Formal Syntax and Grammar Engineering

Stephan Oepen

Universitetet i Oslo & CSLI Stanford

oe@csli.stanford.edu

Lilja Øvrelid

Göteborgs Universitet

lilja.ovrelid@svenska.gu.se

http://www.delph-in.net/courses/04/fs/

Morpho-Syntactic Categories (1 of 2)

Number — Person — Case — Gender

That dog barks. — Those dogs bark. I bark. — You bark. — They bark. — Sam shaved himself. We bark. — You bark. — Those dogs bark. I saw her. — She saw me. — My dog barked.

How many distinct verb forms according to number and person?

Tense — Aspect — Mood

The dog barks. — The dog barked — The dog will bark. The dog has barked. — The dog is barking. If I were a carpenter, ...



- GOTHENBURG — 3-NOV-04 (oe@csli.stanford.edu)

Formal Syntax and Grammar Engineering (18)

Morpho-Syntactic Categories (2 of 2)

Parts of Speech (PoS)

| cat, dog, neighbours, | noun (N) |
|------------------------------|-----------------|
| barks, chased, was, | verb (V) |
| fierce, angry, black, young, | adjective (Adj) |
| quickly, probably, not, | adverb (Adv) |
| a, the, my, that, | determiner (D) |
| of, by, on, at, under, | preposition (P) |
| she, mine, those, what, | pronoun (Pro) |
| and, neither nor, because, | conjunction (C) |

- **Paradigm** set of word forms ('units'), e.g. {*bark*, *barks*, *barked* };
- **Unit Categories** dimensions structuring a paradigm *internally*;
- Paradigm Categories properties *common* to all paradigm units.



GOTHENBURG — 3-NOV-04 (oe@csli.stanford.edu)

Grammatical Functions

Licensing — Government — Agreement

The dog barks. — *The dog a cat barks — *The dog barks a cat. Kim depends on Sandy — *Kim depends in Sandy The class meets on Thursday in E230 at 9:00.

- Head licenses additional constituents and can govern their form;
- **Specifier** precedes head, singleton, nominative case, agreement;
- **Complement** post-head, licensed and governed, order constraints;
- Adjunct 'free' modifier, optional, may iterate, designated position;
- **Government** *paradigm* category of c_1 determines the form of c_2 ;
- Agreement bi-directional: concord of *unit* categories of c_1 and c_1 .



Our Initial Type Hierarchy

- Linguistic and non-linguistic objects organized below 'top' type (*top*);
- fundamental division of constituents: word (lexicon) vs. phrase (rules);
- root is the 'start symbol' parts of speech as separate sub-hierarchy.



Structured Categories in a Unification Grammar

- All (constituent) categories in the grammar are typed feature structures;
- SPR and COMPS lists to encode licensing, government, and agreement;
- specific TFS configurations may correspond to 'traditional' categories;
- labels like 'S' or 'NP' are mere abbreviations, not elements of the theory.



The Linguistic Knowledge Builder (LKB)

Compiler and Interactive Debugger

- Grammar definition errors identified at load time by position in file;
- inheritance and appropriateness tracked by type and attributes;
- batch check, expansion, and indexing of full lexicon on demand;
- efficient parser and generator to map between strings and meaning;
- visualization of main data types; interactive stepping and unification.
- Main developers: Copestake (original), Carroll, Malouf, and Oepen;
- implementation: Allegro CL, Macintosh CL, (LispWorks, CMU CL);
- available in open-source and binary form for common platforms.



GOTHENBURG — 3-NOV-04 (oe@csli.stanford.edu)

The Format of Grammar Rules in the LKB



