Giving Prosody a Meaning

Christian Lieske[1], Johan Bos[1], Martin Emele[2], Björn Gambäck[3], CJ Rupp[1]

(1) Computational Linguistics, University of Saarland; Postfach 151150; D-660 41 Saarbrücken

(2) Institute of Computational Linguistics, University of Stuttgart; Azenbergstrasse 12; D-70174 Stuttgart

(3) Centre for Speech Technology, Royal Institute of Technology; Box 70014; S-100 40 Stockholm

1 Introduction

Systems for spoken-language understanding can use prosodic information on the speech recognition side as well as the linguistic processing side. In the former case, prosody improves recognition accuracy and speed. In the latter case, it contributes to the computation of meaning. The following paragraphs discuss this meaning-related use of prosody in the spoken-language machine translation system Verbmobil (VM).

The overall goal of the Verbmobil system is to provide speech-to-speech translations from both German and Japanese to English. The scenario is that of appointment scheduling dialogues between businessmen, and the overall system comprises some 50 modules which use a specifically designed architecture and protocol to communicate with each other [1].

2 Language and Prosody in VM

The prosody module of the Verbmobil system connects to the recording unit and the speech recognizer, on the input side, and the morphology module, on the output side. The data structure for communication (input from the recognizer and output) are *word lattices* whose edges are annotated with recognition probabilities and so-called *infostrings* which amongst others may encode three different kinds of prosodic information: sentence modality, phrase boundaries, and stress [2].

In a very wide sense, Verbmobil comprises five modules related to linguistic processing: syntactic-semantic processing (SynSem), semantic evaluation, transfer (TR), generation, and dialogue. Of those, currently only SynSem and TR make use of prosodic information, so we restrict our sketch of the system to these two compoments.

Syntactic-semantic processing is based on a parser for a unification-based grammar which interleaves syntactic analysis and semantic construction [3]. Semantic construction compositionally builds representations called VITs (Verbmobil Interface Terms) which include semantic, syntactic, pragmatic and prosodic information. VITs allow the representation of ambiguities such as the relative scope of quantifiers, and are very flat (minimal recursive) structures [4].

Verbmobil adheres to the idea of transfer-based machine translation (transfer based on semantic representations to be specific), i.e. there is a non-trivial mapping between the structures resulting from the analysis of a source-language (SL) utterance and the structures used for the generation of the corresponding target-language (TL) expression. The approach is a compositional one, meaning that the semantic predicates for the SL are mapped in chunks onto the semantic predicates for the TL [5]. Two features are worth mentioning: Firstly, since pragmatic information is found in the VITs, the rules are even able to do some local anaphora resolution. Secondly, since VITs are underspecified semantic representations, transfer can preserve ambiguities, e.g., related to scopal relationships, when mapping from SL to TL. Transfer rules can utilize all of the information found in the VIT (e.g., values of syntactic features).

3 Prosody-Language Interface

Coupling the prosody module with the module for syntactic-semantic processing, requires some efforts, since a statistics-based tradition (prosody) and a logic-based one (computational linguistics) have to be reconciled [6]. Problems arise because mainstream unification-based linguistic processing 1) uses symbols but the prosody module yields numbers (e.g., *there is a ninety percent change that this utterance is a declarative sentence*), 2) attaches information to lexical items whereas the prosody module delivers information related to larger units (e.g., *here we have got a fifty percent chance of a rising progredient*), and 3) leaves relationships underspecified (and thus meaning representations ambiguous) but the prosody module does not allow for ambiguities (e.g., one cannot find prosodic information of the forme *if this word hypothesis is stressed than the utterance is a question whereas otherwise it is a declarative statement*).

In terms of implementation, two things had to be done to deal with this in Verbmobil: Firstly, the numeric values for prosodic features from the word lattices had to be transformed into a set of predefined symbols (since this is what a unification-based grammar can handle). Secondly, information from these symbols had to be translated into feature values. The solution was the following: The parser for our gram-

mar for each node in the word lattice creates an additional node/symbol encoding information about accent, boundary and sentence modality (called *prosodic wordform*). For this, the information in the infostring is used, and the symbols look like *ak2_b3gr_prsprog*. For each syntactic category c, we then write grammar rules like `c -> c p`, i.e., require each word of category c to be followed by a an input symbol belonging to the category p of prosodic wordforms. The relationship between the categories c and actual input words as well as the relationship between the generic representation p and feature values related to prosody is established in the lexicon (where, e.g., the input symbol *ak2_b3gr_prsprog* is mapped onto the category symbol p with accent feature *ak* being assigned the value *2*).

4 Prosody and Meaning

As mentioned, two Verbmobil modules related to linguistic processing make use of prosodic information: syntactic-semantic processing (SynSem) and transfer (TR). This section concentrates on SynSem, the following section with examples in addition describes the utilization during transfer. SynSem makes *conservative use* of prosodic information in the sense that it is only used if syntactic and semantic evidence do not override it. If, e.g., the grammar does not allow for a clause boundary but prosody indicates that there is one, then the grammar takes precedence.

Currently, prosodic information is dealt with in three areas of SynSem:

segmentation Since grammars are clause-oriented but utterances may contain more than one clause, the overall utterance has to be segmented into clauses (or other meaningful segments) as described in [6, ?].

sentence mood Evidence from syntax or semantics (e.g., occurence of wh-words) often is sufficient to determine the whether a sentence is declarative, imperative, or a question. For an isolated German verb initial sentence like *treffen wir uns am montag*, however, only prosody reveals if it is a question, an imperative, or even a declarative with topic drop.

focus The interpretation of particles like *auch* and *nur* depends so-called *semantic focus*. Prosody can help to determine the concept in focus, since it is usually a stressed word or phrase in its scope [7]. For instance, *nur am Montag Vormittag* may mean either only in the morning, on Monday, or only on Monday, in the morning.

5 Lexical Choice in Transfer

Also transfer makes use of prosodic information when processing focus related adverbs and adjectives (e.g., *noch*). An example, is *lexical choice* or *transfer equivalency*: Rule (1) says that the predicate related to `noch` and the predicate `indef` related to the indefinite article `a` should be mapped to a predicate for `another` if the predicate `noch` is stressed. The relevant test is checked in the conditional part of the transfer rule via the predicate `pros_accent` which refers to the symbolic information about prosodically stressed predicates in the prosdoy slot of the VIT.

(1)
```
[L:noch_fadv(F,S), L1:indef(I,G,S1)],
    [pros_accent(L)] -> [L:another(I,G,S1)].
```

Otherwise in case of absence of stress for `noch` this rule will not trigger and the default rule (2) is applied which deletes `noch`.

(2)
```
[L:noch_fadv(F,S)],
    [L:indef(I,G,_)] -> [].
```

For example an utterance like *Können wir noch einen Termin ausmachen?* is translated either as *Could we arrange* **another** *appointment?* vs. *Could we arrange* **an** *appointment?* depending whether `noch` is stressed or not.

6 Further Work

Our research agenda for the coupling of prosodic processing and linguistic processing addresses two issues: Firstly, we want to identify additional cases where prosody can add constraints pertaining to syntax or semantics. Secondly, we are aiming at refining the calculations that are involved in the addition of constraints due to prosodic information.

References

[1] M. Kay, J. M. Gawron, and P. Norvig, *Verbmobil: A Translation System for Face-to-Face Dialog.* No. 33 in Lecture Notes, Stanford, California: CSLI, 1994.

[2] W. Hess, A. Batliner, A. Kiessling, R. Kompe, E. Nöth, A. Petzold, M. Reyelt, and V. Strom, "Prosodic modules for speech recognition and understanding in VERBMOBIL," in *Computing Prosody: Approaches to a Computational Analysis and Modelling of Prosody of Spontaneous Speech* (Y. Sagisaka, N. Campbell, and N. Higuchi, eds.), ch. 23, pp. 363–384, New York, New York: Springer, Nov. 1996.

[3] H. U. Block and S. Schachtl, "Trace and unification grammar," in *Proceedings of the 14th International Conference on Computational Linguistics*, vol. 2, (Nantes, France), pp. 658–664, ACL, July 1992.

[4] J. Bos, B. Gambäck, C. Lieske, Y. Mori, M. Pinkal, and K. Worm, "Compositional semantics in Verbmobil," in Bos *et al* 1996 [8], pp. 131–136.

[5] M. Dorna and M. C. Emele, "Semantic-based transfer," in COLING96 [8], pp. 316–321.

[6] G. Bakenecker, U. Block, A. Batlinger, R. Kompe, E. Nöth, and P. Regel-Brietzmann, "Improving parsing by incorporating 'prosodic clause boundaries' into a grammar," in *Proceedings of the 3rd International Conference on Spoken Language Processing*, vol. 3, (Yokohama, Japan), pp. 1115–1118, Sept. 1994.

[7] J. Bos, A. Batlinger, and R. Kompe, "On the use of prosody for semantic disambiguation in VERBMOBIL," VM Memo 82, Universität des Saarlandes, Saarbrücken, Germany, July 1995.

[8] ACL, *Proceedings of the 16th International Conference on Computational Linguistics*, (København, Denmark), Aug. 1996.