# A PROSODIC RESOLUTION OF GERMAN CASE AMBIGUITIES

Tina Bögel
University of Konstanz

**Abstract**

Previous work shows that speakers and listeners can use prosodic information to clarify the meaning of syntactically ambiguous sentences. For German case ambiguities, earlier experiments reveal a significant effect of fundamental frequency, segmental duration and pause duration. This paper reports on an experiment with fully ambiguous sentences caused by case ambiguities in the German determiner system. Its purpose is to find generalisations over the different prosodic cues, thus providing the attributes and their values, which are responsible for the prosodic disambiguation of the German case phenomenon. The resulting data is integrated into the p-diagram, which allows for a fine-grained representation of the speech signal and corresponds with the syntactic module via LFG's parallel projection architecture. The information stored in the p-diagram is extracted in the respective syntactic rules via OT-like phrase structure constraints, which allow for a detailed interpretation of the rather complex results.

# 1 Introduction

Discourse context usully calls for a single reading and previous work shows that speakers and listeners can use prosodic information to clarify the meaning of syntactically ambiguous sentences. Based on the assumption that the prosodic and the syntactic component are parallel, but interacting modules of grammar (O'Connor 2005, Bögel et al. 2009), I investigate how the interaction between the syntactic and the prosodic component can be formalized to account for the exchange between syntactically ambiguous structures and the corresponding prosodic cues that can resolve this ambiguity.

In their paper, Gollrad et al. (2010) investigate how and to which degree different prosodic cues are used by speakers to disambiguate German genitive and dative case constructions. Their findings reveal a significant effect of the fundamental frequency, of segmental duration and pause duration. However, their experimental material was only partly ambiguous in that the ambiguity was always resolved by the choice of the clause final verb. Furthermore, their data set involved three DPs for each sentence, adding a neutral object before the clause final verb. This paper, on the other hand, reports on an experiment that extends and modifies the experiment conducted by Gollrad et al. (2010), in that the data set also includes fully ambiguous sentences containing only two DPs as in (1).

(1) Überraschend   antwortete   [der Diener]             [der Gräfin]
    Surprisingly    answered      the.MASC.NOM servant    the.FEM.GEN/DAT duchess

   'Surprisingly, the duchess's servant answered // the servant answered the duchess.'

---

The aim of this experiment was to identify relevant prosodic cues that could, in principle, be used from a machine processing perspective. Speakers usually apply several prosodic cues in combination to express a certain case construction. For a grammar to cover as many speakers as possible, the less frequent cues (i.e., the ones usually not prominent enough to have a significant result over all speakers of a sample) should be taken into consideration as well.

The results give a clear indication as to which attributes of a speech signal are relevant for a thorough interpretation. In section 4, these attributes are integrated into the p-diagram, first introduced by Bögel (2012), which allows for a fine-grained representation of the speech signal and provides an easily accessible structure from which the relevant cues can be extracted via LFG's parallel projection architecture (e.g. Kaplan 1987, Asudeh 2006) (section 5). This information extraction is accomplished via OT-like phrase structure constraints, which allow for a detailed and satisfying interpretation of the rather complex results (section 6).

## 2 German case ambiguities

German case is mainly encoded by the determiner system. In Table 1, the determiners for the three genders and four cases of German are listed.

|  | **masculine** | **feminine** | **neuter** |
|---|---|---|---|
| **nominative** | der | die | das |
| **genitive** | des | ***der*** | des |
| **dative** | dem | ***der*** | dem |
| **accusative** | den | die | das |

Table 1: The German determiner system.

Note that there is a syncretism between the feminine form of the dative and the genitive[1], which leads to the type of ambiguity illustrated in (2), where the ambiguity in the subordinate clause is caused by the ambiguous feminine article of the second DP, which could either be dative or genitive.
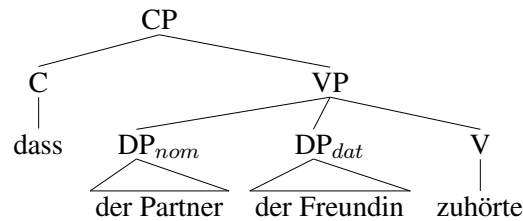
(2)  Alle                waren        überrascht        dass
     Everyone            was          surprised         that

     [der Partner]$_{DP1}$            [der Freundin]$_{DP2}$          zuhörte
     the.ART.MASC.NOM partner         the.ART.FEM.GEN/DAT friend      listened

     'Everyone was surprised that [the friend's partner listened // the partner listened to the friend].'
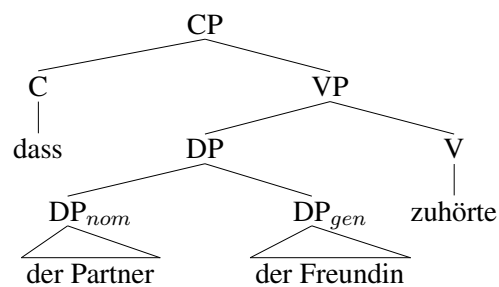
The sentence in (2) is fully ambiguous, because the final verb is intransitive, but optionally allows for a dative object as well. This full-sentence ambiguity results in two possible c-structures: In (3a), the two DPs are independent daughters of the VP, while in (3b), the two DPs form a possessive construction.

---

[1]The syncretism is also true for the masculine nominative form. However, this fact has no impact on the ambiguity of the construction analysed in this paper.

(3) a. **Dative**: The partner listened to the friend

```
                        CP
                   ┌─────┴─────────┐
                   C               VP
                   │        ┌───────┼────────┐
                  dass   DP_nom   DP_dat      V
                        ┌──┴──┐  ┌───┴───┐    │
                     der Partner der Freundin zuhörte
```

b. **Genitive**: The friend's partner listened

```
                        CP
                  ┌──────┴────────┐
                  C               VP
                  │          ┌─────┴─────┐
                 dass        DP          V
                       ┌─────┴──────┐    │
                    DP_nom        DP_gen zuhörte
                   ┌──┴──┐       ┌──┴──┐
                der Partner   der Freundin
```

While the purely syntactic analysis of ambiguous structures leads to multiple representations, the structures can be distinguished with the help of prosodic cues. One major cue that has been mentioned in previous studies (Gollrad et al. 2010, a.o.) is the position of the phonological phrase boundary. With the dative construction, this phrase boundary would be after the first DP (4a), while its position would be after the second DP in the genitive construction (4b).

(4) a. ... dass der Partner)$_{\text{PhP}}$( der Freundin ...

b. ... dass der Partner der Freundin)$_{\text{PhP}}$(...

The following experiment aims to evaluate the importance of the phonological phrase boundary and to establish further prosodic indicators for each condition.

## 3 An experiment on German case ambiguities

In their paper, Gollrad et al. (2010) identified prosodic cues which can be used for disambiguating between genitive and dative constructions. However, their study did not involve completely ambiguous structures; instead, the sentences consisted of three determiner phrases, always involving an object and a disambiguating final verb. While their results stated the overall significant indicators for each condition in their data set, they did not report on subgroups, i.e., prosodic cues only used by a subset of the speakers, which are not significant for all the speakers and are thus usually not reported on. From an automatic processing perspective, however, those cues are of interest as well, as they can be integrated via ranked constraints, weighting the single prosodic cues in accordance to their overall frequency of occurence.

In order to investigate these subgroups, to gain information on fully ambiguous sentences consisting of only two noun phrases and to gather the absolute values of the different prosodic cues, an experiment similar to the one in Gollrad et al. (2010) was conducted.

Allbritton et al. (1996) state that subjects will not consistently use prosodic cues to indicate a certain interpretation of syntactically ambiguous sentence, even if the context disambiguates the intended meaning. However, they find that if the speakers were aware of the ambiguity and were asked to pronounce the sentence according to a certain interpretation, the prosodic cues were much more distinct for each condition. In order to control for these findings as well, the experiment involved several types of constructions:

1. Ambiguous and unambiguous constructions hidden in a larger text, where the ambiguous structures were disambiguated by the context.

2. Unambiguous structures, where the two DPs were masculine.

3. Fully ambiguous structures, where the speaker was made aware of the intended meaning.

The sentence types were presented to the participants in this order; full awareness of the ambiguity and a demand for a distinct pronunciation, however, was only given with group three sentences. While I agree with Allbritton et al. on the fact that speakers do not consistently use one set of prosodic cues (this being a statement which is supported by the findings in this paper), for the here presented experiment I could not find any significant differences between the three groups and thus for the influence of speaker awareness on prosodic pronunciation.[2]

For the experiment, 15 female participants produced a total of 480 sentences. Recordings took place in a sound proof chamber.[3] All resulting 480 sentences were hand-annotated using the Praat software (Boersma and Weenink 2013) for the fundamental frequency, the duration of the syllables and for the length of pauses. Statistical analysis was done with a linear mixed effects regression model (LMER) with subject and item as crossed random factors and the two conditions (genitive and dative) as fixed factors. The statistical analysis for the whole group of speakers showed the following results:

- A significant drop in fundamental frequency from the first DP to the determiner of the second DP in the dative condition (DAT: 32.5 Hz, GEN: 23.2, $\beta$=-9.31, SE=2.64, t=-3.53).

- A significant pause between the first and the second DP in the dative condition: ($\beta$=-2.35, SE=0.92, t=-2.55).

- The duration of the last syllable of the first DP was significantly longer in the dative condition than in the genitive condition ($\beta$=-2.8, SE=0.79, t=-3.58).

While these indicators are well known (Gollrad et al. 2010, a.o.), other prosodic cues are only significant for a smaller subgroup of speakers. For this reason, the statistical analysis was also performed for each individual – only then is it possible to identify smaller subgroups and to make sure that there are not two opposing groups of prosodic cues (i.e., one group shows a drop and one group a rise within the same position and in the same condition).

Figure 1 shows a 'prototypical dative' as viewed in Praat. The top part shows an oscillogram of the signal, which takes time and amplitude (sound pressure) into account. The middle part is a spectrogram in which energy is displayed by time (horizontally) and various frequencies (vertically). The darker a phasis in the spectrogram, the higher the energy density. Below the spectrogram, another level has been added. This consists of hand-annotated reference syllables (the gloss) and a GToBI annotation (Grice and Baumann 2002), indicating High and Low pitch accents and boundary tones (for more on GToBI see section 4.3).
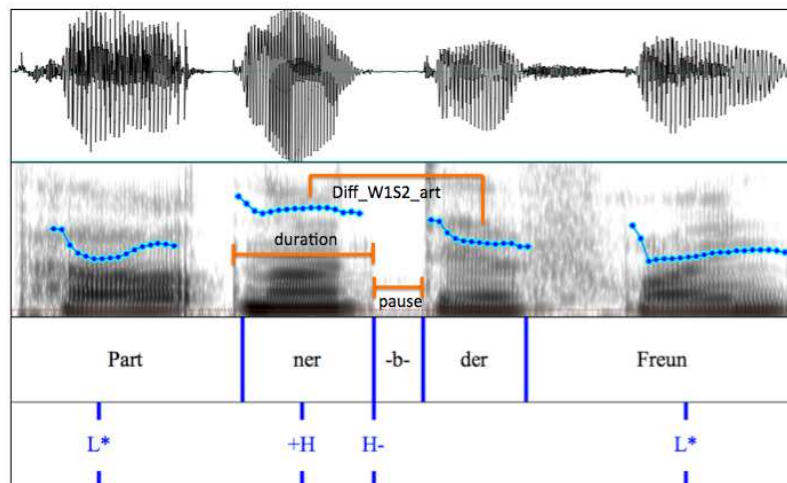


Figure 1: A prototypical **dative**.

The blue dotted lines in Figure 1 represent the (discontinuous) fundamental frequency, which shows the speech melody of the speaker. The orange solid lines have been added by hand and indicate the three most frequently used prosodic cues for the dative construction.[4]

1. **Pause**: 40% of the speakers use a statistically significant pause to indicate a prosodic break between the two DPs.

2. **Duration**: 47% of the speakers lengthen the last syllable of the first DP in a dative construction.

---

[4]The labels are abbreviations for the specific prosodic cue they indicate. **diff_W1S2_art**, for example, stands for the **diff**erence of the fundamental frequency between the second syllable **S2** of the first noun **W1** and the **art**icle.

3. **Diff_W1S2_art**: 40% of the speakers have a significant drop in the fundamental frequency from the last syllable of the first DP to the following determiner of the second DP.

While all of these indicators are significant if measured for all participants, they are certainly not true for each individual. Participants used different indicators and combinations thereof to indicate the dative construction. The same is true for the genitive construction.
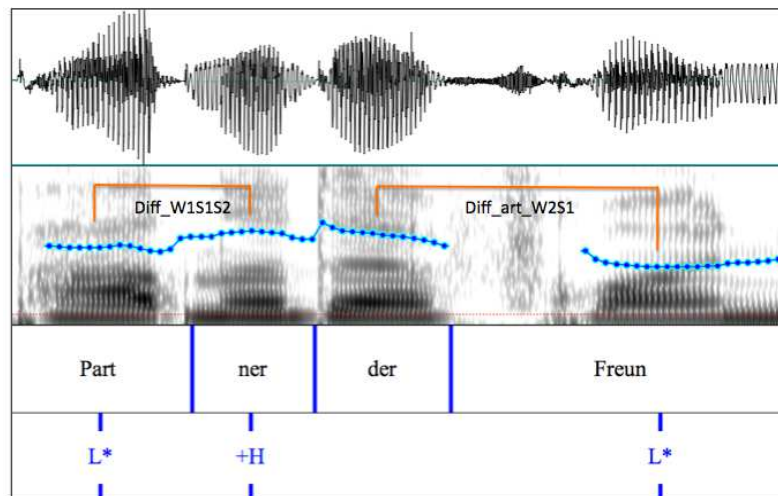


Figure 2: A prototypical **genitive**.

First of all, all of the above mentioned ('dative') indicators are significant for the genitive as well in that they are *not* present. Indicators visible in the genitive speech signal are

1. **Diff_W1S1S2**: 27% of the participants show a smaller difference in fundamental frequency between the first and the second syllable of the first noun.[5]

2. **Diff_art_W2S1**: 20% show a drop from the fundamental frequency value of the determiner *der* to the first syllable of the following noun.

The results found in the study show more prosodic cues, but these failed to satisfy the following constraints which were introduced to ensure a certain significance in the overall strategies on the one hand and to avoid strategies which excluded each other on the other hand. Thus, for a certain speech signal phenomenon to be considered as a subgroup, the following constraints had to be met:

1. The phenomenon must be statistically significant for at least 20% of the speakers. As a consequence, prosodic cues that were only applied by, e.g., 10% of the speakers were ruled out.

---

[5]It is not clear if the L*+H GToBI annotation is justified here, but I left it in to indicate that there still is a rise in fundamental frequency from the first to the second syllable.

2. No phenomenon that is contrasted by the phenomenon of another subgroup may be used. This excludes subgroups, who for example used a High tone on syllable X in the genitive condition if another subgroup used a Low tone on the same syllable X in the genitive condition as well.

From a machine processing perspective, all of the above described phenomena should be included in the grammar in order to cover as many speakers as possible. After an introduction to the p-diagram (Bögel 2012), which allows for a detailed analysis of the speech signal, I will introduce the necessary attributes used to indicate the different prosodic cues. The speech signal information encoded in the p-diagram can then be retrieved by c-structure annotation rules via LFG's correspondence architecture (section 5). In order to measure up to the frequency of occurence of the different prosodic cues discussed in this section, OT-like phrase structure constraints will be applied (section 6).

# 4   The p-diagram

Building on the proposal made by Bögel (2012), the p-diagram allows for a fine-grained representation of the speech signal. Following Dalrymple and Mycock (2011) with the additions in Bögel (2012), I assume that each string has (at least) two representations: The p-string is the abstract (IPA-encoded) representation of the speech signal which is, by definition, a sound wave and thus not visible to the eye. The s-string is the functional-morphological representation. The p-string/the speech signal is encoded in the p-diagram within p-structure and the s-string is part of the syntactic module.

The information provided by the p-diagram for a certain string can be further processed by the syntactic module, allowing, for example, for a distinction between syntactically ambiguous structures based on prosodic cues. The lexicon functions as a look-up instrument and includes syntactic, phonological and semantic information on each word and is thus at the interface between s-string and p-string

Figure (3b) shows part of the p-diagram representation of a typical dative construction with the respective c-structure representation in (3a):

a.

b.

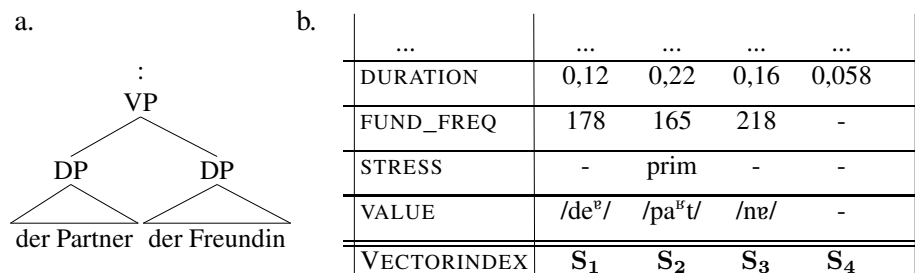| | ... | ... | ... | ... | ... |
|---|---|---|---|---|---|
| DURATION | 0,12 | 0,22 | 0,16 | 0,058 |
| FUND_FREQ | 178 | 165 | 218 | - |
| STRESS | - | prim | - | - |
| VALUE | /deɐ/ | /paʁt/ | /nɐ/ | - |
| VECTORINDEX | $S_1$ | $S_2$ | $S_3$ | $S_4$ |

Figure 3: c-structure and (reduced) p-diagram of
*der Partner der Freundin* (dative).

The speech signal is interpreted horizontally syllable by syllable, which allows a linear representation of phonemes (forming the p-string), pauses, duration and other attributes. However, the signal is also represented vertically attaching the different layers of the speech signal to each syllable, e.g. the duration and the fundamental frequency.

Each syllable and each pause receives a vector S which encodes the speech signal information on the one hand, i.e., the different aspects of the speech signal and their respective value at the time the syllable in question is uttered. On the other hand, the lexical information (phonological information and information on lexical stress) are also stored in the vector. For example, for the first syllable of the word 'Partner', the corresponding vector is as in (5):[6]

(5)

$$
\text{S:} \begin{pmatrix} value \\ stress \\ F_0 \\ duration \\ ... \end{pmatrix} (n) \quad \text{which would yield, e.g.,} \quad \text{S:} \begin{pmatrix} /pa^{ʁ}t/ \\ prim \\ 165Hz \\ 0,22s \\ ... \end{pmatrix} (2)
$$

These vectors are then merged into the p-diagram (Figure 4), enabling a fine-grained representation of the original speech signal. Once encoded in the p-diagram, the information can be easily extracted; for example, the function (S2 DURATION) = 0,22 refers to the second vector's value for the attribute 'duration' which is 0,22 seconds.

| ... | ... | ... | ... | ... | INTERPRETATION |
|---|---|---|---|---|---|
| SEMIT_DIFF | .. | -1.3 | 4.8 | - | ↓ |
| GToBI | - | L* | +H H- | - | |
| BREAK_IND | - | - | - | 3 | |
| PHRASE | - | - | - | PhP | |
| DURATION | 0,12 | 0,22 | 0,16 | 0,058 | SIGNAL |
| FUND. FREQ. | 178 | 165 | 218 | - | ↓ |
| STRESS | - | prim | - | - | LEXICON |
| VALUE | /de$^ɐ$/ | /pa$^ʁ$t/ | /nɐ/ | - | ↓ |
| VECTORINDEX | $S_1$ | $S_2$ | $S_3$ | $S_4$ | ... → |

Figure 4: The p-diagram of /de$^ɐ$'pa$^ʁ$tnɐ/ ('the partner')

The p-diagram itself is divided into several parts, which draw on different aspects of the speech signal as explained in the following sections.

---

[6]Note that for a pause vector, most values will be empty, except for the duration attribute.

## 4.1 The lexical level

This paper follows the work by Levelt et al. (1999), who divide lexical access into three aspects: concept, lemma and form. The lexical entry as proposed by this paper also divides the lexical string into three aspects: concept, s-form and p-form.

| concept | s-form | p-form |
|---------|--------|--------|
| [partner] | Partner | /ˈpaʁt.nɐ/ |
| ... | ... | ... |

(6) shows the lexical entry associated with the s-form 'Partner' as it is encoded in the lexical section of an (XLE) LFG grammar.

(6)  Partner  N  (↑ PRED) = 'Partner'
                  (↑ GEND) = masc

While this is a standard LFG entry, the information associated with the p-form requires further explanation. If dealing with isolated words, the syllables are clearly set in our mind. However, if the context changes, the syllables' clear boundaries may be lost during runtime. Resyllabification might take place, where the coda of the previous syllable is drawn to the onset of the next syllable.

This is one of the reasons why Levelt et al. (1999) propose that the lexicon stores the segments of a morpheme, but does not group these segments into syllables. Instead, a metrical frame is stored along with the segments,[7] which contains information on the number of syllables and the distribution of stress. This stress distribution can change with a morphological process triggering lexical phonology, e.g., if a specific derivational affix is added.[8]

During runtime, the metrical form and the segments are merged, respecting postlexical phonological rules like, e.g., syllabification. For the word 'Partner', the p-form entry would thus look like the following:

(7)  <partner>  SEGMENTS       /p aʁ t n ɐ/

                METRICAL FRAME   $\overset{x}{(\sigma\sigma)}$

Depending on the context (i.e. the following segments and metrical frames) and according to the rules which form a syllable, the segments are then merged into the frame. For the word 'Partner', the respective p-form would be /ˈpaʁt.nɐ/. Since *tn is an illegal consonant cluster in the onset of a German syllable, the /t/ becomes part of the coda of the first (stressed) syllable while /n/ occupies the onset position of the second syllable.

---

[7]Note that the segments here merely represent phonological feature bundles as described by Lahiri and Reetz (2010). However, for the purpose of this paper, each segment is represented as an IPA representation.

[8]A widely known example, e.g. discussed in Gussenhoven (1991), is the stress difference between 'Japan' and 'Japanese'.

The information gathered in the lexicon is then encoded within the lowest levels of the p-diagram; VALUE represents the syllable in runtime[9] and STRESS the lexical stress assignment, which can be either primary or secondary (or unspecified).

| STRESS | - | prim | - | - | ... |
|---|---|---|---|---|---|
| VALUE | /deᵉ/ | /paᵘt/ | /nɐ/ | - | ... |
| VECTORINDEX | $S_1$ | $S_2$ | $S_3$ | $S_4$ | ...→ |

Figure 5: The lexical level of the p-diagram of /deᵉ ˈpaᵘtnɐ/ ('the partner')

Note that 'stress' here refers to *lexical* stress, i.e., the syllable of a word which carries the stress, in contrast to 'sentence stress' (pitch accent), which refers to high and low points in the fundamental frequency. Traces of lexical stress can be seen in the signal and pitch accents are usually connected to a syllable that also carries primary lexical stress.[10]

## 4.2 The signal information

Information on different aspects of the speech signal (here: DURATION and FUNDAMENTAL FREQUENCY; a possible addition could be INTENSITY) is directly transferred into the p-diagram.

| DURATION | 0,12 | 0,22 | 0,16 | 0,058 | |
|---|---|---|---|---|---|
| FUND. FREQ. | 178 | 165 | 218 | - | |
| ... | ... | ... | ... | ... | |
| VECTORINDEX | $S_1$ | $S_2$ | $S_3$ | $S_4$ | ... |

Figure 6: The signal information of /deᵉ ˈpaᵘtnɐ/ ('the partner')

In Figure 6, DURATION measures the length of a syllable or a pause in seconds. The values under FUNDAMENTAL FREQUENCY refer to the maximum Hertz value

---

[9]The syllable in runtime is, in fact, not really part of the p-diagram's lexical level as the syllabification may change due to postlexical phonological processes. The contrast between lexical and postlexical syllabification can offer valuable clues to the determination of, e.g., the size of the prosodic word. However, for now I leave the depiction of that contrast and the resulting implications to further research.

[10]However, this is not always the case, as the pitch accent can move to a syllable with secondary stress if the primarily stressed syllable is next to another primarily stressed syllable. For English, this phenomena has been described as 'rhythm rule' (e.g., Gussenhoven 1991, Shattuck-Hufnagel et al. 1994) and accounts for the stress shift in words like 'thir_teen_' in the combination 'thir_teen_ _men_'. Note also that an assignment of the pitch accent to an unstressed syllable as in 'Are you thirteen years old?' – 'No, I am thir_ty years old' can be an indication for contrastive focus. For now, the interaction between the GToBI encoding of pitch accents (section 4.3) and lexical stress is left for further research.

of the corresponding syllable, where the Hertz value was strictly measured at the middle part of the syllable nucleus.[11]

## 4.3 The interpretation level

The interpretation level does not include direct information from the speech signal, but rather interprets the information gathered at the lower levels. Since this information is partly calculated on the basis of the neighbouring syllables' values, this level of the p-diagram cannot be part of the initial vector, but is added after the basic p-diagram is created.

| SEMIT_DIFF | .. | -1.3 | 4.8 | - | |
|---|---|---|---|---|---|
| GTOBI | - | L* | +H H- | - | |
| BREAK_IND | - | - | - | 3 | |
| PHRASE | - | - | - | PhP | |
| ... | ... | ... | ... | ... | |
| VECTORINDEX | $S_1$ | $S_2$ | $S_3$ | $S_4$ | ... |

Figure 7: The interpretation level of /de$^ɐ$'pa$^ʁ$tnɐ/ ('the partner')

It is the analyst's decision which information should be gathered and interpreted in the p-diagram. For an adequate interpretation of the speech signal in the genitive-dative variation, four attributes are of use, which are described below:

1. SEMIT_DIFF refers to semitone difference and describes the difference between the fundamental frequency value between two syllables. The semitone scheme is used instead of absolute Hertz values, because the higher two Hertz values are, the smaller is the difference between them from a listener perspective; i.e., the difference between 100 and 150 Hertz is much more significant than the difference between 400 and 450 Hertz. For this reason, semitones are calculated, which measure the relative (and not the absolute) difference between two Hertz values. The formula in (8) calculates the semitone value for a Hertz-value.[12]

(8) $\text{f0max(St)} = 12 * \log_2\left(\frac{f0max(Hz)}{f0min(Hz)}\right)$

---

[11]This way, measuring irregularities which are quite common in Praat at the border of syllables, especially if non-sonorant consonants are involved, can be avoided.

[12]There are several possibilities to calculate semitones: First, the maximum value of the current syllable relative to the minimum value of the fundamental frequency of the utterance. This will apply for all subjects, independently of their pitch range. Second, the semitones can be calculated relative to, e.g., a 100 Hz, in case the minimum value cannot be determined. A 100 Hz will be sufficient for female participants; for male participants, the value should be set to 50 Hz or lower. However, while the absolute semitone values differ with the two formula options, the p-diagram only encodes the difference between two semitone values. This difference value is the same with both options.

12 semitones form an octave. For the above example, the semitone difference between 100 and 150 Hertz will be ca. 7, while the semitone difference between 400 and 450 Hertz will be ca. 2. This difference between semitones gives a better indication of how significant a drop/rise in the fundamental frequency is from a listener perspective. The attribute semit_diff thus does not encode the semitone value for this specific syllable, but the semitone difference with respect to the previous syllable, i.e., the semitone value has to be subtracted from the semitone value of the previous syllable. Consequently, a negative value indicates a drop in the fundamental frequency, while a positive value indicates a rise, a representation that allows for a quick overview on significant drops/rises in the speech melody.

2. GTOBI is a set of conventions for labelling High and Low tones and break indices in German intonation, thus modelling the pitch contour of a speech signal. The GToBI inventory includes two monotonal and four bitonal pitch accents as well as boundary tones (Grice and Baumann 2002).

3. BREAK_IND refers to break indices, which indicate the value of the perceived degree of disjuncture between two words (Beckman et al. 2005). The break indices are connected to the GToBI level in that that they are also part of the annotation conventions. Break indices range from 0 (clitic boundary) to 4 (intonational phrase), thus grouping segments of speech hierarchically. For the current version of the p-diagram, only break indices 3 (phonological phrase) and 4 are taken into account.

4. PHRASE indicates larger (and thus more easily identifiable) prosodic phrasing boundaries, i.e., the phonological phrase and the intonational phrase. These are calculated on different parameters, e.g. the break indices, the boundary tones or the duration of the previous syllable.

While all attributes of the interpretation level could, in principle, be of use for the interpretation of speech signal phenomena by c-structure annotations, this paper will only refer to SEMIT_DIFF and PHRASE to retrieve the necessary information.

## 5 The P-diagram's position within the LFG architecture

Based on the assumption that the prosodic and the syntactic component are parallel, but interacting modules of grammar (O'Connor 2005, Bögel et al. 2009), I assume the string to be the central link between the two components (cf. Dalrymple and Mycock 2011). Figure 8 shows the p-diagram's integration into LFG as exemplified by a dative construction.
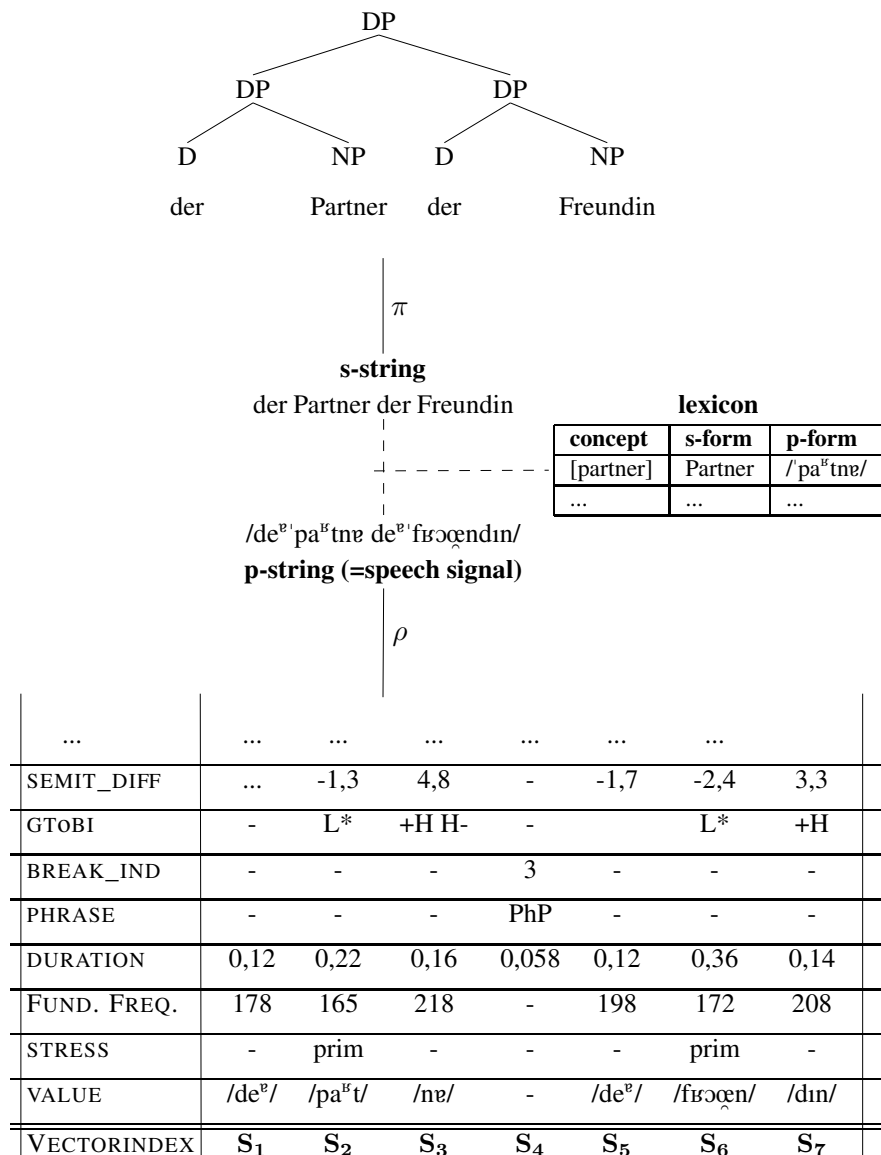
Figure 8: The prosody-syntax interface.

The p-string is the string's abstract representation of the speech signal; the s-string is the string's morphosyntactic representation. P- and s-string are two sides of one coin, connected via the reference to the lexicon, which includes s-form and p-form of each concept unit (see section 4.1).

Two modules are connected to the string: The $\pi$-relation connects c-structure to the (s-)string on the one side, while the $\rho$-relation connects the p-diagram to the (p-)string on the other side.

Following Kaplan (1995), I assume a structural correspondence $\pi$ which maps the string elements to c-structure. The inverse relationship, that is, the mapping from c-structure to the string can thus be defined as $\pi^{-1}$. As Asudeh (2009) states,

the $\pi$-relation projects only from string to terminal nodes. It is thus crucial to formulate further definitions for the inverse mapping from non-terminal nodes to the string.

Here, I follow (with small adjustments) Mycock and Lowe (2013) who define the terminal nodes projected from a current c-structure node $*$ as $T(*)$.[13] The projection from a c-structure node to the corresponding string is thus defined as $\pi^{-1}(T(*))$, which returns an ordered set of s-string-elements. This morphosyntactic string chunk has a corresponding p-string section identified via the reference to the lexicon, where the p- and s-form are stored for every lexical item. The p-string represents the speech signal via an IPA-transcription. It is mapped to the p-diagram in p-structure via the relation $\rho$, which maps the speech signal syllable-wise to the p-diagram (as described in section 4).

In this model, the prosody-syntax interface is the string itself. The correspondence between c-structure and p-diagram can be defined as the composition (Kaplan 1995, Asudeh 2006) of the inverse relation of the s-string and the c-structure and the relation of the p-string and the p-diagram: $\rho(\pi^{-1}(f))$. In order to simplify the reference to this relation within the phrase structure rules, the following abbreviation is used:[14]

$$\rho(\pi^{-1}(f)) \equiv \natural(f)$$

This relation allows c-structure annotation rules to refer to information which is relevant for a specific c-structure analysis, but which is encoded in p-structure in the p-diagram. The relevant values are then retrieved by means of the vector index and the attributes of interest, e.g. *(S2,* STRESS*)* (i.e., the attribute STRESS of the second vector 'S2') returns *primary* (i.e., has the value 'primary').

# 6 Disambiguating c-structure via ranked constraints

As has been shown in section 3, the prosodic indications are numerous, but ranked by frequency in that some of them are applied by almost 50% of the speakers, while others are only used by 20% of the speakers. While theoretical papers usually only consider the prosodic cues used by a *significant* number of speakers, other cues are ignored. However, from a machine processing perspective, we also want to know which strategies are applied by (at least) 20% of the speakers, albeit ranked as less frequent.

The grammar writing platform XLE (Crouch et al. 2013), in which these phenomena are to be implemented, works with 'hard' constraints in that a constraint

---

[13]While Mycock and Lowe focus on the definition of right- and leftmost terminal nodes (driven by their edge-based approach), the function $T(*)$ used in this paper will return all terminal nodes attached to a certain c-structure node.

[14]I chose the musical $\natural$ symbol, because it actually represents the relation between c- and p-structure visually with two horizontal lines and two vertical lines in the middle representing the two sides of the string.

usually either allows or prohibits a certain analysis. For phenomena as described in the previous section, Optimality Theory (Prince and Smolensky 1993, Bresnan 2000) is thus a useful extension to classical LFG theory, because it allows for constraints to be ranked. For this reason, the c-structure annotations in this paper use OT-like constraints when referring to p-structure cues. 'OT-like' in this context means that the notion of constraints is not understood as in the original Optimality Theory and that the implementation into XLE allows for several extensions, as originally proposed by Frank et al. (1998) and extended and modified in the current XLE documentation (Crouch et al. 2013). In contrast to OT, for example, which only allows for negative constraints, XLE also enables the user to mark a certain condition as being preferred via so-called preference marks. This system enables the user to soften the standard constraints provided by XLE and allows for the implementation of phenomena, whose analysis is not easily divided into 'good' and 'bad' as is the case with the prosodic cues described in section 3.

Assuming that we would like to identify a dative analysis, we can rely on the three indicators introduced in section 3: The pause between the two DPs, the duration of the DP's last syllable and the drop in fundamental frequency between the first DP and the following determiner of the second DP. All of these are reliable indicators of a phonological phrase boundary; and, as stated before in section 2, such a boundary is expected to be present between the two DPs in a dative construction. Thus, instead of writing three rules ranking each of the three prosodic cues according to their frequency in speaker production, we can use the p-diagram's interpretation level and its PHRASE attribute, which is calculated on the combination of these three cues.[15] Since the three prosodic indicators for the dative are applied by at least 40% of the speakers, the phrase boundary cue should take up an important position in a (positive) ranking.

On the other hand, if we want to implement the drop in fundamental frequency from the determiner of the second DP to the first syllable of the following noun (diff_art_W2S1), then this prosodic cue should be ranked lower than the phrase boundary cue, because only 20% of the speakers apply it. Again, it is not meaningful to measure the drop in the fundamental frequency value itself for reasons explained in section 4.3; in this case, it is the SEMIT_DIFF value at the p-diagram's interpretation level which is of interest because it normalizes the absolute Hertz values of the fundamental frequency.

In the following implementation, these two attributes (phrase and semitones) are ranked according to their frequency of occurence in spoken data. (9) shows the algorithm to express this relationship where **PHPbreak** is the OT-marker for phonological phrase break (indicating a dative) and **DiffArtW2S1** is the marker for a significant difference in the fundamental frequency between the determiner and the first syllable of the following noun (indicating a genitive).

---

[15]Note that even only one of the indicators is already a reliable cue for a prosodic phrase boundary.

(9)
$$\text{If } \textbf{PHPbreak, then Case=dat}$$
<div align="center">

If **PHPbreak**, then Case=dat
If **DiffArtW2S1** and **not PHPbreak**, then Case=gen
If **not PHPbreak** and **not DiffArtW2S1**, then Case=dat OR Case=gen

</div>

XLE allows a choice between preference and dispreference marks. Dispreference marks coincide with the original idea of OT and are generally used on rare, yet grammatical constructions. Preference marks on the other hand are unique to the XLE implementation and are applied if one reading is preferred.[16]

From the starting point of this paper and the implementation into LFG, it does not really matter if we set preference or dispreference marks. For the implementation here I chose preference marks, which are not in line with classical OT, but save an addition of the prosodically unmarked case (i.e., an expression which has neither distinct prosodic dative nor genitive cues) to the optimality order as the most dispreferred case. This would be necessary because otherwise the unmarked case would always be the preferred case. (10) shows the optimality order via preference marks, indicated by +.[17] Here, the unmarked case would only be the preferred analysis, if none of the constructions with preference marks apply. An extra marking of the unmarked case is thus unnecessary.

(10)  OPTIMALITYORDER   +PHPBREAK   +DIFFARTW2S1
$\rightarrow$ where PHPBREAK is preferred over DIFFARTW2S1, and if none of the marks are present, the unmarked analysis applies.

In (11), an implementation of the phrase structure rule annotation referring to the dative construction with the optimality order as described in (10) is shown.

(11)  XLE implementation for the **dative** construction:

$$
\begin{array}{llll}
\text{VP} \;\rightarrow & \text{DP} & \text{DP} & \text{V} \\
& \{(\natural(T(*))\; S_{max+1}\; \text{PHRASE}) =_c \text{PhP} & & \\
& (\downarrow \text{CASE}) = \text{dat} & & \\
& \text{PHPBREAK} \in \text{o}^* & & \\
& |\, (\natural(T(*))\; S_{max+1}\; \text{PHRASE}) \neq \text{PhP} & & \\
& (\downarrow \text{CASE}) = \text{dat}\} & &
\end{array}
$$

The first DP rule annotation is a disjunction, indicated by the {x | y} annotation. The $\natural(T(*))$ in the first line of part 1 refers to the (set of) terminal nodes connected with the DP and the projection between c-structure and p-structure as described in

---

[16]Note that both methods have problematic issues in that they can lead to preference/dispreference analysis between completely unrelated constructions (for more on this topic see Frank et al. 1998). Thus, an implementation with OT marks should always carefully consider other possible, but unintended, interactions.

[17]The representation is as it would appear in the XLE grammar configuration section.

section 5. $(S_{max+1}$ PHRASE$) =_c$ PhP is a reference[18] to the last vector of the set of DP terminal nodes (i.e., the one with the maximum value, which is the last syllable of the first DP) plus one, i.e., the vector *following* that last vector.[19] That vector applied to the attribute PHRASE must yield the value PhP, indicating a phonological phrase boundary. If this is the case, then the dative case is assigned and the construction receives a preference mark PHPBREAK, indicated by PHPBREAK $\in$ o*[20]. The second part of the rule refers to the case where the attribute PHRASE does not yield the value PhP. In that case, the dative is still assigned. However, if there is another analysis in competition and this analysis carries a preference mark as well, then this default analysis is overruled.

(12) shows an implementation of the phrase structure rule annotation referring to the genitive construction with the optimality order as described in (10).

(12) XLE implementation for the **genitive** construction:

$$
\begin{array}{rll}
\text{DP} & \rightarrow & \text{DP} \qquad\qquad\qquad \text{DP} \\
& & \{(\natural(T(*))\ S_{min+1}\ \text{SEMIT\_DIFF}) < \text{-2} \\
& & \qquad (\downarrow \text{CASE}) = \text{gen} \\
& & \qquad \text{DIFFARTW2S1} \in \text{o*} \\
& & |(\natural(T(*))\ S_{min+1}\ \text{SEMIT\_DIFF}) \geq \text{-2} \\
& & \qquad (\downarrow \text{CASE}) = \text{gen} \}
\end{array}
$$

This rule works similar to the one described in (11). The main difference is that it refers to a different attribute-value pair, that of semitone difference. The first part of the rule assigns a preference mark named DIFFARTW2S1 to the analysis iff the semitones value at the second position (mimimum+1, i.e., the second syllable of the DP, which is the first syllable of the noun) is smaller than -2, indicating a significant drop from the determiner to the following noun. In the second part of the rule, the default is again a genitive assignment.

This rule set in combination with the optimality order stated under (10) leads to the following possibilities:

1. An analysis, which has been assigned a PHPBREAK preference mark will be preferred over all other analyses, whether they carry a DIFFARTW2S1 preference mark or none at all.

2. An analysis, which has been assigned a DIFFARTW2S1 preference mark will be preferred if no PHPBREAK is present in another analysis.

---

[18]Note that LFG's modular architecture and its correspondence functions allows for the interaction between the otherwise separated prosodic and syntactic modules via this constraining equation, which looks up relevant information stored in the prosodic module in order to approve or disapprove of the f-structure CASE annotation in the syntactic module.

[19]The $S_{max+1}$ vector can be either the first syllable of the next word or, as it is the case here, a pause vector.

[20]This rule annotation is translated as 'the mark PHPBREAK is element of the o(ptimality)-structure', thus indicating that this part of the rule is carrying the optimality mark PHPBREAK.

3. If none of the preference marks are present, then the two default options will apply. This will cause the syntactic ambiguity that we first started with, but since in that case there are no prosodic cues disambiguating the structure, it is an ambiguity that should be present.

Figure 9 shows an example for a dative speech signal encoded into a p-diagram (a reduced version of the p-diagram in Figure 5, with only the relevant information present).

| ... | ... | ... | ... | ... | ... | ... | |
|---|---|---|---|---|---|---|---|
| SEMIT_DIFF | ... | -1,3 | 4,8 | - | -1,7 | -2,4 | 3,3 |
| PHRASE | - | - | - | PhP | - | - | - |
| DURATION | 0,12 | 0,22 | 0,16 | 0,058 | 0,12 | 0,36 | 0,14 |
| ... | ... | ... | ... | ... | ... | ... | |
| VALUE | /deɐ̯/ | /paʁt/ | /nɐ/ | - | /deɐ̯/ | /fʁɔœ̯n/ | /dɪn/ |
| VECTORINDEX | $S_1$ | $S_2$ | $S_3$ | $S_4$ | $S_5$ | $S_6$ | $S_7$ |

Figure 9: Parts of the p-diagram for the dative version of
*der Partner der Freundin.*

This p-diagram shows how important it is to rank the different constraints, as it is quite common for speakers to mix indicators. While the speaker indicates the dative with the strong prosodic cue of a phonologcial phrase break, in this case calculated on the basis of a very long pause between the two DPs (0,58s), she also applied a prosodic cue for the genitive, the semi-tone difference between S5 and S6. However, since the phrase boundary is ranked higher than the semitones difference, the optimal analysis will be the (correct) dative one.

# 7 Conclusion

In this paper I have introduced the possibility to include ranked prosodic cues into phrase structure annotations to help resolving c-structure ambiguities caused by the German dative/genitive case alternation. The case alternation is caused by syncretism between the feminine article of the dative and the genitive, leading to two possible c-structures. As previous studies have shown (Gollrad et al. 2010), German speakers disambiguate dative and genitive constructions by means of prosody. The relevant prosodic cues for each condition were established with the help of an experiment conducted for this paper. These cues were then ranked according to the overall percentage of speakers that applied that specific strategy to express one of the two conditions, dative or genitive.

These findings require an interaction between the syntactic and the prosodic component. The formalization of this interaction and the p-diagram approach first

introduced by Bögel (2012) allows for a thorough and compact depiction of the speech signal and an interpretation via c-structure annotation rules. In this paper, the p-diagram was extended by further attributes (e.g., the semitone difference) encoding the relevant prosodic cues in an easily accessible way.

The ranking between the different prosodic cues has been satisfied by means of OT-like constraints as they are used in the current XLE documentation.

# References

Allbritton, David W., McKoon, Gail and Ratcliff, Roger. 1996. Reliability of Prosodic Cues for Resolving Syntactic Ambiguity. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22(3), 714–735.

Asudeh, Ash. 2006. Direct Compositionality and the Architecture of LFG. In Miriam Butt, Mary Dalrymple and Tracy Holloway King (eds.), *Intelligent Linguistic Architectures: Variations on themes by Ronald M. Kaplan*, CSLI Publications.

Asudeh, Ash. 2009. Adjacency and Locality: A Constraint-based Analysis of Complementizer-Adjacent Extraction. In *Proceedings of LFG09*, CSLI Publications.

Beckman, Mary E., Hirschberg, Julia and Shattuck-Hufnagel, Stefanie. 2005. The original ToBI system and the evolution of the ToBI framework. In S.-A. Jun (ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*, Chapter 2, pages 9–54, Oxford University Press.

Boersma, Paul and Weenink, David. 2013. Praat: doing phonetics by computer [Computer program], Version 5.3.56, retrieved 15 September 2013 from http://www.praat.org/.

Bögel, Tina. 2012. The P-diagram – a Syllable-based Approach to P-structure. In *Proceedings of LFG12*, CSLI Publications.

Bögel, Tina, Butt, Miriam, Kaplan, Ronald M., King, Tracy Holloway and Maxwell III., John T. 2009. Prosodic Phonology in LFG: A New Proposal. In *Proceedings of LFG09*, CSLI Publications.

Bresnan, Joan. 2000. Optimal Syntax. In Joost Dekkers, Frank van der Leeuw and Jeroen van de Weijer (eds.), *Optimality Theory: Phonology, Syntax, and Acquisition*, pages 334–385, Oxford University Press.

Crouch, Richard, Dalrymple, Mary, Kaplan, Ron, King, Tracy Holloway, Maxwell III, John T. and Newman, Paula. 2013. XLE documentation, on-line documentation.

Dalrymple, Mary and Mycock, Louise. 2011. The Prosody-Semantics Interface. In *Proceedings of LFG2011*, CSLI Publications.

Frank, Anette, King, Tracy Holloway, Kuhn, Jonas and Maxwell, John. 1998. Optimality Theory Style Constraint Ranking in Large-scale LFG Grammars. In *Proceedings of LFG98*, CSLI Publications.

Gollrad, Anja, Sommerfeld, Esther and Kügler, Frank. 2010. Prosodic cue weighting in disambiguation: Case ambiguity in German. In *Proceedings of Speech Prosody*.

Grice, Martine and Baumann, Stefan. 2002. Deutsche Intonation und GToBI. *Linguistische Berichte* 191, 267–298.

Gussenhoven, Carlos. 1991. The English Rhythm Rule as an Accent Deletion Rule. *Phonology* 8(1), 1–35.

Kaplan, Ronald M. 1987. Three Seductions of Computational Psycholinguistics. In Peter Whitelock, Mary McGee Wood, Harold L. Somers, Rod Johnson and Paul Bennett (eds.), *Linguistic Theory and Computer Applications*, pages 149–181, London: Academic Press.

Kaplan, Ronald M. 1995. The Formal Architecture of Lexical-Functional Grammar. In Mary Dalrymple, Ronald M. Kaplan, John T. Maxwell III and Annie Zaenen (eds.), *Formal Issues in Lexical-Functional Grammar*, pages 7–27, Stanford: CSLI Publications.

Lahiri, Aditi and Reetz, Henning. 2010. Distinctive Features: Phonological Underspecification in Representation and Processing. *Journal of Phonetics* 38, 44–59.

Levelt, Willem J.M., Roelofs, Ardi and Meyer, Antje S. 1999. A Theory of Lexical Access in Speech Production. *Behavioral and Brain Sciences* 22, 1–75.

Mycock, Louise and Lowe, John. 2013. The Prosodic Marking of Discourse Functions. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of LFG13*, CSLI Publications.

O'Connor, Rob. 2005. Information Structure in Lexical-Functional Grammar: The Discourse-Prosody Correspondence in English and Serbo-Croation. In *Proceedings of LFG05*, CSLI Publications.

Prince, Alan and Smolensky, Paul. 1993. Optimality Theory: Constraint interaction in Generative Grammar. Technical Report, Rutgers University Center for Cognitive Science.

Shattuck-Hufnagel, Stephanie, Ostendorf, Mari and Ross, Ken. 1994. Stress Shift and Early Pitch Accent Placement in Lexi cal Items in American English. *Journal of Phonetics* 22, 357–388.