6th European Conference on
Speech Communication and Technology
(EUROSPEECH'99)
Budapest, Hungary, September 5-9, 1999

A COMPARISON BETWEEN SYNTACTIC AND PROSODIC PHRASING

Marcus L. Fach
Institute of Natural Language Processing, University of Stuttgart
Azenbergstr.12, 70174 Stuttgart, Germany
Marcus.Fach@ims.uni-stuttgart.de
http://www.ims.uni-stuttgart.de/Phonetik

ABSTRACT

This study presents a comparison between syntactic and prosodic phrasing. A parser is used to calculate the syntactic structures from the orthographic text the prosodic structures of which are given by means of ToBI label files. For the automatic evaluation the prosodic break indices "3" (intermediate phrase boundary) and "4" (intonation phrase boundary) are compared with the terminals extracted from the extensive syntactic structures generated by the parser. These terminals are assumed to be the carriers of the phrase boundaries.

Keywords: syntax, parsing, prosody, phrase boundaries, prosodic labelling.

1. INTRODUCTION

This study aims to show that there is a strong correspondence between syntactic phrasing and prosodic phrasing. It is based on the comparison of ToBI-labelled break indices and the output of a parser.

There has been a lot of discussion about the correspondence between syntactic and prosodic phrasing ([9], [10], [11], [13]). Most of the criticism concerns syntactic analysis. While some analyses of syntactic structure are heavily recursive, prosodic representation is not. Thus, it is not possible to directly compare syntactic and prosodic structure. Figure 1, taken from [1], shows the differences between syntactic and prosodic structure.

In [1], Abney argues for a modification of the standard approaches to phrase structure (as shown in Figure 1) in order to achieve a much closer correspondence with the units of prosodic structure. The units resulting from these modifications are called *chunks*. Chunks are defined as tree fragments in which so-called "problematic" segments (e.g., prepositional phrase attachments) are left unattached. Figure 2. shows the resulting structure.

This method of analysis will finally lead to a very similar representation of prosodic and syntactic structure (see Figure 1).

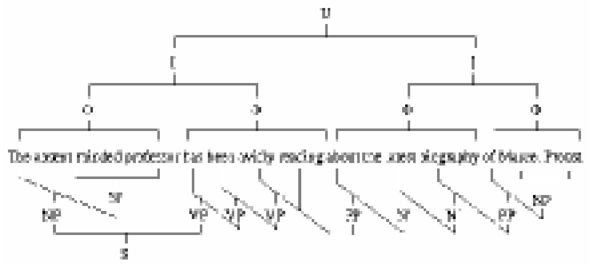

Figure 1. Syntactic and prosodic structure according to Abney [1].

The utterance is divided into sequences of simplex clauses, clauses consist of sequences of chunks, and chunks are made up of sequences of words.

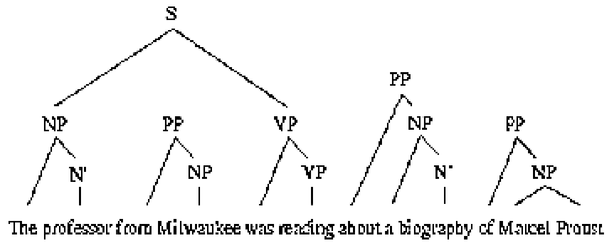

Figure 2. Chunk representation of the example of Fig1.

There are different labelling conventions for the prosodic boundaries as depicted in Figure 1. In [2] the prosodic boundaries are marked by break indices ranging in value from 0 to 6, a break index of 0 being assigned between words where no prosodic break is perceived and a break index of 6 marking sentence boundaries. Index 1 is assigned to prosodic word boundaries and index 2 groups several words together. The break indices 0-3 are called "minor" prosodic boundaries below the level of the intermediate phrase.
Finally the break indices 4 and 5 are assigned to intermediate and intonation phrases. The break indices 4-6, all marking boundary tones constitute the "major" prosodic boundaries, cf. ([4],[5],[7]).

A subset of this labelling scheme is found in [3], where the break indices have range from 1 to 4. This subset is used in the study presented here. An example of this labelling system is shown in Fig 3. The

10.21437/Eurospeech.1999-136

sample utterance consists of two intonation phrases (IPs), as marked by break index 4. In the first IP there are no intermediate phrases, while the in the second IP there is an intermediate phrase (ip), as marked by break index 3.

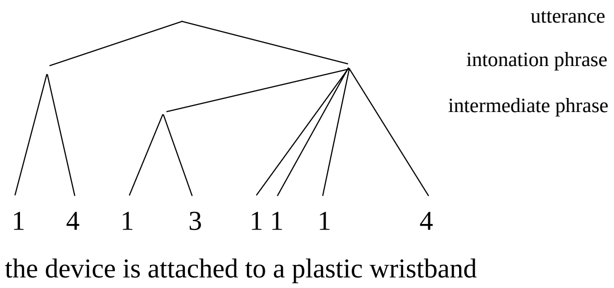

Figure 3. A sample utterance from the Boston Radio Newscorpus [3] with break indices assigned to each word.

Based on the hypothesis that only the major prosodic boundaries have syntactic correlates, the break indices 1 and 2 are not treated in this study.

2. DATA

The data for the experiment described here is provided by the Boston Radio News Corpus [3] which is a database of spoken American English recorded from WBUR radio studio during broadcast. The corpus consists of news stories read by seven speakers (three female and four male). Each news story consists of several utterances which are annotated with orthographic transcription, phonetic alignments, part-of-speech tags and prosodic labels. The phonetic alignment is done by a segmentation algorithm using stochastic segment models (see [3] for further details on the alignment).

In accordance with the ToBI system [7] the description of American English prosody the different label types are structured in different tiers. Thus, tonal structure (accents and phrase tones) is represented on a tone tier and phrasing structure on a break tier, etc. The speech files are digitized at a 16kHz sample rate using a 16 bit A/D.

3. THE EXPERIMENT

For the subset of the Boston Radio News Corpus used in the experiments one female announcer (speaker f2b) was chosen. This database contains 12613 words (counting contractions like „it's“ or „I'm“ as one word) and 3902 breaks (counting only the break indices "3" and "4") in all in all 165 news stories. From these 165 stories the syntactic structure was generated by the parser described in the following subsection 3.1.

3.1. The Parser

The chunk parser employed in this study was implemented at our institute by [8]. The basic concept of chunks follows the above cited principles [1]. Categories are interpretable in terms of a feature decomposition, but are treated as atomic in this formalism. The grammar includes complementation rules for verbs, nouns, and adjectives. Complements are attached at a level above the chunks, called the phrasal level. For further details on grammar and formalism see [8].

The input to the parser has to be ascii text. There are only a few constraints on the format of the input text, e.g., the punctuation marks have to be separated by blanks and each sentence must to correspond to one line.

The output also is ascii text. The tree structure is coded by brackets. Figure 4 shows the sample utterances from Figure 3 in a typical tree representation.

3.2. Analysis

In order to compare the different boundary information the tree-like output of the parser has to be converted into a format which contains only the boundary information. This information is coded in the nonterminal nodes of the tree.

As can be seen in Figure 4, the level below the sentence node (*S_C-attached*) represents the chunks (*NC1-device, VFC1-attached*) which should correspond to the prosodic boundaries at the level of intonation phrases. On the next lower level (N_C-device, VFP_C-attached, PC1-to) the intermediate phrase boundaries are made explicit. The one to one correspondence can be seen when comparing Figure 3 and Figure 4.

Two different experiments are conducted. In the first experiment all chunks which consist of only one word are treated as carriers of boundary information. In the second experiment these one-word-chunks are not treated as chunks and consequently not treated as carriers of boundary information. Thus the one-word-chunks in first experiment are denoted as "true" (T) and in the second experiment as "false" (F).

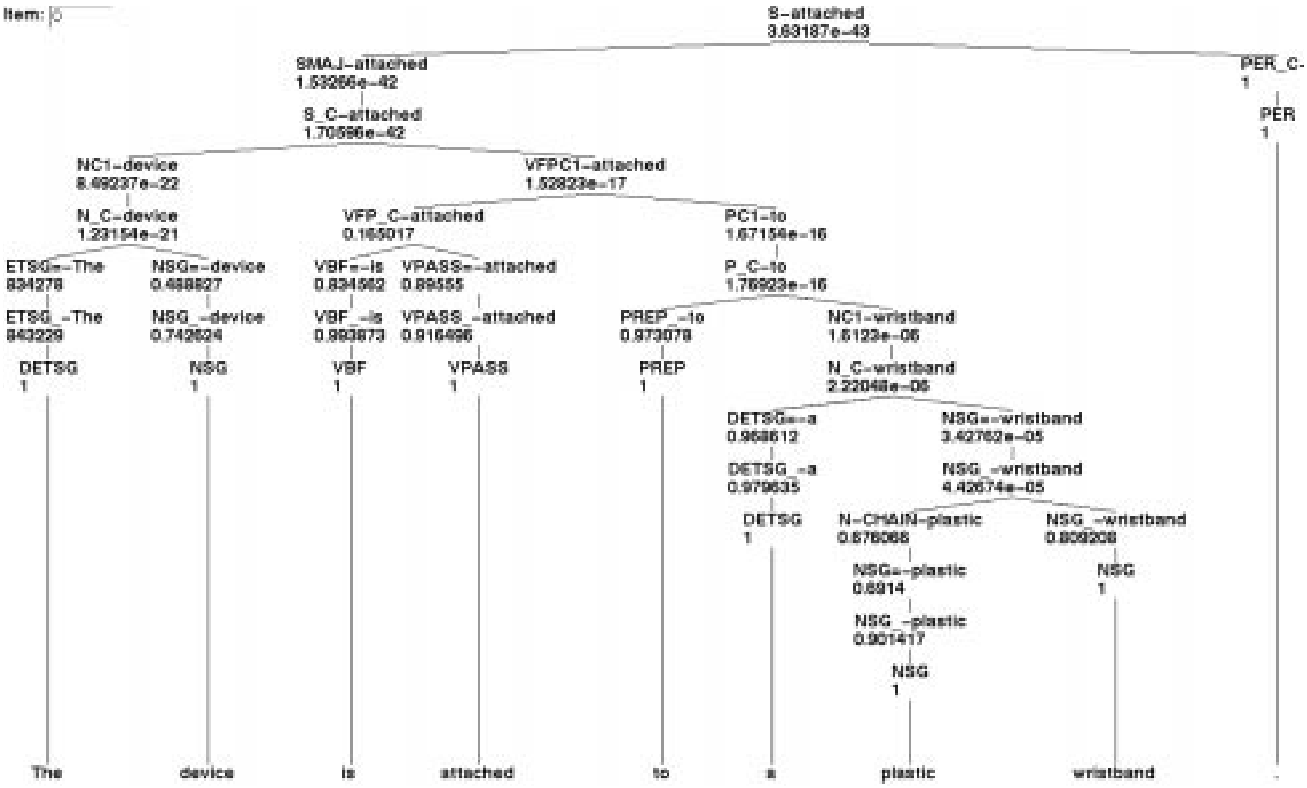

Figure 4. Sample sentence from Fig1 in the representation of the chunk parser.

The comparison on the level of the intermediate phrase boundaries is the same for both experiments. This comparison is made treating the rightmost terminal in a chunk as the carrier of the boundary information denoted with T. All other terminals in the same chunk are denoted with F.

Figure 5 illustrates the mapping between the phrasing levels and the indices to be compared. The top line indicates the syntactic phrasing and the bottom line shows the prosodic phrasing.

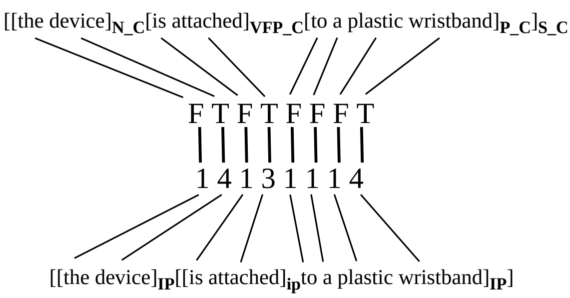

Figure 5. Mapping of prosodic structure onto syntactic structure.

The syntactic phrasing is mapped onto boolean variables T and F, using the corresponding break tier labelfile from the database is used.

4. RESULTS

The results of the experiments are given in terms of "Recall" and "Precision" as defined in [6]. The results reflect the syntactic distinction between boundary/no boundary, overall results are also given.

Experiment 1: one-word-chunks are boundary carriers.

	Recall	Precision
chunk boundaries	94%	47%
no chunk boundaries	52%	95%
both	65%	65%

Table 1. Results for experiment 1

Experiment 2: one-word-chunks are no boundary carriers.

	Recall	Precision
chunk boundaries	70%	77%
no chunk boundaries	90%	87%
both	84%	84%

Table 2. Results for experiment 2

The difference between experiment 1 and experiment 2 is caused by the specific speaking style of the respective radio announcer. The overall results in experiment 2 are significantly higher because not all possible one-word-chunk boundaries are realized by the announcer. Therefore the experiment 2 constitutes an optimization for the read speech of radio announcers. For spontaneous speech, however, the algorithm from experiment 1 is the adequate method.

The experiments' overall results are a good base for further studies. A larger database may lead to even better results. Experiments on spontaneous speech will also be interesting, but there are a lot of problems still to be solved concerning the corresponding parser. Further development in parsing will enable future studies comparing syntactic and prosodic phrasing.

5. CONCLUSION

While prosodic boundaries do mirror syntactic boundaries, they are also influenced by other factors like rhythmic constraints and speaker-specific style. Nevertheless, the results of the experiment show that at least for read speech 65% of syntactic boundaries are coded in the prosodic boundaries. An optimizatition for the specific radio announcer style will lead up to 84%.

These results are encouraging considering their significance for the generation process in speech synthesis. The text to be synthesized can pass a parser without any problems, but there are a lot difficulties to generate adequate prosodic information in the stage of pre-processing. However, the results of the experiment lead to the assumption that the required prosodic boundary information can be derived from the syntactic structure calculated by the parser. The results of this study have to be evaluated by further experiments in speech synthesis systems.

For the speech recognition systems, on the other hand, the results are less satifying. Nevertheless, prosodic phrase boundary marking seems to be the only way to disambiguate the syntactic structure of speech (cf. [6], [12])

6. REFERENCES

[1] Abney, S. Chunks and Dependencies: Bringing Processing Evidence to Bear on Syntax. *University of Tuebingen (1991).* http://www.sfs.nphil.uni-tuebingen.de/~abney.

[2] Wightman, C.W., Shattuck-Hufnagel, S., Ostendorf, M., Price, P. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of Acoustical Society of America, Vol. 91, No.3, 1707-1717.*

[3] Ostendorf, M., Price, P.J., Shattuck-Hufnagel,S. (1995). The Boston University Radio News Corpus. *Boston University.*

[4] Batliner,A., Kompe, R., Kiessling, A., Mast, M., Niemann, H., Noeth, E. M = Syntax + Prosody: A syntactic-prosodic labelling scheme for large spontaneous speech databases. *Speech Communication 25 (1998) 193-222.*

[5] Taylor, P., Black, A. W. Assigning phrase breaks from part-of-speech sequences. *Computer Speech and Language (1998) 12, 99-117.*

[6] Batliner, A., Feldhaus, A., Geissler, S., Kiessling, A., Kiss, T., Kompe, R., Noeth, E. Integrating Syntactic and Prosodic Information for the Efficient Detection of Empty Categories. *In: Proc. International Conference on Computational Linguistics, Copenhagen, Vol. 1, pp. 71-76.*

[7] Silverman, K., Beckman, M., Pitrelli, J., Osterdorf, M., Wightman, C., Price, P., Pierrehumbert, J., Hirschberg, J. ToBI: a standard for labelling english prosody. *Proceedings, 2nd International Conference on Spoken Language Processing, Banff/Canada, 1992, 867-870.*

[8] Carrol, G., Rooth, M. Valence Induction with a Head-Lexicalized PCFG. *Emperical Methods in NLP workshop, Granada, 1998.*

[9] Cinque, G. A null theory of phrase and compound stress. *Linguistic Inquiry 24:239-297. 1993.*

[10] Selkrik, E. Phonology and syntax: The relation between sound and structure. *MIT Press, Cambridge, Mass., 1984.*

[11] Bader, M. Sprachverstehen: Syntax und Prosodie beim Lesen. *Dissertation, University of Stuttgart, 1995.*

[12] Steinhauer, K., Alter, K., Friederici, A.D. Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature neuroscience, vol. 2, no.2, 1999.*

[13] Tuckenbrodt, H. Phonological phrases: Their relation to syntax, focus, and prominence. *Doctoral dissertation, MIT, Cambridge, Mass., 1995.*