

Grammar Writing in LFG Introduction

Victoria Rosén and Annie Zaenen

University of Bergen and Xerox Research Centre Europe

Proceedings of the LFG99 Conference

The University of Manchester

Miriam Butt and Tracy Holloway King (Editors)

1999

CSLI Publications

<http://www-csli.stanford.edu/publications/>

Although theories of generative grammar in principle have made it possible for decades, grammar writing is not a common activity among generative linguists. As a theoretical syntactician one may concentrate on one or a few grammatical phenomena, ignoring all aspects of the grammar of a language that are not directly relevant to one's special interests. If, however, one tries to write even a small grammar of a language, one is immediately confronted with all kinds of decisions about how various linguistic issues should be dealt with.

Grammar writing is interesting from at least three quite distinct perspectives: it may have a purely theoretical motivation, it may be motivated by the need for a grammar as part of an application, or it may be done to establish the basis of linguistic engineering.

From a theoretical point of view, grammar writing should be an important way of testing linguistic hypotheses. Once an analysis involves more than two or three rules, it can become quite difficult to mentally or manually compute whether or not the rules actually produce the analysis argued for. Writing even a small grammar with the help of a grammar writer's workbench (for example, the [LFG Grammar Writer's Workbench](#) or [XLE](#)) can assist the linguist in testing hypotheses and developing more satisfactory analyses. This type of grammar writing can also help to determine whether analyses of various phenomena are compatible, in other words, whether it is actually possible to implement them within the same grammar. It is regrettable that so few linguistics departments understand the need to train their students in using this type of tool.

Grammar writing from an application perspective is becoming increasingly important. The demand for grammars with realistic coverage of authentic texts is making itself felt in many different areas, among them machine translation, information extraction, and multilingual authoring. LFG, being a declarative, unification-based grammar formalism, is well suited for applications. The importance of theory-based approaches in industrial applications is often contested, but all non-theory-based applications eventually run into a wall: no further development is possible because the whole system has become totally untransparent. It is not clear that current theory-based applications will be able to avoid this pitfall, but if not, it is not because of too much but rather because of too little theory: whereas there are well developed parts, for example core syntax, there are many necessary building blocks which are less well understood and formalized, for example pragmatics and discourse structure.

A third use of grammar writing might turn out to be the most important for some time to come. Linguistics is presently recognized as a scientific endeavor and the need for applications is clearly perceived, but what is less clear is that between theory as it exists now and applications there is a need for engineering and for reflection on the characteristics of good engineering. This may not be a science, but it is an academic subject, or at least it should be: in the same way that the engineering of tangible artifacts as an academic discipline develops and transmits standards and codifies and tests practice, document engineering should have its standards and codified practice. A good example, but not the only one, of where this academic engineering might be useful at this point in time and where it will be dependent on large scale grammar writing is in translation. There is a lot of anecdotal literature on the problems of translation, but there is no clear understanding of the importance of the various problems and of their interaction. For example, ambiguity is certainly a big problem, but how much of it can be alleviated by carrying over ambiguities from language to language, by interactive disambiguation, by comparing translations for more than one language pair? These are all interesting engineering questions, the importance of which cannot be judged without substantial grammar development. In the course of solving these engineering questions, new more run-of-the-mill theoretical questions will be raised because these pursuits will help highlight where theoretical understanding is lacking. Unfortunately, there is at this point in time very little action on the front of academic document engineering: very few universities in Europe or in the United States have a coherent approach to linguistic engineering. Stuttgart and Saarbruecken are exceptions in Germany. Clermont-Ferrand in France is trying to elaborate a program focused on controlled language authoring. Stanford has some projects that go in that direction but no consistent teaching activities.

The LFG ParGram project ([ParGram](#)) intends to provide a basis for these various approaches. The ParGram project originally involved three languages, English, French, and German, and researchers from Xerox PARC in Palo Alto, the Xerox Research Centre Europe in Grenoble, and the University of Stuttgart.

The goal was to write large-scale grammars for these languages based on common linguistic principles and a common set of grammatical features. In the summer of 1999, the University of Bergen joined the project, and new grammars will be written within the [NorGram](#) project for the two standard varieties of written Norwegian, Bokmål and Nynorsk. Various aspects of the ParGram project will be discussed in the next presentations.

The workshop "Grammar Writing in LFG" at LFG99 in Manchester was organized by Victoria Rosén. At the workshop the following presentations were made:

- Annie Zaenen, "Industrial Grammar Writing: Between Theory and Practice"
- Miriam Butt, Stefanie Dipper, Anette Frank, and Tracy Holloway King, "Writing Large-scale Parallel Grammars for English, French, and German"
- Helge Dyvik, "The Universality of f-structure: Discovery or Stipulation? The Case of Modals"

For these proceedings, the organizer Victoria Rosén and the first speaker Annie Zaenen found it most practical to coauthor a single introduction.

The main body of the workshop was a collaborative presentation on the ParGram grammars by members of the project. Tracy Holloway King gave a general introduction to the ParGram project in which she explained what parallel analysis involves and outlined the coverage of the grammars and the modularity of the XLE system. Stefanie Dipper's presentation focused on the grammar writing process itself, with the treatment of German compound nouns as an example. She discussed the difficult balance between desiderata such as broad coverage, linguistically motivated analyses, and efficient parsing. The third part of the workshop involved two examples of theoretical implications of grammar writing. First Tracy Holloway King presented the proposal for a separate projection for m(orphosyntactic)-structure, and showed how this proposal was motivated by work on German auxiliaries. The second example, presented by Miriam Butt, involved underspecification of grammatical features. She argued that underspecification can help to avoid unwanted ambiguities and overgeneration. Anette Frank completed this section of the workshop with a demonstration of machine translation from English to French using the Xerox Translation Environment (XTE), a system that is an extension of XLE.

Finally, the paper by Helge Dyvik demonstrated how the introduction of a new language in a multilingual project can present new perspectives on universality. Dyvik took up the way in which auxiliaries and modals have been analyzed in the ParGram project and showed why a similar analysis would not be well motivated for Norwegian. This in turn led to deeper foundational questions

about the theoretical status of f-structure as a universal level of syntactic representation.