A new approach to prosodic grouping

Gerrit Kentner & Caroline Féry

Institut für Linguistik, Goethe-Universität Frankfurt a.M.
kentner@lingua.uni-frankfurt.de, caroline.fery@gmail.com

1 Introduction

Coordinated names, like *Anna and Bill or Mary*, form a syntactically ambiguous structure, in the same way as an arithmetic procedure like *3 – 2 + 1*, which can be resolved as 2 or as 0, depending on the order of the operations. In the case of coordinated names, the ambiguity concerns the branching direction and the level of syntactic embedding of the construction: either all three names may be on the same level of embedding (1-a), or two adjacent names may be grouped together to form a complex constituent that figures at the same level of syntactic embedding as the remaining simplex name ((1-b) and (1-c)). Depending on the kind of conjunction used, the different groupings may impinge on the truth value of a sentence the conjoined names are part of.

(1) a. [Anna or Bill or Mary]
b. [[Anna and Bill] or Mary]
c. [Anna and [Bill or Mary]]

Researchers have examined how different groupings of coordinated elements are realized prosodically (as for instance Ladd (1992) and Wagner (2005) for English, Schubö (2010) and Féry and Truckenbrodt (2005) for German). All authors have investigated phonetic differences in duration or pitch at conjunct boundaries and found a strong dependency between the prosodic realization and the syntactic place of the conjuncts in the coordination structure.

According to the results of previous research (e.g. Cooper and Paccia-Cooper, 1980; Lehiste, 1983; Gee and Grosjean, 1983), it may be considered verified that the prosodic boundary between adjacent constituents tends to be stronger the stronger the syntactic boundary between these constituents is. Correspondingly, prosodic boundaries are said to reflect syntactic structure. However, it is open to debate how close the match between syntactic and prosodic structure is.

We present results of a production and a perception experiment on various structures with coordinated names in German. It turns out that the expression of prosodic boundaries, as evidenced by pitch and duration, signals the depth of syntactic embedding of the constituents as well as the branching direction of the coordination structure. The results of the production experiment inspire a model of syntax-prosody mapping which assumes that the strength of a prosodic boundary after a given constituent is a function of a) the syntactic relation to the following constituent and b) the depth of its syntactic embedding. This model is compatible with accounts that allow a recursive representation of prosodic constituent structure at the level of the phonological phrase and above (Féry and Schubö, 2010; Ito and Mester, 2012; Ladd, 1996/2008; Wagner, 2005). A perception experiment with the same material indicates that listeners recognize embedded coordination structures on the basis of the prosodic form of the sentence, confirming that listeners are able to decode recursive syntactic structures on the basis of prosodic cues.

In section 2, we review previous experimental and theoretical work on the prosodic expression of syntactic structure, and we introduce a new model which accounts for the prosodic expression of syntactic boundaries. The production experiment is reported on in section 3. Based on the results of the production experiment, we evaluate our model and compare it with the predictive success of two existing models of prosodic boundary strength in section 4. Section 5 presents the results of a perception experiment on the coordination structures. We conclude with a general discussion in section 6, where we take up the issue of recursion in prosody as well.

2 Background and new proposal

2.1 Previous experimental work

There has been a keen interest in the psycholinguistic and phonetic literature as to how prosodic boundaries correlate with syntactic structure, especially in the case of structurally ambiguous sentences. Cooper and Paccia-Cooper (1980), Gee and Grosjean (1983), and Ferreira (1993) examine the placement as well as the strength of prosodic and intonational breaks in relationship to syntactic structure in speech production; Clifton et al. (2002, 2006) discuss the interpretation of prosodic boundaries with respect to sentence processing. See Watson and Gibson (2004) and Frazier et al. (2006) for summaries of previous research.

Speakers mark prosodic boundaries with characteristic acoustic cues: the duration of pre-boundary words is typically increased and there may be a period of phonetic silence; also, prosodic boundaries are characterized by deflections of pitch on the preceding syllable(s) (e.g. Cooper and Paccia-Cooper, 1980; Ferreira, 1993; Lehiste, 1983; Pierrehumbert, 1980; Price et al., 1991; Selkirk, 1984).

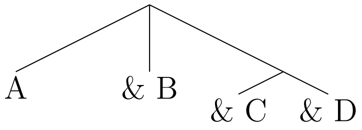
As for the relation between syntactic and prosodic boundaries, Watson and Gibson (2004) provide a model of prosodic boundaries called the Left hand side/ Right hand side Boundary hypothesis (LRB), in (2), in which the sizes of the preceding and the following syntactic constituents are the predictors for the likelihood of intonational phrase boundaries. Intonational phrases are defined in Watson and Gibson (2004) as prosodic constituents of indeterminate length ending in a boundary tone and containing at least one syllable that receives a pitch accent (cf. Pierrehumbert and Hirschberg (1990)). Watson and Gibson's motivation for a model making reference to the size of constituents is related to processing demands: within a larger utterance, speakers need time to recover after particularly long constituents, and they need planning time for long upcoming constituents. The time needed for recovery and planning is provided by intonational phrase boundaries. Therefore, according to the LRB, the likelihood of an intonational break at any given word increases with the size of the surrounding constituents. The size of the left and right constituent are predicted to have an equal share in predicting the likelihood of an intonational boundary.

(2) The Left hand side/ Right hand side Boundary Hypothesis (LRB, Watson and Gibson (2004))
The likelihood of an intonational boundary at a word boundary is a function of:
a. the size of the most recently completed constituent and
b. the size of the upcoming constituent if it is not an argument of the most recent head.

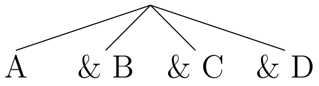
The LRB is shown to predict intonational phrase boundary location at least as well as, or even better than, more complex boundary strength models like Cooper and Paccia-Cooper (1980), Gee and Grosjean (1983) and Ferreira (1993). Watson and Gibson's own experiments, however, suggest that the LRB is too simplistic: their results show that the size of the preceding constituent has a much stronger influence on the likelihood of a boundary than the size of the upcoming one (see also Kentner (2007), who confirms this asymmetry for German).

Also, as Wagner (2005) observes, the LRB only predicts effects of adjacent constituents but cannot account for non-local effects of syntactic structure on boundary strength. In a production experiment, he found that simplex constituents such as A and B within a coordination structure like (3), which have a branching sister, are produced with longer duration than comparable simplex constituents that have only simplex sisters (4). Importantly, this also holds for simplex sisters that are non-adjacent to the complex constituent.[1]

(3)



(4)



Accounting for such non-local effects, Wagner (2005) proposes an alternative model which relates the strength of prosodic boundaries to syntactic levels of embedding rather than the size of adjacent constituents. This is the Scopally Determined Boundary Rank (SBR) in (5).

(5) Scopally Determined Boundary Rank (SBR, Wagner (2005)):
If Boundary Rank at a given level of embedding is *n*, the rank of the boundaries between constituents of the next higher level is *n+1*.

Although the predicted non-local increase in prosodic boundary strength due to embedding has been confirmed in Wagner's (2005) experiments, the SBR cannot easily account for the finding that the boundary strength also increases with the size or complexity of the surrounding constituents as predicted by the LRB and confirmed by the results of both Watson and Gibson (2004) and Wagner (2005). Moreover, as Wagner (2005) acknowledges, the SBR's success crucially depends on the use of different normalizing procedures depending on the various conditions tested.

Wagner's (2005) experiment on structures like (3) and (4) reveals another prosodic effect, which, however, neither the LRB nor the SBR succeed in predicting: the prosodic boundary after constituent C, if embedded as in (3), is significantly shortened relative to the boundary at the same position in the baseline pattern (4).

Given these problems of the LRB and SBR algorithms, we propose a new approach to the prediction of boundary strength based on two general principles that we call *Proximity* and *Similarity*.

[1] Concurring with Wagner (2005), we consider coordinations of like categories (in this case: NPs and coordinations thereof), i.e. symmetric coordination. Correspondingly, *n*-ary branching trees are assumed to be appropriate syntactic representations when there are more than two conjuncts at the same level.

2.2 The Proximity/Similarity model

We propose two general principles responsible for the interface between syntactic constituent structure and prosodic structure. These principles shape the expression of prosodic boundaries for the syntactic domain under consideration, i.e. a sentence or part thereof.

First, Proximity is inspired by a principle with the same name that Lerdahl and Jackendoff (1983) formulated in the context of musical grouping.[2] In Lerdahl and Jackendoff (1983), this principle is perception-oriented and amounts to the observation that two adjacent musical notes are perceived as belonging to different groups if the interval between them is large relative to other intervals in the vicinity. Here, Proximity operates on syntactic constituent structure, reflecting syntactic boundaries in prosodic structure. According to this principle, adjacent elements which are syntactically grouped together into one constituent should be realized in close proximity. Proximity between two elements is achieved by substantially weakening the prosodic boundary cues (segmental lengthening or boundary tone) on the first element. A corollary of Proximity is the opposite effect: adjacent elements not grouped together into one constituent should be realized with prosodic distance. As for Anti-Proximity, longer duration (final lengthening) and a higher boundary tone increase the distance to adjacent material to the right that is not part of the same immediate constituent. These effects are formalized in (6).

(6) Proximity

a. The prosodic boundary at the terminal constituent x is weakened if the following terminal constituent y is the sister of x or dominated by the sister of x – unless x is immediately dominated by the root node of the domain under consideration.

b. (Anti-Proximity): The prosodic boundary at the terminal constituent x is strengthened if the following terminal constituent y is not a sister of x.

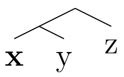
Note that (6) implies directionality because it is always the realization of the left of two elements that reflects whether the element to its right belongs to the same constituent or not. In other words, the prosodic expression of Proximity or Anti-Proximity on a lexical item only mirrors its syntactic relation to constituents to the right and not to those to the left.

There are four ways in which (6) may impinge on a lexical item:

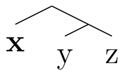
(7) A lexical item x may be subject to

[2] Lerdahl and Jackendoff's grouping principles are inspired by works in the tradition of Gestalt psychology (e.g. Wertheimer, 1938).

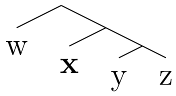
a. Proximity (P) in



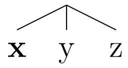
b. Anti-Proximity (A) in



c. both P and A in



d. neither P nor A (baseline) in



Proximity is fulfilled in (7-a), where x and y belong to the same constituent to the exclusion of z. Anti-Proximity is shown in (7-b), where x does not belong to the same constituent as y and z. Proximity and Anti-Proximity have contradictory effects; a single lexical item may be subject to both when it is the left element of a larger embedded constituent, but the following terminal element is not its sister (7-c). In this case, we assume that the two effects cancel each other out.

The baseline representation in (7-d) corresponds to a list of lexical items with no hierarchical ordering. Here, all constituents are at the same level of embedding and are directly dominated by the root node. According to (6), the default prosodic break is neither strengthened nor is it weakened; instead, simple list intonation is predicted to apply.[3]

The second principle, Similarity, operates on the depth of syntactic embedding. It claims that constituents at the same level of embedding should be realized in a similar way, that is, they should be similar in pitch and duration, irrespective of their inherent complexity.

Similarity predicts prosodic adjustment of simplex elements as compared to complex constituents at the same level of embedding. More specifically, simplex elements are lengthened to approximate the duration of the complex constituent. This also holds for simplex elements that are non-adjacent to complex constituents if they are at the same level of syntactic embedding.

(8) Similarity
The prosodic boundary at the terminal constituent x is strengthened if a sister constituent of x is complex.

The two principles are predicted to interact to shape the prosody of syntactic structures.

[3] We suggest that the characteristics of the default prosodic break depend on the structures under scrutiny. In the current case, the string of conjoined names makes up an intonational phrase that is separated by prosodic phrase boundaries after each name, where prosodic phrase is understood as a prosodic unit that contains one pitch accent.

While previous research has provided evidence for effects that may be explained in terms of Proximity and Similarity (e.g. Hunyadi, 2006; Wagner, 2005; Watson and Gibson, 2004), it is as yet unclear whether these principles overcome the aforementioned shortcomings of the LRB or SBR algorithms; if so, it is not obvious what the relative contribution of the two principles is, i.e. how much of the prosodic surface structure is attributable to the workings of Proximity and how much is due to Similarity. Moreover, a syntax-prosody mapping model that makes use of Proximity and Similarity has to be clear about how these factors interact given that syntactic structures are subject to both.

To answer these questions, we conducted a production experiment designed to test the effects of (recursive) syntactic grouping on prosodic structure. Assuming that speakers do produce prosody that signals recursive syntactic embedding, it then remains to be verified whether listeners are able to deduce such nested syntactic structure from the prosodic form. This will be examined in a perception experiment.

In this paper, we aim at developing a model with Proximity and Similarity as main predictors. On the basis of the observed prosodic patterns we show that the performance of the Proximity/Similarity model is superior to that of the LRB, the SBR and a model combining both the LRB and SBR.

3 Production experiment

3.1 Method and material

The production experiment is based on Wagner's (2005) very similar experiment on the prosody of coordinate structures in English. The material consisted of different groupings of three or four conjoined proper names, all disyllabic and trochaic, like *Mila, Nino and Willi.* All groupings tested in the experiment are illustrated in (9) and (10), where N1 stands for the first name, N2 for the second name and so on. The conjunction *und* ('and') was always used inside of a bracket, and the conjunction *oder* ('or') outside of a bracket.[4] The structures 4.4 and 4.5 include embedded groupings, which are right-branching in the case of 4.4 and left-branching in the case of 4.5. As a result, we have three right-branching structures, 3.2, 4.2 and 4.4, and three left-branching structures, 3.3, 4.3 and 4.5.

[4] We are aware that the use of different conjunctions may have had additional confounding effects (see Ladd (1992) and also Féry and Truckenbrodt (2005) for the effect of different sentence conjunctions in a sequence of three coordinated sentences). However, using only one type of conjunction would have led to very dull sentences. Given that the speakers were provided with explicit bracketing to mark the respective conditions, we think any nuisance effects stemming from the different conjunctions will be minor.

(9) 3.1 N1 N2 N3 *Nino oder Willi oder Mila*
3.2 (N1 N2) N3 *(Nino und Willi) oder Mila*
3.3 N1 (N2 N3) *Nino oder (Willi und Mila)*

(10) 4.1 N1 or N2 or N3 or N4 *Nino oder Willi oder Mila oder Susi*
4.2 N1 or N2 or (N3 and N4) *Nino oder Willi oder (Mila und Susi)*
4.3 (N1 and N2) or N3 or N4 *(Nino und Willi) oder Mila oder Susi*
4.4 N1 or (N2 or (N3 and N4)) *Nino oder (Willi oder (Mila und Susi))*
4.5 ((N1 and N2) or N3) or N4 *((Nino und Willi) oder Mila) oder Susi*
4.6 (N1 and N2) or (N3 and N4) *(Nino und Willi) oder (Mila und Susi)*

Participants were presented altogether 4 items from each of the nine conditions. The items were presented on screen one by one in randomized order. The grouping condition was made explicit by brackets and by a logical form. To trigger the target structure, a context plus a question was presented (a screen display is exemplified in (11)). Additionally, the context and question were presented auditorily over headphones once the screen display was shown.

(11) Context: Susi and Lena always go to the pool together, and Willi also does a lot of swimming.
Question: With whom do you want to go for a swim tomorrow?
Target: With (Susi and Lena) or Willi.
Logical Form: $(a \wedge b) \vee c$

The participants were 21 female students from the University of Potsdam, monolingual speakers of German in their twenties, coming from the Northern area of Germany. They were paid 6 Euros or got credit points for their participation. Recordings were made in an unechoic chamber on a DAT recorder. The participants were instructed to read the context carefully and to pay attention to the best way of realizing the groupings. They were given as much time as they wanted to utter the answer, and had the opportunity to correct themselves. If corrections were made, the last production of the item in question was taken. Altogether, 756 sentences were recorded and analyzed, 252 with three names (21 subjects x 3 conditions x 4 contexts), and 504 sentences with four names (21 subjects x 6 conditions x 4 contexts).

3.2 Measurements

An example of the realization is given in Fig.1.

The recordings were re-digitized from DAT at a sampling frequency of 44.1 kHz and 16 bit resolution. Every name as well as every conjunction were labeled and delimited by a boundary set manually in an annotation tier in praat (Boersma and

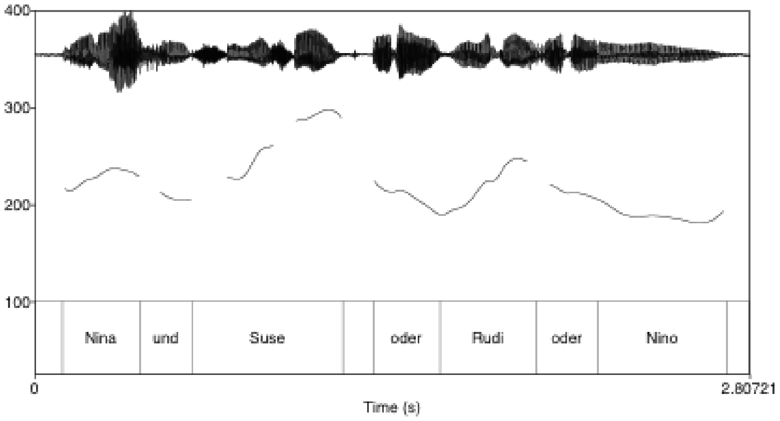


Figure 1: Pitchtrack for condition 4.3.

Weenink, 2009). We measured the duration of every name, of the pauses between names and of the conjunctions. As a measure of prosodic boundary strength, we summed the duration of each name and the following pauses, i.e. we considered the pauses part of the boundaries (see also Gee and Grosjean, 1983; Wagner, 2005; Wightman et al., 1992). A comparison with measurements without pauses did not reveal any relevant difference in the results. The analysis of pitch was conducted in praat, applying the smoothing algorithm (frequency band 10 Hz) to diminish microprosodic perturbations. Time-normalized contours were created by dividing up each constituent into five equal-sized intervals and by interpolating the aggregated mean F0 (in Hz) over speakers and sentences for each interval. All measurements were checked post hoc, and corrected manually when necessary (e.g. in the case of octave errors). Statistic analyses were performed using the statistical computing environment R.

3.3 Predictions

Based on earlier results from prosody research in German (Grabe, 1998; Féry and Kügler, 2008; Truckenbrodt, 2002, and others), some assumptions about the production of the expressions can be formulated. The realizations without grouping, 3.1 and 4.1, are taken as baselines and all other patterns are compared in relation to these baselines. In the baseline patterns without groupings, all names are expected to be of equal prominence and separated by boundaries of the same strength. Each name gets a pitch accent, which is expected to be rising (L*H) in non-final position and falling (H*L) in final position. L* and H* are the pitch accents, and the trailing tones H and L are the boundary tones of their respective domain. Pitch and duration of the final constituent are expected to be identical in all cases. In other words, we expect neutralization of the prosodic boundary at the end of all patterns, due to a final low boundary tone at the end of a declarative sentence. Another prediction is that, in the baseline, every high tone is downstepped relative to the preceding one, and no difference in duration occurs among the names.

If syntactic groupings are reflected in prosody, this is expected to happen by means of changed pitch accents, boundary tones and duration, the main intonational events. We derive our hypotheses about the prosodic realization of different syntactic groupings from the two general principles Proximity and Similarity.

As an example, the structures in 4.2 and 4.3 in (12) display one simple grouping of two elements into one constituent each.

(12) a. 4.2: *Nino oder Willi oder (Mila und Susi)*
b. 4.3: *(Willi und Mila) oder Susi oder Nino*

As a result, there are three constituents at the top level in these conditions, two simplex ones and a complex one. The simplex names are predicted to be lengthened and thus adjust to the duration of the complex constituent in order to achieve similarity across constituents at the top level. In addition, as predicted by Anti-Proximity, the element outside of but left-adjacent to a grouping should exhibit a stronger prosodic boundary (cf. *Willi* in (12-a)). The same applies to the rightmost name of a grouping (cf. *Mila* in (12-b)). The left elements of groupings are expected to show weaker prosodic boundary cues in order to fulfill Proximity (*Mila* in (12-a), *Willi* in (12-b)). To sum up, Proximity and Anti-Proximity should have local effects: weakening of the left and strengthening of the right element within a grouping, as well as strengthening of simplex elements that are left-adjacent to a grouping. Similarity implies that syntactic grouping has non-local effects as well: compared to the baseline, all simplex elements that have a complex sister should be lengthened (even those that are not adjacent to groupings). The different effects of Proximity (P), Anti-Proximity (A) and Similarity (S) are tabulated for each condition and each non-final name in Table 1 for the conditions with three names, and in Table 2 for the conditions with four names.

	N1	N2
3.1 N1 or N2 or N3	–	–
3.2 (N1 and N2) or N3	P	A
3.3 N1 or (N2 and N3)	A,S	P

Table 1: Non-final names subject to Proximity (P), Anti-Proximity (A) and Similarity (S) in conditions with three names.

	N1	N2	N3
4.1 N1 or N2 or N3 or N4	–	–	–
4.2 N1 or N2 or (N3 and N4)	S	A,S	P
4.3 (N1 and N2) or N3 or N4	P	A	S
4.4 N1 or (N2 or (N3 and N4))	A,S	P,A,S	P
4.5 (N1 and N2) or N3) or N4	P	A	A,S
4.6 (N1 and N2) or (N3 and N4)	P	A	P

Table 2: Non-final names subject to Proximity (P), Anti-Proximity (A) and Similarity (S) in conditions with four names.

3.4 Results for three names

The results for duration and pitch are shown simultaneously in Figure 2. In the description of the pitch contours, we concentrate on the high tones on the names themselves, and largely ignore the conjunctions, which behave as transitions between the names. The low tones are also discarded in the discussion. The baseline pattern 3.1 (light grey) presents downstep between N1, N2, and N3. However, N3, the final name, is neutralized in all patterns, and will not be considered any further. Pattern 3.2 (black) shows an important difference compared to the baseline: N1's high tone is much lower than in the baseline, while N2 has a higher pitch value (upstep), reaching a level comparable to N1 of the baseline condition. By contrast, the tonal pattern of 3.3 (dark grey), a right-branching structure, is very similar to that of the baseline 3.1. They both have a high N1 and subsequent downstep on the further two names. N1 in 3.3. is not significantly higher than N1 in the baseline condition. However, the N2 of pattern 3.3 is slightly lowered as compared to the baseline condition 3.1. As a result the difference in pitch (i.e. the amount of downstep) between N1 and N2 is larger in 3.3 than in 3.1. Comparing the high tones across conditions, a mirror-image relation between the left-branching condition 3.2 and the other conditions is apparent: the upstepped H-tone of N2 in 3.2 approximates the height of N1 in the other conditions. Conversely, the height of N1 in condition 3.2 closely resembles the height of the downstepped H-tones on N2 in the other conditions.

As for duration, the three names of the baseline pattern 3.1 (light grey columns) display small differences; the slightly longer duration of N2 (mean difference compared to N1 is about 40ms) is significantly different from N1 (t=3.8, p<0.001). We return to this effect in the discussion section. Compared to the baseline, pattern 3.2 (black) has a significantly shorter N1 (a group-initial element) and a significantly longer N2 (a group-final element). In contrast, in pattern 3.3 (dark grey), N1 (simplex element, left-adjacent to a grouping) is longer while N2 (group-initial element) is shorter than the baseline. We also see that N3's duration is neutral-

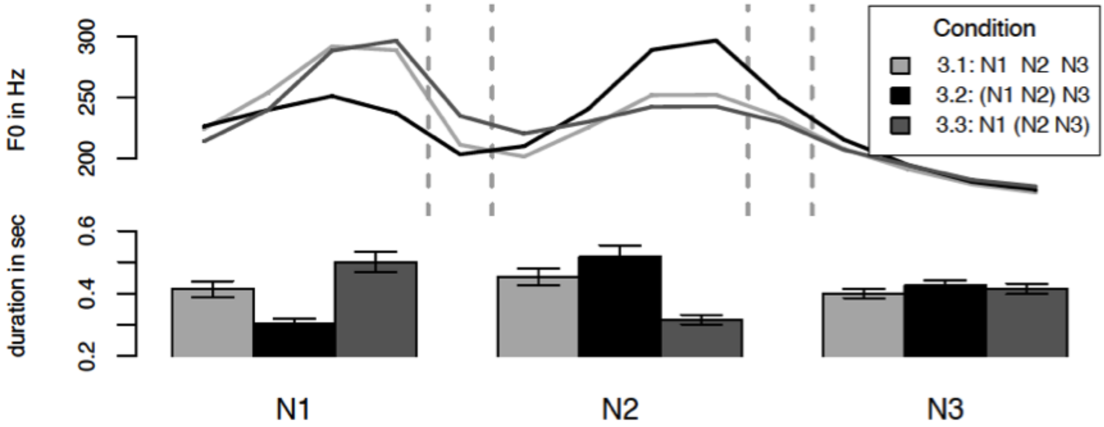


Figure 2: mean pitch in Hz and mean duration in ms of the conditions with three names.

ized. Indeed, this neutralization of the last name is persistent in all conditions, as we will see, both in duration and in pitch.

To sum up the three-name conditions, pitch and duration deliver equivalent results in that higher pitch on non-final names generally coincides with longer duration and lower pitch patterns with shorter duration. The pitch tracks reveal an interesting asymmetry: The right-branching pattern (3.3) has a striking resemblance to the baseline – both have a downstep pattern. But the left-branching pattern (3.2) has a different shape, namely a lower pitch on N1 and a clear upstep on N2. Both patterns with groupings clearly differ from the baseline with respect to duration. The first element of a grouping is always shorter than in the baseline, and the last element of a grouping is always longer than in the baseline (except in N3 because of final neutralization). These results are in line with the general principles of Proximity, Anti-Proximity and Similarity: names that are affected by Anti-Proximity and Similarity express a stronger prosodic boundary while the ones that are subject to Proximity are clearly shortened and lowered in pitch compared to the baseline.

3.5 Results for four names

In this section, we compare the realizations of the baseline 4.1 to the various conditions with groupings 4.2 to 4.6. An overview of all results on pitch and duration is given in the plots depicting difference scores between the baseline and other conditions with 95% confidence intervals in Figures 8 and 9 below.

First, the Figures 3 and 4 show the results for the right-branching conditions 4.2 and 4.4 as compared to the baseline condition 4.1. As was the case for the three-name patterns, the discussion for pitch concentrates on the relationship between the high tones of names. In the right-branching structures, 4.2 and 4.4, and in the baseline 4.1, there is downstep throughout. The general impression is that 4.2 and 4.4 have roughly the same shape as the baseline. However, in 4.2 and 4.4, N3 is somewhat lower than in the baseline. Correspondingly, the downstep between N2 and N3 is also larger than in the baseline, due to the fact that N3 is the first element of a grouping in these conditions and is thus compressed in pitch. A similar enhancement of downstep due to tonal compression was observed in the right-branching condition 3.3. In 4.2 and 4.4, the elements preceding a grouped constituent bear higher tones than the corresponding names of the baseline. Turning to duration, the baseline (grey) presents an unexpected pattern with N2 clearly longer, and N3 clearly shorter than N1. This durational effect is not accompanied by a similar effect in pitch. We will come back to this effect in the discussion section below. N1 of 4.2, a simplex element, is longer than in the baseline. Similarly, N1 of 4.4, which is in front of a left parenthesis, is also significantly lengthened, even more so than N1 of 4.2. This difference is explained by the fact that N1 in 4.2 is subject only to Similarity, whereas it is subject to both Anti-Proximity and Similarity in 4.4. In contrast, N3 in 4.2 and 4.4 are realized much shorter than in the baseline, but they do not significantly differ from each other (see also Figure 8 and Figure 9 for comparison). These are first elements of groupings and as such subject to Proximity. N2 is in both patterns located before a left parenthesis, but in 4.4, it is at the same time the first element of a recursive grouping. In the latter condition, it has a similar duration as in the baseline. Neutralization at the end of the sentence is once again observed in all patterns.

The left-branching structures in 4.3 (Figure 5) and 4.5 (Figure 6) differ from the baseline in several respects. Except if it is the last one in the sentence, the rightmost element of a grouping is higher in pitch than in the baseline. This explains why N2 in 4.3 and 4.5 as well as N3 in 4.5 are the highest points in these sentences. In all three patterns, N1, the first element of the groupings, is then realized at a much lower level. The N2s do not present very large differences in their absolute values as compared to the baseline, but an upstep from N1 to N2 can be observed (whereas in the right-branching conditions, downstep was the rule). The duration relations of left-branching structures in 4.3 and 4.5 differ from the baseline in several respects. In both 4.3 and 4.5, N2 is located in front of a (non-final) right parenthesis. These names are significantly lengthened compared to the baseline. Moreover, N3 of 4.5, again preceding a right parenthesis, is much longer than all other third names. In contrast, N1 in 4.3 and 4.5 is realized significantly shorter than in the baseline. Neutralization at the end of the sentence is once

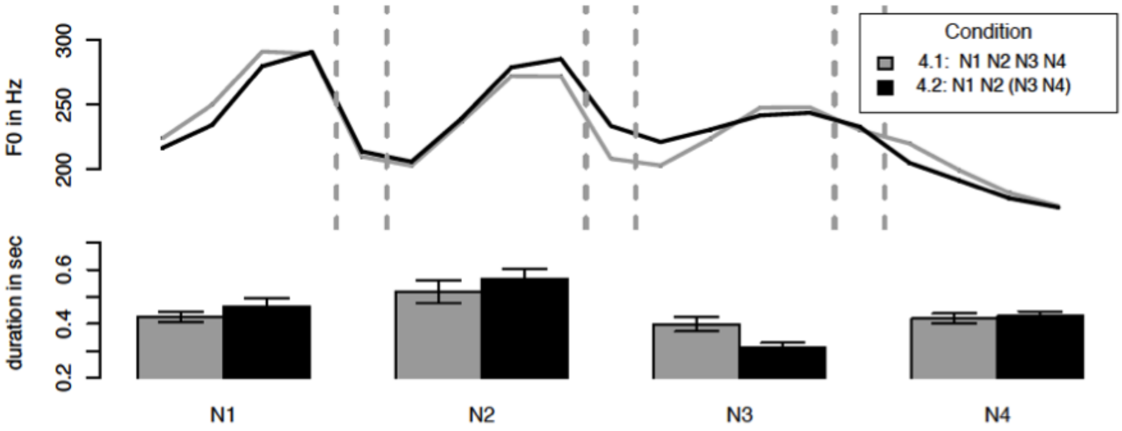


Figure 3: Comparison of simple right-branching condition (black) with baseline (grey).

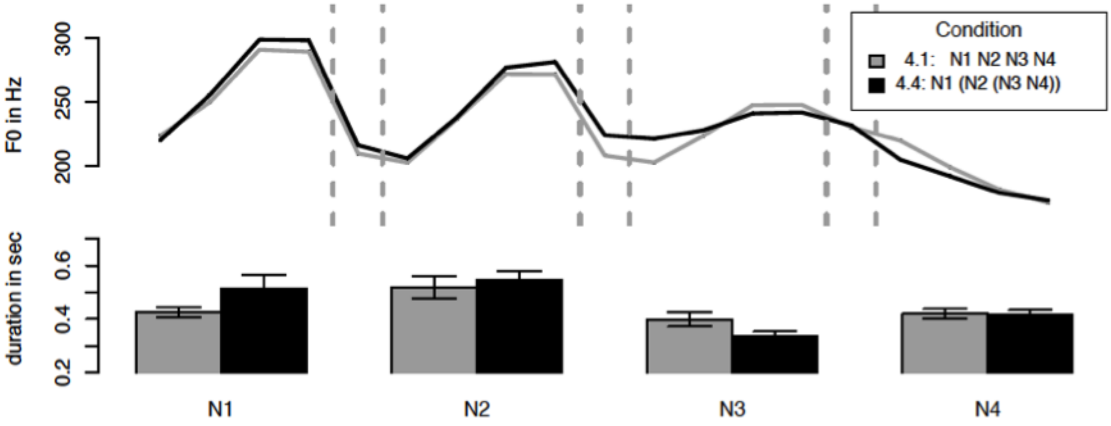


Figure 4: Comparison of embedded right-branching condition (black) with baseline (grey).

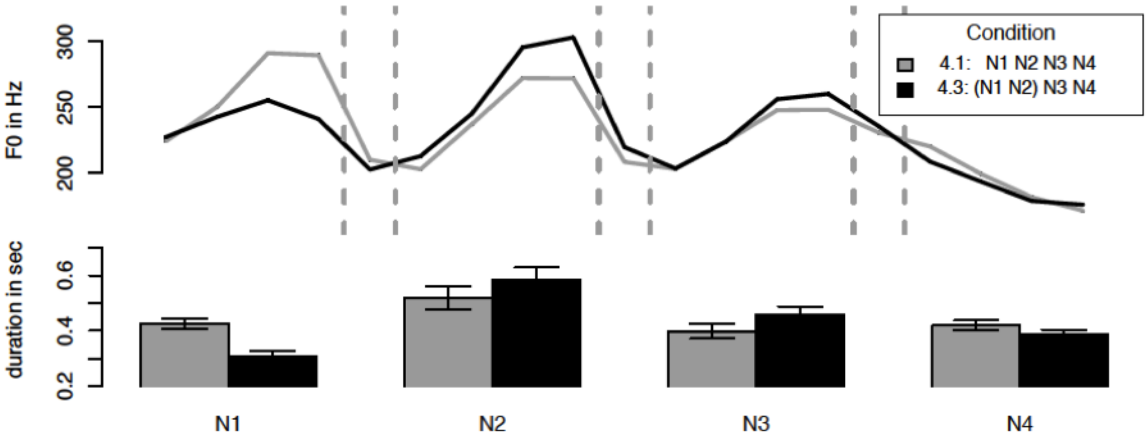


Figure 5: Comparison of simple left-branching condition (black) with baseline (grey).

again observed in all patterns.

Finally, 4.6 with a double grouping is also compared to the baseline. In this pattern, we observe once more that the rightmost element of a grouping is higher and longer than in the baseline. This is the case for N2. N1, the first element of the grouping, is then much shorter and is realized at a much lower level, and an upstep from N1 to N2 can be observed. N3 is lower than in the baseline due to the fact that it is the first element of a grouping, and it is also shorter. As was observed in the three-name patterns, the downstep between N2 and N3 is larger than in the baseline.

Again, we generally find a strong correlation of duration and pitch.

As predicted, names that are subject to Proximity are shortened and compressed in pitch, while names that are subject to Anti-Proximity are lengthened and show upstep.

3.6 Discussion

In sum, the predictions of the Proximity/Similarity model are largely borne out. Each of the syntactic conditions appears to have a unique prosodic rendition, and the Proximity/Similarity model correctly predicts the prosodic effects that were observed: Names that are subject to Anti-Proximity are lengthened and show upstep, thereby strengthening a prosodic boundary. In contrast, names that are subject to Proximity are shorter and lower in pitch compared to the baseline, reflecting the cancellation of a prosodic boundary. The effect of Similarity appears

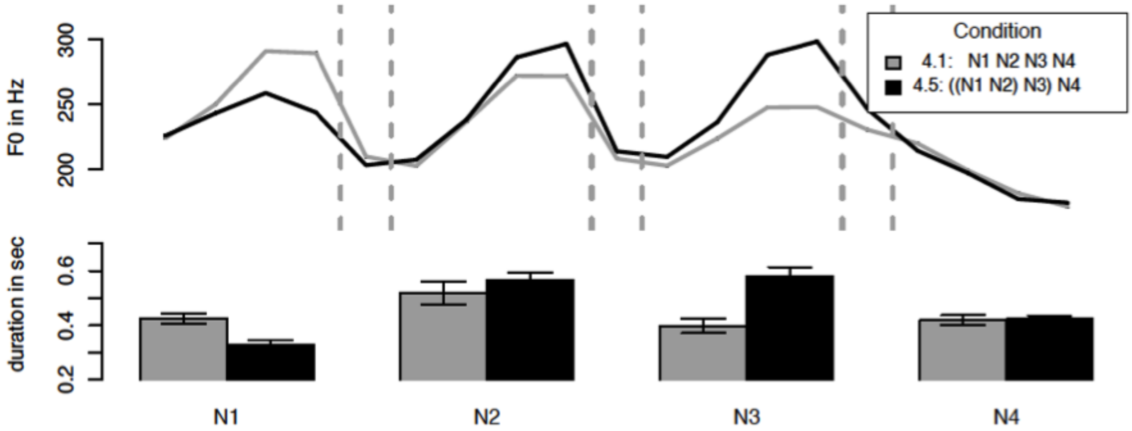


Figure 6: Comparison of embedded left-branching condition (black) with baseline (grey).

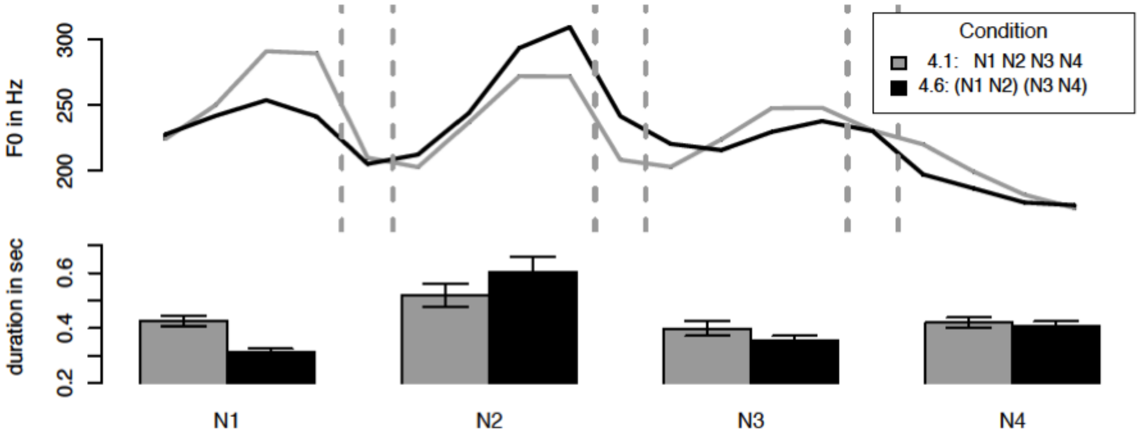


Figure 7: Comparison of double grouping condition (black) with baseline (grey).

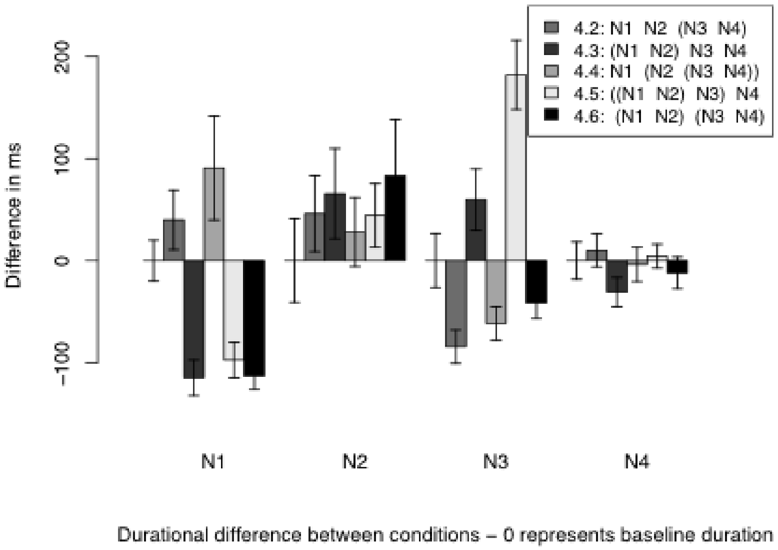


Figure 8: Differences in duration between baseline and other conditions broken down by name.

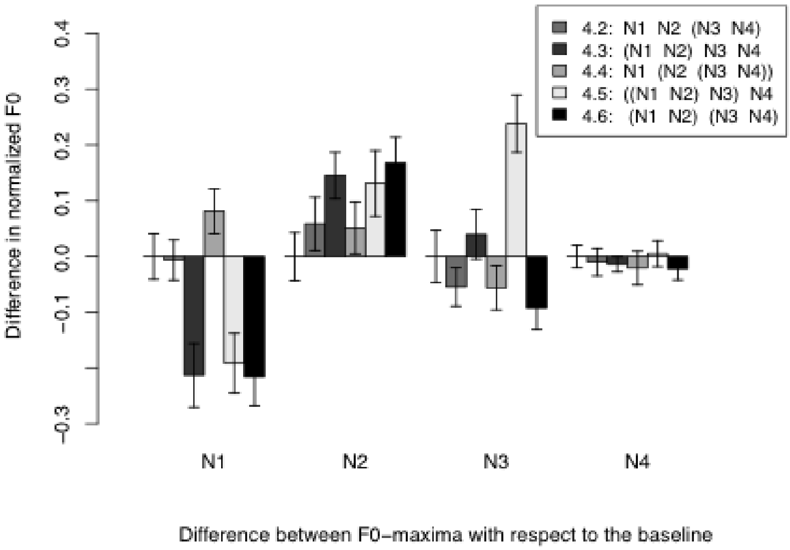


Figure 9: Differences in normalized F0 between baseline and other conditions broken down by name.

to be weaker than that of Proximity or Anti-Proximity, but it still accounts for significant effects in duration (e.g. N1 of 4.1, N3 of 4.3).

A deviance in the parallelism regarding pitch and duration concerns the baselines 3.1 and 4.1. In 3.1, N2 was clearly longer than N1. Similarly, 4.1 displays a conspicuous lengthening of N2 and shortening of N3 compared to N1, but no comparable effect in pitch. According to the flat syntactic structure without grouping, the names were expected to be equivalent in duration across positions. We take the lengthening of N2 and shortening of N3 in the baseline 4.1 to be a reflection of abstract or 'inherent' grouping: even in the absence of syntactic motivation for grouping, speakers may favor a binary branching structure, which corresponds to an abstract grouping of N1 with N2 and N3 with N4. Independent evidence for such rhythmic grouping in the absence of explicit syntactic motivation comes from the prosodic rendering of telephone numbers: Baumann and Trouvain (2001) show that speakers preferably chunk a string of numbers into groups of two. Hunyadi (2006) reports a similar effect in a non-linguistic task: he presented Hungarian speaking participants with visual stimuli (4 equal-spaced dots in a row) and asked them to represent the visual display by mouse clicks. Measuring the time between clicks, Hunyadi found that participants needed more time between the second and third click than between the first and second. This effect of abstract grouping, however, was not confirmed in a speech production experiment in which participants read out a row of four letters. In any case, the tendency for abstract binary grouping without bracketing has a much weaker effect than the explicit boundaries in the binary branching structure of condition 4.6.

Overall, the right-branching structures (4.2 and 4.4) appear to be prosodically less articulate than the left-branching structures (4.3 and 4.5) and, correspondingly, right-branching structures are much more similar to the baseline. The prosodic markedness of the left-branching structures is considered to be due to the preponderance of upstep of boundary tones in these structures. Upstep is predicted for constituents that are subject to Anti-Proximity and is particularly strong if a non-final element that is subject to Anti-Proximity is preceded by an element that is subject to Proximity and thus compressed in pitch. The sequence of names which are subject to Proximity followed by names that are subject to Anti-Proximity is found in left-branching structures only. Correspondingly, the Proximity/Similarity model accounts for this specific prosodic markedness of left-branching structures as opposed to right-branching ones.

4 Model comparison

While the general predictions of Proximity and Similarity seem to be largely confirmed by the production data, we have yet to show how this model compares to other models of prosodic boundary likelihood or strength.

4.1 Method

In what follows, we evaluate the goodness of fit of the Proximity/Similarity model with the competing SBR and LRB algorithms. To do this, the boundary strength values that each theory predicts are calculated for each name of the structures 4.1 to 4.6.

For the Proximity/Similarity model, this is done as follows: The first factor Proximity has three levels: the baseline level is 0, i.e. all constituents of the baseline receive this predictor value. Names that are subject to Proximity are predicted to be shorter than the baseline; the corresponding predictor value is -1. For constituents that are subject to Anti-Proximity, the value 1 serves as the predictor. N2 in the right-branching condition with double embedding 4.4 is subject to both Proximity and Anti-Proximity. In this case, the two predictor values are simply summed, yielding 0 as the predictor for these constituents.

The second factor, Similarity, has two levels, 1 for names that are subject to Similarity and 0 for other names. The coding of the Proximity/Similarity model is summarized for the conditions with four names in Table 3.

Proximity/Similarity	N1	N2	N3
4.1 N1 or N2 or N3 or N4	Prox: 0, Sim: 0	P: 0, S: 0	P: 0, S: 0
4.2 N1 or N2 or (N3 and N4)	Prox: 0, Sim: 1	P: 1, S: 1	P: -1, S: 0
4.3 (N1 and N2) or N3 or N4	Prox: -1, Sim: 0	P: 1, S: 0	P: 0, S: 1
4.4 N1 or (N2 or (N3 and N4))	Prox: 1, Sim: 1	P: 0, S: 1	P: -1, S: 0
4.5 ((N1 and N2) or N3) or N4	Prox: -1, Sim: 0	P: 1, S: 0	P: 1, S: 1
4.6 (N1 and N2) or (N3 and N4)	Prox: -1, Sim: 0	P: 1, S: 0	P: -1, S: 0

Table 3: Coding scheme for the Proximity/Similarity model.

As for the SBR, the predictions are taken directly from Wagner (2005) and summarized in Table 4.

To adapt the LRB for our case, we need to make two deviations from Watson and Gibson's original algorithm: First, note that the LRB differs from the Proximity/Similarity model and the SBR in that it was designed to predict the likelihood of an intonational/intermediate phrase boundary in terms of the ToBI system (Beckman and Ayers, 1997) rather than the strength of a phrase break in terms of duration. Here, we consider the boundary strength to be reflected by the duration of the preceding constituent plus the following pause as dependent measure. As Wagner (2005) notes, "the advantage of this measure is that the annotation does not presuppose a theory of phrasing, and no labeling of prosodic categories (such

SBR: boundary strength after	N1	N2	N3
4.1 N1 or N2 or N3 or N4	1	1	1
4.2 N1 or N2 or (N3 and N4)	2	2	1
4.3 (N1 and N2) or N3 or N4	1	2	2
4.4 N1 or (N2 or (N3 and N4))	3	2	1
4.5 ((N1 and N2) or N3) or N4	1	2	3
4.6 (N1 and N2) or (N3 and N4)	1	2	1

Table 4: Coding scheme for the SBR model.

as intonational phrase or intermediate phrase as in a ToBI-labeling) is necessary." Concurring with Wagner (2005), we will assume that the likelihood of an intonational/intermediate phrase boundary is strongly correlated to the duration of a prosodic break at any given position. In fact, Watson and Gibson themselves also use the term 'boundary weight,' which does justice to the gradient nature of prosodic boundaries. The second difference to Watson & Gibson's original approach is related to the nature of the materials used in the experiments. Compared to the sentences used in Watson and Gibson (2004), our structures are relatively short.[5] Therefore, IP boundaries are not necessarily expected. Correspondingly, we measure the complexity of the left-hand side and right-hand side in terms of phonological words rather than phonological phrases. At each word boundary, the boundary strength is calculated in accordance with (13) (cf. Watson & Gibson 2004).

(13) The LRB weight at a word boundary between w1 and w2 is defined to be the sum of

a. the size of the left-hand side (LHS) constituent terminating at w1, measured in terms of phonological words (p-words);

b. the projected size of the right-hand side (RHS) constituent in p-words starting at w2, if this is not an argument of w1;

c. 1, if w1 marks the end of a phonological phrase.

The predictions of the modified LRB model are summarized in Table 5.

We compare the predictions of the Proximity/Similarity model with the predictions of the SBR and the LRB. Specifically, we evaluate the experimental results against the predictors of the three models. The duration of the individual items in each condition was averaged for each speaker. All models are mixed effects models that evaluate the log-transformed durations[6] of the names against the specific

[5] Watson and Gibson (2004) used sentences including relative clauses, such as *The director who the critics praised at a banquet insulted an actor from an action movie during an interview.*

[6] log transformation is applied because the raw duration data is necessarily distributed in non-

LRB: boundary likelihood after	N1	N2	N3
4.1 N1 or N2 or N3 or N4	1+1=2	1+1=2	1+1=2
4.2 N1 or N2 or (N3 and N4)	1+1=2	1+2=3	1+1=2
4.3 (N1 and N2) or N3 or N4	1+1=2	2+1+1=4	1+1=2
4.4 N1 or (N2 or (N3 and N4))	1+3=4	1+2=3	1+1=2
4.5 ((N1 and N2) or N3) or N4	1+1=2	2+1+1=4	3+1+1=5
4.6 (N1 and N2) or (N3 and N4)	1+1=2	2+2+1=5	1+1=2

Table 5: Coding scheme for the adapted LRB model. Each predictor is the sum of the LHS (first addend), the RHS (second addend) and – where applicable – the addend 1 reflecting the end of the phonological phrase (cf. (13-c)).

model predictors with speaker as random effect.

4.2 Results

Table 6 displays the modeling results for the Proximity/Similarity model. The formula in the upper row of each panel in the table represents the linear model, which evaluates the dependent variable (logarithmized duration values) against the fixed effects (coded as described above). In the first model (upper panel), the single effect of the Proximity predictor (Prox) is evaluated; the second model evaluates the Similarity (Sim) predictor; in the third model (lower panel), the model estimates for the two fixed effects and the interaction are given. The variance that is due to the different speakers from the production experiment is accounted for in these models by including the variable "speaker" as a random effect term. As may be seen, the two fixed effects and the interaction account for significant portions of the distribution of the dependent variable (absolute t-values >2 indicate significance at α=0.05).

The SBR and LRB models are summarized in Table 7, which also displays a combined model with main effects of SBR and LRB plus the respective interaction. These three models confirm that LRB, SBR and the corresponding interaction have significant effects on the dependent variable.

That is, the predictors of all the models under consideration may each explain significant portions of the variance; however, we still need to determine which of the models (and which of the fixed factors) best explains the variance in the data. To this end, a comparison of the fit of the models is in order.

As a measure of model fit, we take the R^2 value, i.e. the proportion of variability in the data set that the statistical model accounts for.[7] The R^2 values and the

normal fashion, as there are only positive durations. Non-normal distribution would possibly violate the assumptions of the statistical model.

[7] R^2 is the squared correlation of i) the fitted values of the model under consideration and ii)

Formula: log(duration) ~ Prox + (1|speaker)

	Estimate	Std. Error	t-value
Prox	0.2742	0.00925	29.63

Formula: log(duration) ~ Sim + (1|speaker)

	Estimate	Std. Error	t-value
Sim	0.24108	0.02697	8.94

Formula: log(duration) ~ Prox × Sim + (1|speaker)

	Estimate	Std. Error	t-value
Prox	0.28928	0.01052	27.49
Sim	0.10853	0.01969	5.51
Prox:Sim	-0.16881	0.02683	-6.29

Table 6: Parameters for models evaluating the Proximity factor (upper panel), the Similarity factor (middle panel), and the combined Proximity/Similarity factors and interaction.

Formula: log(duration) ~ SBR + (1|speaker)

	Estimate	Std. Error	t-value
SBR	0.26628	0.01520	17.51

Formula: log(duration) ~ LRB + (1|speaker)

	Estimate	Std. Error	t-value
LRB	0.16651	0.009532	17.47

Formula: log(duration) ~ LRB × SBR + (1|speaker)

	Estimate	Std. Error	t-value
LRB	0.36630	0.03499	10.469
SBR	0.52232	0.04895	10.670
LRB:SBR	-0.13394	0.0158	-8.478

Table 7: Parameters for models evaluating the predictions of SBR (upper panel), of LRB (middle panel), and a combined model.

respective number of parameters (only fixed effects and interactions) are listed for each model under consideration in Table 8.

Model	R^2	# of fixed effects
SBR	0.50	1
LRB	0.50	1
SBR × LRB	0.63	3
Sim	0.24	1
Prox	0.74	1
Prox × Sim	0.77	3

Table 8: Model comparison.

Evidently, the best model in terms of model fit is the Proximity/Similarity model, which clearly outperforms the combined SBR/LRB model. Note that both models make use of three fixed parameters (two main effects plus interaction term).[8] Therefore, the success of the Proximity/Similarity model is not simply due to the model's complexity. A model with Proximity as sole predictor fares second best, still outperforming the combined SBR/LRB model. However, the inclusion of Similarity is justified in that it significantly improves model fit, as determined by an analysis of variance comparing the simple Proximity model with a combined Proximity/Similarity model (χ^2=43.923, df=2, p<0.001).

The success of the Proximity/Similarity model is probably due to the fact that it accounts for the weakening of a prosodic boundary between two names that are grouped together. Neither the SBR nor the LRB covers this effect. Instead, these models predict that the boundary after a left member of a grouped constituent is equivalent to the boundaries in the flat baseline structure.

All in all, the model comparison approach taken here suggests that the formulation of the Proximity/Similarity model has proven to be valuable. However, whether this model can account for the prosodic rendering of other syntactic environments is an open issue.

5 Perception experiment

As observed in the production experiment, the different syntactic groupings are reflected in different prosodic renderings.

The following perception experiment is conducted to answer the question whether listeners make use of the prosodic differences between the conditions, i.e. whether

the actual duration values. R^2 can take values between 0 and 1 with 1 indicating a perfect fit.

[8] All models also include the random effects parameter "speaker," so no difference in model fit is attributable to this parameter.

the appropriate syntactic structure is recoverable from the prosodic form. Specifically, we wanted to find out whether listeners recognize the syntactic structure that is determined by (recursive) syntactic embedding and the branching direction on the basis of prosodic information.

5.1 Predictions

The production experiment has revealed that each of the six syntactic conditions has a unique prosodic signature. Uniqueness of prosodic rendition, however, does not guarantee that the different conditions are easily discernable. How well the conditions can be recognized in perception depends for one thing on how strongly the conditions differ from each other in terms of prosodic rendition. Conditions that are marked by striking prosodic features are certainly more easily discernable compared to conditions that more closely resemble other conditions. That is, the higher the prosodic markedness, the better a certain syntactic structure may be recognized.

On the other hand, it may be more difficult for listeners to recognize syntactically complex structures, as these require higher processing costs. Accordingly, structures with recursive embedding should be more difficult to recognize than simply embedded structures.

Since the different left-branching structures (conditions 4.3 and 4.5) were marked by a very distinct upstep of boundary tones, it is hypothesized that these structures are more easily discernable than the right-branching structures (4.2 and 4.4), which all show a regular downstep pattern and more closely resemble the baseline pattern (condition 4.1).

Furthermore, we hypothesize that the conditions with recursive embedding (4.4 and 4.5) are more difficult to recognize than simply embedded structures (4.2, 4.3, and 4.6) or the baseline (4.1).

5.2 Methods

For each of the six conditions with four names, one sentence per speaker was arbitrarily chosen for the perception experiment. Correspondingly, the 21 speakers each contributed one sentence per condition (21 speakers × 6 conditions). The 126 resulting sentences were distributed over 3 blocks (each with 42 items) with speaker and conditions counterbalanced across blocks. In each block, the order of items was pseudo-randomized such that sentences of the same condition or the same speaker had a minimal distance of three items.

For each block, the individual sound files were pasted into a single sound string in the order determined by the randomization procedure. Each sentence was preceded by the auditory presentation of the sequence number spoken by the first author. The inter-stimulus interval was set to 4 seconds. The record level of the individual sounds was adjusted to 70db using an automated normalization procedure in praat. Forty-five listeners (15 per block) were equipped with an answer sheet and listened to the sequence of 42 experimental sentences over headphones. On the answer sheet, the six conditions were presented as abstract groupings with parentheses next to the corresponding item number. The format of the grouping is exemplified in (14) for condition 4.4.

(14) N1 (N2 (N3 N4))

While listening, the participants were asked to note on the answer sheet for each item which of the six conditions it belonged to by ticking the respective answer box. The presentation speed was determined by the recording. Listeners could not stop the presentation to listen again.

5.3 Results

Of the total 1890 presented items, 28 (1.5%) received no or no clearly identifiable response. These items were excluded from further analysis. For the 1862 (98.5%) valid responses, the confusion matrix in Table 9 shows the distribution with the presented condition tabulated against the condition chosen by the listeners.

	Chosen condition							Recognition
	4.1	4.2	4.3	4.4	4.5	4.6	total	precision
4.1 N1 N2 N3 N4	**260**	8	11	6	15	11	311	.84
4.2 N1 N2 (N3 N4)	15	**190**	16	20	23	44	308	.62
4.3 (N1 N2) N3 N4	5	14	**231**	11	33	17	311	.74
4.4 N1 (N2 (N3 N4))	10	127	20	**113**	10	28	308	.37
4.5 ((N1 N2) N3) N4	3	8	21	7	**264**	7	310	.85
4.6 (N1 N2) (N3 N4)	2	24	19	3	1	**265**	314	.84
total	295	371	318	160	346	372	1862	

Table 9: Confusion matrix tabulating the presented condition (rows) against the condition chosen by the listeners (columns).

The conditions were recognized overall with an accuracy of 71%, which is well above chance level (16.67%). The recognition precision for the presented conditions 4.1, 4.5 and 4.6 exceeds 80%; conditions 4.2 and 4.3 were recognized correctly less often (62% and 74% respectively).

As for the baseline 4.1 (84% recognition precision), the few misclassifications (n=51) are relatively equally distributed across the competing conditions. The precision for the complex right-branching condition 4.4 is by far the lowest with only 37%. When presented with condition 4.4, listeners chose the simple right-branching structure 4.2 more often than the target structure (n=127, 41%). That is, while listeners often recognized the branching direction correctly, they had problems identifying the depth of embedding in the right-branching structures. The confusion between 4.2 and 4.4 is asymmetric, however: if the simple right-branching structure 4.2 was presented, listeners correctly recognized it in 62% of the cases and most confusion occurred with condition 4.6 which was incorrectly chosen in 44 cases (14%). Note that, like 4.2, condition 4.6 also involves a grouping of the last two names.

Compared to the right-branching structures, the left-branching conditions 4.3 and 4.5 are not as prone to confusion with 74% and 85% correct classifications respectively. As for 4.3, most of the few incorrect classification answers concern condition 4.5; conversely, when listeners misclassified 4.5, they chose the simple left-branching structure 4.3 most often. That is, if listeners were presented with a left-branching structure (simple or complex) they recognized a left-branching structure in 88% of cases.

When presented with condition 4.6 (recognition precision 84%), most of the few misclassifications concerned the simple left-branching or the simple right-branching structure. Note that, just as 4.6, both 4.2 and 4.3 show strengthening of the prosodic boundary on N2; compared to N2, N3 is downstepped and significantly shorter in these conditions. This prosodic similarity might well explain the pattern of confusion.

For the statistical model, which evaluates the effects of syntactic embedding and branching direction on the recoverability of the structures, the following coding scheme was applied (cf. Table 10): For the first factor, syntactic embedding, the condition without embedding (baseline 4.1) was coded as 0, conditions with simple grouping (conditions 4.2, 4.3 and 4.6) were coded as 1 and conditions with multiple embedding (conditions 4.4 and 4.5) were coded as 2. For the second factor, branching direction, the left-branching conditions 4.3 and 4.5 were coded as 1, and the right-branching conditions 4.2 and 4.4 were coded as -1. Conditions 4.1 and 4.6, which lack a clear branching direction, were coded as 0.

	Condition	Embedding	Branch. Dir.
4.1	N1or N2 or N3 or N4	0 (flat)	0 (neutral)
4.2	N1or N2 or (N3 and N4)	1 (simple)	1 (right)
4.3	(N1and N2) or N3 or N4	1 (simple)	-1 (left)
4.4	N1or (N2 or (N3 and N4))	2 (double)	1 (right)
4.5	(N1and N2) or N3) or N4	2 (double)	-1 (left)
4.6	(N1and N2) or (N3 and N4)	1 (simple)	0 (neutral)

Table 10: Coding scheme for evaluation of perception experiment.

	Estimate	Std. Error	z value	p value
Embed	-0.6156	0.1297	-4.747	<0.001
Branch	0.7420	0.3250	2.283	0.0224
Embed × Branch	-1.1200	0.2093	-5.351	<0.001

Table 11: Results of the GLMM on the perception data

A generalized linear mixed model (GLMM) with item, speaker and listener as random effects yields significant main effects for the fixed predictors embedding and branching direction as well as for the interaction. The results of this model, shown in Table 11, confirm that i) left-branching structures are more easily recognized than right-branching structures and ii) that increasing depth of embedding hampers recognition. The significant interaction reflects the fact that embedded left-branching structures are much less prone to confusion than embedded right-branching structures. Note that the doubly nested left-branching structure has the highest recognition precision of all conditions, while the doubly nested right-branching structure was recognized worst.

5.4 Discussion

As predicted, the left-branching conditions were better recognized than the right-branching conditions. Also, conditions with deeper embedding are more difficult to recognize than those with flatter structure, unless the former are clearly left-branching ones. The high recognition precision on the doubly nested, left-branching condition suggests that syntactic complexity does not hamper recognition if appropriate prosodic cues are provided. In contrast, the overall low precision on the right-branching structures reflects the shortage of adequate cues in these conditions.

Correspondingly, these results are best explained with recourse to the prosodic realization of the various conditions in the production experiment. The left-branching structures exhibit a distinct upstep pattern and clear pauses, which mark constituent boundaries. Such strong prosodic markedness is absent in the right-branching structures, which show regular downstep and thus resemble the baseline. As discussed above, upstep is particularly clear on a constituent that is subject to Anti-Proximity when it is preceded by a constituent that is subject to Proximity. We suggest it is the specific upstep patterns and the corresponding boundary cues that make the left-branching structures easily recognizable. The depth of embedding has additional prosodic effects, namely the lengthening of simplex constituents in structures with grouped constituents (effect of Similarity). Although significant, this effect turned out to be rather weak in production and it might therefore only have had little effect on recognition in the perception experiment.

6 General discussion

6.1 The effects of Proximity and Similarity

Our experiments confirm that speakers use prosody for the rendition of syntactic grouping and embedding of coordinated names, thus disambiguating otherwise ambiguous structures. Conversely, listeners use prosody to retrieve the configuration intended by the speaker.

The two principles, Proximity and Similarity, account for the specific prosodic structure of the various grouping conditions in our experiment. The first principle, Proximity, accounts for the lower pitch and shorter duration observed on the left member of groupings compared to the flat structure of the baseline. Anti-Proximity has the opposite effect and strengthens the boundary between two constituents not grouped together. Such a boundary is expressed by longer duration and a greater hight of the high boundary tone. The second principle, Similarity, accounts for the observation that simplex elements in an expression containing groupings are lengthened. Arguably, this increased duration of simplex elements serves to achieve similar prosody to complex elements at the same level of embedding. The two principles guarantee that both branching direction and the depth of embedding have prosodic correlates.

A comparison of the Proximity/Similarity model with other models of prosodic boundary strength attests the P/S model's predictive power, at least for the structures tested in this experiment. The model comparison also reveals that the Proximity principle accounts for a much greater portion of the variance compared to the Similarity factor.

Although all conditions under scrutiny are distinguishable by virtue of prosody, the results show that prosodic cues are distributed asymmetrically: while right-branching structures are more similar to the flat baseline, left-branching structures are marked extensively by upstep and pauses at grouping boundaries. Accordingly, left-branching structures are more easily discernable in perception and significantly less prone to confusion than right-branching structures.

6.2 Recursion in prosodic structure

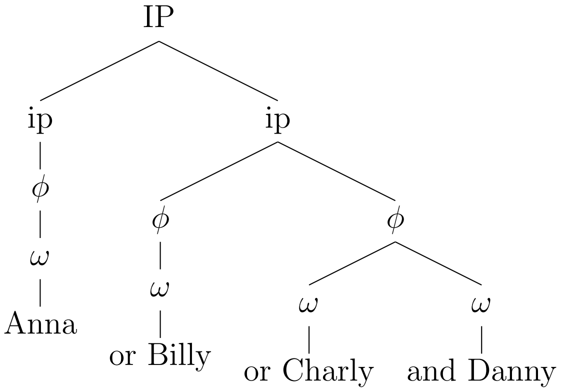
Recursion is understood as the property of grammatical constituents of being embedded in constituents of the same kind. A sentence can be embedded in another sentence, or a noun phrase in another noun phrase. This property is uncontroversial for the syntactic structure of most languages. Traditional accounts of prosodic phonology explicitly deny that the same is true of prosodic structures, and the Strict Layer Hypothesis (SLH) of Selkirk (1984) and Nespor and Vogel (1986/2007) forbids recursion in prosody. In such a model, prosodic constituents can only iterate, that is, constituents of the same level can appear in a row but they cannot be organized hierarchically.

Based on the results of the production experiment, we claim that recursion in prosodic phrasing is a necessity if we do not want to allow uncontrolled profusion of additional prosodic levels.

The fine gradation of prosodic boundary strength, which systematically reflects the branching direction and the level of embedding, makes it difficult to interpret the results in terms of a strictly layered prosodic hierarchy that disallows recursion. Especially problematic is the ban on merging unlike prosodic categories, which the SLH imposes. If we conform to the SLH, in order to represent the prosody of a doubly nested coordinated NP made up of simple names (conditions 4.4 and 4.5 of the experiment), at least 4 prosodic categories are necessary. For demonstration, we may use the widely adopted categories ω (phonological word), ϕ (phonological phrase), ip (intermediate phrase) and IP (intonational phrase). Assuming that the IP, which wraps the complex NP, is part of a sentence and thus embedded within a larger prosodic domain, at least one additional larger prosodic category is needed. There is, however, no obvious category which could do this job – at least none for which there is independent evidence.[9] Therefore, the consequence of the ban on recursion is the uncontrolled and undesired profusion of stipulated prosodic categories.

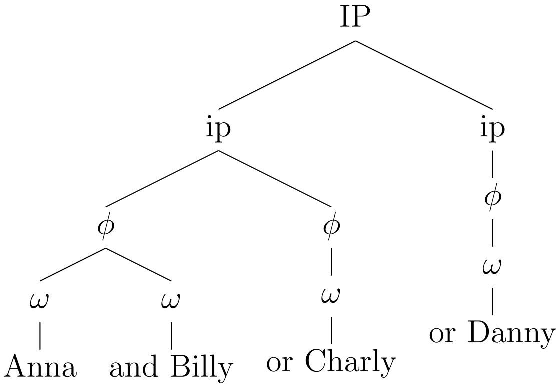
Moreover, according to the SLH, the first name in (15) would be equivalent to an intermediate phrase, even though it comprises only two syllables. The tension between the shortness of the name and its high status in the prosodic hierarchy is certainly contra-intuitive.

(15)



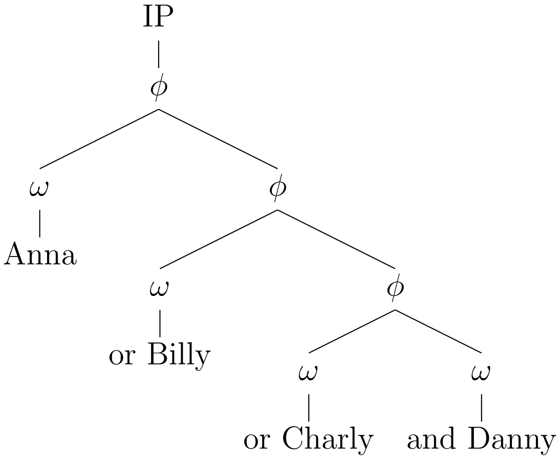
[9] Clearly the 'Clitic Group' proposed by Nespor & Vogel (1986/2007) is not an adequate prosodic domain in this context. The proper names comprise at least a prosodic foot and thus cannot be subject to cliticization.

(16)

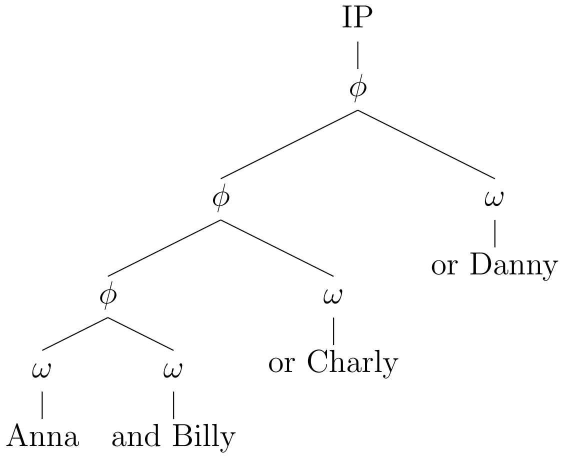


An alternative approach, which is in line with proposals by Ito and Mester (2012), Ladd (1986, 1996/2008), and Wagner (2005), explicitly allows recursion in prosodic structure. Recursively embedded syntactic NPs may thus be rendered as recursively embedded prosodic phrases. The device of recursion allows the generation of hierarchically ordered prosodic layers, without assuming different prosodic categories for each nesting level (cf. Ito and Mester, 2012). Also, in contrast to the SLH, prosodic constituents of different categories may be adjoined to form a prosodic constituent of a higher level. We assume that, in our case, each name corresponds to a prosodic word and grouped constituents form p-phrases of a higher order. The root node (or maximal prosodic projection) is represented as an intonational phrase. That way, the prosodic structure of the doubly embedded conditions can be represented much more economically (cf. (17), (18)).

(17)



(18)



An approach allowing recursion and merging of unlike prosodic categories predicts the prosodic differences between left-branching and right-branching structures that were attested in the experiment – differences that are not predicted within the SLH approach. Consider the representations that conform to the SLH. For both the right-branching (15) and the left-branching structure (16), the SLH predicts one ip-boundary, one ϕ-boundary and one ω-boundary between the four names (albeit in different orders); this would suggest that the prosodic structures should be equally complex – irrespective of the branching direction. In contrast, the recursive representation rightly predicts a difference in prosodic complexity between the two conditions: while (17) features no internal right boundary of a ϕ-phrase, (18) features two right edges of ϕ (after the 2nd and the 3rd name, respectively); in line with this representation, the left-branching structure proved to be prosodically more articulate in the experiment.

Given these considerations, we take our results to support the notion of recursion in prosodic structure. To sum up, we suggest that recursion of prosodic structure is clearly visible in German, and that speakers use it to disambiguate complex syntactic structure. The presence of prosodic recursion may be a feature of German (and other intonation languages), and does not need to be universal. Indeed, in an identical experiment with Hindi, reported in Féry and Kentner (2010), we showed that Hindi does not reveal the same prosodic features that have led us to assume recursion in German.[10]

[10] An additional difference between German and Hindi is the robust head-final nature of Hindi as opposed to head-initiality in part of the syntax of German. It remains to be tested whether the 'articulate' prosody of German left-branching structures as opposed to the apparently inflexible prosody in Hindi is due to the difference with respect to head directionality between the two languages.

7 Conclusion

In this paper, we have shown the results of a production experiment with German speakers uttering sequences of three and four coordinated names, with different syntactic groupings. Our experiment was inspired by Wagner's (2005) work on English. The names were grouped in right- and left-branching structures, and two (of six) conditions for four names showed embedding of a group of names into a larger one. Groupings of names were always binary. A follow-up perception experiment was also performed in which other German speakers listened to the structures of the production experiment and had to decide which exact structure they had just heard. The results of both experiments were straightforward. German speakers and listeners heavily rely on prosody to disambiguate syntactic structure. Right-branching structures resemble the baseline, a sequence of names without any grouping, whereas left-branching patterns had different, more articulate realizations. Each single pattern had its own prosodic contour, although some patterns were more similar to each other than others.

We propose that the prosodic patterns are best accounted for by two principles called Proximity/Anti-Proximity and Similarity. Proximity claims that the default prosodic boundary separating each name from the next one is weakened when both names are grouped together. Anti-Proximity predicts strengthening of the boundary between two names that are not syntactic sisters. And Similarity requires that elements at the same level of syntactic embedding be separated by similar prosodic boundaries. While the Similarity component alone has relatively little predictive power, the Proximity/Similarity model as a whole is superior to both the Left hand side/ Right hand side Boundary Hypothesis (LRB) of Watson and Gibson (2004) in which the size of the preceding and of the following syntactic constituents are the predictors for the likelihood of intonational phrase boundaries, and the Scopally Determined Boundary Rank (SBR) of Wagner (2005), which relates the strength of prosodic boundaries to syntactic levels of embedding rather than to the size of adjacent constituents.

As for the prosodic structure of German, the conclusion presenting itself is that recursion has to be assumed. The traditional Strict Layer Hypothesis of Selkirk (1984) cannot account for the kind of embedded structure exemplified in the paper. This confirms results of Féry and Schubö (2010) that showed the necessity of recursive prosodic structures in German.

References

Baumann, S., Trouvain, J., 2001. On the prosody of German telephone numbers. In: 7th European Conference on Speech Communication and Technology. Aalborg, Denmark, pp. 557–560.

Beckman, M., Ayers, G., 1997. Guidelines for ToBI labelling. The OSU Research Foundation.

Boersma, P., Weenink, D., 2009. Praat: Doing phonetics by computer (version 5.1.12). Software developed at the Institute of Phonetic Sciences, University of Amsterdam.

Clifton, C., Carlson, K., Frazier, L., 2002. Informative prosodic boundaries. Language and Speech 45 (2), 87–114.

Clifton, C., Carlson, K., Frazier, L., 2006. Tracking the what and why of speakers' choices: Prosodic boundaries and the length of constituents. Psychonomic Bulletin & Review 13 (5), 854–861.

Cooper, W., Paccia-Cooper, J., 1980. Syntax and speech. Vol. 3. Harvard University Press.

Ferreira, F., 1993. Creation of prosody during sentence production. Psychological Review 100 (2), 233–253.

Féry, C., Kentner, G., 2010. The prosody of embedded coordinations in German and Hindi. In: Proceedings of Speech Prosody, 5th International Conference. Chicago, Illinois, pp. 1–4.

Féry, C., Kügler, F., 2008. Pitch accent scaling on given, new and focused constituents in German. Journal of Phonetics 36 (4), 680–703.

Féry, C., Schubö, F., 2010. Hierarchical prosodic structures in the intonation of center-embedded relative clauses. The Linguistic Review 27 (3), 293–317.

Féry, C., Truckenbrodt, H., 2005. Sisterhood and tonal scaling. Studia Linguistica 59 (3), 223–243.

Frazier, L., Carlson, K., Clifton, C., 2006. Prosodic phrasing is central to language comprehension. Trends in Cognitive Sciences 10 (6), 244–249.

Gee, J., Grosjean, F., 1983. Performance structures: A psycholinguistic and linguistic appraisal. Cognitive Psychology 15 (4), 411–458.

Grabe, E., 1998. Pitch accent realization in English and German. Journal of Phonetics 26 (2), 129–143.

Hunyadi, L., 2006. Grouping, the cognitive basis of recursion in language. Argumentum 2, 67–114.

Ito, J., Mester, A., 2012. Recursive prosodic phrasing in Japanese. In: Borowsky, T., Kawahara, S., Sugahara, M. (Eds.), Prosody matters. Equinox Publishing, London, pp. 280–303.

Kentner, G., 2007. Length, ordering preference and intonational phrasing: Evidence from pauses. In: Proceedings of the 8th Annual Conference of the International Speech Communication Association. Antwerp, Belgium, pp. 1385–1388.

Ladd, D., 1986. Intonational phrasing: The case for recursive prosodic structure. Phonology 3 (1), 311–340.

Ladd, D., 1992. Compound prosodic domains. Occasional Papers from the Linguistics Department.

Ladd, D., 1996/2008. Intonational phonology. Cambridge University Press.

Lehiste, I., 1983. The many linguistic functions of duration. The Journal of the Acoustical Society of America 73, S65.

Lerdahl, F., Jackendoff, R., 1983. A generative theory of tonal music. The MIT Press.

Nespor, M., Vogel, I., 1986/2007. Prosodic phonology. Mouton de Gruyter.

Pierrehumbert, J., 1980. The phonology and phonetics of English intonation. Ph.D. thesis, MIT, distributed by Indiana University Linguistics Club, Bloomington.

Pierrehumbert, J., Hirschberg, J., 1990. The meaning of intonational contours in the interpretation of discourse. In: Cohen, P., Morgan, J., Pollack, M. (Eds.), Intentions in communication. The MIT Press, pp. 271–312.

Price, P. J., Ostendorf, M., Shattuck-Hufnagel, S., Fong, C., 1991. The use of prosody in syntactic disambiguation. Journal of the Acoustical Society of America 90 (6), 2956–2970.

Schubö, F., 2010. Recursion and prosodic structure in German complex sentences. Unpublished MA Thesis.

Selkirk, E., 1984. Phonology and syntax. MIT Press Cambridge, Mass.

Truckenbrodt, H., 2002. Upstep and embedded register levels. Phonology 19 (1), 77–120.

Wagner, M., 2005. Prosody and recursion. Ph.D. thesis, MIT.

Watson, D., Gibson, E., 2004. The relationship between intonational phrasing and syntactic structure in language production. Language and Cognitive Processes 19 (6), 713–755.

Wertheimer, M., 1938. Laws of organization in perceptual forms. In: Ellis, W. (Ed.), A source book of Gestalt psychology. London: Routledge & Kegan Paul, pp. 71–88.

Wightman, C., Shattuck-Hufnagel, S., Ostendorf, M., Price, P., 1992. Segmental durations in the vicinity of prosodic phrase boundaries. Journal of the Acoustical Society of America, 1707–1717.