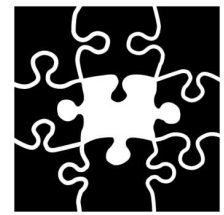
Revisiting competence & performance

Andreas van Cranenburgh



Huygens ING
Royal Netherlands Academy of Arts and Sciences

Institute for Logic, Language and Computation
University of Amsterdam

June 30, 2015

25 years of DOP, Amsterdam, 2015

Introduction

Chomsky (1965):

Competence system of rules describing idealized knowledge of language

Performance language behavior affected by ambiguity, errors, reaction times, frequency effects

Introduction

Chomsky (1965):

Competence system of rules describing idealized knowledge of language

Performance language behavior affected by ambiguity, errors, reaction times, frequency effects

Scha (1990):

- Difficult to write descriptively adequate grammar by hand.
- Problem of ambiguity;
  need to know relative plausibility of analyses.

Ergo, we need "performance-models of language (...), "which take into account statistical properties of actual language use."

Traditional parsing approach

1. Pick a grammar with the right linguistic & computational properties (competence)
2. Add a probabilistic disambiguation component (performance)
3. Apply pruning if necessary (performance)
4. Evaluate quality of model (performance)

Formal language theory

Definition

A *formal grammar* characterizes a language as a set of sentences and their structures.

Chomsky hierarchy:

Type 0: Unrestricted: Model-Theoretic Syntax, e.g., HPSG

Type 1: Context-Sensitive: Mildly Context-Sensitive, e.g., TAG, CCG, LCFRS

Type 2: Context-Free: PCFG, proj. dependency grammar

Type 3: Regular: finite-state technology

Grammar transformations

Capabilities of grammar formalisms can be extended, e.g.:

- Encode information in labels
- Apply pre- and postprocessing
- Intersect multiple grammars

Grammar transformations

Capabilities of grammar formalisms can be extended, e.g.:

- Encode information in labels
- Apply pre- and postprocessing
- Intersect multiple grammars

Examples:

- TSG or TIG $\Rightarrow$ CFG + backtransform table
- Dependency grammar $\Rightarrow$ PCFG
- Discontinuous constituents $\Rightarrow$ non-projective dependencies
- ...

Psycholinguistic Evidence

Do humans exploit hierarchical structure during processing?

- No Frank & Bod (Psy. Sci. 2011): Insensitivity of the human sentence-processing system to hierarchical structure
- Yes van Schijndel & Schuler (NAACL 2015): Hierarchic syntax improves reading time prediction

Psycholinguistic Evidence

Do humans exploit hierarchical structure during processing?

- No: Frank & Bod (Psy. Sci. 2011): Insensitivity of the human sentence-processing system to hierarchical structure
- Yes: van Schijndel & Schuler (NAACL 2015): Hierarchic syntax improves reading time prediction

Center-embedding:

- Hard for humans, natural for CFG
- Karlsson (2007): only occurs up to depth 3 in written language, depth 2 in spoken lang.

Psycholinguistic Evidence

Do humans exploit hierarchical structure during processing?

No Frank & Bod (Psy. Sci. 2011): Insensitivity of the human sentence-processing system to hierarchical structure

Yes van Schijndel & Schuler (NAACL 2015): Hierarchic syntax improves reading time prediction

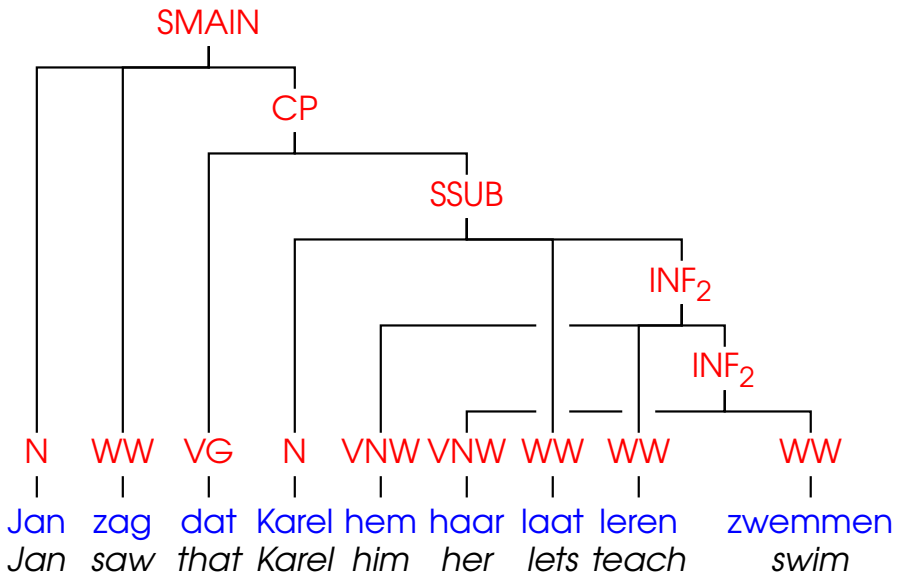
Center-embedding:

- Hard for humans, natural for CFG
- Karlsson (2007): only occurs up to depth 3 in written language, depth 2 in spoken lang.

Cross-serial dependencies:

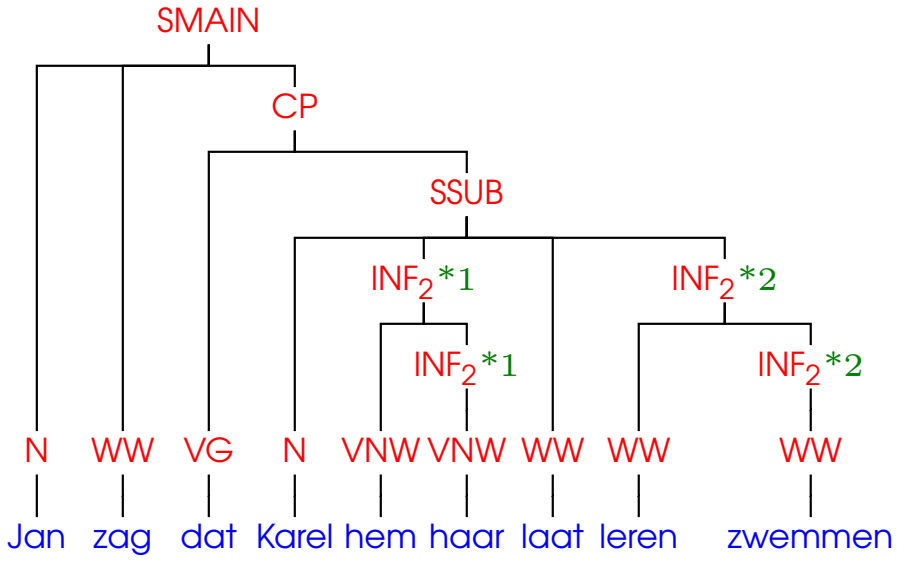
- Cross-serial dependencies not possible with CFG, but easier for humans than center-embedding: Bach et al. (1986) Cross and nested dependencies in German and Dutch: A psycholinguistic study.

Long-Distance Dependencies



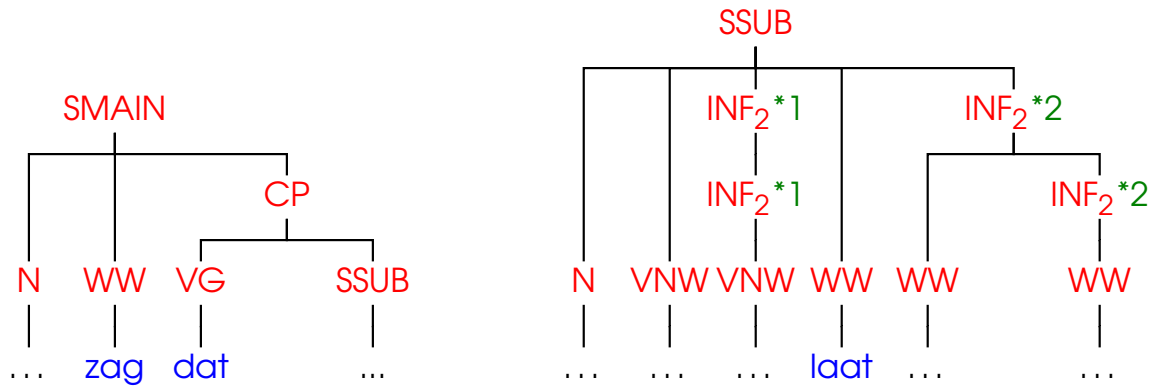
- Cross-serial dependencies are beyond context-free
- Can be captured by mildly context-sensitive grammars

CFG approximation



- Alternatively, long-distance dependencies can be encoded in the labels

DOP fragments



- With DOP tree fragments,
  complex linguistic phenomena can be captured
  statistically instead of formally

Conclusion

- Performance phenomena play an important role in computational models of language
- Instead of searching for the right formal grammar, consider how system as a whole copes with
  - ambiguity
  - cognitive limitations
  - linguistic complexity

References

- Noam Chomsky (1965). Aspects of the Theory of Syntax, MIT press.
- Remko Scha (1990). Language theory and language technology; competence and performance, in Q.A.M. de Kort and G.L.J. Leerdam, editors, *Computertoepassingen in de Neerlandistiek*, pp. 7–22. English translation: `http://iaaa.nl/rs/LeerdamE.html`