A corpus study on the anaphoricity of the German pronominal adverb *danach*[1]

Mareile Knees

Inst. Germanist. Sprachw., Friedrich-Schiller-Universität Jena, Germany

`mareile.knees@uni-jena.de`

The relational anaphor *danach*

It is assumed that a discourse model contains representations of extra-linguistic entities (discourse referents) that have been specified by linguistic expressions in the ongoing discourse. It also stores attributes of and relations between these discourse referents. Following Webber et al. (2003) and Miltsakaki et. al (2003), adverbial connectives like *however*, *therefore*, etc. express (as subordinate and coordinate conjunctions) binary predicate-argument relations. Nonetheless, they only get one of their two arguments structurally, namely the matrix clause. The other argument has to be anaphorically derived from the discourse context. In German there are so-called pronominal adverbs like *danach* (*after that*) and *dadurch* (*thereby*) which consist of an anaphoric element (e.g. *da-*) and a relational element (e.g. *-nach* or *-durch*) (cf. Fraurud 1992, Rüttenauer 1978). Due to their anaphoric element these pronominal adverbs refer to a referent previously introduced into the discourse model. Besides their anaphoric reference they also establish a temporal, causal or any other discourse relation between the referent of the antecedent and the referent of the matrix clause of the anaphor (s. (1)). Thus, they can function as relational anaphora and adverbial connectives (Webber et al. 2003, Miltsakaki et al. 2003).

(1)   *Eine Abordnung des Münchner Polizeipräsidiums legte an der Gedenktafel einen Kranz für den 1972 getöteten Polizeibeamten Anton Fliegerbauer nieder*. **Danach** *begaben sich die Mitglieder der Deutsch-Israelischen Gesellschaft und der anderen Gruppen zu Fuß ins 17 Kilometer entfernte Dachau.*
(Tigercorpus, 1201-1202)

    'A delegation of the police headquarters of Munich lay down a chaplet at the commemorative plaque of the police officer Anton Fliegerbauer killed 1972. **After that** the members of the German-Israelic Society and the other groups went to the 17 kilometers away Dachau.'

In order to find out how the resolution of the relational anaphor *danach* works, I first consider and classify a variety of occurrences of *danach* by means of a corpus study. There are approximately 100 anaphorically used occurrences of *danach* in the Tigercorpus. Syntactically (with respect to the syntactic properties of the antecedent), these occurrences can be divided into two main classes: either the antecedent is an NP or PP or it is a clause-based phrase (a sentence, a clause, a VP etc.). From a semantic perspective the occurrences can be classified with respect to the relation between the referent of the antecedent and the referent of the clause containing the anaphor. Surprisingly, only in half of the cases *danach* expresses temporal succession. In the Tigercorpus *danach* is often used in order to introduce the content of a previously mentioned ‚documentation'-noun like *study*, *agreement* or *report*. As the corpus consists of newspaper articles this finding might correlate with the text sort. Considering the ontological status of the discourse referent, we can distinguish between situational reference (the referent is a situation, cf. Consten and Knees forthc.), temporal reference (the referent is an entity of time, like *1995* or *Weimarer Republic*) or ‚document'-reference (the referent is a ‚document').

The resolution process

As I am interested in the mechanisms underlying the cognitive process of anaphora resolution, this classification needs further specification. Thus, I define an annotation schema which is influenced by Müller and Strube's (2001) and Poesio's (2004). They aim to develop data for training and evaluation systems which automatically resolve anaphoric relations (co-reference and bridging relations). In contrast, I annotate data in order to get a more fine-grained distinction for the variety of the *danach* occurrences (as different forms of anaphoric reference might involve different resolution processes) and to capture those cognitive aspects which influence the process of anaphora resolution. I assume that different sources of knowledge are important for the process of anaphora resolution and I want to document how syntactic, semantic, discourse-structural features and world knowledge interact in this process. Thus, syntactic and structural features like the distance between the anaphor and its antecedent or the position of the anaphor within the sentence (Vorfeld vs. rest) are annotated. Moreover, tense and modus of the antecedent- and anaphor-context-clause are recorded. With respect to semantic and conceptual issues the temporal distance between the antecedent and the anaphor-context situation is reflected. Furthermore, the anaphoric relation is judged as single (one antecedent for one anaphor), chain (for anaphoric chains), ambiguous (the anaphor has more than one plausible potential antecedent) or vague (no antecedent phrase could be determined (s. Eckert and Strube 2001)). The schema also reflects which of the knowledge sources is the main factor in the resolution process. So I distinguish between the following factors: syntactic (e.g. antecedent and anaphor are part of a coordination), temporal-info (the antecedent is temporally modified), focus-time (e.g.

the antecedent is the previous main clause), causal (the antecedent situation is causally connected to the anaphor-context situation) or world knowledge (the anaphor can only be resolved by world knowledge). Finally, I plan to integrate the aspects of the annotation schema into an anaphora resolution model following Schwarz-Friesel's (forthc.) account of text world models, Cristea et al.'s (2002) approach of an Anaphora-Resolution-engine and the resolution model proposed by Consten and Knees (forthc.) as some of the features account for the relation between the linguistic expressions and their representations in a discourse model.

References

Consten, M. and M. Knees (forthc.). Complex Anaphors in Discourse. In A. Benz and P. Kühnlein, eds., Constraints in Discourse.

Cristea, D., O.-D. Postolache, G.-E. Dima and C. Barbu (2002). AR-Engine – a framework for unrestricted co-reference resolution. Proc. of the LREC 2002 – Third International Conference on Language Resources and Evaluation, Las Palmas de Gran Canaria, Spain, 29-31 May 2002, vol. VI, pp. 2000-2007.

Eckert, M. and M. Strube (2001). Dialogue Acts, Synchronising Units and Anaphora Resolution. Journal of Semantics, **17(1)**, pp. 51 –89.

Fraurud, K. (1992). Situation Reference. What does ‚it' refer to?. In: Fraurud, K.: Processing Noun Phrases in Natural Discourse. PhD thesis. Depart. of Linguistics, Stockholm University.

Rüttenauer, M. (1978). Vorkommen und Verwendung der adverbialen Proformen im Deutschen. Buske, Hamburg.

Miltsakaki, E., C. Creswell, K. Forbes, A. Joshi and B. Webber (2003). Anaphoric Arguments of Discourse Connectives: Semantic Properties of Antecedents versus Non-antecedents. Proc. of the Computational Treatment of Anaphora Workshop, EACL 2003, Budapest.

Müller, C. and M. Strube (2001). Annotating Anaphoric and Bridging Relations with MMAX. Proc. of the 2nd SIGdial Workshop on Discourse and Dialogue. Aalborg, Denmark, September, pp. 90-95.

Poesio, M. (2004). The MATE/GNOME Scheme for Anaphoric Annotation, Revisited. Proc. of SIGDIAL, Boston, April.

Webber, B., A. Joshi, M. Stone, and A. Knott (2003). Anaphora and Discourse Structure. Computational Linguistics, **29(4)**, pp. 545-587.

Schwarz-Friesel (forthc.). Text comprehension as the interface between verbal structures and cognitive memory processes: the case of resolving direct and indirect anaphora. In C. Zelinsky-Wibbelt, eds., Memory and Language. Benjamins, Amsterdam.