Towards an LFG Syntax-Semantics Interface for Frame Semantics Annotation

Anette Frank[1] and Katrin Erk[2]

[1] DFKI GmbH, Language Technology Lab, Stuhlsatzenhausweg 3,
66123 Saarbrücken, Germany,
Anette.Frank@dfki.de,
WWW home page: http://www.dfki.de/~frank
[2] Universität des Saarlandes, Computational Linguistics Department,
66123 Saarbrücken, Germany
erk@coli.uni-sb.de,
WWW home page: http://www.coli.uni-sb.de/~erk

Abstract. We present an LFG syntax-semantics interface for the semi-automatic annotation of frame semantic roles for German in the SALSA project. The architecture is intended to support a bootstrapping cycle for the acquisition of stochastic models for frame semantic role assignment, starting from manual annotations on the basis of the syntactically annotated TIGER treebank, with smooth transition to automatic syntactic analysis and (semi-)automatic semantic annotation of a much larger corpus, on top of a free-running LFG grammar of German.
Our study investigates the applicability of the LFG formalism for modeling frame semantic role annotation, and designs a flexible and extensible syntax-semantics architecture that supports the induction of stochastic models for automatic frame assignment. We propose a method familiar from example-based Machine Translation to translate between the TIGER and LFG annotation formats, thus enabling the transition from treebank annotation to large-scale corpus processing.

1 Introduction

This paper is a first study of an LFG syntax-semantics interface for frame semantic role assignment. The architecture is intended to support semi-automatic semantic annotation for German in SALSA – the Saarbrücken Semantics Annotation and Analysis project[3] – which is based on Frame Semantics and is conducted in cooperation with the Berkeley FrameNet project [1, 15].

The aim of SALSA is to create a large lexical semantics resource for German based on Frame Semantics, and to develop methods for automated assignment of corpora with frame semantic representations.

In the first (and current) phase of the SALSA project, semantic annotation is fully manual, and takes as its base the syntactically annotated TIGER treebank

[3] See [8] and the SALSA project homepage http://www.coli.uni-sb.de/lexicon