DELPH-IN Summit 2006

— Formalism Extensions Discussion —

Stephan Oepen

Universitetet i Oslo and CSLI Stanford

`oe@csli.stanford.edu`

Fefor – June 16, 2006

The DELPH-IN Reference Formalism

Foundations

- Strongly typed, conjunctive, closed world typed feature structure logic;

- blend of [Carpenter, 1992], [Copestake, 1992], and [Krieger, 1995];

- fixed-arity, fixed-order phrase structure rules $\rightarrow$ parser and generator;

- no FS disjunction, macros, implications, relational constraints, et al.

Processing Environments

- LKB: grammar development environment (Lisp) [Copestake, 2002];

- PET: efficient, industry quality runtime engine ($C^{[++]}$) [Callmeier, 2000];

- LiLFeS (Tokyo), ACE (Woodley Packard), LIGHT AM (Liviu Ciortuz), ...

Where To Go?

- Originally rooted in LinGO experience $\rightarrow$ Need for formalism evolution?

A Shopping List of Candidate Devices (1 of 2)

Default Unification

- Available in LKB (Lascarides & Copestake, 1999); only used in KRG.

Relational Constraints

- Difference list manipulation tricky: maybe provide `cons()` & `append()`;

- what about additional list operations: `member()`, `remove()`, `shuffle()`?

- And then what about other data types (e.g. sets) and general relations?

Linearization or Discontinuous Parsing

- Daniels (2005) extends PET for linearization-based HPSG (*GIDLP*);

- LKB generator operates ($\sim$ efficiently) without adjacency conditions.

A Shopping List of Candidate Devices (2 of 2)

'Type Resolution' (Feature Co-Occurrence)

```
a := top & [ FOO bool, BAR bool ].
b := a & [ FOO -, BAR + ].
c := a & [ FOO +, BAR - ].
```

- Should feature structures $\begin{bmatrix} \text{FOO} + \end{bmatrix}$ and $\begin{bmatrix} \text{BAR} + \end{bmatrix}$ be allowed to unify?

- Likewise, what should be the effect of unifying $_a\begin{bmatrix} \ \end{bmatrix}$ and $\begin{bmatrix} \text{FOO} - \end{bmatrix}$?

Inter-Constituent Dependencies

- PET implements 'chart dependencies', a 'lexical' filter, e.g. *take* + *back*;

- LKB idiom co-occurence constraints could be enforced prior to parsing.

One Item in the LOGON Shopping Basket

Motivation

- In generation, enumerating all paraphrases not necessarily desirable:

> Abrams, Browne, and Chiang arrived on the second Thursday.
> Abrams, Browne and Chiang arrived on the 2nd Thurs.
> Abrams Browne and Chiang arrived on the 2nd Thursday.
>
> $\cdots$

$\rightarrow$ notion of (genre-specific) 'canonical' surface forms and punctuation.

Realization

- Introduce additional notion of 'per-constituent' roots (or maybe filters);

- only allow edges into the generator chart that satisfy, say, $\left[\text{CAN} +\right]$.