The prosody-syntax interface: A computational implementation

Tina Bögel
University of Konstanz/Goethe University Frankfurt

Proceedings of the LFG'22 Conference

Miriam Butt, Jamie Y. Findlay and Ida Toivonen (Editors)

2022

CSLI Publications

http://csli-publications.stanford.edu/

Abstract

This paper introduces a new approach for the integration of prosodic structure into the computational LFG grammars. Based on the theoretical assumptions made in Bögel (2015), the implementation includes the processing and categorization of speech signal information, the automatic generation of a p-structure representation, and shows how this information can ultimately be used to disambiguate syntactically ambiguous structures in German.

1 Introduction

Computational grammars in LFG (henceforth compLFGs) have long been established and used for a multitude of purposes with a strong focus on syntactic and semantic processing (a.o., Butt et al. 1999, Bobrow et al. 2007, Sulger et al. 2013, Dalrymple et al. 2020, Crouch et al. 2022).[1] The input to all of these grammars is the s(yntactic)-string, which consists of a string of words that make up a written sentence (or a fragment thereof) as in *Ravi ate a banana.* In a standard compLFG this string is tokenized into single words whose lexical morphosyntactic information is accessed and made available for further processing of the string into c- and f-structures and semantic representations. This basic structure (including variations or extensions thereof) has been the established core structure of all compLFGs since the start of the ParGram project (see Butt et al. (1999) for details).

While the parsing of written text into compLFGs has been well established for decades, the inclusion of spoken language was not pursued and as a consequence, any linguistic phenomenon whose analysis requires prosodic information cannot be interpreted by the traditional compLFGs. This includes the prosodic disambiguation of syntactically ambiguous structures in numerous languages (as discussed in this paper, but also see Butt and Biezma (2022), Butt et al. (2020)) or the prosodic indication of (contrastive) focus in examples like *Amra ate the RED apple* (implying that Amra did not eat the green or the yellow apple; see Xu and Xu (2005), Gussenhoven (2008) for related work), to name two of many possible examples where prosodic information is crucial for the overall interpretation.

There are only two previous (and more limited) attempts to include prosodic information into comp-LFGs. Both Butt and King (1998) and Bögel et al. (2010) did not process the actual speech signal, but relied on experimental findings from previous studies with respect to the placement of prosodic boundaries in order to discuss a very specific problem. In the proposal made by Butt and King (1998) p-structure is projected off from c-structure to explain syntactically ambiguous structures in Bengali. In their analysis, they distinguished between p(rosodic)-structure,

[†]I gratefully acknowledge funding from the DFG (German research foundation). The work presented here was done as part of the DFG-funded Walter Benjamin Stipend *Integrating prosodic structure into computational grammars*.

[1]See also the XLE-Web Interface, implemented by Paul Meurers, which features a number of different compLFGs: https://clarino.uib.no/iness/xle-web.

which includes prosodic phrasing and information on tones in an AVM representation, and a 'phonological component' which is not further defined, but which is assumed to include prosodic restructuring rules (i.e., postlexical phonology). While the analysis provided by Butt and King works well for the problem at hand, the approach is problematic in that p-structure is positioned between c-structure and s-structure in the LFG architecture. This stands in contrast to assumptions generally made in models of language processing (see, e.g., Section 3, Figure 7) and excludes the possibility of processing a speech signal per se.

Bögel et al. (2010) picks up on this issue and explains second position clitics in Serbian/Bosnian/Croatian by distinguishing between a p(rosodic)-string, where the clitic is (prosodically) placed after the first constituent, and an s-string, where the clitic inhabits the (syntactically but not prosodically valid) first position. While from an architectural perspective, this approach is in line with models of language processing in that it assumes that the prosodic component precedes the syntactic component, it focussed solely on the matching between p- and s-string and to this end required prosodic bracketing to become part of syntactic structure. Apart from larger prosodic units, the model did not allow further relevant information from the speech signal (e.g., pitch patterns, accent types, smaller prosodic units, segmental phonology) to be processed.

In conclusion, no previous computational approach has included detailed information on spoken language or the speech signal itself. Among the previous *theoretical* proposals to the interface between prosody and syntax, only Bögel (2015) develops a model which includes the speech signal and all the information provided therein.[2] This paper takes up this theoretical approach to the interface and shows how the information from the speech signal can be processed automatically and how the gained information can be made available to the compLFGs. The result is an extension of the compLFGs which makes any relevant information present in the speech signal available for the overall analysis of a linguistic phenomenon.

The paper is structured as follows: Section 2 describes the data, which consists of syntactically ambiguous structures that can be resolved via prosody. The type of data is a traditional challenge to the compLFGs, as the grammars overgenerate possible solutions without the constraining prosodic structures. The section also reports on a production experiment which establishes which prosodic cues are able to disambiguate the syntactic structures. Section 3 describes the theoretical analysis of the data at the prosody-syntax interface. Section 4 describes the computational integration of the speech signal and p-structure into the compLFGs. Section 5 concludes the paper.

[2]See Bögel (forthcoming) for a detailed discussion of all previous theoretical and computational approaches in LFG, their architectural assumptions and their advantages and disadvantages.

2 The data: syntactically ambiguous structures

Consider the syntactically ambiguous structure in example (1):

(1) Sie sahen,
They saw

dass [der Partner]$_{NP1}$ [der Freundin]$_{NP2}$ fehlte
that the.MASC.NOM partner the.FEM.GEN/DAT friend was.missing

a) "They saw that the friend's partner was missing."
b) "They saw that the friend missed the partner."

The ambiguity in example (1) is caused by the syncretism of the determiner *der* 'the' in combination with the verb's valency. The determiner is ambiguous in this position as it can be interpreted either as a feminine dative or a feminine genitive (Table 1) which allows for the complete second NP *der Freundin* 'the friend' to be interpreted as either dative or genitive.

case	**masc**	**fem**	**neut**
gen	des	*der*	des
dat	dem	*der*	dem

Table 1: German determiner system (for the singular genitive and dative)

In addition, the verb *fehlen* 'missing' is ambiguous in its valency: it can either be intransitive or transitive, in which case it requires a dative object. As a result, the second NP can either be interpreted as a dative object to the verb or as a possessor phrase to the first NP *der Partner*, as indicated by the two translations given in example (1). This full syntactic ambiguity is reflected in the corresponding c-structures in Figure 1.

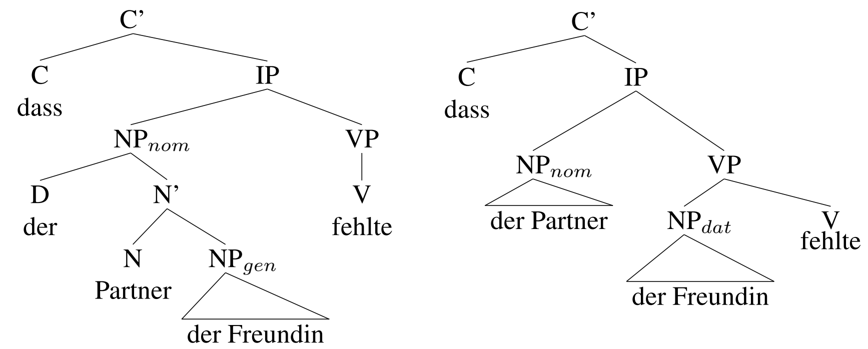


Figure 1: C-structures for example (1), genitive on the left and dative on the right.

For some syntactically ambiguous structures, prosody can be used to clarify the meaning (e.g., Price et al. 1991); in particular prosodic boundaries can be used to indicate syntactic constituency. This possibility is discussed in the next section.

2.1 Prosodic phrasing predictions

The current main approach to the syntax-prosody interface is Selkirk (2011)'s MATCH THEORY, which assumes a phonological phrase (hence PhP/φ) for every syntactic XP (NP, PP, ...) and which is usually combined with Truckenbrodt (1999)'s WRAP, which assumes that a recursive XP/PhP is merged ('wrapped') into a single PhP. For the syntactic structures given in Figure 1 and the string *der Partner der Freundin*, MATCH THEORY predicts a PhP boundary for every NP, resulting in two PhPs for the dative structure, and one nested PhP in the genitive structure. WRAP then assumes that the nested PhP in the genitive is wrapped into a single PhP. The resulting prediction for this sequence is that there is a PhP boundary after the first NP in the dative, but not in the genitive. Table 2 illustrates:

Dative	Syntax		[der Partner $]_{NP}$ [der Freundin $]_{NP}$	
	Prosody	MATCH	$_{\varphi}$(der Partner $)_{\varphi}$$_{\varphi}$(der Freundin $)_{\varphi}$	↕
		WRAP	**$_{\varphi}$(der Partner $)_{\varphi}$$_{\varphi}$(der Freundin $)_{\varphi}$**	
Genitive	Syntax		[der Partner [der Freundin $]_{NP}$ $]_{NP}$	
	Prosody	MATCH	$_{\varphi}$(der Partner $_{\varphi}$(der Freundin $)_{\varphi}$$)_{\varphi}$	↕
		WRAP	**$_{\varphi}$(der Partner der Freundin $)_{\varphi}$**	

Table 2: Prosodic phrasing predictions for the c-structures in Figure 1.

In order to confirm these predictions with respect to the placement of the PhP boundaries, a production experiment was conducted. This experiment was described in detail in previous work (Bögel 2020), so the following description will only highlight the most important findings.

2.2 Experimental findings

Material: The stimuli consisted of nine fully ambiguous structures similar to example (1), where the first NP was always masculine and the second one feminine, followed by a verb with an ambiguous valency. All nouns had a disyllabic, trochaic foot structure (i.e., the first syllable carried lexical stress and the second one was unstressed (X –)).

Participants and procedure: The participants were fifteen female native speakers of German. Each participant was presented with a context and a target sentence. Participants were asked to read the context silently and to 'mentally understand' the sentence, before producing the sentence as naturally as possible. Participants were recorded in the soundproof booth of the phonetic laboratory at the University of Konstanz. Each participant produced 18 sentences (9 genitive and 9 dative constructions), resulting in 270 sentences.

Statistical analysis: A linear mixed effects regression (lmer) with items and subjects as random factors yielded the following results:

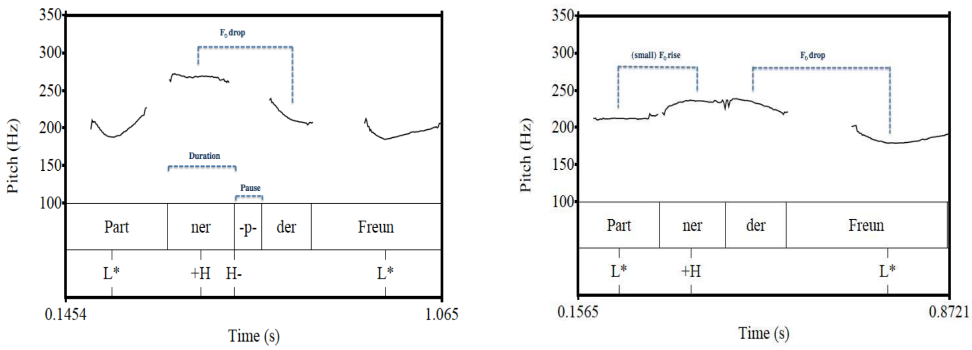
- A significantly steeper **drop in F0** (‘Reset’) between NP1 and NP2 (as measured at the final syllable of NP1 and the determiner of NP2) in the dative as compared to the genitive condition (β = -9.31, SE = 2.64, t = -3.53, p $<$ 0.01).

- A **pause** between the first and the second NP in the dative as compared to the genitive condition: (β = -2.35, SE = 0.92, t = -2.55, p $<$ 0.05).

- The **duration** of the last syllable of the first NP was significantly longer in the dative condition compared to the genitive condition (β = -2.8, SE = 0.79, t = -3.58, p $<$ 0.01).

These findings confirm the placement of a PhP boundary after the first NP in the dative. The following Figure 2 illustrates both, a ‘prototypical’ dative with a strong F_0 reset, a longer duration on the last syllable before the boundary, and a pause between the two NPs, and a ‘prototypical’ genitive, where all of these acoustic representatives of a PhP boundary are less prominent or not given at all.

Prototypical dative : Prototypical genitive:



... dass der Partner)$_\varphi$(der Freundin ... ... dass der Partner der Freundin ...

Figure 2: Prototypical representations of a dative and a genetive speech signal.

While these results are in line with the predictions given in Table 2, the question remains as to how these findings can be used to prosodically disambiguate syntactically ambiguous structures in LFG.

3 Theoretical anaysis in LFG

For the theoretical analysis, the paper follows the approach to the prosody–syntax interface proposed in Bögel (2015) which allows for the integration of the speech signal and is based on underlying assumptions made in the models of language architecture as proposed in Jackendoff (2002), see Figure 3.

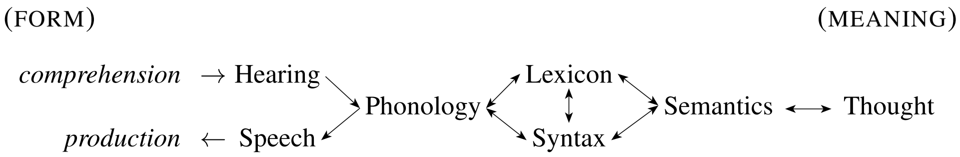


Figure 3: The language processor (cf. Jackendoff 2002, 197, modified)

The model assumes a mostly linear order of the modules between the two poles of 'form' and 'meaning' (see also Kaplan (1987)) and distinguishes between two processes: *comprehension* (i.e., parsing, listening: prosody→syntax) and *production* (i.e., generation, speaking: syntax→prosody). The following brief analysis focusses on the comprehension perspective – but see Bögel (2020) for a more detailed theoretical analysis of both, comprehension and production of the ambiguous structures discussed in this paper.

The approach to the prosody-syntax interface proposed in Bögel (2015) assumes the exchange of information at the interface on two levels: a) the **transfer of vocabulary** exchanges phonological and morphosyntactic information of lexical elements via the multidimensional lexicon, and b) the **transfer of structure** (♮) exchanges information on syntactic and prosodic phrasing, and on intonation.

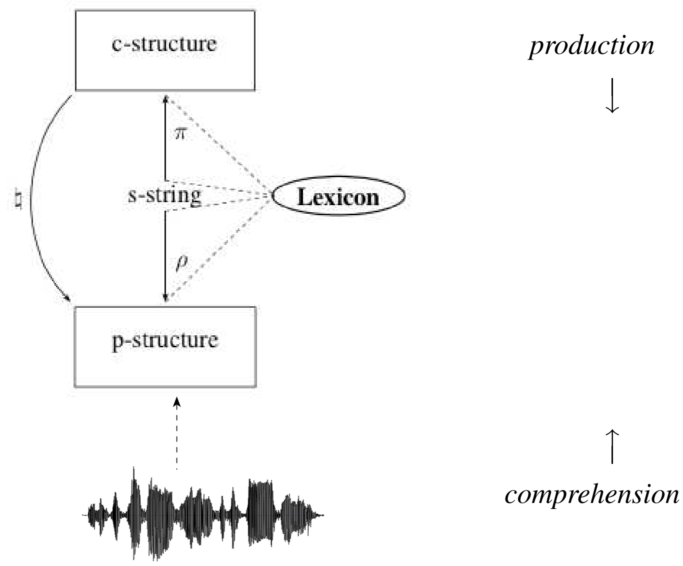


Figure 4: The model of the prosody-syntax interface as proposed by Bögel (2015)

P-structure is represented via the p-diagram, a linear syllablewise representation of the speech signal over time. During *comprehension*, acoustic information from the speech signal feeds into p-structure and is stored at the *signal* level. Each syllable in the signal receives a vector (S_n) which contains information, e.g., on the segments[3], the duration, or the mean fundamental frequency (F_0) of that syllable.[4] Figure 5 shows the p-diagram fragment for the six syllables related to the string *der Partner der Freundin*.

...	...	...	...	...	...	...	signal
DURATION	0.15	0.25	0.25	0.13	0.31	0.19	↓
FUND. FREQ.	192	181	269	209	188	218	
SEGMENTS	[de:6]	[pa6t]	[n6]	[de:6]	[fROYn]	[dIn]	
VECTORINDEX	$\mathbf{S_1}$	$\mathbf{S_2}$	$\mathbf{S_3}$	$\mathbf{S_4}$	$\mathbf{S_5}$	$\mathbf{S_6}$	

Figure 5: The signal level of the p-diagram for *der Partner der Freundin*.

The 'raw' signal information given in Figure 5 encodes patterns which can be interpreted in categorical terms at the *interpretation* level. For example, a strong rise in F_0 and a following drop (S_2–S_4) and a comparatively long duration on the last (unstressed) syllable of *Partner* (as seen at S_3: [n6]) are strong indicators for a phonological phrase boundary. As a result, PHRASING = $)_\varphi$ is added to the syllable's vector at the interpretation level (Figure 6).

...	...	...	...	...	...		interpretation
PHRASING	-	-	$)_\varphi$	$(_\varphi$	-	-	↓
SEMIT_DIFF	...	-1	6.8	-4.3	-1.9	2.6	
GTOBI	-	L*	+H H-	-	L*	+H	
DURATION	0.15	0.25	0.25	0.13	0.31	0.19	signal
FUND. FREQ.	192	181	269	209	188	218	↓
SEGMENTS	[de:6]	[pa6t]	[n6]	[de:6]	[fROYn]	[dIn]	
VECTORINDEX	$\mathbf{S_1}$	$\mathbf{S_2}$	$\mathbf{S_3}$	$\mathbf{S_4}$	$\mathbf{S_5}$	$\mathbf{S_6}$	

Figure 6: The interpretation level of the p-diagram: interpreting acoustic information in categorical terms

Further possibilities at the interpretation level include, e.g., a GTOBI (Grice and Baumann 2002) analysis of the pitch by means of high and low tones, or the difference between adjacent semitones (SEMIT_DIFF), that allows for an interpretation of the slopes leading to and from the accent (i.e., the scaling of the tones).

The *transfer of vocabulary* associates morphosyntactic and phonological information on lexical elements via the multidimesional lexicon (Figure 7). Following proposals made by, e.g., Levelt et al. (1999), the lexicon distinguishes between several dimensions: the (semantic) *concept* (not further discussed here), the

[3]Segments are represented in SAMPA, a computer-readable phonetic alphabet (Wells 1997).

[4]Mean F_0 is calculated based on the complete syllable and serves as a quick orientation for the researcher and not as a basis for the computational calculation discussed below.

s(yntactic)-form which contains the traditional morphosyntactic information, and the *p(honological)-form* which contains the segments and the metrical information: the number of syllables, the lexical stress pattern, and the prosodic status (e.g., whether the element is a clitic, underspecified, or a prosodic word).

concept	**s-form**			**p-form**	
FREUNDIN	N	(↑ PRED)	= 'Freundin'	SEGMENTS	/f R OY n d I n/
		(↑ NUM)	= sg	METRICAL FRM	$('\sigma\sigma)_\omega$
		(↑ GEND)	= fem		
DETERMINER	D	(↑ PRED)	= 'der'	SEGMENTS	/d e 6/
		(↑ NUM)	= sg	METRICAL FRM	σ
		(↑ GEND)	= fem		
		(↑ CASE)	= {gen | dat}		

Figure 7: (Simplified) lexical entries for *der* and *Freundin*.

The lexicon is modular in that there is a strict separation of module-related information: Each lexical dimension can only be accessed by the related module, i.e., p-structure can only access p-forms, and c-structure can only access s-forms. At the same time, the lexicon has a translating function: Once a dimension is triggered, the related dimensions can be accessed as well: If p-structure accesses a p-form, the related s-form becomes available and the morphosyntactic information is instantiated to c- and f-structure (=comprehension) – and vice versa, if c-structure accesses an s-form, the related p-form information becomes available to p-structure (=production).

During the *Transfer of structure*, information on prosodic and syntactic constituents are exchanged. The annotation below checks whether there is a (left) phonological phrase boundary associated with the left edge of the second NP's corresponding prosodic unit in p-structure. If this is the case, an object with dative case is projected to f-structure.

NP

$(\natural(T(*))\mathrm{S}_{min}$ PHRASING) $=_c$ $(_\varphi$ [5]

(↑ OBJ) = ↓

(↓ CASE) = dat

PHRASING	...	...	$(_\varphi$	...	...	...	...		
SEGMENTS	...	...	[de:6]	[fROYn]	[dIn]	...	...		
VECTORINDEX	...	...	$\mathbf{S_4}$	$\mathbf{S_5}$	$\mathbf{S_6}$	...	...		

Figure 8: The *transfer of structure*: prosodic and syntactic phrasing

[5]For all terminal nodes T of the current node *, for the syllable with the smallest index (S_{min}) in this set of terminal nodes, there must be a (left) phonological phrase boundary $(_\varphi$. See Bögel (2015, Ch. 3) for details and definitions.

The following figure shows the complete analysis of a dative structure at the prosody-syntax interface during comprehension, where the transfer of structure disambiguates the syntactically ambiguous structures by means of prosodic information, in this case a phonological phrase boundary between the two NPs.

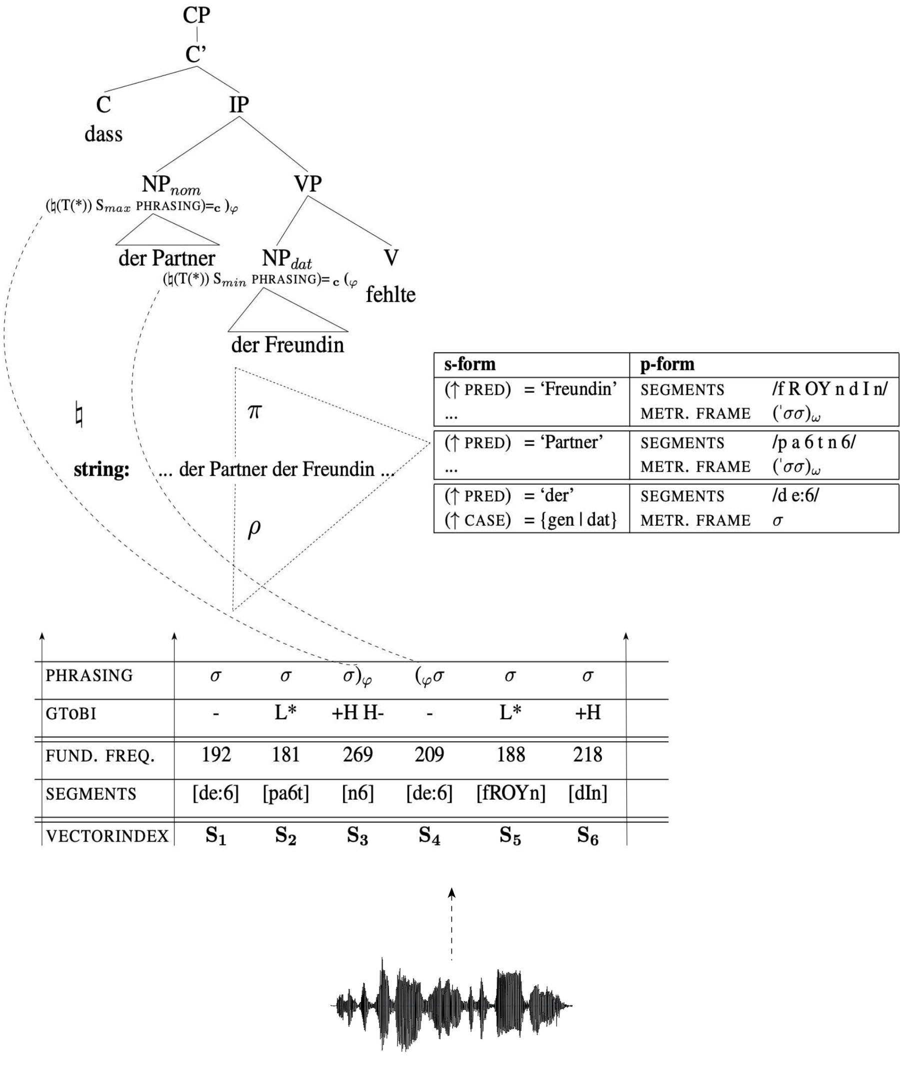


Figure 9: A dative construction at the prosody-syntax interface: comprehension

This section showed how syntactically ambiguous structures can be disambiguated with the approach to the prosody–syntax interface proposed in Bögel (2015). The following section uses this theoretical analysis as a blue-print for the computational implementation, extending the existing compLFGs to include information from the speech signal as well.

4 Computational Implementation

The following implementation of the theoretical analysis presented in Section 3 is a new approach that includes the integration of the speech signal itself, categorizes the gradient information gained from the signal and organizes it within p-structure's p-diagram. It then matches the information against a lexicon containing p-form and s-form information. The matching process allows for the creation of the s(syntactic)-string which is the linear concatenation of all matching s-forms and thus corresponds to the string that was originally used as input to the compLFG grammars. The s-string (and the lexical morphosyntactic information associated with each word in the string) allows for c- and f-structure to be parsed with XLE (Crouch et al. 2022). In a final step, the implementation allows for c-structure to be disambiguated based on the automatically determined prosodic phrase boundaries in p-structure. The implementation is in perl, with added scripts from Praat (Boersma and Weenink 2013), xfst (Beesley and Karttunen 2003), and R (R Core Team 2016), all of which are open-source and commonly used software.

4.1 Extracting information from a speech signal

The input used for the computational implementation is a Praat sound-file annotated with syllables in SAMPA.[6]

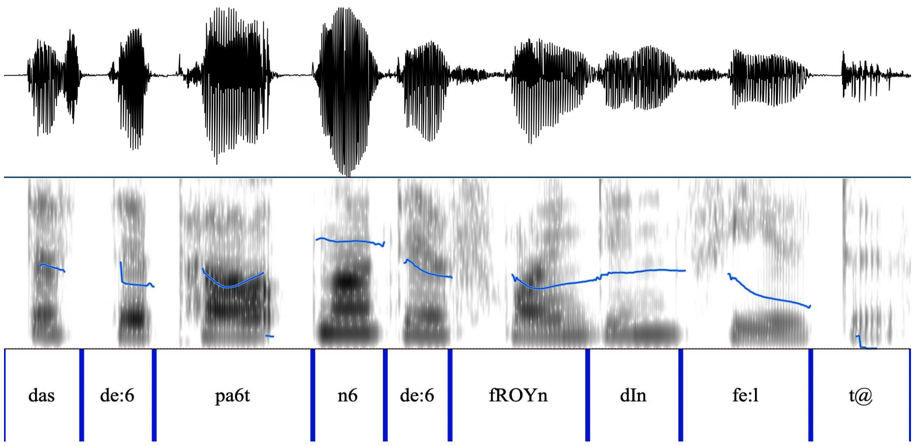


Figure 10: Input: a sound file annotated with syllables (here: example (1))

In a first step, information is gathered from the speech signal with a Praat script (Boersma and Weenink 2013). This script extracts the syllable segments, the duration of each syllable, and the mean F_0-values for each syllable for the p-diagram's

[6]For German, the annotation of the syllables could be automatized as well, e.g., by a combination of MAUS (Kisler et al. 2017, Schiel 1999) and CELEX (Baayen et al. 1995). However, the main focus of this paper is an implementation of the prosody-syntax interface - and not automatic speech recognition.

signal level (Figure 5). It also divides the syllable into five even-spaced sub-intervals, which allows for a more fine-grained analysis of the pitch and effectively time-normalizes each syllable. In order to normalize the pitch, all F_0-values are automatically turned into semitones.

4.2 Interpreting the pitch

In a second step, the raw values from the speech signal are interpereted in terms of categories in order for them to become 'meaningful' for other modules of grammar. Different measures are used for the interpretation of the pitch: Besides the semitones that already normalize the F_0 values, the implementation also makes use of the residuals of a linear regression. In that case, a linear regression line is calculated based on the pitch values of a given speech signal. As the overall pitch of a speech signal tends to get lower over time, this regression line is also descending towards the end of the utterance. The residuals return the distance each value has from this line and are thus a good measure to describe deviations from the average while at the same time including the natural decline in the overall signal.

These measures are then used to a) determine the minimums (L) and maximums (H) in a signal, and b) to determine the slopes between these categories in terms of whether the rises/falls are strong or weak. In order to mark both categories, type of accent and type of slope, in one representation, the following system was devised, where each level of the L or H is characterised by a particular height and shape of the slopes leading to it (lead) and following it (tail). Boundary tones in form of H- and L- are also included

Cat.	Min/Max	lead	tail	Cat	Min/Max	lead	tail
H4	Max	strong	strong	L4	Min	strong	strong
H3	Max	strong	normal	L3	Min	strong	normal
H2	Max	normal	strong	L2	Min	normal	strong
H1	Max	normal	normal	L1	Min	normal	normal
H-	Max	normal/strong		L-	Min	normal/strong	

Table 3: System of pitch accents and slopes in the computational implementation

H4 and L4 thus represent accents where the lead and the tail show a strong rise/fall respectively, while H1 and L1 have a relatively flat lead and tail. L2/L3 and H2/H3 are positioned between these two extremes, with each having a slightly different shape depending on the slopes. These tone values are then stored in the interpretation level of the p-diagram (Figure 6), where they replace the traditional GToBI values in order to facilitate (and simplify) the automatic interpretation by other modules of the grammar.

4.3 Lexical matching: The transfer of vocabulary

During the transfer of vocabulary, the input from the speech signal is matched against the p-forms of the multidimensional lexicon, which then allows for the associated s-forms to become available for syntactic parsing. During this transfer process, the p(honological)-string ... *de:6.pa6t.n6.de:6.fROYN.dIn* is matched exhaustively until all material is accounted for. The output of the lexicon is the corresponding s-string ... *dass der Partner der Freundin ....*, which in turn is the input for c-structure.

Input (p-string)	**Lexicon**		**Output (s-string)**
	p-form	**s-form**	
... de6.fROYn.dIn ... →	de:6	der	→ ... der Freundin ...
	fROYn.dIn	Freundin	
	...	...	

Table 4: The transfer of vocabulary: from p-string to s-string

In addition to making the s-string and the associated morphosyntactic information available to c- and f-structure, the matching of p-forms against the lexicon also allows for the p-form information stored with each item to become available for further processing, for example information on lexical stress or on the prosodic word status.[7] As the lexicon in Figure 8 shows, *Freundin* is a prosodic word and consists of two syllables with a trochaic foot. The determiner *der* is not a prosodic word, and has only one, unstressed syllable. This information is encoded by means of a finite-state transducer using the xfst technology (Beesley and Karttunen 2003).

4.4 Creating the p-diagram

The production experiment reported in Section 2.2 showed which acoustic factors can be relevant for the determination of a PhP boundary: A rise followed by a drop in F_0, a pause, and a relatively long unstressed syllable. Based on the information gathered in the p-diagram so far, a PhP boundary can thus be assumed following the last syllable of *Partner* in a speech signal with a dative construction. Figure 11 shows the automatically created p-diagram, where σ indicates syllables, unmarked brackets indicate prosodic words (e.g., enclosing *partner*: (pa6t n6)) and PhP boundaries are marked with `pp (` and `) pp`.

[7]This information is especially relevant for production (not discussed here), because it allows the modelling of a prosodic baseline that can later be 'translated' into phonetic terms. But it is also relevant for comprehension, in that it is generally assumed that pitch accents are only associated with lexically stressed syllables in German. Due to vowel quality differences and other reasons, however, the machine might also determine the local maximum or minimum to be on the previous or following syllable (see also Figure 11 where the L2 accent should be placed on the first syllable of *Partner* and not the preceding determiner). Lexical stress indication can then, in principle, be used to slightly shift the accents in the representations.

pros_phrase	pp(σ	σ	(σ	σ))pp	pp(σ	(σ	σ)	(σ	σ))pp
pitch_tones		L2		H4		L2	H1		
lex_stress	–	–	x	–	–	x	–	x	–
F0_mean	225.62	193.49	198.90	267.53	219.35	194.02	213.77	176.27	85.71
duration	0.17	0.16	0.33	0.18	0.14	0.30	0.20	0.28	0.22
syllables	das	de:6	pa6t	n6	de:6	fROYn	dIn	fe:l	t@
Vector_index	1	2	3	4	5	6	7	8	9

Figure 11: P-diagram for a dative construction

However, the information on prosodic phrase boundaries in p-structure does not automatically disambiguate c-structure. For this, c-structure has to determine that there is an ambiguity in the first place, and furthermore 'understand' where this ambiguity is situated.

4.5 Disambiguation and the fchart: The transfer of structure

Parsing the s-string as created in Section 4.3 will result in two possible c-structure parses. The syntactic ambiguity leading to these parses can be made visible by printing out the 'fchart', a prolog representation of all choices, constraints, c-structure relations, and more, in one file.

(2) `print-prolog-chart-graph filename.pl`

The command in (2) will return a prolog-file *filename.pl*. The following description discusses only the relevant parts of the (otherwise rather extensive) prolog-representation and how they can be used to determine the actual linear position of the ambiguity.

The fact that there are two possible structures (A1 and A2) is encoded in the section "Choices".

(3)
```
% Choices:
  [
choice([A1,A2], 1)
  ],
```

These two choices refer to the ambiguity in the verb's valency in the section "Constraints".

(4)
```
% Constraints:
  [
cf(A1,eq(var(3),semform('fehlen',4,[var(4),var(2)],[]))),
cf(A2,eq(var(3),semform('fehlen',4,[var(4)],[]))),
  ],
```

As indicated in (4), the verb *fehlen* in choice A1 has two arguments (var(4) and var(2)) and in choice A2 only one argument (var(4)). With respect to the linguistic data discussed in this paper, choice A1 thus refers to the (transitive) dative, and choice A2 to the (intransitive) genitive.

In the section “C-structure”, the fspans of the arguments (i.e., over which elements/s-forms the argument ‘spans’) are encoded with indexing numbers, where the first number (17 for var(4) and 29 for var(2)) indicates the start of the span, and the second number (41 for var(4) in A2, and var(2) in A1, and 28 for var(4) in A1) the end of the span.

(5)
```
% C-Structure:
  [
cf(A2,fspan(var(4),17,41)),
...
cf(A1,fspan(var(4),17,28)),
cf(A1,fspan(var(2),29,41)),
...
```

These numbers are related to the surface forms (i.e., the s-forms or terminal nodes in c-structure). As shown in (5), index number 17 is the starting position of the first argument var(4) in both, option A1 and A2. This number is associated with the start of the determiner *der* of the first NP [der Partner].

(6) `cf(1,surfaceform(9,`**'der',17,20**`))`
$\rightarrow$ start of the first argument var(4) in both options

In option A1, the span of the first argument var(4) is terminated with the indexing number 28, which also indicates the end of the surface form *Partner* in example (7). The first argument var(4) in option A1 (but not A2) is thus the NP [der Partner].

(7) `cf(1,surfaceform(11,`**'Partner',21,28**`))`
$\rightarrow$ end of the first argument var(4) in option A1 (subject in the dative construction)

The surfaceform of the determiner of the second NP starts with index number 29. As seen in (5), this is also the start of the second argument var(2) in option A1.

(8) `cf(1,surfaceform(13,`**'der',29,32**`))`
$\rightarrow$ start of second argument var(2) in option A1

Finally, the surfaceform *Freundin* ends with index number 41. This is also the terminating index number of the second argument (var(2)) of option A1, and of the first and only argument (var(4)) of option A2.

(9) `cf(1,surfaceform(15,`**'Freundin',33,41**`))`
$\rightarrow$ end of second argument var(2) in option A1 (object of the dative)
$\rightarrow$ end of first argument var(4) in option A2 (subject of the genitive)

By considering these different fspans with respect to the arguments given in the two options, it can be concluded, that the ambiguity arises at the end of the first NP [der Partner], where option A1 concludes the first argument, and option A2 does not.

Since the edges of syntactic NPs are associated with PhP boundaries, the algorithm now needs to check whether there is a PhP boundary after the last syllable of *Partner* in p-structure. If this is the case, then option A1 (the dative) should be selected. If there is no PhP boundary then option A2 (the genitive) is more likely. The selected option can be encoded in the prolog file.

(10)
```
[
select(A1, 1)
]
```

The new fchart is then reparsed (`read-prolog-chart-graph filename.pl`) and only returns the selected option, thus effectively disambiguating syntactic structure by means of prosodic information.

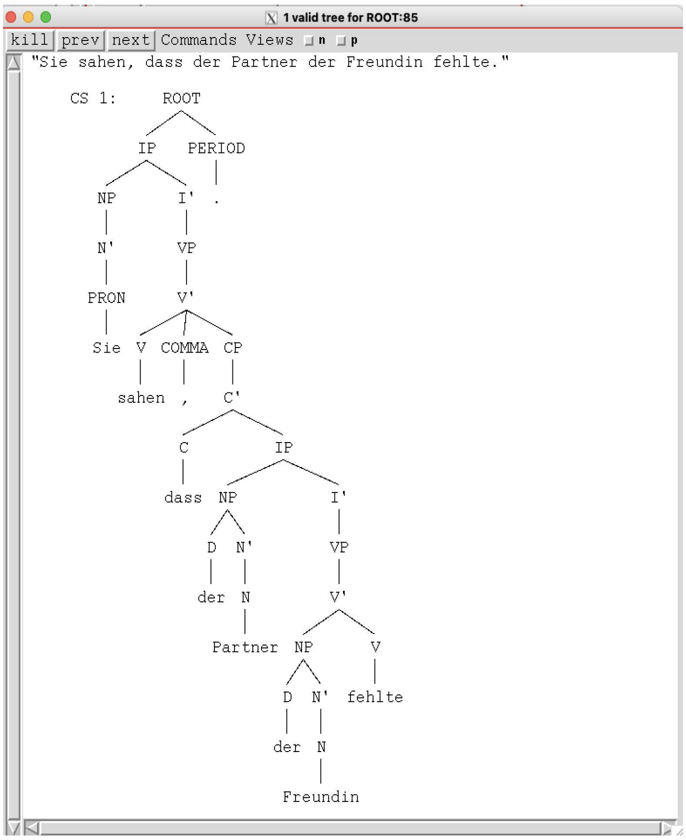


Figure 12: Disambiguated c-structure representation of a dative structure

5 Conclusion

This paper introduced a computational implementation of the prosody-syntax interface which enables the traditional compLFGs to process spoken language and to integrate the speech signal information into the analysis of linguistic phenomena, thus extending far beyond the previous approaches discussed in Section 1. The implementation presented in this paper provides a detailed analysis of a given speech signal, interprets the raw speech signal input in categorical tems by, e.g., determining prosodic phrase boundaries or pitch accents in the signal, and creates a representation of p-structure that can be accessed by other modules as well, while at the same time following current models of language processing. As a consequence, in addition to syntactic and semantic analyses, the compLFGs now can in principle process and interpret any phenomena indicated by prosody alone including the prosodic disambiguation of syntactically ambiguous structures, or the prosodic indication of broad and narrow focus.

The approach presented in this paper is work in progress and future research includes the evaluation of the existing system and the extension to other phenomena and languages. Challenges are manifold, and foremost is the problem that prosody is always gradient and includes a lot of variation (within and between speakers, but also within and between different dialects, etc.). Syntax and semantics, in contrast, are less prone to variation and are mostly relying on categorical information, which makes the communication between these modules and p-structure more difficult.

Nevertheless, the system introduced in this paper proves that an integration of spoken language and p-structure into the existing compLFGs is possible and desirable in order to allow for a complete end-to-end analysis between *form* (the speech signal) and *meaning* (the semantic interpretation), thus enabling the automatic analysis of linguistic phenomena from all relevant angles.

References

Baayen, R. Harald, Piepenbrock, Richard and Gulikers, Leon. 1995. The CELEX Lexical Database (CD-ROM). In *Linguistic Data Consortium*, Philadelphia: University of Pennsylvania.

Beesley, Kenneth R. and Karttunen, Lauri. 2003. *Finite State Morphology*. Stanford, CA: CSLI Publications.

Bobrow, Daniel G., Cheslow, Bob, Condoravdi, Cleo, Karttunen, Lauri, King, Tracy Holloway, Nairn, Rowan, de Paiva, Valeria, Price, Charlotte and Zaenen, Annie. 2007. PARC's Bridge and Question Answering System. In *Proceedings of the Grammar Engineering Across Frameworks Workshop (GEAF 2007)* .

Boersma, Paul and Weenink, David. 2013. Praat: doing phonetics by computer [Computer program, Version 5.3.56], available at http://www.praat.org/ [retrieved 15.09.2013].

Bögel, Tina. 2015. *The Syntax–Prosody Interface in Lexical Functional Grammar*. Ph.D. thesis, University of Konstanz, Konstanz.

Bögel, Tina. 2020. German case ambiguities at the interface: production and comprehension. In Gerrit Kentner and Joost Kremers (eds.), *Prosody in Syntactic Encoding*, Linguistische Arbeiten, No. 573, pages 51–84, Berlin: De Gruyter.

Bögel, Tina. forthcoming. Prosody and its interfaces. In Mary Dalrymple (ed.), *The Handbook of Lexical Functional Grammar*, Language Science Press.

Bögel, Tina, Butt, Miriam, Kaplan, Ronald M., King, Tracy H. and Maxwell III, John T. 2010. Second position and the prosody-syntax interface. In *Proceedings of LFG10*, Stanford, CA: CSLI Publications.

Butt, Miriam and Biezma, María. 2022. The Syntax and Interpretation of Urdu-Hindi Polar kya. In *Proceedings of LFG22*, University of Konstanz/CSLI Publications.

Butt, Miriam, Jabeen, Farhat and Bögel, Tina. 2020. Ambiguity resolution via the syntax-prosody interface: the case of *kya* 'what' in Urdu/Hindi. In Gerrit Kentner and Joost Kremers (eds.), *Prosody in Syntactic Encoding*, volume 573 of *Linguistische Arbeiten*, pages 85–118, Berlin: De Gruyter.

Butt, Miriam, King, Tracy H., Niño, María-Eugenia and Segond, Frédérique. 1999. *A Grammar Writer's Cookbook*. Stanford, CA: CSLI.

Butt, Miriam and King, Tracy Holloway. 1998. Interfacing phonology with LFG. In *Proceedings of LFG98*, Stanford, CA: CSLI Publications.

Crouch, Richard, Dalrymple, Mary, Kaplan, Ronald M., King, Tracy H., Maxwell III, John T. and Newman, Paula. 2022. *XLE documentation*. Palo Alto, CA: Palo Alto Research Center, online documentation.

Dalrymple, Mary, Patejuk, Agnieszka and Zymla, Mark-Matthias. 2020. XLE+Glue – A New Tool for Integrating Semantic Analysis in XLE. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of LFG20*, Stanford, CA: CSLI Publications.

Grice, Martine and Baumann, Stefan. 2002. Deutsche Intonation und GToBI. *Lin-*

guistische Berichte 191, 267–298.
Gussenhoven, Carlos. 2008. Types of focus in English. In C. Lee, M. Gordon and D. Büring (eds.), *Topic and Focus*, volume 82 of *Studies in Linguistics and Philosophy*, pages 83–100, Springer.
Jackendoff, Ray. 2002. *Foundations of Language: Brain, Meaning, Grammar, Evolution*. New York: Oxford University Press.
Kaplan, Ronald M. 1987. Three seductions of computational psycholinguistics. In P. Whitelock, M. McGee Wood, H. L. Somers, R. Johnson and P. Bennett (eds.), *Linguistic Theory and Computer Applications*, pages 149–181, London [a.o.]: Academic Press.
Kisler, Thomas, Reichel, Uwe and Schiel, Florian. 2017. Multilingual processing of speech via web services. *Computer Speech and Language* 45, 326–347.
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.
Price, Patti, Ostendorf, Mari, Shattuck-Hufnagel, Stefanie and Fong, Cynthia. 1991. The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America* 90(6), 2956–2970.
R Core Team. 2016. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
Schiel, Florian. 1999. Automatic Phonetic Transcription of Non-Prompted Speech. In *Proceedings of ICPhS*, pages 607–610, San Francisco.
Selkirk, Elisabeth O. 2011. The syntax-phonology interface. In J. Goldsmith, J. Riggle and A. Yu (eds.), *The Handbook of Phonological Theory*, pages 435–484, Malden, MA: Blackwell.
Sulger, Sebastian, Butt, Miriam, King, Tracy Holloway, Meurer, Paul, Laczkó, Tibor, Rákosi, György, Dione, Cheikh Bamba, Dyvik, Helge, Rosén, Victoria, De Smedt, Koenraad, Patejuk, Agnieszka, Cetinoglu, Ozlem, Arka, I Wayan and Mistica, Meladel. 2013. ParGramBank: The ParGram Parallel Treebank. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 550–560, Sofia, Bulgaria: Association for Computational Linguistics.
Truckenbrodt, Hubert. 1999. On the relation between syntactic phrases and phonological phrases. *Linguistic Inquiry* 30(2), 219–255.
Wells, John C. 1997. SAMPA computer readable phonetic alphabet. In Dafydd Gibbon, Roger Moore and Richard Winski (eds.), *Handbook of Standards and Resources for Spoken Language Systems*, pages 684–732, Berlin, New York: Mounton de Gruyter.
Xu, Yi and Xu, Ching X. 2005. Phonetic realization of focus in English declarative intonation. *Journal of Phonetics* 33, 159–197.