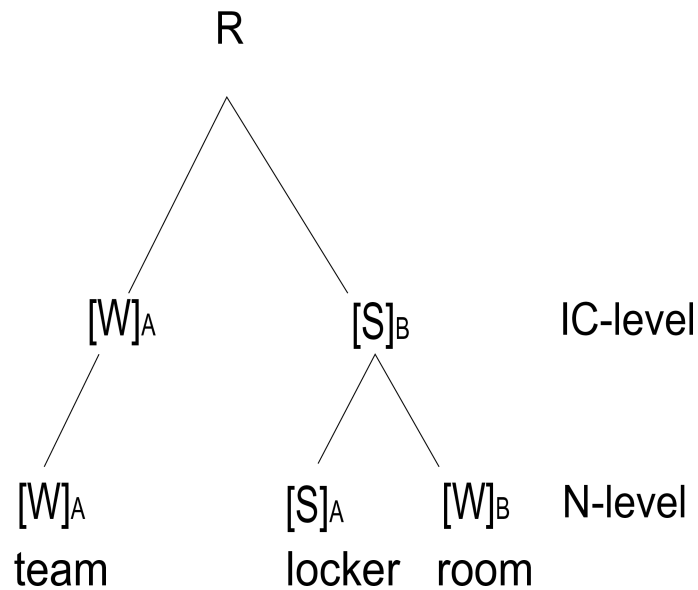team   locker room

Figure 2: Right-branching compound

For instance, Kvam (1990) investigated 40 noun-noun-noun constructions in a production experiment. Even if his methodology remains somewhat unclear, Kvam found that the majority of the investigated compounds, namely 30 out of 40, was either exclusively or by the majority of the experimental subjects prominent on constituent N2. Yet, Kvam points out that only 10 of these compounds were also clearly right-branching, i.e. the group of compounds that should indeed have this stress pattern. Hence, only in 10 cases stress assignment could be directly related to the branching direction of the compound.

Apart from Kvam's study, additional evidence towards variation in the prominence assignment of NNN constructions is provided by Berg (2009). Taking an explorative approach by looking at a total of 642 noun-noun-noun combinations taken from the British National Corpus (Berg, 2009, p. 87), Berg finds that 57.2% of the combinations are prominent on constituent N2, and 26.5% on N1. Thus, Berg's findings go in the same direction as Kvam's results, revealing a general tendency for triconstituent compounds to be in their majority most prominent on the second constituent, be they left- or right-branching. In addition to that, Berg also provides information about a number of right-branching compounds with prominence on constituent N1 and N3, as well as left-branching compounds

with prominence on N3. Similar counter-examples are also provided in a more recent approach by Giegerich (2009). His study investigates the traditional English Compound Stress Rule, and, in contrast to Kvam and Berg, explicitly argues against the LCPR and its predictions.

Finally, a considerable number of counter-examples to the LCPR were also documented in a recent corpus study by Kösling and Plag (2009), who tested the predictions of the LCPR by analyzing about 500 compounds taken from the Boston University Radio Speech Corpus (Ostendorf et al., 1996). Their analysis of the corpus data revealed a general trend for left-and right-branching compounds to behave as predicted by Liberman and Prince's LCPR, i.e. left-branching compounds tended to be prominent on constituent N1, and right-branching compounds on constituent N2. Crucially, however, their corpus data also revealed a significant proportion of compounds violating the rule, both at the level of the immediate constituents (IC-level) and at the level of the final nodes of the tree (N-level). Table 1 lists some these N-level and IC-level violations. Note that some of the examples may appear to allow different branching interpretations than the one given in the table. Yet, the context in which the triconstituent compounds occurred in the corpus clearly show that the given interpretation of the internal structure is the most plausible one.

Table 1

LCPR violations

Branching direction	Violation at N-level	Violation at IC level
left-branching	[*science fíction*] *shocker* [*capital gáins*] *tax*	[*Boston area*] *commúnities* [*weekend*] *séries*
right-branching	*state* [*health prógram*] *school* [*drug úse*]	*commúnity* [*meeting hall*] *crédit* [*scoring system*]

The examples listed in Table 1 raise the question of which factors are responsible for their aberrant prominence behavior.[2] Based on their data, Kösling and Plag (2009) argue that the violations at the N-level and the IC-level may both be explained by the presence of right-prominence in NN compounds. Previous studies have shown that there is a host of right-prominent NN constructions, and their frequency has been estimated to roughly 30 percent of all NN constructions (cf. Sproat, 1994, Plag, 2010, Kunter, 2011, Bell and Plag, 2012).

Kösling and Plag (2009) propose that the violations at the N-level arise due to this presence of embedded right-prominent NN compounds: for example, *science fiction shocker* has prominence on the second constituent because the complex left constituent consists of the right prominent *science fiction*; the same assumption is also put forward by Giegerich (2009). This prominence pattern is not predicted by the LCPR or the *np* system, because the existence of right-prominent compounds is not possible under the formulation of the LCPR: in Liberman and Prince's (1977) approach, right-prominent NNs (apparently including *science fiction*) are simply assumed to be syntactic phrases, and the prominence pattern of phrases is assumed to follow a different prominence rule than the LCPR. The same problem is present in the *np* system, which also hinges on a clear distinction between compounds and phrases. For example, Sproat (1994) claims that there is a structural difference between *Párk Street* (which is considered a compound) and *Madison Ávenue* (which is considered not a compound but a

---

2   These examples were obtained from the Boston University Radio Speech Corpus. The prominence pattern given in Table 1 reflects the pronunciation by at least one speaker from the corpus. Of course, other speakers might prefer different prominence patterns, as prominence assignment in compounds is known to show a certain degree of variability (e.g. Kunter, 2011). Yet, the number of violations found in Kösling and Plag (2009) is so large that they cannot be attributed to individual variations only.

phrase, Sproat, 1994: 83). His explanation of this difference appears to be based on the prominence pattern alone, thus leading to a circular argument: his *np* system will only yield right prominence for *Madison Avenue* if it is a phrase, but the claim that *Madison Avenue* has to be a phrase seems to be derived from the stress pattern alone. Indeed, it has been shown (e.g. in Bauer, 1998, Giegerich, 2009) that the distinction between compounds and phrases is very difficult to draw on formal grounds, and *Park Street* and *Madison Avenue* are a case in point.

If we apply the *np* system to triconstituent compounds, the patterns in Table 1 involving right-prominent NN compounds are unpredicted by the *np* system. For *science fiction shocker*, the algorithm would deaccentuate *fiction* (unless it is considered a phrase, but there is no independent evidence for this), and also deaccentuate *shocker*, because it is not complex, yielding *scíence fiction shocker* with prominence on the first constituent, not on the second as in the corpus data. Likewise, if we assume *drug use* to be a compound, *use* would be deaccentuated because it is not complex, so the resulting prominence pattern would be *school drúg use*, which conflicts with the attested *school drug úse*.

With reference to the IC-level violations, Kösling and Plag (2009) suggest that the same factors responsible for prominence variation in NN compounds are also responsible for prominence variation in NNN compounds. The literature on prominence variation in NN compounds usually discusses three groups of factor, namely structure, semantics and analogy. Experimental and corpus studies testing these factors (e.g. Plag, 2006, Plag et al., 2008, Kunter, 2011, Arndt-Lappe, 2011) found that especially semantics and analogy play a role with structure being less significant.[3] Hence, right-prominent NN compounds are responsible

---

3   A more detailed account of these factors can be found below in the section on target compound construction.

for many instances of aberrant prominence patterns at the IC-level, because the same mechanisms operating in biconstituent compounds, i.e. at the N-level, may also operate at the IC-level of a given compound.

The present analysis remains agnostic about this ongoing debate on the factors affecting the prominence patterns of compounds at the IC-level, and concentrates on the N-level violations instead, for which we can formulate explicit predictions. We want to test Kösling and Plag's (2009) and Giegerich's (2009) hypothesis regarding the violations at the N-level with a new methodological instrument. We focus on the question as to whether the prominence pattern of the embedded NN compound affects prominence assignment in triconstituent compounds as suggested by these authors. In particular, we want to test whether left-branching compounds with embedded right-prominent NN compounds are prominent on constituent N2, and whether right-branching compounds with embedded right-prominent NN compounds are prominent on constituent N3. These two cases would be unpredicted by the LCPR. Furthermore, we want to test whether compounds with embedded left-prominent NN compounds are prominent on constituent N1 (in the case of left-branching compounds), or on constituent N2 (in the case of right-branching compounds), as it is predicted by the LCPR. We may reformulate these assumptions as two related hypotheses, which, taken together, will be referred to as the Embedded Prominence Hypothesis (EPH, proposed in Kösling, 2011) for the rest of this paper.

(3)   Embedded Prominence Hypothesis (EPH):

   a. In left-branching and right-branching NNN compounds with embedded left-prominent NN compounds, highest prominence is assigned to the left member of the complex constituent.

b. If the embedded NN compound is right-prominent, left-branching and right-branching compounds have highest prominence on the right member of the complex constituent.

So, in effect, the Embedded Prominence Hypothesis claims that in a triconstituent compound that has highest prominence on the complex constituent, the highest prominence falls on the constituent that is also prominent in the embedded compound. Accordingly, the hypothesis makes predictions for four different types of compounds: on the one hand, left-branching compounds with either an embedded left-prominent NN compound or an embedded right-prominent NN compound, and on the other hand, right-branching compounds with either an embedded left-prominent NN compound or an embedded right-prominent NN compound. These four types will be referred to as 'L/N1', 'L/N2', 'R/N2' and 'R/N3' in the remaining sections of this paper. The labels 'L' (left) and 'R' (right) indicate the branching direction of the NNN compound at the IC-level. The labels 'N1', 'N2' and 'N3' refer to the constituent in the compound that the EPH predicts to receive highest prominence in the compound. For example, right-branching compounds with a left-prominent complex constituent are expected to have highest prominence on the left constituent of the embedded complex constituent, which is, as seen above in Figure 2, constituent N2. This type of compound is therefore labeled 'R/N2'. In contrast, 'R/N3' compounds are right-branching compounds with embedded right-prominent NNs, as the EPH predicts that the right-prominence is preserved under embedding, and highest prominence should therefore fall to constituent N3. Similarly, 'L/N1' and 'L/N2' compounds are left-branching compounds with left- and right-prominent complex constituents, respectively. Table 2 summarizes the four NNN compound types.

Table 2

Classification of different types of branching direction and their embedded prominence

patterns predicted by the EPH

Label	Branching direction	Prominence pattern of embedded NN	Expected prominence pattern of NNN	Example
L/N1	left	[ŃN]	[ŃN]N	[*háy fever*] *treatment*
L/N2	left	[NŃ]	[NŃ]N	[*science fíction*] *book*
R/N2	right	[ŃN]	N[ŃN]	*business* [*crédit card*]
R/N3	right	[NŃ]	N[NŃ]	*family* [*Christmas dínner*]

In order to empirically test the EPH, one does not only need large amounts of pertinent compounds, but also a reliable method of determining the actual prominence pattern for each of these compounds. This is not a trivial task, and this paper will put forward a new methodology that is able to overcome certain shortcomings of previous approaches.

Pitch and prominence

The autosegmental-metrical framework of intonational phonology (for a summary, see e.g. Ladd, 1996) proposes that prominence differences in multi-word constructions are linked to the presence of pitch accents on some of the words in such a construction. In principle, pitch accents are tonal targets which are aligned with stressed syllables and that shape the pitch contour of an intonational phrase. In English, these tonal targets can be either high (H*) or low (L*), or in the case of bitonal pitch accents, a combination of both (L*+H, L+H*). Perceptually, words that are accented are perceived as more prominent than unaccented words, and the last accent in a phrase (the nuclear accent) has highest prominence. In addition to pitch accents, tonal targets can also be aligned with the right edge of phrases. These phrase accents and boundary tones are not assumed to be prominence-lending, but provide acoustic cues to the phrasing of larger utterances, as well as the pragmatic coherence of the phrases (see Pierrehumbert, 1980, Beckman and Pierrehumbert, 1986, Beckman and Edwards, 1994,

Gussenhoven, 2004 for detailed descriptions of the intonational system of English, and Hirschberg, 2004 for a summary of the link between intonation and pragmatic functions).

The autosegmental-metrical account assumes a particular relation between perceived prominence and notable changes in pitch height, and the temporal distribution of these changes (see Arvaniti et al., 1998, and Dilley et al., 2005 for discussions of the timing of pitch contours): listeners are attuned to the shape of the pitch contour, and perceive those elements as particularly prominent during which the pitch contour shows particular events. This link between pitch and prominence has been supported by the findings in studies such as Rietveld and Gussenhoven (1985) and Terken (1997): syllables received increasingly higher prominence ratings with increasing pitch excursion sizes. Rietveld and Gussenhoven (1985) have shown that for Dutch, a pitch excursion of 1.5 ST is a sufficient cue to prominence; Terken (1997) accordingly concludes that higher prominence ratings indeed appear to be proportional with the size of pitch excursions. It is noteworthy, though, that this proportional relation relates only to single pitch peaks, i.e. to situations in which there is just a single pitch accent, but it does not account for the relative perceptual prominence in sequences of pitch accents.

Indeed, it has been shown (e.g. in Rietveld and Gussenhoven, 1985, Gussenhoven and Rietveld, 1988, Ladd et al., 1994, Gussenhoven et al., 1997) that in a sequence of two pitch peaks, the second peak is perceived to be as prominent as the first peak even if the second pitch excursion is lower than the first one, which accounts for the claim that the nuclear accent in an intonational phrase is perceived as most prominent.

Thus, in sum, there is empirical support for the claim made in intonational phonology that accentuation plays a crucial role in determining the distribution of prominence in multi-word sequences in languages such as English or Dutch. By extension, this claim can be assumed to apply not only to phrases, but also to compounds. Gussenhoven (2004), for

instance, proposes an OT-based description of accentuation in English that he assumes to apply to both phrases and compounds. However, until recently there was only weak empirical support for the assumption that prominence patterns in compounds can also be attributed to accentuation differences.

Farnetani et al. (1988) present an early investigation of the acoustic correlates that distinguish left-prominent and right-prominent NN compounds in English. The authors show that duration, intensity, and maximum pitch show significant differences for the two prominence patterns. With respect to pitch, the difference between maximum pitch measurements in the first and second element is found to be only small in right-prominent compounds, but large in left-prominent compounds. Farnetani et al. interpret this in terms of pitch accent distribution and conclude that in the former case both elements are accented, while in left-prominent compounds, only the first element carries an accent. Later, Kunter (2011) explicitly tests this interpretation in an analysis of pitch, intensity, duration, and spectral balance in NN compounds. His analysis provides further robust support for the hypothesis that prominence patterns in NN compounds are distinguished by the presence or absence of a pitch accent on the second element, while the first element is always accented.

Turning to the four different types of triconstituent compounds described in Table 2, we would predict the following distribution of pitch accents. L/N1 compounds should have only one accent on the first constituent, while the other constituents should be unaccented. L/N2 should have an accent on N1 and on N2, because the complex constituent has the same distribution of accents as a right-prominent NN compound. The same pattern should be found in R/N2 compounds, but for slightly different reasons: the complex right constituent is a left-prominent NN compound, and therefore, N2 should be accented. At the same time, N1 should also be accented, because at the IC-level, R/N2 compounds are expected to be right-prominent, and in right-prominent compounds, the left constituent should be always accented.

Finally, R/N3 compounds should have an accent on all three constituents: N2 and N3 together form a right-prominent NN compound (with accents on both constituents), and N1 should be accented in R/N3 compounds due to the same reasons as in R/N2 compounds.

In sum, previous studies have shown that there is indeed a relation between pitch and prominence in compounds, and that this relation can be described in terms of accentuation. As the relation between prominence and pitch can be quantified, it offers an opportunity to test statistically whether different types of compounds are produced with different prominence patterns. Using acoustic data obtained from speech corpora and production experiments, this methodology has been employed in recent studies of NN compounds (Farnetani et al., 1988, Plag, 2006, Štekauer et al., 2007, Plag et al., 2008) and NNN compounds (Kösling and Plag, 2009, Kösling, 2011). However, the way that pitch is measured in all of these studies suffers from an identical weakness, namely that the researchers looked at rather crude abstractions of the pitch contour in the investigated compounds. For instance, Farnetani et al. (1988) measure the peak pitch in each compound constituent, Plag (2006) measures pitch at the mid-point of the syllables with primary stress, while Štekauer et al. (2007), Plag et al. (2008), Kösling and Plag (2009), and Kösling (2011) measure pitch averages in each constituent. Any information not captured by these measurements is ignored in the analysis, which is why these measurements run the risk of ignoring potentially systematic variation in the pitch contour that may provide perceptually important cues to prominence.

The movement direction of the pitch contour, i.e. whether pitch is falling or rising, may be one of these cues. Terken and Hermes (2000), who have investigated in how far pitch contributes to the perceived prominence of syllables, have found that a fall is perceived as more prominent than a rise of the same excursion size. Yet, all pitch measurements employed in the earlier literature on compound prominence are one-dimensional representations of the pitch contour in the sense that they describe only localized pitch events, and as such incapable

of registering any kind of pitch movement.[4] Furthermore, only parts of the pitch contour are considered in these studies, for instance the nucleus (Štekauer et al., 2007) or the rime (Plag et al., 2008) of the stressed syllables in each compound constituent. However, as Arvaniti et al. (1998) have shown, the tonal targets of bitonal pitch accents are not always aligned with the accented syllable itself. Thus, it might be possible that the second tonal target of L+H* accents does not fall within the measurement intervals used in the aforementioned studies. In such a case, the potentially prominence-lending high pitch of this accent type would be missed in the analysis.

These complications show that a methodology that considers the complete pitch contour in compounds may be more successful in detecting prominence differences than previous approaches which were restricted to portions of the contour. Accordingly, the present paper presents a statistical model of the pitch contour in the four different types of English triconstituent compounds described in the previous section. In this context, we understand the intonation contour of a compound as the combination of the contours of the involved pitch accents (plus any boundary tones and phrase accents). For the task of modeling this pitch contour, we will use generalized additive models, a flexible and powerful statistical technique that is especially suited for the problem at hand. This new method circumvents the limitations of previous studies, and at the same time, allows one to detect statistically significant

---

4   A notable exception can be found in Morrill (2012), who, in her study of cues to prominence in adjective-noun compounds, investigates both pitch minimum/maximum, the pitch change within a constituent, and the pitch change across constituent boundaries, thus aiming at a more complex view on pitch than the earlier studies. Yet, even this analysis simplifies the pitch contour of a compound, as secondary peaks or the shape of transitions between tonal targets cannot be described by these measures.

differences between the contours that may be interpreted as different prominence patterns. The details of the statistical analyses will be explained below.

METHODOLOGY

Target compounds

For the production experiment, a total number of 40 NNN compounds was devised, ten of each type described in Table 2. Thus, in order to test whether left-branching compounds have highest prominence on constituent N1 and right-branching compounds on constituent N2 if the complex constituent is left-prominent, we constructed 10 compounds of the L/N1 type and 10 compounds of the R/N2 type. Furthermore, in order to test whether a right-prominent embedded NN compound causes left-branching compounds to have highest prominence on constituent N2 we constructed 10 compounds of the L/N2 type. Finally, in order to test whether a right-prominent embedded NN compound causes right-branching compounds to have highest prominence on constituent N3, we constructed 10 compounds R/N3 type, respectively. The appendix lists all 40 target compounds.

The prominence pattern of the embedded NN compounds was controlled by means of various American English dictionaries (*Oxford Student's Dictionary of American English* (Hornby, 1983), *Longman Dictionary of American English* (Longman, 2002), *Longman Advanced Dictionary of American English* (Summers, 2000), *Oxford Advanced Learner Dictionary CD* (Hornby, 2000), and *The Oxford Spanish-English dictionary* (Carvajal and Horwood, 1996), as provided in Teschner and Whitley, 2004). Only those NN compounds whose prominence pattern was attested in at least one of these sources qualified as potential complex constituents for our triconstituent compounds. NN compounds that were attested in more than one source but with different prominence patterns were not considered as potential complex constituents either. Furthermore, compounds which have been reported in the

18

literature to exhibit variable prominence patterns (e.g. *ice cream* vs. *ice créam*, Bloomfield, 1933) were also not considered.

The left- and right-prominent two-word compounds were combined with another noun to form an NNN compound. If the third noun was attached to the right, this resulted in a left-branching NNN, if the third noun was attached to the left of the two-word compound, this resulted in a right-branching NNN. The choice of the third noun was constrained by several factors. First, all resulting NNN compounds had to be unambiguously left-branching or unambiguously right-branching in order to exclude the possibility that potential prominence variation would be caused by structurally ambiguous compounds. The problem with structural ambiguous compounds such as *kitchen towel rack* or *police dog trainer* is that they are argued to be prominent either on constituent N1 or on constituent N2, depending on their interpretation (e.g. Warren, 1978, Visch, 1999). For example, the compound *kitchen towel rack* may be interpreted as either 'a rack for kitchen towels' (left-branching) or 'a towel rack located in the kitchen' (right-branching), and the compound *police dog trainer* may be interpreted as either 'a trainer of police dogs' or 'a dog trainer working for the police'. In case of a right-branching interpretation, the LCPR predicts the compound to be prominent on constituent N2, whereas a left-branching interpretation will lead to prominence on constituent N1. Given this variable prominence behavior of ambiguous compounds and given the fact that it is difficult to control which interpretation the speaker may choose, compounds were constructed in such a way that they left as little room as possible for different interpretations of their branching structure.

In order to test the role of the embedded compound's prominence it is necessary to control for potentially intervening prominence relations at the IC-level. The study by Kösling and Plag (2009), for example, found left-branching compounds being prominent on constituent N3 and right-branching compounds prominent on constituent N1, which may be

the result of prominence determinants working at the IC-level. For example, it has been claimed that semantic categories trigger the occurrence of right prominence in NN compounds (e.g. Fudge, 1984, Ladd, 1984, Liberman and Sproat, 1992, Sproat, 1994, Olsen, 2000, 2001, Bell, 2008). Examples of these semantic subgroups are for instance the group of copulative compounds (e.g. *poet-translator*), compounds with temporal (*summer níght*) or locative modifiers (*Boston márathon*) or compounds with an ingredient and material as N1 (*silk tíe, chocolate púdding*) (cf. Kösling and Plag, 2009). Another such factor in compound prominence assignment is analogy (e.g. Schmerling, 1971, Liberman and Sproat, 1992, Spencer, 2003, Plag, 2006, 2010, Arndt-Lappe, 2011). Compounds that share the same left or right constituent tend to have the same prominence pattern. An example often mentioned in this context is that of *street* vs. *avenue* compounds: while *street* compounds are generally left prominent (*Óxford Street*), *avenue* compounds (*Madison Ávenue*) are consistently right-prominent (e.g. Plag, 2003).

In order to ensure that the above-mentioned factors would not influence the present analysis, we avoided all semantic relations and categories at the IC-level of the compounds that have been found to trigger rightward prominence in NN compounds in earlier empirical studies (e.g. Plag et al., 2007, Plag et al., 2008, Kunter, 2011). In addition to that, we also avoided lexical items as heads which are generally mentioned in the literature to trigger rightward prominence in NN compounds (e.g. *pie*, *avenue*). By applying these different criteria, the role of the embedded NN compound on the prominence pattern of the entire compound could be tested without any of these factors potentially intervening at the IC-level of the compounds. Other factors, such as frequency or the degree of lexicalization, were not taken into consideration.

For the production experiment, the 40 target compounds were embedded in short declarative sentences that were adapted to be compatible with the semantics of the respective

compounds. In order to avoid any potential effects that sentence position may have on the prominence pattern of compounds, all targets were placed in object position of the sentence, followed by a two-word adverbial. The adverbial was added to keep both the nuclear accent of the sentence and the sentence-final boundary tones out of the target compound, as these tonal events may be expected to be assigned to an utterance due to intonational regularities that are outside of the scope of prominence assignment in compounds. Furthermore, we avoided potential environments in which contrastive meaning was evoked, such as *He read about a coffee TABLE designer, not a coffee MUG designer*. As indicated in the example, contrastive emphasis can quite freely change the canonical prominence pattern of a given compound (cf. Bauer, 1998). The compounds were presented in the sentence in such a way that they always expressed new information, as to avoid any effect of information structure on the placement of tonal events in the target compounds (cf., for instance Hirschberg, 2002). Examples of carrier sentences are given in (3).

(4)   a.   He started *hay fever treatment* last week.

        b.   He was sentenced to *prisoner community service* last month.

        c.   He sold a *cotton candy maker* last month.

Procedure

For the experiment, 7 female and 6 male undergraduate and graduate students of the University of Toronto were recorded, aged between 18 and 27 years. All were monolingual native speakers of North American English. The majority grew up in the province of Ontario; 3 speakers originated from other Canadian provinces, and one speaker was originally from Massachusetts, and had been living in Toronto for two years at the time of the recordings. None of the participants, who were paid for their participation, reported any speaking or hearing disorder, and none of them was aware of the purpose of the experiment during the recordings.

21

Before each recording session, the participants were asked to read the instructions provided to them on a sheet of paper. In addition, the subjects were instructed to read aloud five training sentences that were used to adjust the loudness level of the recorder to the subjects' individual voices and to give the subjects an opportunity to familiarize themselves with the test situation and the process of reading out loud. The training sentences differed from the target sentences of the experiment in that they contained no compounds.

In addition to the 40 target sentences containing the constructed compounds, the same number of filler sentences was added in order to distract the participants from the actual purpose of the experiment. The filler sentences differed from the target sentences in that they contained no triconstituent compound. Instead, the object slot was filled by a long noun phrase. To control for order effects, the total set of 80 sentences were pseudo-randomized to create three lists, each list with a different order of the 80 sentences. Pseudo-randomization was done on the basis of blocks of 10 sentences and systematically varying the order of these blocks. The sentences were presented to the participants on four separate sheets of papers.

The recordings were taken in a sound proof booth at the University of Toronto using a Marantz PMD660 portable solid state audio recorder at a sampling rate of 44.1 kHz. An external microphone was placed a distance of about 50 cm in front of the participants. Depending on the subjects' reading speed, each recording took between 15 to 25 minutes. A researcher was present in the sound proof booth during each recording, and asked participants to repeat sentences during which mispronunciations, reading disfluencies, or noisy interferences (such as rustling papers or loud chair movements) occurred.

Acoustic pitch measurements

Using the phonetic software *Praat* (Boersma and Weenink, 2007), the beginning and end of each clause in which a target compound was embedded were manually annotated, as were the beginning and end of the compound itself, and the two boundaries between the three

constituents. The usual criteria for segmentation were applied (see, for instance, Ladefoged, 2003 and Turk et al., 2006) to decide on the location of these boundaries.

For each sound file containing one target clause, the *Praat* auto-correlation algorithm (Boersma, 1993) was used to extract a pitch object. Pitch floor and pitch ceiling were set according to the sex of the speakers: for female speakers, the pitch range was 100–500 Hz, and for male speakers, it was 75–300 Hz. All other parameters of the pitch tracker were left at their default values.

Any gap in the contour of the resulting pitch object was filled by interpolating linearly between the closest available pitch measurements to the left and the right of the gap. This approach assumes that the pitch contour of an utterance is, in principle, uninterrupted, and that the overall shape of a pitch contour in an utterance is not affected by breaks introduced for instance by unvoiced speech segments. This assumption is in line with other representations of pitch contours such as that in Hermes (2006), who argues that pitch contours are perceived as continuous, and that interruptions in the contour caused by unvoiced speech segments are not noticed by listeners as affecting the overall shape of the pitch contour.

By applying a smoother to the derived pitch contour, the transitions between interpolated and measured segments of the pitch contour receive a more natural shape, and localized misreadings of the pitch tracker algorithms are eradicated. In addition, the effect of microprosodic changes in the pitch contour, such as the lowered F0 that is frequently associated with voiced obstruents (see, for instance, Kingston and Diehl, 1994) is reduced, which is considered advantageous as the present study focuses on the macroscopic scale of the pitch contour that relates to prosodic prominence.
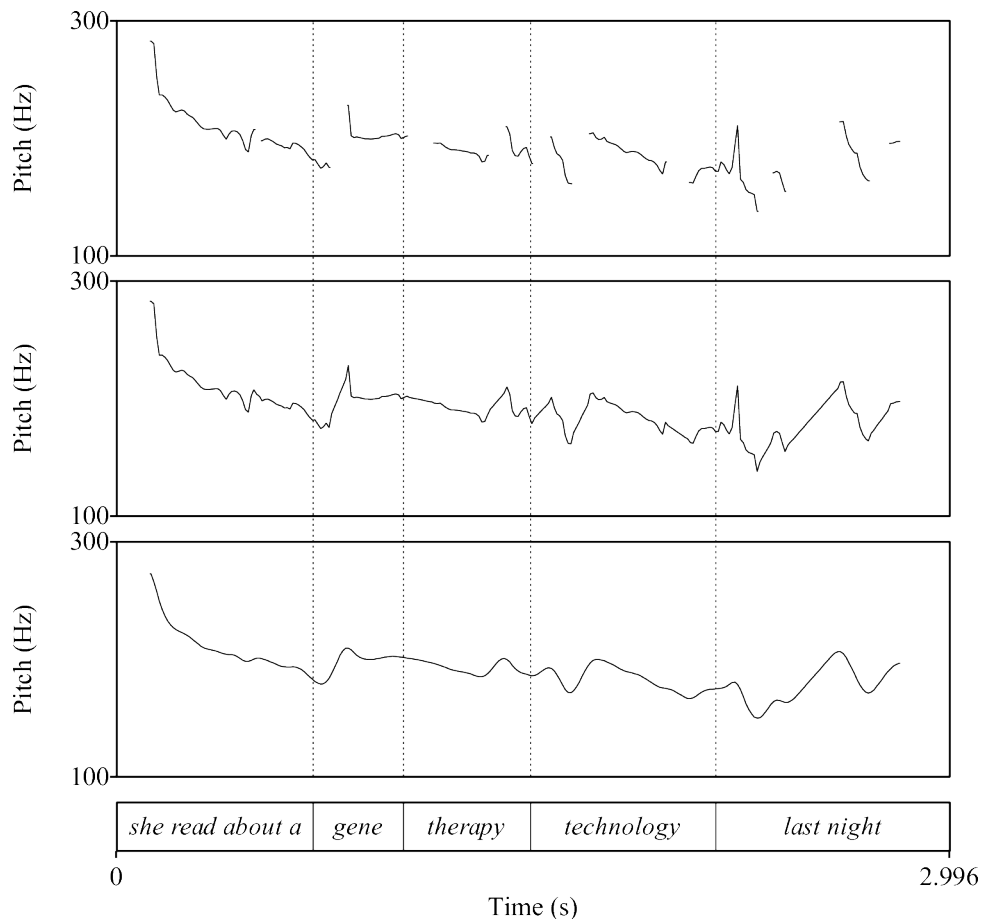
Figure 3: From top to bottom: unmodified, interpolated, and smoothed pitch contours.

The effects of these modifications are shown in Figure 3. The top panel of Figure 3 shows the unmodified pitch contour of the sentence *She read about a gene therapy technology last night* as produced by female speaker SP49. There are numerous gaps in the contour that can be associated with unvoiced segments; for instance, the two interruptions in *therapy* are linked to the voiceless segments /T/ and /p/. In the middle panel, these interruptions are filled by interpolating between the surrounding portions of the pitch contour. The bottom panel shows the contour after application of *Praat's* 'Smooth' function with a bandwidth of 10.0 Hz. The unnatural spikes that were created by the interpolation, for instance that in *gene*, are flattened. Likewise, small changes in the pitch contour such as the tiny spikes in the middle portion of *technology*, which presumably are too small to be noticed by listeners and thus are probably not affecting the perception of the overall intonation contour, are removed.

24

Only the interval of the pitch contour associated with the target compound (indicated by the left-most and right-most dotted line around *gene therapy technology* in Figure 3) is considered for the present analysis. This target interval is split up into 100 parts of equal length, and the pitch is obtained from the smoothed contour. To account for the non-linear nature of the Hertz scale, all measurements are transformed to the linear semitone (ST) scale relative to 1 Hz, by using the conventional formula . Every pitch measurement is stored together with the number of that part of the interval at which it was taken (ranging from 0 to 99). This number will be used as the time variable in the different smooth terms.

Very rarely, the automatic pitch tracker failed to obtain pitch measurements for parts of the compound, even after interpolation and smoothing. This could occur, for instance, if the end of the respective target sentence was produced with creaky voice. Non-modal phonation has been shown to occur with very irregular F0 pulses (see, for instance, Blomgren et al. (1998) for an analysis of the properties of non-modal phonation), which is why automatic pitch trackers frequently fail to detect a periodic signal. The interpolation process described above attempts to compensate for this failure by supplying linear approximations of the pitch contour during problematic parts of the recordings, but for 426 out of the 1340×100 measurement points (0.8 percent), no pitch information could be obtained. Thus, the total number of observations was 51,574.

What is notable here is that an equal number of measurements is taken for each compound token, irrespective of any duration differences either between different compound types, or between tokens of the same compound type. For example, the compound type *gene therapy technology* shown in Figure 3 has, on average across all speakers in the experiment, a longer duration than the compound type *day care center* (the average durations are 1.413 s and 0.784 s, respectively). Speaking rate differences between individuals may lead to duration

differences between tokens of the same type: for instance, the duration of the token in Figure 3 is 1.434 s, but observed duration of this type ranged from 1.277 s to 1.599 s.

The underlying assumption for using the same number of measurements for compounds of various lengths is that the pitch contour of different compound types is predominantly determined by their branching type. For instance, the LCPR will associate specific tonal events with specific constituents of the NNN compound, but this association is expected to be indifferent to the length of each constituent, or the time a speaker requires to produce the compound. Thus, by taking the same number of pitch measurements from each compound, irrespective of the absolute duration of the item, durational differences are normalized, resulting in a time-averaged series of 100 measurements from each item.[5] Yet, in order to relate the subsequently obtained model of the pitch contour to the linguistic structure of the compound types used as stimuli, each pitch measurement was stored together with information on whether the measurement was taken from the left, middle, or right constituent of the compound.

Generalized additive models

Generalized additive models GAMs (Wood, 2006, 2011) are an extension of the generalized linear model that provides flexible tools for modeling nonlinearities. GAMs take the form

$$y = X\beta + f_i(x_1, x_2, \ldots) + \ldots + \varepsilon,$$

---

5   Note that this normalization process is applied to each token of a compound separately, but not on a by-constituent basis. This means that any durational differences that are caused by differences in the distribution of pitch accents in a particular type of compound will be reflected by alignment differences in the modeled pitch contour for that type. Thanks to one of the reviewers of this paper for pointing this out.

where y is the response variable, $X\beta$ is a linear predictor, and f_i are smooth functions of the covariates x_i. The central concept motivating GAMs is that the dependence of the response variable is decomposed into two submodels. The first submodel, represented above by $X\beta$, is a parametric model exactly as in classical linear regression, with β representing a vector of coefficients, one for each of the predictors in the X matrix. The second submodel is non-parametric, and provides one or more smooth functions for one, or several predictor variables. A smooth function for a single predictor variable makes it possible to model wiggly lines in the plane spanned by predictor and response variable. A smooth function for two or more predictors defines a wiggly surface in the (hyper)space spanned by the predictor variables and the response variable. By combining the two submodels, it is often possible to obtain a more parsimonious and insightful model compared to a fully general model stating that the response variable is some smooth function of all covariates jointly. In other words, the aim of generalized additive modeling is to provide the simplest possible yet adequate model for the data, allowing nonlinearities where necessary, but using standard linear terms wherever possible.

In what follows, we make use of the `mgcv` package (Wood, 2012) for R (R Development Core Team, 2011). For smooths of single predictors, we used restricted cubic splines, henceforth referred to in tables with the abbreviation `s()`. For smooths in more than one dimension, we made use of tensor products (abbreviated in tables to `te()`). Tensor product smooths estimate a wiggly surface (or hypersurface) from two (or more) basis functions. The smoothness of the basis functions (typically cubic splines) determines the amount of wiggliness of the regression surface. The mathematics of tensor products are complex, an informal way of describing a tensor given its marginal basis functions is that the tensor surface is constructed in such a way that in each dimension it is as faithful as possible to the shape of its marginal function in that dimension. Random effect factors (such as subject

and item) can be brought into the model as well, leading to a generalized additive mixed model. The problem of estimating the appropriate smoothing parameters can be solved in various ways, we have used the default, generalized cross-validation. Similar results were obtained with relativized maximum likelihood. Importantly, the optimal degree of smoothness is not determined by the user, but is estimated as part of the model fitting, along with the parametric coefficients and random intercepts for subject and item.

Significance of parametric terms is evaluated by means of the usual *t*-tests, and/or by means of analysis of deviance tests. Significance of nonparametric terms is evaluated by means of the Bayesian p-values recommended by Wood for smooths. The evaluation of significance for GAMs is more difficult than for standard linear models, hence, it should be kept in mind that probability values are approximate.

Of central interest to the following analysis is how pitch develops over time for the four BRANCHING conditions (L/N1, L/N2, R/N2, R/N3). We therefore included in our model a restricted cubic spline for time (*t*) which we allowed to be optimized for each branching condition separately. A series of model comparisons (see below for further details) clarified that models with less than four branching conditions (obtained, for instance, by collapsing L/N2 and R/N2), provided significantly inferior fits. BRANCHING was also included in the parametric submodel in order to test for (time-independent) differences between the intercepts for the four branching conditions.

Various other predictors were included as controls. First, random intercepts were included for the random-effect predictors SPEAKER and ITEM (i.e. compound). Second, the SEX of the speaker was included, in order to control for potential differences between female and male speakers. As these differences are expected primarily over time, we included separate temporal smooths for females and males.

Third, POSITION was included as numeric predictor, as different average pitch measurements may be expected for the left, middle and right constituent (encoded as 0, 1, and 2, respectively) due to the overall declination of pitch.[6]

Fourth, the length of a constituent (in phonemes) was included, to control for possible consequences of LENGTH for the planning of a constituent's pitch declination. We also considered an interaction of constituent length by position, to allow for systematic changes in pitch declination as a function of position in the compound. An alternative measure that might be considered is constituent length in milliseconds. In the present study, we have not added such an additional measure to our set of predictors, for two reasons. First, a durational measure would be highly correlated with length in phonemes, and increase multicollinearity. More importantly, as our measure of time in the analysis is a relative one, normalized to the same interval for all compounds, measures of constituent duration would require further normalization with respect to the durations of the other constituents. We leave this issue to future research.

Finally, the lemma frequency of each constituent was also included as a predictor, as frequency of occurrence has been found to have an influence acoustic durations Bell et al. (2003). We therefore sought to rule out the possibility that the effects of BRANCHING might be

---

6   Pitch declination, i.e. the tendency of speakers to start an utterance with a comparatively high pitch that decreases throughout the utterance is a fairly well-researched area (see, for instance, Strik and Boves, 1995 or Gussenhoven, 2004, ch. 6). Therefore, all things being equal, pitch measurements taken from a compound that occurs early in the beginning of a sentence will be higher than if the compound occurs late in a sentence. Inclusion of information that a pitch measurement was taken from the left, middle, or right constituent will help to reduce such prominence-independent pitch variation.

confounded with an effect of FREQUENCY on pitch. For each constituent in our test items we obtained its lemma frequency from the British National Corpus, and this frequency was added as a covariate to the pitch measurements taken for this constituent. Thus we could explore whether the effect of frequency might vary over time by means of a tensor product smooth.

RESULTS

A first mixed-effects GAM was fitted with the 51,574 pitch measurements as the dependent variable. However, as the model residuals revealed, there was one male speaker among the participants for whom the average difference between the predicted and the observed pitch measurements was much larger than for any other speaker. Subsequent auditory inspections of the recordings from this speaker showed that he had used intonation in a strongly pronounced way that was strikingly unusual in comparison to recordings from other speakers.[7] Obviously, his use of intonation was different from the speaking mode of the other speakers. Large model residuals for this speaker implied that the GAM had problems accounting for the variance introduced by this particular speaker. It was therefore decided to discard the pitch measurements obtained from this speaker, and the model was refitted with the data from the remaining 12 speakers. The total number of observations was thus reduced to 47,574 data points.

A comparison of a sequence of generalized additive mixed models is summarized in Table 3. Each new row compares two nested models, where the second model has one predictor or interaction term more than the first, and evaluates whether there is a reduction in deviance, and if so, whether this reduction in deviance is significant given the number of parameters invested in obtaining this reduction. Significance is evaluated with the help of an F-test. The

7   Incidentally, the speaker had remarked during pre-experimental conversation that he was an actor.

baseline model (not shown in the table), included an intercept (grand average) only. The first row, labeled 'speaker', indicates that including random intercepts for speaker (for which the GAM invests 11 degrees of freedom) reduces the AIC by 92699.33.[8] The *F*-test for this predictor comes with an extremely small p-value, indicating that inclusion of speaker as a predictor leads to a significantly better fit of the model to the data. Table 3 clarifies that for most predictors, inclusion is well motivated, because there is a statistically significant reduction in AIC. Note that although BRANCHING by itself has weak support, positing different smooths for the four branching conditions find strong support (reduction in AIC 1871.06, *p*<0.0001). In other words, even though there are no significant differences in the pitch for the four branching conditions at *t*=0 (i.e., at the intercept), significant differences between the conditions develop as time progresses.

Table 3

Sequential model comparison.

predictor	*Number of parameters*	reduction AIC	*F*	*p*
speaker	11.0	92699.33	34552.19	<0.0001
item	38.3	1989.99	70.07	<0.0001
LENGTH	1.1	79.49	98.45	<0.0001
POSITION	1.1	8741.81	9533.90	<0.0001
LENGTH:POSITION	1.0	10.56	13.23	0.0003

8   The AIC (Akaike Information Criterion, cf. Akaike, 1974) allows model comparison by taking into account both the maximum likelihood and the number of degrees of freedom that is invested by each model. In practice, the optimal model for a data set is the one that minimizes the AIC.

BRANCHING[9]		-0.03		
s(t, BRANCHING)	34.6	1869.75	57.97	<0.0001
SEX		0.00		
s(t, SEX)	6.8	150.15	24.48	<0.0001
FREQUENCY	1.0	12.72	14.75	0.0001
te(t, FREQUENCY)	19.1	498.52	28.40	<0.0001

Table 4 presents the coefficients and associated statistics for the parametric submodel, as well as the *F* tests for the smooth terms in the non-parametric submodel.

The R/N3 condition seems to have, on average, a pitch that is half a semitone higher than that of the L/N1 condition, which acts as the reference level in the model and is expressed in the intercept term. But in the absence of a main effect of BRANCHING, this difference is best ignored, at least for the present data. However, the number of estimated degrees of freedom (edf) invested in the four smooth terms by BRANCHING ranges from 6.13 to 8.3, and all are well supported by the *F* tests. As the estimated degrees of freedom are far above 1, the pitch contours cannot be modeled precisely with just a linear term, nor with the combination of a linear and a quadratic term (which would require two degrees of freedom).

Table 4

Summary of final mixed-effects general additive model (R^2(adj) = 0.893, *n*=47,574)

Parametric coefficients				
	Estimate	Std. Error	*t*	*p*
intercept	86.08	3.27	26.32	<0.0001

[9] The addition of BRANCHING and SEX to the model specification does not lead to an improved model. In fact, for BRANCHING, the model becomes slightly worse, as indicated by the negative value for the reduction in AIC. Closer inspection of the relevant models reveals that BRANCHING does not have any added value above what is already captured by the by-item random intercepts. Likewise, all variation captured by SEX is entailed by the by-speaker random intercepts.

length	0.09	0.02	4.71	<0.0001
position	0.22	0.06	3.61	0.0003
branching L/N2	1.72	0.73	2.35	0.0187
branching R/N2	1.32	0.75	1.77	0.0775
branching R/N3	3.11	0.75	4.17	<0.0001
sex male	-9.86	1.78	-5.53	<0.0001
length:position	-0.06	0.01	-7.56	<0.0001

Approximate significance of smooth terms

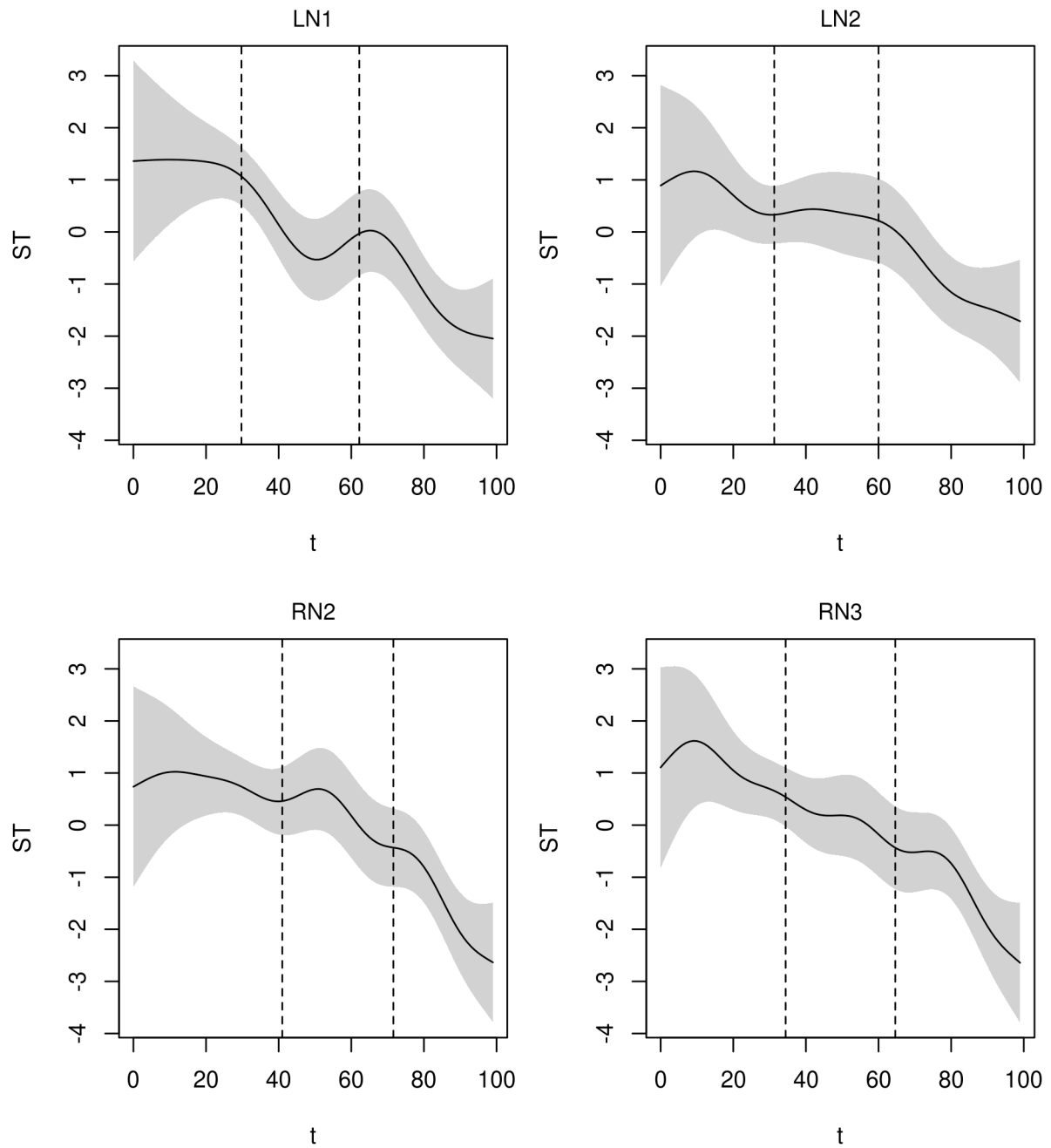|	*edf*	Ref.*df*	*F*	*p*
speaker	10.00	10.00	9238.27	<0.0001
item	35.36	36.00	106.28	<0.0001
s(t):L/N1	8.29	8.71	18.28	<0.0001
s(t):L/N2	7.63	8.31	9.53	<0.0001
s(t):R/N2	8.32	8.73	9.41	<0.0001
s(t):R/N3	8.37	8.75	10.20	<0.0001
s(t):sex female	4.96	6.01	24.78	<0.0001
s(t):sex male	0.00	0.00	0.03	0.996
te(t,Freq)	19.96	20.80	25.43	<0.0001

Figure 4: Estimated pitch contours by branching type.

Figure 4 illustrates the shapes of the pitch contours for each branching type that are captured by each of the four smooths. Each panel shows the smooth term for one of the four branching types. The horizontal axis is the normalized time scale; the vertical axis indicates pitch measurements in semitones relative to 1 Hz. The curve thus illustrates the estimated changes

of the pitch contour in relation to the time dimension. Each curve presents the *partial* effect of the smooth, which is why each curve is centered around zero.

In order to be able to relate the shape of the estimated pitch contours to the linguistic components of the target compounds, average constituent boundaries are displayed by the dashed vertical lines. Due to different lengths of the constituents that were used in the construction of the target compounds, these boundaries are not the same for all four types. For example, the second constituents of L/N2 compounds (indicated by the interval between the two dotted lines in the top right panel) are, on average, clearly shorter than the first constituents of R/N2 compounds (bottom left panel). While the average constituent boundaries do not, of course, correspond to the actual boundaries of any specific compound, they are nevertheless useful in identifying the alignment of the pitch contour with the three constituents of the different compound types, as they reflect the average point of time at which the pitch contour moves from one constituent to the next.

Across all four panels, the expected pitch declination is clearly visible. In every smooth, pitch starts high and shows a general downward trend toward the end of the compound. The high start is an indication of the presence of a pitch accent. For L/N1 and R/N2 compounds, there is a nearly linear decline in pitch up to the beginning of the second constituent. For L/N2 and R/N3 compounds, the pitch function is nonlinear, with an initial nearly steady state followed by a steep decline that levels off before the second constituent. This suggests that the high tone in the first constituent of L/N2 and R/N3 compounds occurs later in the first constituents.

The contour for L/N1 compounds (top-left panel) drop further by a semitone during the second constituent. This implies that for the vast majority of L/N1 compounds, there is no second accent with a high tone on N2. However, the contour shows a small rise by about 0.25 ST at the boundary between N2 and N3. This secondary peak appears to be too small to be

interpreted as a further high tone (recall that pitch excursions of 1.5 ST and more have been found by Rietveld and Gussenhoven, 1985 as a sufficient cue to prominence; the peak here is clearly smaller), but as it seems to be aligned with the right edge of the complex constituent, we may interpret this as a phrase accent that speakers use to indicate the internal structure of this type of triconstituent compounds.

There is no similarly steep drop of the pitch contour during the second constituent in L/N2 and R/N2 compounds (top-right and bottom-left panel, respectively). Instead, there is a clear rise in the contour of N2, which suggests that there is a high target for these compound types. In other words, L/N2 and R/N2 compounds are accented both on N1 and on N2.

There are further differences between the pitch contour of L/N2 and R/N2 with respect to the boundary between second and third constituent, and within the third constituent. In L/N2 compounds, the pitch appears to spread out longer toward the boundary than in R/N2 compounds, and drops steeply at the beginning of the third constituent. In R/N2, the pitch drop occurs earlier, before the constituent boundary. This may be an indication that L/N2 compounds, just as L/N1 compounds, have a phrase accent at the boundary between N2 and N3, which prevents an immediate drop of the pitch contour after the pitch peak in N2.[10]

The remaining compound type R/N3 (bottom-right panel) reveals a similar hint of a pitch target on the third constituent, that is located further towards the end of the word. In fact, the pitch function for these compounds is suggestive of three falls, one in each constituent, followed by a plateau. This would suggest three accents, with clearly decreasing peak heights. Apparently, speakers prefer downstepped pitch accents on N2 and N3 in this type of compound. Using conventional notations, the accentuation pattern of R/N3 compounds thus

---

10   Thanks to one of the reviewer for suggesting this interpretation of the different pitch contours of L/N2 and R/N2.

may be described as H* !H* !H*. The finding that there is a pitch accent on every constituent is to be expected if we assume that right-prominent NN compounds also have pitch accents on both constituents (see, for example, Kunter, 2011, ch. 5 for evidence and detailed discussion).

Thus, in sum, the smooths produced by the mixed-effects GAM for the four different compound types reveal obvious differences of the pitch contour, and by extension, for the prominence patterns of these types. L/N1 compounds tend to have a single pitch accent on N1, which may be expected to be the most prominent constituent of the overall compound. In perhaps L/N2 and clearly in R/N2 compounds, there is also a pitch accent on the first constituent, but compounds of these types tend to have another pitch accent on the second constituent, as predicted by the EPH. Finally, R/N3 compounds frequently have accents on the third constituent (in addition to any preceding pitch accent).

The reliability and accuracy of these pitch contours for the four types of compounds depends in part on how well other factors have been controlled for. A first question that should be addressed is whether these contours are due to a conflation of different patterns for females and males. A model incorporating a main effect of sex, with by-speaker random intercepts, does not reveal a significant differences between males and females: any differences are absorbed by the random intercepts for speaker. However, a smooth for time in interaction with sex does increase the goodness of fit (see Table 3). For males, there is no significant trend, but females show a small effect (1 semitone) to have lower pitch near the end of the compound. In the present model, this effect for females is the assumed to be the same for all compound types. When this assumption is relaxed, by considering a smooth for each individual combination of sex and branching type, an improved fit is obtained at the cost of some 30 additional degrees of freedom. The general pattern of results remains similar to that shown in Figure 4. For females, the pitch contours tend to be slightly more wiggly, but

key inflections in these contours are retained. We therefore conclude that although differences between females and males were detectable, they do not invalidate our conclusions.

A second control variable is position, a numeric variable specifying which of the three constituents is under consideration at a given point in time. Position entered into a significant interaction with length (our third control variable), such that at the first constituent (position 1) a greater length (in phonemes) elicited a slightly higher pitch. At the third constituent, this effect reversed such that long words elicited a substantially reduced pitch. This suggests that long third constituents afford speakers the opportunity of a prolonged continuation of a downward pitch movement. Importantly, this effect of length is part of our model, and hence the pitch movements for the branching types cannot be due to a confound with word lengths. In other words, by incorporating these two control variables, our model reduces the potentially obscuring effect of position and length on the linguistically interesting factors affecting the shape of the pitch contours.

Our third control variable, a word's frequency of occurrence, also emerged as significant. As a simple main effect, it emerged with a significantly negative slope, indicating that higher-frequency words had a lower pitch. Further inspection with a tensor product suggests that this effect of frequency is restricted mainly to the first constituent. Just as higher-frequency words tend to have shorter acoustic durations, due to their relatively high degree of informational redundancy (Bell et al., 2003, Gahl and Garnsey, 2004), higher-frequency words appear to have somewhat attenuated initial pitch targets. This finding fits well with the smooth signal redundancy hypothesis (Aylett and Turk, 2004). Crucial for the present study is that frequency of occurrence, which was not controlled for beforehand in the selection of materials, is under statistical control through a tensor product in our generalized additive model.

Finally, there is a potential confound in our data, namely the position of the lexically stressed syllable in a given constituent. For example, if a constituent has two syllables and is stressed on the first syllable, we expect an earlier pitch peak than if we had a disyllabic constituent that is stressed on the second syllable. These differences in the temporal location of the main stressed syllable of a constituent were not controlled for in the present analysis, since we are using data that were collected with a different methodology in mind, in which differences in the temporal location of lexical stresses within constituents would not have mattered due to the principled insensitivity of the measurements concerning this variable. Furthermore, the distribution of lexical stresses is actually quite uniform across the test items. For first constituent, we have mostly disyllabic words with initial stress. There is only one polysyllabic word that is not stressed initially (*security*). Similarly, there is only one polysyllabic second constituent that is not stressed on the first syllable from the left, *community*. For the first two constituents we can safely assume that the pitch contour is only very weakly affected by the variability of lexical stress placement. Only in the third constituent do we find more variability. Eight out of 40 constituent are not stressed on the initial syllable (e.g. *designer, removal, delivery*. This means that the contours for the third constituent may be interpreted in such a way that for about three quarters of the items the peaks are actually slightly more to the left, and that for the minor of eight items the peaks are actually later than it appears in the plots. With regard to the decisive question whether the contours give evidence for the presence of a pitch accent, these complications do not seem to play a decisive role, and with only so few constituents that did not carry their lexical stress on the initial syllable, there appears to be little need for including further statistical controls to counter this potentially confounding influence of stress position.

DISCUSSION AND CONCLUSION

The shape of the pitch contours and the corresponding statistical models invite an interpretation according to which branching direction is not the sole determinant of NNN prominence. To the contrary, the analysis of the pitch contours of the pertinent compounds in the experiment has shown that the prominence pattern of the embedded compound has to be taken into account. Left-branching compounds with a left-prominent embedded compound behave in accordance with the LCPR, and so do right-branching compounds with a left-prominent embedded compound. However, if the embedded compound is right-prominent, the predictions of the LCPR fail. The LCPR predicts the same prominence pattern as before, but the empirical facts run counter to that expectation. Left-branching compounds with an embedded right-prominent compound have an accent on N2, and right-branching compounds with a right-prominent embedded compound have an accent on N3. It was also shown that all triconstituent compounds have an accent on the first constituent. Overall, the predictions of the EPH turned out to be in accordance with the statistical analysis of the pitch contours. The pitch contours strongly suggest that those constituents which were predicted to be accented do indeed appear to receive a pitch accent.

Based on these findings, an accent-based classification of the different compound types is given in Table 5.

Table 5

Typology of prominence

Branching direction	Embedded prominence	Accents	Example with highest prominence indicated
left	left	Ń N N	[*háy fever*] *treatment*
left	right	Ń Ń N	[*science fiction*] *book*
right	left	Ń Ń N	*business* [*crédit card*]
right	right	Ń 'N Ń	*family* [*Christmas dínner*]

Our findings are in line with other empirical studies that have found problems with the

predictions of the LCPR. Giegerich (2009) gives a number of examples that show effects in accordance with the EPH, while the extent of the phenomenon remains open, he concludes that the LCPR is wrong. Similarly, Kösling and Plag (2009) find violations of the LCPR in their speech corpus data that suggest an effect of embedded prominence, and these authors state that these patterns merit "further empirical testing with more carefully controlled data" (p. 229). The present study provides such data and strong empirical support for the idea that embedded prominence, and not branching, determines the prominence of triconstituent noun compounds. The present paper has also shown that mixed effects generalized additive models can be fruitfully employed to model prominence in triconstituent compounds. Very similar results, yet with a lower level of descriptive detail, have been found in Kösling (2011), who analysed the same data set using a different methodology that has been established in many previous studies (e.g. Plag 2006, Plag et al. 2008, Kösling and Plag 2009, Kunter 2011). Furthermore Kunter and Plag (2007) have shown that this methodology is capable of approximating the perception of prominence patterns in compounds to a large degree. The compatibility of both approaches shows that the models used in the present paper provide a valid way of describing pitch contours, and the pitch contours as predicted by the model lend themselves to straightforward interpretations in terms of autosegmental-metrical phonology, as the theoretically predicted pitch accents can indeed be traced in the contours. Future research that links these types of models to perception of prominence by speakers will show the validity of these conclusions.

A natural extension of the present research program would deal with the question of what happens at the IC-level. Recall that in the present experiment, with left-branching compounds, IC-level prominence was carefully controlled for, and only IC-left-prominent NNNs were produced by the participants. Given that rightward prominence can also occur at the IC level (contrary to the assumptions of the LCPR), we would expect that this could also

lead to prominence patterns that are not in accordance with the branching direction-based predictions of the LCPR. Preliminary acoustic evidence presented by Kösling and Plag (2009) as well as Giegerich's (2009) small selection of pertinent words point in this direction, but a systematic study is called for that investigates these patterns in more detail. A reliable method for such an investigation is now available.

The usefulness of the type of statistical models employed in this study, however, goes beyond the analysis of compounds. Generalized additive models are an accessible way of representing pitch contours in clearly defined data sets, and they allow a principled, statistical comparison between the different contours, which makes them a very suitable tool for the empirical investigation of intonation patterns in larger utterances. One case in point may be the link between intonation contours and the pragmatic meaning of utterances. Hirschberg (2004) associates declarative sentences and *wh*-questions in standard American English with an H* L-L% sequence, and *yes-no*-questions with L* H-H%. These mappings of intonation patterns on specific sentence types appears to be mostly uncontroversial. However, Hirschberg notes that there may be more links between specific pitch contours and certain pragmatic meanings of the utterance, but she considers these links to be "both more controversial and more elusive" (Hirschberg, 2004, p. 533). For instance, it has been proposed that downstepped contours such as H* !H* !H* L-L% are frequently used in sentences that introduce a new topic in a didactic context, but to our knowledge, this proposal has never been investigated. The models described in this article might be employed to examine empirically whether such an effect of context is indeed traceable in different intonational contours.

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19(6), 716-723.

Arndt-Lappe, S. (2011). Towards an exemplar-based model of stress in English noun-noun compounds. *Journal of Linguistics* 47(3), 549-585.

Arvaniti, A., Ladd, D. R., and Mennen, I. (1998). Stability of tonal alignment. The case of Greek prenuclear accents. *Journal of Phonetics*, 26(1), 3–25.

Aylett, M. and Turk, A. (2004). The smooth signal redundancy hypothesis: a functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, 47, 31–56.

Bauer, L. (1998). When is a sequence of two nouns a compound in English? *English Language and Linguistics*, 2(1), 65–86.

Beckman, M. E. and Edwards, J. (1994). Articulatory evidence for differentiating stress categories. In Keating, P. A. (Ed.), *Phonological structure and phonetic form: papers in laboratory phonology III* (pp. 7–33). Cambridge, UK: Cambridge University Press.

Beckman, M. E. and Pierrehumbert, J. B. (1986). Intonational structure in English and Japanese. *Phonology Yearbook*, 3, 255–309.

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., and Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America*, 113, 1001–1024.

Bell, M. (2008). Noun noun constructions and the assignment of stress. Paper presented at the 1st Conference of the International Society for the Linguistics of English, Freiburg, 8-11 October.

Bell, M. and Plag, I. (2012). Informativeness is a determinant of compound stress in English. *Journal of Linguistics*.

Berg, T. (2009). *Structure in language. A dynamic perspective*. New York: Routledge.

Blomgren, M., Chen, Y., Ng, M. L., and Gilbert, H. R. (1998). Acoustic, aerodynamic, physiologic, and perceptual properties of modal and vocal fry registers. *Journal of the Acoustical Society of America*, 103(5), 2649–2658.

Bloomfield, L. (1933). *Language*. New York: Holt.

Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proceedings of the Institute of Phonetic Sciences*, 17, 97–110. University of Amsterdam.

Boersma, P. and Weenink, D. (2007). Praat: doing phonetics by computer (version 4.5.11). Computer program. Retrieved 30 January 2007, from http://www.praat.org/.

Carvajal, C. S. and Horwood, J. (1996). *The Oxford Spanish-English dictionary: New international edition*. Oxford, UK: Oxford University Press.

Dilley, L. C., Ladd, D. R., and Schepman, A. (2005). Alignment of L and H in bitonal pitch accents: testing two hypotheses. *Journal of Phonetics* 33(1), 115–119.

Farnetani, E., Torsello, C. T., and Cosi, P. (1988). English compound versus non-compound noun phrases in discourse: an acoustic and perceptual study. *Language and Speech*, 31(2), 157–180.

Fudge, E. (1984). *English word-stress*. London, UK: George Allen & Unwin.

Gahl, S. and Garnsey, S. (2004). Knowledge of grammar, knowledge of usage: syntactic probabilities affect pronunciation variation. *Language*, 80(4), 748–774.

Giegerich, H. J. (2009). The English compound stress myth. *Word Structure*, 2(1), 1–17.

Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge, UK: Cambridge University Press.

Gussenhoven, C., Repp, B. H., Rietveld, A. C. M., Rump, H. H., and Terken, J. (1997). The perceptual prominence of fundamental frequency peaks. *Journal of the Acoustical Society of America*, 102(5), 3009–3022.

Gussenhoven, C. and Rietveld, A. C. M. (1988). Fundamental frequency declination in Dutch: testing three hypotheses. *Journal of Phonetics*, 16(3), 355–369.

Hermes, D. J. (2006). Stylization of pitch contours. In Sudhoff, S., Lenertová, D., Meyer, R., Pappert, S., Augurzky, P., Mleinek, I., Richter, N., and Schließer, J. (Eds.), *Methods in empirical prosody research* (pp. 29–61). Berlin, Germany: de Gruyter.

Hirschberg, J. (2002). Communication and prosody: functional aspects of prosody. *Speech Communication*, 36(1), 31–43.

Hirschberg, J. (2004). Pragmatics and intonation. In Horn, L. R. (Ed.), *Handbook of pragmatics* (pp. 515–537). Malden, MA: Blackwell.

Hornby, A. S. (Ed.) (1983). *Oxford Student's Dictionary of American English*. Oxford, UK: Oxford University Press.

Hornby, A. S. (2000). *Oxford Advanced Learner's dictionary, CD-Rom*. Oxford, UK: Oxford University Press.

Kingston, J. and Diehl, R. L. (1994). Phonetic knowledge. *Language*, 70(3), 419–454.

Kochanski, G., Grabe, E., Coleman, J. and Rosner, B. (2005). Loudness predicts prominence: fundamental frequency lends little. *Journal of the Acoustical Society of America* 118(2), 1038–1054.

Kösling, K. (2011). *Prominence assignment in English triconstituent compounds.* PhD thesis, Universität Siegen.

Kösling, K. and Plag, I. (2009). Does branching direction determine prominence assignment? *Corpus Linguistics and Linguistic Theory*, 5(2), 201–239.

Kunter, G. (2011). *Compound stress in English. The phonetics and phonology of prosodic prominence*. Berlin, Germany: de Gruyter.

Kunter, G. and Plag, I. (2007). What is compound stress? In *Proceedings of the 16th International Congress of Phonetic Sciences, Saarbrücken* (pp. 1005–1008).

Kvam, A. M. (1990). Three-part noun combinations in English, composition – meaning – stress. *English Studies*, 71, 152–160.

Ladd, D. R. (1984). English compound stress. In Gibbon, D. and Richter, H. (Eds.), *Intonation, accent and rhythm* (pp. 253–266). Berlin, Germany: de Gruyter.

Ladd, D. R. (1996). *Intonational phonology*. Cambridge, UK: Cambridge University Press.

Ladd, D. R., Verhoeven, J., and Jacobs, K. (1994). Influence of adjacent pitch accents on each other's perceived prominence: two contradictory effects. *Journal of Phonetics*, 22(1), 87–99.

Ladefoged, P. (2003). *Phonetic data analysis*. Malden, MA: Blackwell.

Liberman, M. and Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, 8, 249 –336.

Liberman, M. and Sproat, R. (1992). The stress and structure of modified noun phrases in English. In Sag, I. A. and Szabolcsi, A. (Eds.), *Lexical matters* (pp. 131–181). Stanford, CA: Center for the Study of Language and Information.

Longman (2002). *Longman dictionary of American English*. White Plains, NY: Longman.

Morrill, T. (2012). Acoustic correlates of stress in English adjective-noun compounds. *Language and Speech*, 55(2), 167–201.

Olsen, S. (2000). Compounding and stress in English: a closer look at the boundary between morphology and syntax. *Linguistische Berichte*, 181, 55–69.

Olsen, S. (2001). Copulative compounds: A closer look at the interface between syntax and morphology. In Booij, G. & J. v. Marle. (Eds.), *Yearbook of morphology 2000* (pp. 279–320). Dordrecht, Netherlands: Kluwer.

Ostendorf, M., Price, P., and Shattuck-Hufnagel, S. (1996). *Boston University Radio Speech Corpus*. Philadelphia: Linguistic Data Consortium

Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation*. PhD thesis, Massachusetts Institute of Technology.

Plag, I. (2003). *Word-Formation in English Linguistics*. Cambridge, UK: Cambridge University Press.

Plag, I. (2006). The variability of compound stress in English: structural, semantic, and analogical factors. *English Language and Linguistics*, 10(1), 143–172.

Plag, I. (2010). Compound stress assignment by analogy: The constituent family bias. *Zeitschrift für Sprachwissenschaft*, 29(2), 243–282.

Plag, I., Kunter, G., and Lappe, S. (2007). Testing hypotheses about compound stress assignment in English: a corpus-based investigation. *Corpus Linguistics and Linguistic Theory*, 3(2), 199–233.

Plag, I., Kunter, G., Lappe, S., and Braun, M. (2008). The role of semantics, argument structure, and lexicalization in compound stress assignment in English. *Language*, 84(4), 760–794.

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

Rietveld, A. C. M. and Gussenhoven, C. (1985). On the relation between pitch excursion size and prominence. *Journal of Phonetics*, 13(3), 299–308.

Schmerling, S. F. (1971). A stress mess. *Studies in the Linguistic Sciences*, 1, 52–66.

Spencer, A. (2003). Does English have productive compounding? In Booij, G. E., DeCesaris, J., Ralli, A., and Scalise, S. (Eds.), *Topics in Morphology. Selected papers from the 3rd Mediterranean morphology meeting* (pp. 329–341). Barcelona, Spain: Institut Universitari de Lingüística Applicada.

Sproat, R. (1994). English noun-phrase accent prediction for text-to-speech. *Computer Speech and Language*, 8, 79–94.

Štekauer, P., Zimmermann, J., and Gregová, R. (2007). Stress in compounds: an experimental research. *Acta Linguistica Hungarica*, 54(2), 193–215.

Strik, H. and Boves, L. (1995). Downtrend in and . *Journal of Phonetics*, 23(1-2), 203–220.

Summers, D. (Ed.) (2000). *Longman Advanced American Dictionary*. Harlow, UK: Longman.

Terken, J. (1997). Variation of accent prominence within the phrase: models and spontaneous speech data. In Sagisaka, Y., Campbell, N., and Higuchi, N. (Eds.), *Computing prosody. Approaches to a computational analysis and modelling of the prosody of spontaneous speech* (pp. 95–116). New York: Springer.

Terken, J. and Hermes, D. J. (2000). The perception of prosodic prominence. In Horne, M., (Ed.), *Prosody: theory and experiment. Studies presented to Gösta Bruce* (pp. 89–127). Dordrecht, Netherlands: Kluwer.

Teschner, R. V. and Whitley, M. S. (2004). *Pronouncing English*. Washington, D.C.: Georgetown University Press.

Turk, A. E., Nakai, S., and Sugahara, M. (2006). Acoustic segment durations in prosodic research: a practical guide. In Sudhoff, S., Lenertová, D., Meyer, R., Pappert, S., Augurzky, P., Mleinek, I., Richter, N., and Schließer, J., (Eds.), *Methods in empirical prosody research* (pp. 1–28). Berlin, Germany: de Gruyter.

Visch, E. (1999). The rhythmic organization of compounds and phrases. In Van der Hulst, H. and Plank, F. (Eds.), *Word Prosodic systems in the Languages of Europe* (pp. 161–221). Berlin, Germany: de Gruyter.

Warren, B. (1978). *Semantic patterns of noun-noun compounds*. Gothenburg: Acta Universitatis Ghotobugenis.

Wood, S. N. (2006). *Generalized Additive Models. An introduction with R*. Boca Raton, FL.: Chapman & Hall/CRC.

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)*, 73(1), 3–36.

Wood, S. N. (2012). *Mixed GAM Computation Vehicle with GCV/AIC/REML smoothness estimation*. Version 1.7-20. URL: http://cran.r-project.org/web/packages/mgcv/index.html. Last accessed Sept 4, 2012.

Appendix

Table 6

List of compounds

L/N1		L/N2	
	coffee table designer		*city hall restoration*
	day care center		*cotton candy maker*
	field hockey player		*cream cheese recipe*
	hay fever treatment		*diamond ring exhibition*
	kidney stone removal		*family planning clinic*
	lung cancer surgery		*gene therapy technology*
	money market fund		*maple syrup production*
	security guard service		*science fiction book*
	sign language class		*silicon chip manufacturer*
	weather station data		*silver jubilee gift*
R/N2		R/N3	
	adult jogging suit		*baby lemon tea*
	business credit card		*company internet page*
	celebrity golf tournament		*family Christmas dinner*
	conference time sheet		*pilot leather jacket*
	passenger test flight		*pizza home delivery*
	piano sheet music		*prisoner community service*
	restaurant tourist guide		*student string orchestra*
	student season ticket		*tennis grass court*
	team locker room		*tennis group practice*
	visitor name tag		*woman fruit cocktail*