

# Unification-based Processing Underway to Dot Com

**Dan Flickinger**

CSLI, Stanford University    YY Software Corporation, 110 Pioneer Way  
Stanford, CA 94305, USA    Mountain View, CA 94041, USA  
danf@csli.stanford.edu

**Stephan Oepen**

Computational Linguistics, Saarland University  
Postfach 15 11 50, 66041 Saarbrücken, Germany  
oe@coli.uni-sb.de

In this tutorial we will review the state of the art in the development and application of broad-coverage declarative grammars built on sound linguistic foundations (the ‘deep’ processing paradigm) and present several aspects of an international research effort—a consortium involving Saarbruecken (Germany), Stanford (USA) and the University of Tokyo (Japan)—to produce comprehensive, re-usable grammars and efficient technology for parsing and generating with such grammars. While statistical methods, often described as ‘shallow’ processing techniques, can bring real advantages in robustness and efficiency, they do not provide the precise, reliable representations of meaning that more conventional symbolic grammars can supply for natural language. We will illustrate the benefits and viability of the declarative approach both in multilingual grammar development (for English, German, and Japanese), and in commercially relevant applications including machine translation, speech prosthesis, and automated email response. The topics we will discuss and demonstrate will include:

- Descriptive formalism: balancing linguistic adequacy and processing requirements
- Linguistic resources: production and maintenance of large-scale grammars and lexicons
- Development tools: specialized environments and techniques for grammar engi-

neering, inspection and debugging

- Diagnostics and measurement: quality control and regression testing as a focal point in the development cycle
- Processing efficiency: recent advances in unification, parsing and ambiguity packing with large typed feature structures
- Semantic engineering: defining the appropriate level of semantic abstraction and underspecification
- Re-usability and exchange: convergence on a joint set of basic assumptions enables the synthesis of knowledge and software
- Practical applications: unification-based grammars and parsers as core components in commercial products