

**THE P-DIAGRAM – A SYLLABLE-BASED APPROACH
TO P-STRUCTURE**

Tina Bögel
University of Konstanz

Proceedings of the LFG12 Conference

Miriam Butt and Tracy Holloway King (Editors)

2012

CSLI Publications

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

Abstract

This paper introduces a formal account of p-structure via a new conception, the p-diagram. It is based on the phonological unit of the syllable, which constitutes the phonological string. On top of this basic entity, prosodic layers are added ‘as needed’ and tied to the respective syllables via vectors. The result is a theory-independent and compact description of the speech signal itself, which allows for easy extraction of relevant information for other correspondence structures and can be adjusted to language- and theory-specific needs. An example of how this approach works is given by describing a solution to the syntactically ambiguous, but prosodically unambiguous phenomenon of constituent grouping in coordination.

1 Introduction

Current prosodic research is very much driven by the prosodic hierarchy theory, originally proposed by Selkirk (1984), which consists of hierarchically ordered prosodic units, e.g., *intonational phrase*, *phonological phrase*, *prosodic word* etc. However, these prosodic units are only descriptions of events in the speech signal, like stress, intensity, duration of elements, or breaks. The literature on the question of how these units should be defined and which information of the speech signal should be used for their calculation is huge and quite controversial. A majority of the community (e.g. Selkirk, 1984; Nespor and Vogel, 1986) takes, e.g., the prosodic word to *minimally* be of the same size than the the morphosyntactic stem or even the syntactic word. While some researchers of this group only allow for prosodically deficient items (clitics/function words (e.g. Selkirk, 1984)) to be included into the prosodic word, others believe that, under certain circumstances the prosodic word can contain the whole sentence (Kleinhenz, 1998). On the other hand, there are also researchers who propose that a prosodic word starts with the stressed syllable of a trochaic foot encountered in the speech signal (Lahiri and Plank 2010, see also Dalrymple and Mycock 2011). A consequence of this approach is that the prosodic word is not necessarily identical to a syntactic word; instead, the morphosyntactic stem can in principle be split by a prosodic word boundary, separating the syllables of one morphosyntactic stem into two different prosodic words.

Apart from the controversial discussion on the exact size and nature of the different prosodic units, a further problem is posed by the fact that the status of the prosodic speech units might change in relation to external factors like speech tempo. Thus, the same (written) string might be realized by several variants of prosodic grouping.

In order to avoid these problems, my approach is based on the largest ‘reliable’ unit, the syllable. Syllables are rhythmic units, which consist of at least a vowel or

[†]I would like to thank the audience of LFG 2012, especially Ron Kaplan and Mary Dalrymple, and Louise Mycock for comments on a preliminary version of this approach. Furthermore, I would like to thank Melanie Seiss for her math coaching.

a syllabic consonant (the nucleus), and mostly include consonants grouped around this nucleus according to certain rules. Although syllables are considered to be part of the prosodic hierarchy, the approach introduced in this paper is not built upon the general notion of prosodic units and their respective grouping according to hierarchical restrictions, but encodes the events encountered in the speech signal in a linear order. Since prosodic units are determined by different indicators within the speech signal, this syllable-based approach is in principle able to account for them as well, because the required information (on, e.g., breaks or boundary tones) is encoded in the p-structure representation. The syllable as the basic unit also allows for independence from external factors like speech tempo, since the basic unit of the syllable remains the same (in contrast to a time-based or a prosodic grouping approach, for example).¹

Apart from the notion of the syllable as the basic unit, this paper also introduces a new way of representing prosodic information: The p-diagram. This representation allows, in principle, for all relevant elements of the speech signal to be encoded. It is thus broadly structured in the sense of being able to encode the speech signal's structure independently of the researcher's theoretical assumptions, but allows for a much finer-grained analysis of the speech signal in comparison to other approaches. The information stored in the p-diagram is not conveyed to the overall grammar in general; instead, its content is accessible via the correspondence structure of LFG.

In order to demonstrate this idea concretely, the phenomenon of constituent grouping in coordination is analyzed within this system. Coordination grouping is interesting for the syntax-prosody interface in that syntax alone provides no basis for a decision between several possible constituent groupings, but that a consideration of prosody yields the information necessary for disambiguation.

This paper is structured as follows: First, an account of the p-diagram and the different elements involved in the speech signal representation, i.e., the lexicon and the strings and the exact shape of these representations, are given. In the second part, the implementation of grouping in coordination shows the implementation of a phenomenon, where the prosody-syntax interface is at its best.

2 The Lexicon

Following (Levelt et al., 1999, p. 4), I assume that a lexical entry consists of several parts.

1. The *concept*: The concept describes the semantic concept of a lexical item; i.e., the *idea* of an entity in the world, which we have in our minds before we add a lemma (a morphosyntactic item) and a form (a phonological representation) to this idea.

¹There might be some syllable reduction or deletion depending on the respective speech tempo, but either a syllabic consonant remains or the syllable is completely deleted. This does not affect the overall representation.

2. The *lemma*: The lemma is the morphosyntactic representation of the concept.
3. The *form*: The form describes the phonological representation of the concept.

Following Dalrymple and Mycock (2011), I will refer to the lemma as *s(yntactic)-form* and to the form as *p(honological)-form*. The s-form represents the morphosyntactic representation of a concept; its realisation is a terminal node of the c-structure. It carries morpho-syntactic information, e.g., word-class, tense or sub-categorization frames, which is subsequently processed by the grammar. The p-form, on the other hand, encodes information about syllable structure, word stress (if applicable) and an IPA transcription of its phonetic representation.² In (1), (') indicates the nucleus of the primarily stressed syllable; syllable structure is indicated by periods.

- (1) The p-form of the lexical entry for 'übersetzen' (German, 'to translate')

concept	p-form
<i>translate</i>	/y:.bʊ.z'ɛ.tʃn/

S-form and p-form are two different aspects of the same concept; they represent two sides of one coin, and while they may look very different and encode very different information, they still represent the same core – their concept. This close relationship has been similarly described before, recently in LFG by Dalrymple and Mycock (2011). I would like to build on that work and show how these two-dimensional lexical entries are essential (but not omnipotent) for the disambiguation of ambiguous p-forms and s-forms and how this relationship works in both directions.

2.1 Disambiguating s-forms via word stress

As can be seen in (1), the phonological form includes information on word stress. Word stress is especially helpful for the disambiguation of elements, which have an identical morphosyntactic representation, but belong to different concepts. In English, for example, the word accent can differentiate between lexical categories (e.g., *p'ermit* (noun) vs. *perm'it* (verb)); thus, depending on the position of the word stress, a specific lexical entry is chosen and further processed by syntax as either a noun or a verb with its respective c- and f-structure representations. The grammar does not rely on a specific representation of phonological structure to differentiate

²Note that one lexical entry can include many p-forms, as they may differ according to dialect, speech register and other external reasons. I follow Lahiri and Reetz (2002, 2010), who assume the p-form of the lexical entry to be an abstract and underspecified representation of phonological features, which allows them a flexible treatment of different phonetic representations of the same concept. However, in order to simplify the p-form representation in the lexicon, I will depict the IPA 'standard' pronunciation.

between the two concepts, because this distinction is already represented by the syntactic structure, which encodes information on the respective lexical category extracted from the lexicon.

However, consider German, where the difference in word stress does not necessarily represent a difference in lexical categories. Here, a (phonological) representation of word stress in a grammar projection is essential in order to recognize the exact meaning of the sentence. Consider the German verb *übersetzen*, which can mean ‘to translate’ or ‘to cross over’. The first verb is a standard verb in German; the second one, on the other hand, is a particle verb, which can be split in certain syntactic environments. However, there are also constructions (verb-final sentences), which will prevent a disambiguation by means of syntax. At this point, the phonological information is essential: While the s-forms of the two concepts are identical, the respective p-forms show a difference ((2)). If the word stress is on the first syllable, the associated concept would be ‘to cross over’. If, on the other hand, the word stress is on the third syllable, the concept would be ‘to translate’.

- (2) **S-string:** Lass uns übersetzen
 Let us translate / cross over

Lexical entry:

concept	s-form	p-form
<i>translate</i>	übersetzen (V)	/y:.bɛ.z'ɛts.ŋ/
<i>cross over</i>	übersetzen (Particle-V)	/'y:.bɛ.zɛts.ŋ/

If this information on word accent is available ‘outside’ of the lexicon (e.g., in form of a speech signal representation), a disambiguation of meaning can take place. Syntactic analysis on its own, on the other hand, cannot differentiate between the two verbs.

2.2 Disambiguating p-forms via spelling and syntax

While the example in (2) shows disambiguation via information from the p-form of an ambiguous s-form, syntax is, on the other hand, often needed to disambiguate ambiguous p-forms. This can be seen in (3), where the p-form of German /ʃpɪnən/ (‘spiders’ / ‘to be crazy’ / ‘to spin’ (with a spinning wheel)) refers to several concepts.

- (3)
- | | concept | s-form | p-form |
|----|-----------------|-------------|----------|
| a. | <i>spiders</i> | Spinnen (N) | /ʃpɪnən/ |
| b. | <i>be crazy</i> | spinnen (V) | /ʃpɪnən/ |
| c. | <i>spin</i> | spinnen (V) | /ʃpɪnən/ |

On the basis of the different word category, (3a) can be disambiguated from (3b) and (3c) with the help of the s-form in that the position of the word in a sentence gives an indication of its word-class and as a consequence, of its concept. The

c-structure helps to disambiguate and tokenize the output of the speech signal on the one hand, and p-form entries (and p-structure as we will see in section 5) help to disambiguate syntactically ambiguous concepts/constructions on the other hand.

Although the representation of the p-string provided here already conveys information about the phonological/prosodic side of a string, it must be understood that the p-string is merely a partial visualization of the utterance’s sound wave. The speech signal itself carries much more information, e.g., the specific intonation of an utterance, the length of the breaks or the rhythmic grouping of the units. This information is captured in p-structure.

4 P-structure

Prosody is the aspect of grammar which is connected to the speech signal. This includes intonation, rhythmic grouping and stress, and the information connected with each of these subcategories of prosody (e.g., information on intensity and duration of single elements). Each of these aspects has received growing attention among researchers, but the most debated topic within prosodic research is probably the theory of the Prosodic Hierarchy, as proposed by Selkirk (1984, 2011), which separates the elements of the speech signal into (rhythmic) groups according to indications given by phonological processes, intonation patterns or other indications in the speech signal. These rhythmic groups are ordered hierarchically (Figure 2).

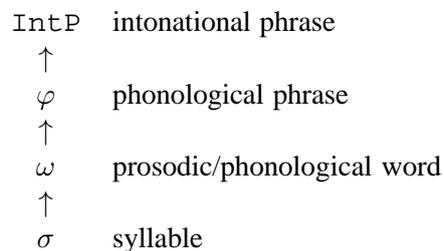


Figure 2: The Prosodic Hierarchy

As mentioned in the introduction, the exact nature, number and definition of the prosodic units is quite controversial. The problem is that the speech signal mostly consists of a continuous string. Boundaries to indicate prosodic units (be they prosodic words or phrases) might exist, but are not necessarily reliable. If, for example, the sentence in Figure 1 (“lass uns übersetzen”) is spoken with a certain speed and intonation, the signal will give no indications of where to put the prosodic boundaries.

The approach pursued in this paper therefore shifts the focus of prosodic grouping to the information received from the speech signal and ties this information to the basic rhythmic unit, the syllable. However, it retains the ability of encoding prosodic grouping if need be, because the relevant information (depending on the

theory: the foot, the pauses, the F_0 -pattern, or a mixture of the three) is still available and can still be interpreted as grouping information.

4.1 Previous approaches within LFG

Within LFG, the prosodic aspect of grammar has not (yet) received much attention, although some attempts of encoding prosodic information within LFG have been undertaken, the first one being Butt and King (1998), who encoded prosodic structure in an attribute value matrix (AVM) projected from c-structure on the basis of prosodic units.

This was followed by O'Connor (2004), who combines the higher units (IntP and φ) of the Prosodic Hierarchy with a rough description of the speech signal's fundamental frequency by means of the ToBI framework.³ In contrast to the AVM-approach of Butt and King (1998), he chooses a tree-like representation in order to avoid an intermediate projection between the AVM and the string itself on the one hand and to encode hierarchical structures between the two prosodic rhythmic units, based on the annotated high and low tones, on the other hand. While Butt and King (1998) view prosodic structure as being projected from syntactic structure, O'Connor treats syntax and prosody as independent projections. Bögel et al. (2009) follow this view of parallelism and discuss a range of mismatches between syntactic and prosodic grouping. They do not encode a separate prosodic representation per se, but include prosodic bracketing into the syntactic string.

Dalrymple and Mycock (2011) (building on Mycock 2006) develop an elaborate prosodic representation to account for comma intonation and question intonation. They view the string as being at the heart of the projection architecture and use it as the intermediate step between a prosodic tree (based on the prosodic units as described in Figure 2) and the c-structure tree. In addition, they project two further structures, which contain the bracketing information relevant for the alignment of the syntactic and the prosodic string, information on the intonational contour, etc.

All of these approaches encode a specific set of information relevant for a specific phenomenon. While some approaches have a broader potential (e.g., Dalrymple and Mycock 2011), others are too narrow for a wider range of prosodic phenomena (Bögel et al. 2009). Most use either the tree-representation or an AVM-approach to encode prosodic representations. As we will see, both representations are suboptimal when it comes to representing prosodic structure.

4.2 The representational problem

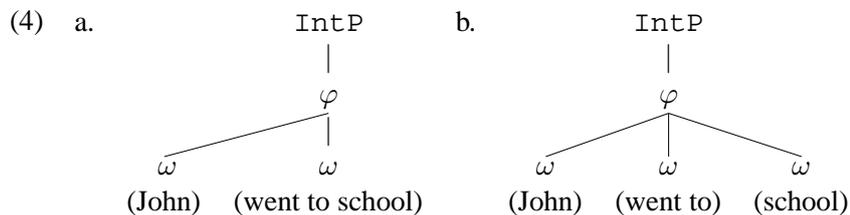
Within LFG, several ways of representing prosody have been proposed. These representations can be divided into two categories: The tree-based representations

³ToBI represents conventions for assigning High and Low tones to the fundamental frequency of a speech signal, thus describing relevant aspects of the intonational contour (Silverman et al., 1992).

(Dalrymple and Mycock 2011, O'Connor 2004) on the one hand, and the AVM-approach proposed by Butt and King (1998) on the other hand.

The tree-based representation relies on hierarchical structures and allows only for a single aspect of the prosodic information to be encoded. O'Connor (2004) uses the tree to encode the intonation of a sentence via the TOBI annotation scheme. On the basis of this annotation, the prosodic grouping is represented in the string via bracketing. Further information given by the speech signal is not encoded within this approach. Dalrymple and Mycock (2011), on the other hand, use the hierarchical structure of the tree to represent the units of the prosodic hierarchy. Further information (e.g., on boundaries, intonation, discourse functions) has to be projected into a separate (AVM) structure, which serves as an intermediary for the remainder of the grammar.

While syntactic structure is hierarchical, the inherent nature of the speech signal is linear; thus, a hierarchically organised representation is not necessarily the right representation. There is no doubt of a certain rhythmic grouping of units, but these mostly apply to more complex constructions and are but one aspect of the speech signal. For the majority of four-word sentences, this grouping is rather irrelevant, as can be seen in (4). The respective tree can look very different depending on either the theory of phrasing (i.e., which elements are phrased together on the basis of which reasons) or the speech tempo and quality (casual, slow, ...). (4a) and (4b) show two possible encodings.



Furthermore, a tree structure is always bound to one possible aspect out of all possible aspects of the speech signal. Additional information on, e.g., intonation has to be stored in an extra structure projecting away from (and thus depending on) the elements of the tree-based structure. Thus, a representation of p-structure should not be built on the less fine-grained and highly variable phrasing of hierarchically organised prosodic units, but should provide a fine-grained approach to the utterance by enabling the description of various aspects and layers of the speech signal in combination with the smaller prosodic units encountered in a sentence (in the p-diagram approach, this would be the syllable).

In contrast to the tree-based representations, the AVM-structure provided by Butt and King (1998) allows for the representation of a broad spectrum of information in that the attribute value pairs can encode several aspects of a speech signal in one structure. However, typical LFG AVMs, for example f-structure, do not represent information in a linear order. If an AVM-approach is pursued, the AVM must include information on precedence relations; otherwise, an extra ordering instance between string and structure is needed. Apart from these ordering issues,

the AVM has another drawback: If the speech signal and all its values are to be encoded into an AVM and its inherent attribute-value pairs, the AVM would grow to an enormous size. For two single syllables, the corresponding AVM would already be quite large (Figure 3).

	→	VALUE: IPA representation of the syllable.
$\left[\begin{array}{l} \text{syll}_1 \\ \text{syll}_2 \end{array} \right.$	$\left[\begin{array}{l} \text{VALUE} \quad [\text{ra}] \\ \text{STRESS} \quad + \\ \text{TONE} \quad \text{H} \end{array} \right]$	→ STRESS: present + or absent –.
		→ RHBT: <i>right hand break time</i> - indicates the length of a break following this syllable in relation to other breaks in the sentence (1= short break, 3= long break).
	$\left[\begin{array}{l} \text{VALUE} \quad [\text{vi}] \\ \text{STRESS} \quad - \\ \text{RHBT} \quad 3 \end{array} \right]$	→ TONE: H indicates a high tone in the pattern of the fundamental frequency (the “melody”).

Figure 3: A possible AVM-representation of /ravi/

Thus, while the AVM is, in principle, able to encode the relevant information, the representation is not the most desirable one. With more information added (e.g., on syllable length or intensity), the AVM representation would quickly grow in size and, in parallel, become less clear and thus less interpretable.

For this reason, I have developed a new representation, which a) meets the desideratum of a fine-grained representation of the speech signal, b) allows for easy extraction of relevant information and c) provides a compact representation.

4.3 A new approach: The p-diagram

The approach presented in this paper does not view p-structure as an attribute-value matrix or a prosodic tree. Instead, the speech signal (i.e., the sound wave) is transformed into a human-readable way by describing different aspects of the signal in relation to the syllables and the breaks in between these syllables. These calculations depend on the nature of intonation patterns and stress behavior, e.g., the fact that the general level of the fundamental frequency F_0 (the “melody”) will decrease towards the end of the sentence. Thus, different layers of information are extracted from the speech signal, e.g.,

- The basic rhythmic unit (the syllable)
- The stress pattern of the syllables (word accent, in combination with lexicon)
- The stress of the overall sentence (sentence accent)
- Possible pauses and their time frames
- The fundamental frequency F_0 indicating the intonation pattern

- ... and other relevant aspects as they are needed for the analysis of a specific prosodic phenomenon

All of these variables are connected with the each syllable via the following function, which describes the relevant information as an ordered list, i.e., a vector.

$$(5) \quad \mathbf{S}(\mathbf{u}(\mathbf{n}), \mathbf{n})$$

(Where \mathbf{S} is the syllable, \mathbf{u} is the vector and \mathbf{n} is the index of the syllable)

That is, each syllable of the p-string and each pause between two strings of syllables receives a vector, which includes the relevant information as an ordered list. The result is a set of vectors, which include ordered information of different speech signal dimensions in relation to the syllable of the p-string. These vectors can be generalized as

$$S: \begin{pmatrix} \text{value} \\ \text{stress} \\ F_0 \\ \dots \end{pmatrix} (\mathbf{n}) \quad \text{which would yield, e.g.,} \quad S: \begin{pmatrix} /y:/ \\ \text{prim} \\ H^* \\ \dots \end{pmatrix} (1)$$

for a specific syllable S_1 , which would be the first syllable in a string with the value /y:/, and which carries primary stress and a high tone. With this vector formula, the speech signal is transformed into a bundle of syllable-dependent vectors, encoding the prosodic information as it is related to the specific syllable in the speech signal.

This set of vectors is then read into a representation, which is similar to a diagram, where the syllables and the pauses encountered in the signal are the basic entities of the prosodic representation (the ‘x-axis’). On top of these basic entities, layers which include different prosodic information (e.g., stress patterns or intonation) are added (the ‘y-axis’).

	↑	↑	↑	↑	↑	↑	↑

	Pause	-	-	-	-	3	...
	Fund. Freq.	H*	L	-	-	-	...
	Stress	prim	-	sec	-	-	...
	Syllables	/y:/	/bɐ/	/zɛ/	/tsŋ/	-	...
Vector							
Vectorindex		S₁	S₂	S₃	S₄	S₅	... →

Figure 4: The p-diagram of /y:bɛzɛtsŋ/ (‘to cross over’)

Via the general variable on the y-axis and the respective syllable index on the x-axis, every value of the p-diagram is accessible. For example, the stress value of the third syllable can be extracted directly from the diagram via the relevant function

[S_3 , STRESS], which would return STRESS = SEC (for secondary stress). For an extraction of all syllable-values of the utterance [, SYLLABLES] would return a list of all values found at the position of the SYLLABLES of each vector, that is, all x-axis values for this y-axis variable: SYLLABLES = /y:/ /bʌ/ /zɛ/ /tsɪ/.

A special ‘syllable’ is S_5 in Figure 4. It encodes a pause in the speech signal with the value 3 (long break, as opposed to the value 1, which encodes a short break (Silverman et al., 1992)). For processing reasons, it is much easier to treat the pauses as special ‘syllables’ instead of encoding their presence on the syllable to their left or right. Thus, apart from its pause value, a pause vector contains epsilons (-) for the other variables.

Note that the p-diagram could easily include more layers, e.g., information on the length of each syllable. It is not, per se, hierarchically constructed, but depicts the speech signal in a linear way. However, the possibility of encoding prosodic units and the implied hierarchies is given. For example, a notion like [R PHP] as used by Dalrymple and Mycock (2011) to represent the right edge of a phonological phrase can be easily encoded within a vector and retrieved from the p-structure if an analysis of prosodic phrasing is desired. Depending on the theoretical assumptions of the researcher as to which elements of the speech signal encode prosodic units, these prosodic units can be extracted from the p-diagram and its representation of speech signal elements. The p-diagram is thus a ‘neutral’ representation of a speech event on whose basis the individual theories of prosodic phrasing can be projected.

Figure 5 gives a general overview of the architecture proposed in this paper.

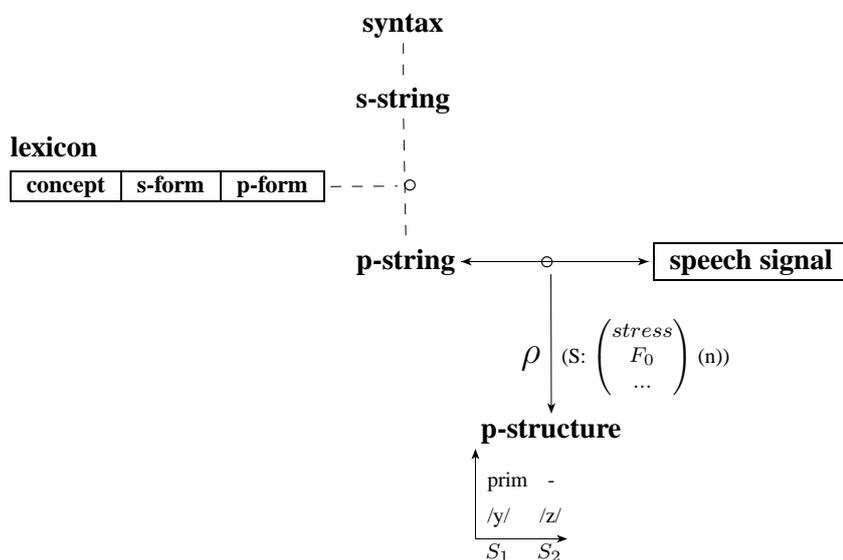


Figure 5: An architectural overview

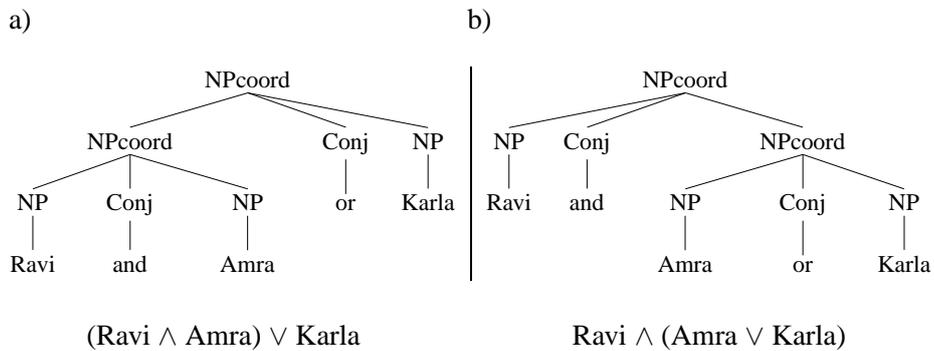
In this architecture, s- and p-string together with the lexicon are at the heart of the grammar. While p-string and speech signal seem to be different structures

in this figure, they are, in fact, one unit, where the p-string merely *represents* a sound wave. A bundle of vectors carrying prosodic information in relation to each syllable is extracted from the speech signal and projected to p-structure via the correspondence relation ρ . P-structure displays the content of these vectors in a compact and easily accessible way via a p-diagram.

5 Disambiguating coordination

There are several aspects of the speech signal which are relevant to structures in the grammar. At this point I will show how the information extracted from the speech signal can help to disambiguate syntactically ambiguous sentences. Consider the syntactic phrasing possibilities of the coordination in (6):

(6) Ravi and Amra or Karla



The two possibilities of phrasing for the coordination in (6) correspond to a fundamental difference in interpretation. The syntactic tree in (6a) groups Amra and Ravi and opposes the two to Karla, while the tree in (6b) takes Ravi and groups him with either Amra or Karla. However, the s-string does not allow for a syntactic (and semantic) distinction between the two choices. While the s-string and syntax are thus not able to disambiguate the sentence, p-structure is able to do so. The difference can, for example, be seen in an oscillogram of the above sentence (Figure 6), which represents (in short) the ‘waveform’ of the signal.

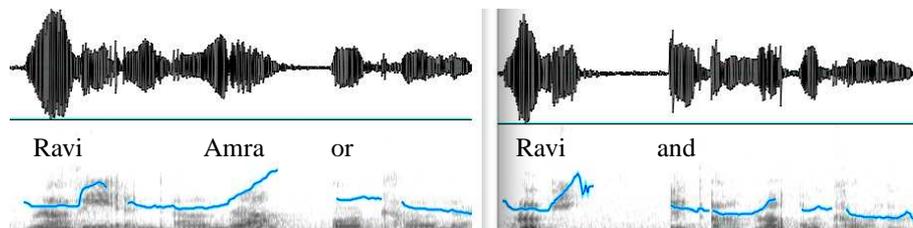


Figure 6: oscillogram for c-str. (6a) on the left and c-str. (6b) on the right

left oscillogram (c-str. (6a)): Clear break after *Amra*
 right oscillogram (c-str. (6b)): Clear break after *Ravi*

These pauses in the speech signal give a clear interpretation of the intended grouping.⁴ In the specific case of grouping, further indicators would be the pitch in the fundamental frequency (below the waveforms) and a possible lengthening of the last syllable of *Amra* on the left and *Ravi* on the right. For c-structure (5b) and the associated speech signal on the right of Figure 6, the following vectors represent the relevant part of the signal (*Ravi + pause*):

$$S: \begin{pmatrix} \text{pause} \\ \text{length} \\ F_0 \\ \text{stress} \\ \text{value} \end{pmatrix} (n) \Rightarrow S: \begin{pmatrix} - \\ 20ms \\ L \\ \text{prim} \\ /ra/ \end{pmatrix} (1); S: \begin{pmatrix} - \\ 25ms \\ H^* \\ - \\ /vi/ \end{pmatrix} (2); S: \begin{pmatrix} 3 \\ - \\ - \\ - \\ - \end{pmatrix} (3)$$

Figure 7: Vectors representing the speech signal (*Ravi + pause*)

These vectors are encoded in the p-diagram in Figure 8.

	↑	↑	↑	
	Pause	-	-	3
	Length	20ms	25ms	-
	F₀	L	H*	-
Vector	Stress	prim	-	-
	Value	/ra/	/vi/	-
	Vectorindex	S ₁	S ₂	S ₃
				... →

Figure 8: The p-diagram of *Ravi + pause*

The relevant information for the syntactic disambiguation can be retrieved from various variables and vectors. The most important factor in this constellation is the break after the string *Ravi* ([S₃, PAUSE]). Further indication comes from the long second syllable ([S₂, LENGTH]) and the high tone on this syllable ([S₂, F₀]). For the specific problem of grouped coordination, I will only refer to the break information.⁵

The information on breaks and the resulting boundaries is not automatically transferred to another structure of the grammar, as it is the case in Bögel et al. (2009),

⁴Such a clear-cut break is not always available, as has been noted by e.g., Allbritton et al. (1996). However, speakers who are aware of the grouping intention produce signals similar to the one in Figure 6.

⁵The tone and length indications have to be calculated in relation to the tones and length of other syllables in the sentence. While the tone information can be interpreted by itself as well, the length information has to be encoded differently for it to be meaningful. I leave this for further research.

and to a certain extent with Dalrymple and Mycock (2011), where the prosodic constituents are matched against and aligned with the syntactic constituents. Instead, the necessary information about possible breaks is *requested* by the respective structure. For the ambiguous noun coordination, this would result in an annotation like the one in example (7).

$$(7) \quad \text{NPcoord} \quad \rightarrow \quad \begin{array}{ccc} \text{NP} & \text{Conj} & \text{NPcoord} \\ (\uparrow_{\pi^{-1}\rho} S_{Nmax+1} \text{ PAUSE}) =c \ 4 \\ \text{OR} & \text{NPcoord} & \text{Conj} \quad \text{NP} \end{array}$$

(7) shows a (simplified) NP-coordination rule, which allows for a choice between the two groupings introduced in (6). The first rule represents [Ravi \wedge (Amra \vee Karla)] (tree 6b) and the second one [(Ravi \wedge Amra) \vee Karla] (tree 6a). Encoded under the first NP node is a restriction, which a) shows the path (the correspondence relation), b) indicates the relevant syllable vector and c) constrains the value of a specific attribute. The path describes the relation between two structures and thus refers to the general idea of correspondence, which allows for the parallel description of different aspects of linguistic information and the resulting dependencies (e.g., Halvorsen and Kaplan, 1995; Kaplan, 1995; Asudeh, 2006). The correspondence relation from this c-structure node to p-structure is described by the composition of the inverse correspondence relation from c-structure to the string (π^{-1}) and the correspondence relation between string and p-structure (ρ).

As discussed in sections 2 and 3, I view the p- and s-string as two parts of the same entity. The two representations are aligned with the help of the lexicon, which encodes both the p-form and the s-form of all lexical entries. This means essentially that any s-string element ‘knows’ the corresponding fragment of syllables of the p-string. Thus, the s-string element *Ravi* and the p-string fragment /ra.vi/ are interlinked. The projection from p-string to p-structure is then managed via the relation ρ ; the vectors and the related p-diagram have been discussed in Figures 7 and 8. The functional correspondence relation between c- and p-structure can thus be described by the following formula:

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

However, in this specific case, it is not the information of one of the corresponding syllables that is of interest to the syntactic rule, but the pause, which follows the last syllable. This fact is captured by the annotation S_{Nmax+1} , which refers to the syllable with the maximum index (the last one in any slice) and adds to this index 1 ($S_2 + 1 = S_3$). If this syllable vector has a pause value of 3, then the first c-structure rule in (7) is parsed. If it is another syllable as would be the case with the reverse grouping in tree (6a), this syntactic rule would not apply and the second one would be parsed.⁶

⁶The second possibility could also carry constraints related to the corresponding speech signal. These have been left out for reasons of simplicity. Furthermore, it would probably be more appropriate to implement an OT-constraint instead of an all-or-nothing condition. I will leave this for further research.

Figure 9 shows the architecture in relation to the (partial) parsing of the string *Ravi and Amra or Karla*, where the associated tree is parsed according to the information on the pause (S_3) in the p-structure.

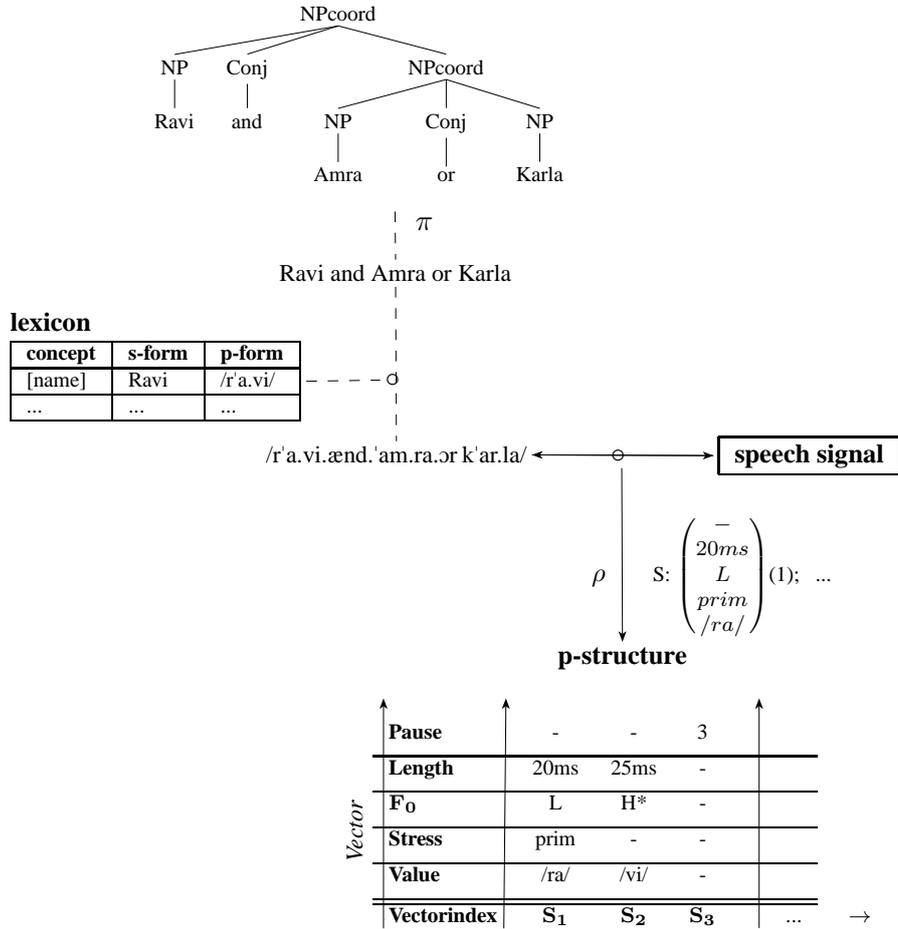


Figure 9: Partial processing of $[Ravi \wedge (Amra \vee Karla)]$

6 Conclusion

This paper presented a new approach to p-structure in several ways: First, the p-diagram presented in this paper is not based on the units of the Prosodic Hierarchy, as it is the case in Bögel et al. (2009) and Dalrymple and Mycock (2011). The reason for this is that the exact hierarchical structure of the different units, the recognition of these units within spoken language and the phonological theories behind these units are very controversial. Instead, the p-diagram approach bases the representation of the speech signal on the notion of the syllable, because syllables (or at least the nucleus of a syllable) is considered to be a very stable and easily recognized unit in the speech signal and is, in contrast to, e.g., phonological phrases, not as sensitive to external factors like speech tempo. Furthermore, the syllable is ‘small’ enough to allow for a fine-grained analysis of the speech signal, i.e., the syllable allows for a closer look at the phonological events within its range.

Second, the notions of the prosodic tree and the prosodic AVM were replaced by a more compact, linearly structured and easily accessible representation, the *p-diagram*. The p-diagram is composed of syllable-based vectors, which contain different aspects of prosodic information connected to their respective syllable in the speech signal itself. The vectors and thus the resulting p-diagram can be constructed according to the specific needs of the prosodic phenomena to be analysed.

The p-diagram approach is theory-independent, in that it does not assume theories of prosodic grouping according to hierarchical approaches (except for the syllable, which is, however, uncontroversial). Furthermore, it does not make predictions about the alignment of prosodic and syntactic units. It is able to encode all of the above aspects on the basis of the speech signal information encoded in the p-diagram.

It is not the basic intention of this approach to align syntactic and prosodic structures or to project prosodic events per se into other structures; instead, the information is depicted in the p-structure component itself and can be retrieved from the relevant projection in the grammar (e.g., the NP coordination rule in c-structure) via the composition of correspondence relations. Thus, the focus shifts from prosody back to other structures, where specific information from the speech signal is of relevance and can be checked in an easy and compact way via the correspondence relation of any structure to p-structure.

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.
- 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.
- Butt, Miriam and King, Tracy Holloway. 1998. Interfacing Phonology with LFG. In *Proceedings of LFG98*, CSLI Publications.
- Dalrymple, Mary and Mycock, Louise. 2011. The Prosody-Semantics Interface. In *Proceedings of LFG2011*, CSLI Publications.
- Halvorsen, Per-Kristian and Kaplan, Ronald M. 1995. Projections and Semantic Description in Lexical-Functional Grammar. In Mary Dalrymple, Ronald M. Kaplan, John T. Maxwell III and Annie Zaenen (eds.), *Formal Issues in Lexical-Functional Grammar*, pages 279–292, Stanford: CSLI Publications.
- 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.
- Kleinhenz, Ursula. 1998. *On Words and Phrases in Phonology*. Ph. D.thesis, University of Tübingen.
- Lahiri, Aditi and Plank, Frans. 2010. Phonological Phrasing in Germanic: The Judgement of History, Confirmed through Experiment. *Transactions of the Philological Society* 108(3), 370–398.
- Lahiri, Aditi and Reetz, Henning. 2002. Underspecified Recognition. In Carlos Gussenhoven, Natasha Werner and Toni Rietveld (eds.), *Labphon 7*, Berlin: Mouton.
- 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. 2006. *The Typology of Constituent Questions: A Lexical-Functional Grammar Analysis of 'WH'-Questions*. Ph.D.thesis, University of Manchester.
- Nespor, Marina and Vogel, Irene. 1986. *Prosodic Phonology*. Dordrecht: Foris.
- O'Connor, Rob. 2004. *Information Structure in Lexical-Functional Grammar: The Discourse-Prosody Correspondence in English and Serbo-Croatian*. Ph.D.thesis, University of Manchester.
- Selkirk, Elisabeth. 1984. *Phonology and Syntax: The Relation between Sound and Structure*. The MIT Press.
- Selkirk, Elisabeth O. 2011. The Syntax-Phonology Interface. In John Goldsmith, Jason Riggle and Alan C. L. Yu (eds.), *The Handbook of Phonological Theory*, pages 435–484, Malden, MA: Blackwell.
- Silverman, Kim, Beckman, Mary, Pitrelli, John, Ostendorf, Mari, Wightman, Colin, Price, Patti, Pierrehumbert, Janet and Hirschberg, Julia. 1992. TOBI: A Standard for Labeling English Prosody. In *Proceedings of the 1992 International Conference on Spoken Language Processing*, Banff.