

Generating Questions with Deep Reversible Grammars

Ivan A. Sag and Dan Flickinger

CSLI, Stanford University

Stanford, California 94305

sag@stanford.edu, danf@stanford.edu

One desirable approach to question generation exploits existing general-purpose grammars which are both linguistically rich and computationally effective. A generator using such a grammar can ensure high-quality output provided that the input semantic representation is reasonably coherent, and the characterization of a well-formed semantic input can be made quite precise. The Linguistic Grammars On-line (LinGO: lingo.stanford.edu) Project, founded at Stanford's Center for the Study of Language and Information in 1994, has been developing the English Resource Grammar (ERG: Flickinger 2000) for generation (as well as parsing) within the context of a series of application-oriented projects, including German-English machine translation in *Verbmobil*; an NSF-funded project to develop a speech prosthesis for physically disabled users; an automated email response product developed with the Silicon Valley start-up YY Technologies; and the Norwegian-English machine translation project called LOGON. We have developed the widely used flat semantic representation called Minimal Recursion Semantics (MRS: Copestake, Flickinger, Pollard and Sag 2005), which is the input accepted by the ERG for generation, and we use the relatively efficient LKB (Linguistic Knowledge Builder: Copestake 2002) generator to produce realizations corresponding to such an input. All of the grammar and processing resources we use are distributed as open-source software through DELPHIN (the Deep Linguistic Processing with HPSG Initiative: www.delph-in.net), a growing international community of linguists and NLP developers.

The basic framework used for the develop-

ment of LinGO/DELPHIN grammars is Head-Driven Phrase Structure Grammar (HPSG; see Pollard and Sag 1994 and related references at hpsg.stanford.edu). While the framework is lexicalist, encoding most syntactic and semantic constraints for a language within its lexicon via a rich hierarchy of types, these implemented grammars are 'construction-based', also employing a hierarchy of types for the relatively small number of syntactic rules used to compose phrases and sentences.

The current ERG includes a basic treatment of interrogative constructions, but a pilot project undertaken in 2003 demonstrated the feasibility of implementing the broader range of interrogative constructions treated in detail by Ginzburg and Sag 2000. We are currently in the process of integrating this extended range of interrogative constructions (sketched in the Appendix) into the ERG. Because the resulting grammar will be purely declarative, equally suited for parsing and generation, integrated into the international DELPHIN community, integrated into ongoing computational work on dialogue by Jonathan Ginzburg (King's College London), Robin Cooper (Gothenburg U.) and their research groups, as well as being freely available as open-source code, we believe it can play an important role in any question-generation effort of the kind being discussed at this meeting.

In addition to the design work on a comprehensive theory of questions, we have focused on implementing robust and relatively efficient generation using the ERG, building on the approach described in Carroll et al. 1999, and benefiting along the way from

steady and substantial improvements in grammar coverage in the intervening years. More recently, we have also incorporated a statistical model for selecting the most probable realization from among those licensed by the grammar, given a semantic input to the generator. This ranking model is trained on a manually constructed treebank, using the maximum-entropy approach presented in Velldal 2007, with his encouraging experimental results based on annotated data from the LOGON development corpus we used for the Norwegian-English machine translation demonstrator. From our experience to date, we view this challenge of realization ranking, the dual of disambiguation in parsing, as central to practical generation systems built on linguistically informed grammars, and we see hybrid systems of the kind used in LOGON as strong candidates for addressing this challenge.

Our experience in implementing grammar-based generation has also led us to view the use of test suites and corpus-based grammar profiling as crucial to the development of effective implementation. The approach we use and advocate, building on Oepen and Flickinger 1998, involves the construction of two kinds of linguistic data sets, one which exhibits each of the linguistic phenomena under study in carefully designed example sentences, and one which is a representative sample of the data found in the target domain(s). For each of these data sets, we have found it very useful to send all of the items through the processing engine, and manually annotate the intended output, then keep these annotations up-to-date as the grammar is further developed. This approach ensures an accurate view of the quality of the analyses/realizations licensed by the grammar, and provides a rich source of information on where the grammar or the generator could benefit from further improvement.

References

- Carroll, J., A. Copestake, D. Flickinger and V. Poznanski 1999. "An efficient chart generator for (semi-)lexicalist grammars," in *Proceedings of the 7th European Workshop on Natural Language Generation (EWNLG'99)*, qToulouse. 86-95.
- Copestake, A. 2002. *Implementing Typed Feature Structure Grammars*. CSLI: Stanford.
- Copestake, A., D. Flickinger, I. A. Sag, and C. Pollard 2005. "Minimal Recursion Semantics. An Introduction," *Journal of Research on Language and Computation* 3:4. Kluwer: Dordrecht.
- Flickinger, D. 2000. "On Building a More Efficient Grammar by Exploiting Types", *Natural Language Engineering* 6:1.
- Ginzburg, J. and I. A. Sag 2000. *Interrogative Investigations: The form, meaning and use of English interrogatives*. CSLI: Stanford.
- Oepen, S. and D. Flickinger 1998. "Towards Systematic Grammar Profiling: Test Suite Technology Ten Years After", in R. Gaizauskas, ed., *Journal of Computer Speech and Language*, special issue on Evaluation in Speech and Language Technology, 12:411-435.
- Pollard, C. and I. A. Sag 1994. *Head-Driven Phrase Structure Grammar*. University of Chicago Press and CSLI, Stanford.
- Velldal, E. 2007. Empirical Realization Ranking. Ph.D. thesis, Department of Informatics, University of Oslo.

Appendix

Polar Interrogatives:

Did the museum close? The museum closed? It was unclear [whether/if the museum closed].

Wh-Interrogatives:

What/Which museum closed? Which exhibit/What did the museum close?

Which painters' masterpieces did the museum show?

It was unclear [which (painter's) exhibit the museum closed].

Multiple Wh-Interrogatives:

(It was unclear) which museum closed which exhibit.

Which museum inquired about which exhibits would be shown when?

'Functional' Readings:

A: What did each museum display most prominently?

B: Its oldest paintings.

Elliptical Interrogatives:

A: Someone removed this picture from the museum.

B: Who?/I wonder who.

Did the museum show paintings or drawings?

Reprise Interrogatives:

[The government said they would close some museum.]

The government said they would close WHICH museum?

[The government said they would close <inaudible>.]

The government said they would close WHAT?

[Did the government said they would close it?]

Did the government say they would close WHAT?

Short Answers:

Yes, No, Fragments as answers to Wh-interrogatives.

Selectional Constraints:

They wondered whether the Museum would open.

They wondered that the Museum would open.

They believed that the Museum would open.

They believed whether the Museum would open.

They knew whether/that the Museum would open.

'Concealed' Questions:

*They discovered/*wondered the museum director's height.*