

PROSODIC CUES FOR INFERRING AND REALIZING DISCOURSE RELATIONS

Ernst Buchberger

Austrian Research Institute for Artificial Intelligence (ÖFAI), Vienna, Austria*

ABSTRACT

This paper reports on investigations into the relationship between prosody and discourse relations. We present the analysis of an example, situated in a formal framework composed of TSM for the representation of prosody and (an extension to) DRT for the representation of discourse, and we argue that the results can be exploited fruitfully for two main areas in the automatic processing of language and speech: analysis (parsing) and synthesis (generation).

INTRODUCTION

In computational linguistics and AI, much research has been performed concerning the structure of discourse, e.g. [1,2,3]. It has been noted that "there may even be a hope of using the various sorts of clues in a program to discover with a reasonable probability of success the underlying rhetorical relations" [3, p.310] and that "rhetorical relations may, though need not, be explicitly signaled by some expression in the text" [3, p.264]. Research specifying this relation between rhetorical relations and cue words includes among others [4,5]. Up to now, the search for cues has been mainly restricted to written text, though. We have investigated the possibility that clues may also be taken from prosody.

Our analyses are based on an adaptation of Fery's [6] application of Pierrehumbert's tone sequence model (TSM) [7] to German with regard to prosodic modelling, and on Asher's [3] extension of Discourse Representation Theory (DRT) [8], which integrates the notion of discourse relations into DRT.

AN EXAMPLE

The example below is one of a number of small discourses taken from German radio news which we have investigated. The text ("Nachrichten39" - see below) has been hand-labelled with ToBI Label-

ling Tools according to the conventions developed at IMS [9] (see figure 1).

Im Berliner Prozeß gegen zwei Polizeibeamte wegen der Mißhandlung eines Vietnamesen sind die beiden Angeklagten heute freigesprochen worden. Das Gericht wertete die Aussagen des geschädigten Vietnamesen als widersprüchlich. Offensichtlich hätten sich die Beamten zwar sehr barsch benommen, eine strafbare Tat sei aber nicht erkennbar.

The corresponding Structured Discourse Representation Structure SDRS (simplified) is depicted in figure 2. Be-

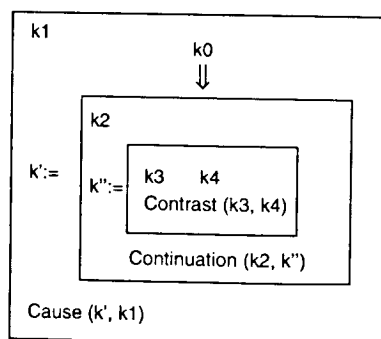


Figure 2. SDRS Representation of Text "Nachrichten39"

fore we start discussing how to arrive at this SDRS representation, we have to bear in mind that we should not expect to derive all elements of its structure purely by means of prosody. We will, however, show where prosodic cues may help us to arrive at a correct interpretation.

The basic constituents of our representation can be found by drawing regression lines through the F0 curve [10]. At a point where a noticeable discontinuity is found, a new line (and with it, a new DRS element) will start. This gives us the four DRSs k1 to k4. k1 corresponds to the first sentence, k2 to the second, k3 to the clause *Offensichtlich ... benommen*, and k4 to the rest of the third sentence (k0 does not correspond to an element in the text, but has to be constructed according

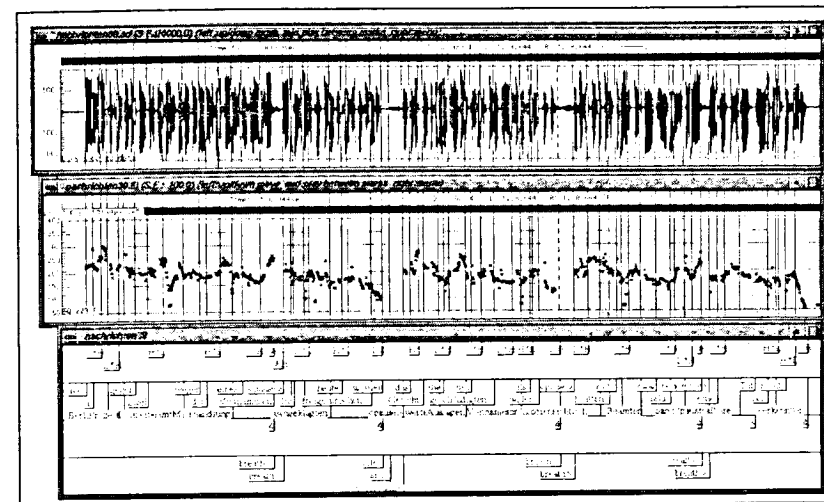


Figure 1. Signal, F0, and Labels for Text "Nachrichten39"

to Asher's topic construction rules). After having processed the first sentence and constructed k1, we find k2 with a compressed pitch range. This compression, and the relatively high ending of k2 suggests a Continuation relation to hold between k2 and what follows. k2 forms part of an SDRS k' which will be constructed only later on. We will, however, ask already at this point whether a clue can be found how this SDRS will be linked to k1. There is indeed one to be found, which is mainly inferential in nature along the lines of Dahlgren's naive semantics [11], but there is also a prosodic marking present, namely the conspicuous pitch accent on *widersprüchlich* ("contradictory"), marking the Cause for k1. In a news text, this prosodic marking is quite advantageous, especially for an inattentive hearer of radio news, (who might be listening to the news while being occupied with other tasks) whose attention is thus drawn to the relevant facts. The connection between k3 and k4 can be clearly seen from the progradient phrase tone % at the end of *benommen*. In the absence of further information we would take this as an indication for a Continuation relation between k3 and k4, but here we have the lexical clue of the word *aber* ("but") from which we infer a Contrastive relation. The pitch accents help to find the contrastive

elements *strafbar* ("punishable") and *nicht erkennbar* ("not to-be-found").

Finally, the final lowering is quite pronounced, which means that closing-off of the yet open discourse structures can be inferred easily. This way, k' will be closed off and the Continuation relation between k2 and k'' we already mentioned above can be constructed now, forming the SDRS k'. What remains to be done is the construction of the Cause relation between k1 and k' which we also had mentioned already, completing the representation.

EXPERIMENTAL VALIDATION

Up to now, recorded speech data from radio news have been hand-labelled and analyzed with regard to the discourse relations involved.

Recently, we have started resynthesizing texts with contours suggesting discourse relations differing from those found in the original texts, using a testing environment which allows for a free modification of pitch and duration of the recorded speech data before the resynthesis based on the PSOLA algorithm is performed [12]. The effects of these resynthesized texts on human listeners are being tested.

*Work reported here was performed while the author was a guest researcher at the Institute of Natural Language Processing (IMS), Stuttgart, Germany

APPLICATIONS

Prosody is considered an important factor for both automatic natural language analysis and synthesis (or parsing and generation). The relation between parsing and generation has often been thematized in Natural Language Processing, specifically regarding grammar [13]. The role of prosody can be summarized as follows: in analysis, it helps to disambiguate, in synthesis, it allows for more natural utterances.

As for analysis, the usefulness of (prosodically derived) boundary information for disambiguating syntactically ambiguous utterances has been shown by other researchers [14,15]. At IMS the effects of pitch accents on pronoun resolution [16,17] are currently investigated, some of which have also an impact on discourse relations. As for ambiguous discourse structure, it seems that discrimination will be mostly relevant not so much for the kind of discourse relation, but mainly for the attachment points of these discourse relations (cf. [18,19]).

With regard to generation, whereas previous research has often focussed on conciseness, based on Grice's maxims [20], everyday utterances show a significant amount of redundancy. Taking prosody into account, it may be advantageous to code discourse relations prosodically and textually, e.g., by not only using cue words, but also marking them prosodically in a consistent way. Thus, even if the lexical clues alone would have allowed inferring the intended meaning, prosodic clues might enhance the textual clues and lead to more natural utterances.

The aspect of unwanted implicature has recently been noted in the generation literature [21]: When in a (written) report an entity is singled out by ascribing to it certain properties, the impression is created that for other non-mentioned entities this property does not hold. The example presented in [21] can be taken over to speech: If we say "Person A worked on task B. HE finished the work in time.", putting (erroneously) a pitch accent on "he", this will create the implication that there were other people involved in the project who did not finish in time.

Finally, during the last years much work in Natural Language Generation has been based on some sort of model for representing discourse relations, most nota-

bly RST [22,23]. With a specification of correspondences between a formal representation of discourse and a model of prosody we can hope to find a way to integrate Natural Language Generation and Speech Synthesis in an elegant way.

PRELIMINARY CONCLUSIONS AND FUTURE WORK

We have found some evidence for prosodic marking of discourse relations and sketched a framework for the investigation of this interdependency. The relation between prosodic phenomena and discourse relations is not a bijective function, rather it has to be seen as clues suggesting a certain structure of discourse, but not enforcing it.

The tone sequence model (TSM) has proven useful for describing some aspects of local discourse structure, and partially also for global discourse structure. What is still lacking is a formal description of other relevant parameters for global discourse structure like pitch range. Also missing is a formal description of the interaction of the various sorts of discourse clues, integrating the contribution of lexical and prosodic cues, among others. Asher has proposed to use nonmonotonic logic to capture the dependency of various discourse relations, but he did not talk about prosody. Maybe prosodic cues could be integrated into such a framework. Another possible candidate might be optimality theory [24], which has recently been shown to be a useful framework for the description of various aspects of prosody, but has – to our knowledge – not yet been applied to the description of interactions between prosody and discourse.

ACKNOWLEDGEMENTS

The research reported here forms part of the multi-national project *Scientific Cooperation in the European Network in Language and Speech* (Contract No. CHRX-CT93-0421). It was performed during a stay at the Institute of Natural Language Processing (IMS), Stuttgart, Germany, and sponsored by the EU's HCM programme. Many thanks to Grzegorz Dogil for fruitful discussions and useful hints, to Jörg Mayer and Stefan Rapp for help with labelling and graphics, and to all other members of IMS / Experimental Phonetics for the cooperative working atmosphere.

REFERENCES

- [1] Grosz B.J., Sidner C.L. Attention, Intentions, and the Structure of Discourse, *Computational Linguistics* 12(3)175-204, 1986.
- [2] Mann W.C., Thompson S.A.: Rhetorical Structure Theory: A Theory of Text Organization, in Polanyi L.(ed.), *The Structure of Discourse*, Ablex, Norwood, N.J., 1987.
- [3] Asher N.: Reference to Abstract Objects in Discourse, Kluwer, Dordrecht, 1993.
- [4] Scott D.R., Souza C.S.: Getting the Message Across in RST-based Text Generation, in Dale R., et al.(eds.), *Current Issues in Natural Language Generation*, Academic Press, New York, 1990.
- [5] Rösner D., Stede M.: TECHDOC: A System for the Automatic Production of Multilingual Technical Documents, in Götz G.(ed.), *KONVENS 92*, Springer, Berlin, 1992.
- [6] Féry C.: German Intonational Patterns, Niemeyer, Tübingen, 1993.
- [7] Pierrehumbert J.B.: The Phonology and Phonetics of English Intonation, Ph.D. Thesis, MIT, Cambridge, MA, 1980.
- [8] Kamp H., Reyle U.: *From Discourse to Logic*, Kluwer, Boston/Dordrecht/London, 1993.
- [9] Mayer J.: A Guide to Labelling within the Stuttgart Labelling System, Technical Report, IMS, Univ. Stuttgart, forthc.
- [10] Huber D.: On the Discourse Function of Intonation, *Proceedings ICPhS, Aix-en-Provence*, 5:191-193, 1991.
- [11] Dahlgren K., McDowell J., Stabler E.P.Jr.: Knowledge Representation for Commonsense Reasoning with Text, *Computational Linguistics* 15(3)149-170; 1989.
- [12] Möhler G., Dogil G.: Test Environment for the Two Level Model of Germanic Prominence, *Eurospeech '95*, Madrid, 1995.
- [13] Strzalkowski T.(ed.): *Reversible Grammar in Natural Language Processing*, University of California, Berkeley, CA, 1991.
- [14] Lehiste I.: Phonetic Disambiguation of Syntactic Ambiguity, *Glossa* 7, 197-222, 1973.
- [15] Ostendorf M., Price P., Bear J., Wightman C.W.: The Use of Relative Duration in Syntactic Disambiguation, in *Proceedings of the DARPA Speech and Natural Language Workshop*, Morgan Kaufmann, 1990.
- [16] Mayer J.: Prosodische Disambiguation von Anaphern im Diskurs, AG 10 *Suprasegmentale Phonologie: Modelle und Prozesse*, DGfS-Jahrestagung Göttingen, March 1-3, 1995.
- [17] Dogil G.: Prosodic Cues to "Alternative" Semantics in DRT, unpublished talk held at the HCM Workshop on Discourse and Dialogue Prosody in Stuttgart, Feb 14-15, 1995.
- [18] Pierrehumbert J.B., Hirschberg J.: The Meaning of Intonational Contours in the Interpretation of Discourse, in Cohen P.R., Morgan J., Pollack M.E.(eds.), *Intensions in Communication*, MIT Press, Cambridge, MA, 1990.
- [19] Hobbs J.R.: The Pierrehumbert-Hirschberg Theory of Intonational Meaning, in Cohen P.R., Morgan J., Pollack M.E.(eds.), *Intensions in Communication*, MIT Press, Cambridge, MA, 1990.
- [20] Grice H.P.: *Logic and Conversation*, in Cole P., Morgan J.L.(eds.), *Syntax and Semantics 3: Speech Acts*, Academic Press, New York, 1975.
- [21] André E., Buchberger E., Cawsey A., Korelsky T., Not E., Rösner D., Teich E.: Report from Working Group 3 - Text Structure and Architecture in Natural Language Generation, in Höppner W., Horacek H., Moore J. (eds.), *Principles of Natural Language Generation*, Dagstuhl Seminar Report, Schloss Dagstuhl, Germany, 1995.
- [22] Hovy E.H.: *Generating Natural Language under Pragmatic Constraint*, Lawrence Erlbaum, Hillsdale, NJ, 1988.
- [23] Moore J.D., Paris C.L.: *Planning Text for Advisory Dialogues: Capturing Intentional and Rhetorical Information*, *Computational Linguistics* 19(4), 651-694, 1994.
- [24] Prince A., Smolensky P.: *Optimality Theory*, Technical Reports of the Rutgers University Centre for Cognitive Science, TR-2, 1993.