

PRAGMATICS IN SPEECH UNDERSTANDING - REVISITED

Astrid Brietzmann
Lehrstuhl fuer Informatik 5 (Mustererkennung)
Univ. Erlangen-Nuernberg, Erlangen, FRG

Guenther Goerz
RRZE
Univ. Erlangen-Nuernberg, Erlangen, FRG
and
Neuropsychiatric Institute
UCLA, Los Angeles, Calif., USA

This paper reflects some thoughts on pragmatics in the context of a Speech Understanding System which is currently developed at the University Erlangen-Nuernberg. After a brief outline of the system's structure with an emphasis on the characteristics of the parser and the knowledge representation scheme we present some of the underlying theoretical considerations. The main part of the paper describes the design criteria for the SEMANTICS, PRAGMATICS, and DIALOG modules, and the structure of their interactions within a general discourse understanding framework, in particular the role of a user/task model.

1. The Erlangen Speech Understanding System

An experimental expert system for understanding continuous German speech is being developed at the Computer Science Department (Lehrstuhl Informatik 5, Mustererkennung), University Erlangen-Nuernberg [8]. Its main characteristics can be summarized as:

- blackboard-oriented architecture (see [5]),
- modularity through separate knowledge sources for ACOUSTICS-PHONETICS, LEXICON, SYNTAX, SEMANTICS, PRAGMATICS, DIALOG, RETRIEVAL, and STRATEGY,
- ease of reconfiguration through clearly specified interfaces, so that modules can easily be exchanged,
- parallelism (currently simulated),
- ability to conduct flexible, adaptive dialogs featuring mixed initiative, interpretation of indirect answers, resolution of anaphoric references, handling of fragments (ellipses) and application specific dialog schemata and strategies,
- experimental in order to gain data on its performance and on its linguistic and epistemological adequacy which in turn can be used to calibrate the knowledge sources, in particular the strategy involved.

In the following we concentrate on the higher level components, SEMANTICS, PRAGMATICS, and DIALOG, which in principle constitute

parallel, interacting processes. As prerequisites for that we outline briefly the parser and the general underlying knowledge representation scheme.

GLP: a parser. In our view syntactic knowledge plays an important role in natural language understanding. We agree with Bobrow and Webber [2] that there is a significant type of utterance description which is determined by syntactic features and categories, and, partially, also by ordering information. Elements of this description are used to guide semantic, pragmatic, and discourse level recovery processes, which in turn provide a feedback to syntactic analysis. Such processes include interpretation, anaphora resolution, focus tracking, and ellipsis resolution. Syntax gets a first cut in the logical structure of the utterance.

GLP [6] internally provides itself a multiprocessing scheme. It uses two central data structures, the Chart - an active Well-Formed Substring Table -, and the Agenda - a list of processes, which allows task centered scheduling. The whole parsing process is controlled by a monitor, which triggers a grammar rule interpreter. Its linguistic data base consists of a lexicon and a (functional) grammar (see 2). GLP's special features for speech analysis include direction-independent island parsing, the ability to deal with gaps in the input utterance and to handle quality scores for word and phrase hypotheses as well as incremental parsing by tying syntactic and semantic processing closely together. The selection of tasks is controlled by a Scheduler, which realizes a flexible strategy, so that bottom-up or top-down processing are not characteristic for the analysis as a whole, but only for parts of it.

The knowledge representation scheme. The underlying knowledge representation scheme, which will be used throughout the higher level modules of the system, can be characterized by the fundamental distinctions of schema (prototype), actualization (instance), and manifestation (situation-dependent embedding). Basically it is equivalent to an active semantic network with a clear separation of intension (general conceptual taxonomy) and extension (situation descriptions). Its elementary Units or Frames representing Concepts are supposed to cover mainly three aspects for their attributes :

- the role, which designates the attribute's function in the concept,
- restrictions on possible values for the attribute,
- modality, which indicates the importance of the attribute for the concept.

The system itself incorporates reasoning capabilities with an emphasis on property inheritance and default reasoning. Currently we are experimenting with two different approaches: FRL [11], which is already available, and a new system in the spirit of KL-ONE [3], which is currently being implemented.

The application domain. The first domain of discourse to which the speech understanding system will be applied, is travel planning within the West German Intercity train system. This particu-

lar application area was chosen as it can easily be expanded from rather simple question answering on time tables and train connections to more complex aspects of discourse, including planning and problem solving.

2. Some Theoretical Considerations

The general principle of our approach can be characterized as "pragmatics first", i.e. we see the task of natural language understanding from the viewpoint of communication as acting and interacting (see Kambartel [9]). This implies that the underlying grammar model ought to be a functional one [4, 10], i.e., that the recovery of the structure of a natural language utterance must be seen as part of a larger process of analyzing the meaning, intentions, and goals underlying its generation. In particular, we adopt Halliday's taxonomy of the functions of language:

- ideational, as related to the expression of content,
- interpersonal, as related to the purpose of an utterance,
- textual, as related to the coherence of language use.

The structure of the dictionary with regard to these aspects represents

- syntactic information: word classes, morphological information, valencies as structuring syntactic information in relation to functional attributes,
- semantic/pragmatic information: word meanings (based on a system of semantic primitives), case frames (with obligatory and optional attributes like agent, object, etc.), and restrictions (also to be used as expectations).

3. Textual Interpretation: SEMANTICS

Whereas the parser's facilities for mapping structural descriptions into functional attributes are limited to matching operations, interpretation requires reasoning. Based on purely linguistic knowledge, textual interpretation is the genuine task of the SEMANTICS module, which has to build general, situation-independent meaning structures. It provides content analysis by means of inferences using lexical semantic knowledge and applying case grammar rules as well as considering the cotext, i.e., the linguistic environment of the utterance.

We make use of valency properties of the head words, especially the main verb as an intermediary level between surface structure and the underlying case structure, thus following an extended notion of Tesnière's dependency theory [12]. Valency does not only determine a typical syntactic complement-structure for the governing words, e.g., calling for dependent noun groups and prepositional groups in certain surface cases; it also supplies criteria for proper treatment of prepositional phrases and modifier placement.

Besides the revelation of the underlying predicate-argument structure, SEMANTICS' main tasks are word-sense disambiguation

and, in addition, handling quantification and dealing with general spatial and temporal concepts on the level of words, i.e. without referring to factual knowledge. In detail, it has to enforce

- construction of dependency structures and their evaluation by checking their constituents for semantic compatibility,
- analysis of the type and the modality of the utterance,
- transformation of dependency structures into a canonical form, e.g., by completing the proposition in infinitive clauses, or converting passive sentences to active form,
- instantiation of case frames over valency structures by testing the selectional restrictions imposed on the case slots.

The parser's strategy is to be modified in such a way that semantic analysis at the constituent level can be started as soon as a local constituent is syntactically recognized. The results of this interpretation step are semantic hypotheses containing predictions. The parser then has to verify these islands syntactically, to expand them and to concatenate them with other islands.

4. Contextual Interpretation: PRAGMATICS and DIALOG

The PRAGMATICS and DIALOG modules provide the second step in interpreting an utterance. The task of the PRAGMATICS module is to specialize case structures into task specific association structures within the domain of discourse. These in turn are resolved and embedded into the dialog context by the DIALOG module.

As mentioned above, we view language understanding as understanding goal-directed action, in this case speech acts. People in general are capable of forming and executing plans to achieve goals and to infer plans of other agents by observation. Hence, the PRAGMATICS module has to analyze the speaker's intentions, in particular

- to establish points of correspondence between the speaker's and its own knowledge of the world,
- to draw inferences which the speaker intends the hearer to draw, and
- to match those with the particular domain of discourse.

This knowledge on objects, events and abstractions is represented in a group of schemata, which define the concepts of time, space, causality, goals and plans in their pragmatic dimension, i.e. in their relation to acting. In addition, a second group of schemata then provides the necessary domain specific knowledge, largely by specializing the general knowledge and augmenting it by particular knowledge about acting in the application domain. The PRAGMATICS module constructs a task model by starting with a description of the actual situation and an initial goal, which is refined during the following conversation by knowledge about

actions, in particular their (pre)conditions and effects. As the conversation goes on, it builds a plan in terms of a sequence of actions to transform the description of the situation into the desired goal state. There are standard techniques for constructing plans like backward chaining, but they do not provide a solution to a wide class of actions which can be described in natural language (like standing still, preventing something, executing simpler actions in parallel, etc.). To cover these phenomena, a temporal logic must be incorporated into the task model schema [1]. Defining actions by using knowledge about how they can be performed is not sufficient to define their meaning, in particular with regard to the tasks the PRAGMATICS module has to achieve:

- understanding the speaker's intention(s),
- reasoning about its understanding in order to act, in particular by specifying all (including implicit) information which is required to react appropriately (and smartly), and
- situation dependent resolution of references.

Considering what has been mentioned about our general approach on speech acts, PRAGMATICS has to interact closely with the DIALOG module, which incorporates knowledge about communication situations (linguistic-pragmatic context, immediate processing context, psychological context) and standard patterns of discourse (conventions for interactions, reasoning and establishing coherence), augmented by a second level of schemata which specify these with regard to the chosen domain of discourse. Using this knowledge, DIALOG has

- to draw inferences from the context, and
- to draw inferences on the current state of the speaker, including his knowledge,

in order to construct and maintain a user model. This model, starting with a rough idea of standard discourse schemata and techniques tries to understand and to guide the speaker by successive refinements through building discourse plans to achieve a satisfactory conclusion of the dialog. On the other hand, these plans are supposed to influence the overall behavior of the whole system in a larger range of interaction steps, e.g. with respect to its adaptivity and flexibility.

The very similar layout and the proposed close interaction between PRAGMATICS and DIALOG were influenced by results on task-oriented dialogs [7], which state a parallelism between the dialog and the structure of a problem solution. This in turn should allow the resolution of most of the references and a contextual restriction within certain logically and methodologically characterized subdialogs (see the detailed discussion in Webber [13]).

Indeed, the main difference between both components is in the kind of knowledge they represent and use, not in their methods of reasoning. The main contribution of the PRAGMATICS module to the whole understanding process can be paraphrased as a specialization of the general "referential potentiality" (lexical meaning)

of utterances into a particular thematic framework whereas the DIALOG module provides a specialization with regard to a discourse framework, i.e. to knowledge how to conduct a successful dialog.

References

- [1] Allen, J.F., What's Necessary to Hide?: Modeling Action Verbs, in: Proceedings of the 19th Annual Conference of the Association for Computational Linguistics (Stanford, 1981).
- [2] Bobrow, R.J. and Webber, B.L., Some Issues in Parsing and Natural Language Understanding, in: Proceedings of the 19th Annual Conference of the Association for Computational Linguistics (Stanford, 1981).
- [3] Brachman, R., On the Epistemological Status of Semantic Networks, in: Findler, N.V. (ed.), Associative Networks. Representation and Use of Knowledge by Computers (Academic Press, New York, 1979).
- [4] Dik, S., Functional Grammar (North-Holland, Amsterdam 1978).
- [5] Erman, L., The HEARSAY-II Speech-Understanding System: Integrating Knowledge to Resolve Uncertainty, Computing Surveys (12) 1980, 213-253.
- [6] Goerz, G., GLP: A General Linguistic Processor, in: Proceedings of the Seventh International Joint Conference on Artificial Intelligence (Vancouver, 1981).
- [7] Grosz, B., The Structure of Task Oriented Dialogs, in: Proceedings of the IEEE Symposium on Speech Recognition (Pittsburgh, 1977).
- [8] Hein, H.-W., Niemann, H., Expert Knowledge for Automatic Understanding of Continuous Speech, in: Kunt, M. and de Coulon, F. (eds.), Signal Processing: Theories and Applications (North-Holland, Amsterdam, 1980).
- [9] Kambartel, F., Pragmatische Grundlagen der Semantik, in: Gethmann, C.F. (ed.), Theorie des wissenschaftlichen Argumentierens (Suhrkamp Theorie, Frankfurt, 1980).
- [10] Kay, M., Functional Grammar, in: Proceedings of the Fifth Annual Meeting of the Berkeley Linguistics Society (Berkeley, 1979).
- [11] Roberts, R.B., Goldstein, I.P., The FRL Manual, AI Memo 431, MIT (June 1977).
- [12] Tesniere, L., Elements de syntaxe structurale (Klincksieck, Paris, 1965).
- [13] Webber, B.L., Description Formation and Discourse Model Synthesis, in: Waltz, D.L. (ed.), Theoretical Issues in Natural Language Processing-2 (Univ. of Illinois at Urbana-Champaign, 1978).