Chart-based Incremental Semantics Construction with Anaphora Resolution Using λ-DRT

Ingrid Fischer, Bernd Geistert and Günther Görz
University of Erlangen-Nuremberg, IMMD II and VIII (Computer Science)
Am Weichselgarten 9, D-91058 ERLANGEN
Phone: (+49 9131) 85-9909; Fax: 85-9905
E-Mail: goerz@informatik.uni-erlangen.de

In our institute, a student project aiming at the construction of a prototypical natural language processing (NLP) system is being carried out. Its fundamental goal is the modular design and implementation of a simple but complete NLP system which is able to work incrementally and interactively. Due to its pedagogic character, particular attention is paid to a clear and transparent design of all modules in order to demonstrate the basic algorithms, even if this is not sufficient to achieve a high run time efficiency and a very broad linguistic coverage. The system, which is completely implemented in Scheme (SCM), is able to analyze short paragraphs of texts and to answer questions about it with a processing time of a few seconds per sentence. The text sort we chose are logical puzzles taken from the weekly published German newspaper "Die Zeit".

We present an approach for semantic construction and resolution based on an active chart-parser which analyzes syntax and semantics of natural language sentences incrementally. The parser consists of a syntactic component using a unification grammar formalism, an augmented variant of PATR-II, and a semantic component on the basis of Discourse Representation Theory (DRT) [Kamp, Reyle 1993]. In order to provide intersentential anaphora resolution, our incremental construction procedure builds discourse representations in which representations of consecutive sentences are embedded.

Central Design Goals. The design of our parsing system was governed by two main goals: On the one hand syntactic and semantic representations should be constructed *in parallel*, on the other hand input text should be analyzed *incrementally*, which is also extended to anaphora resolution. We chose a *co-descriptive approach* using a different formalism for each level.

Incrementality. A problem of any compositional semantics construction is that the order of construction steps may be disadvantageous for anaphora resolution, because, e.g., the construction for a supposed intrasentential antecedent would have been completed before the anaphoric NP is considered (c.f. [Johnson 1986]). Due to the incrementality of the parsing process it is guaranteed that information progresses from left to right.

λ-DRT. For the semantic formalism a combination of λ-calculus and DRT as developed by [Reyle 1986], [Asher 1993] and [Bos et al. 1994] was used. We adopted the last version, where DRSs are represented as feature structures. A DRS consists of a λ-list, the DRS and a special feature BDR whose value is a list of bindable discourse referents containing discourse referents of unsolved anaphora. Furthermore, we added features which are required for anaphora resolution: An ANCHOR list, an AGRM list, CONTEXT, and WEIGHT (see below). Functional composition of DRSs is realized by unifying parts of the functor and argument feature structures. But in contrast to Pinkal, where for functional composition only one element of the argument's λ-list is satisfied, we abandoned this restriction.

We redesigned the semantic formalism such that a parallel construction of syntactic and semantic structures can be achieved — structures on both levels are built concurrently as opposed to its sequential use in the VERBMOBIL project. There are two generic operators for semantic construction, identity (one-place) and compose (two-place) which are applied during the execution of the fundamental rule: Since in our case chart edges are annotated by categories, syntactic feature structures and λ-DRSs, the applicability of functional composition on the semantic level is an additional condition for the introduction of new edges.

Within this framework, different linguistic phenomena of German have been investigated: processing of proper names, tense (in the style of Reichenbach), negation, verbs as verb valency fillers, scope ambiguities with two newly introduced DRS-operators, plural, and prepositional phrases and other adjuncts.

Resolution of Anaphora. For inter- and intrasentential anaphora resolution, personal pronouns, reflexive pronouns and definite NPs are being considered, for which indefinite NPs and proper names are admissible antecedents. There are basically two means to achieve resolution: restrictions and preference weights (c.f. [Carbonell, Brown 1988]).

Many coreferences of anaphors and antecedents can be excluded by strict linguistic restrictions like gender-number agreement (using AGRM), binding principles, intersentential DRS accessibility constraints and semantic type constraints. In general, only antecedents within the local syntactic domain are candidates for a coreference of reflexive pronouns. For intersentential anaphors, DRS accessibility constraints are modeled by considering only the atomic conditions on the topmost DRS level of the DRSs of the past discourse as possible antecedents.

In establishing reference, anaphors are checked for simple semantic type consistency as soon as the semantic construction has been proceeded so far that it is obvious which valency slot of a verb will be filled by the NP under consideration.

Antecedents obeying those constraints are being scored by preferences, the calculation of which comprises the recency of anaphor and antecedent, syntactic parallelism in coordinated clauses, a simple focus mechanism, and topicalization preference.

References are resolved incrementally by considering inactive chart edges which correspond to anaphoric NPs. Inactive edges are suitable for insertion because they correspond to syntactically completely analyzed parts of the sentence. In the case of unresolved coreferences the respective discourse reference are inserted into the BDR list. If suitable antecedents can be found, for each of them a new inactive edge with a corresponding coreference entry, an empty BDR list, and a WEIGHT entry with the calculated scoring is built. In order to limit the growth of the number of edges, the next steps of the construction procedure are checking for violations (binding principles, semantic type consistency) which could not be recognized before.

Referential ambiguity is handled in the following way: The context entry of the structure from which the antecedent has been taken is inserted into the CONTEXT edge list of the intersentential anaphoric λ-DRSs. A new edge will not be generated if the functional composition of two λ-DRSs, which have been uniquely assigned as the discourse context, fails for the reason of incompatible CONTEXT entries.

Due to its incremental realization, the results of reference resolution are already available before the completion of semantic construction. By multiplying the WEIGHT entries, only the best scored results on the sentence level are used in the following steps.

References

[Asher 1993] Asher, N.: *Reference to Abstract Objects in Discourse.* Dordrecht: Kluwer Academic Press, 1993

[Bos et al. 1994] Bos, J. et al.: *A Compositional DRS-based Formalism for NLP Applications.* Verbmobil Report 59, Saarbrücken 1994

[Carbonell, Brown 1988] Carbonell, J., Brown, R. D.: *Anaphora Resolution: A Multi-Strategy Approach.* In: Proceedings of the 12th Conference on Computational Linguistics, 1988

[Fischer 1993] Fischer, I.: *Die kompositionelle Bildung von Diskursrepräsentationsstrukturen über einer Chart.* Masters thesis. University of Erlangen-Nuremberg, IMMD VIII, 1993

[Geistert 1995] Geistert, B.: *Auflösung anaphorischer Referenzen in der λ-DRT.* Bachelors thesis, University of Erlangen-Nuremberg, IMMD VIII, 1995

[Johnson 1986] Johnson, M., Klein, E.: *Discourse, Anaphora and Parsing.* Proceedings of the 11th Conference on Computational Linguistics, 1986

[Kamp, Reyle 1993] Kamp, H., Reyle, U.: *From Discourse to Logic.* Dordrecht: Kluwer Academic Press, 1993

[Reyle 1986] Reyle, U.: *Zeit und Aspekt bei der Verarbeitung natürlicher Sprache.* PhD Thesis, University of Stuttgart, Institute of Linguistics, 1986