

Mathematical properties of natural languages and mathematical properties of linguistic theories

Anders Søgaard

Center for Language Technology
University of Copenhagen
Njalsgade 80
DK-2300 Copenhagen S
Email: anders@cst.dk

February 8 2007

Introduction

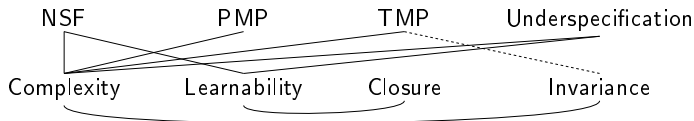
The class of possible natural languages is much smaller than one would expect from our present models of syntax. [Gol67]

Considered here:

- Nonsuperfiniteness,
- polysize model property,
- tree model property, and
- underspecification.

Not considered here:

- (Efficient) recognizability (linear, P, NP), learnability (Gold, PAC, ...), closure properties, and invariance properties (disjoint union, generated substructures, bisimulation).



Nonsuperfiniteness

Definition

A linguistic theory is nonsuperfinite if it does not generate all finite languages.

Theorem

No superfinite class of languages is learnable in the limit from positive data [Gol67].

	NSF	PMP	TMP	Underspecification
Minimalism [Sta97]	-	?	?	?
HPSG	-	?	?	?

Motivation

Learnability and linguistic intuition.

Polysize model property

Definition

A linguistic theory has the polysize model property if any derivation structure for G, ω is at most in $\mathcal{O}(|\omega|^k)$.

- HPSG: $type_1 \rightarrow [FEAT_1 type_1]$

$n = ?$	Space
2	4
12	4096

	NSF	PMP	TMP	Underspecification
Minimalism	-	✓	?	?
HPSG	-	-	?	?

Motivation

Recognizability and cognitive neuropsychology.

Tree model property

Definition

A linguistic theory has the tree model property if any derivation structure for G, ω is equivalent to a (possibly **unordered**) tree structure.

- HPSG: $\left[\begin{array}{l} \text{FEAT}_1 \ \boxed{1} \\ \text{FEAT}_2 \ \boxed{1} \end{array} \right]$

	NSF	PMP	TMP	Underspecification
Minimalism	-	✓	✓	?
HPSG	-	-	-	?

Motivation

Subsumption (recognizability and learnability) and conservativity.

Underspecification (I)

Lexical ambiguity in linguistic theories:

- Minimalism: v /snore/ & n /snore/
- HPSG: $\left[\text{PHON} \langle \text{snore} \rangle \right] \rightarrow \left[\text{CAT } n \vee v \right]$

Motivation

Learnability [Kan98], recognizability [SH07] and cognitive neuropsychology.

Evidence from cognitive neuropsychology for underspecification of

- morphological features [PJE04],
- syntactic category [Far01], and
- argument structure [PF01].

Underspecification (2)

Syntactic ambiguity in linguistic theories:

- Minimalism: $=d + \text{case} = d \vee \text{speaks}$ and ϵ
- HPSG: Application of multiple ID schemata and non-branching extensions.

Motivation

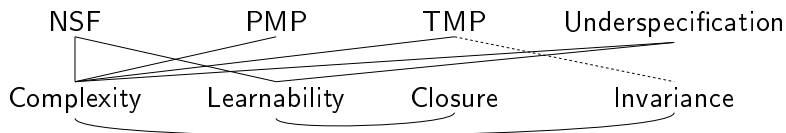
Learnability [Kan98], recognizability [SH07] and cognitive neuropsychology (10^{10} parses for $n = 10$).

Underspecification techniques for

- attachment,
- linking, and
- quantification (see [SH07] and ref. therein).

	NSF	PMP	TMP	Underspecification
Minimalism	-	✓	✓	-
HPSG	-	-	-	✓

Conclusion (1)



Derived $+/-$'s:

	Recognizable	Learnable	Copying	P Verification
Minimalism	+	-	-	+
HPSG	-	-	+	-

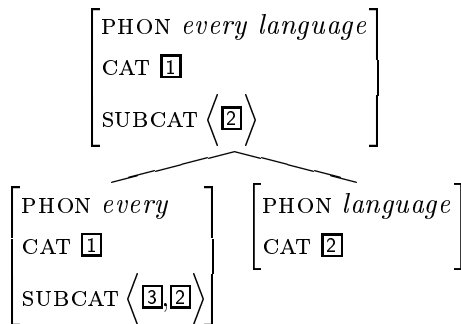
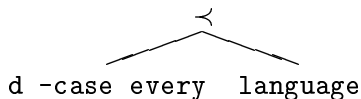
- Minimalism satisfies upper bounds, while
- HPSG satisfies lower bounds.

Conclusion (2)

Observation

HPSG subsumes Minimalism.

=n d -case every + n language \Rightarrow



Conclusion (3)

- A Minimalism $G = \langle (\text{PhonFeat} \cup \text{SemFeat}), (\text{BaseCats} \cup \text{SelfFeat} \cup \text{LiceeFeat} \cup \text{LicorFeat}), \text{Lex}, \{\mathbf{merge}, \mathbf{move}\} \rangle$ translates into an HPSG $\langle \langle \text{Types}, \sqsubseteq \rangle, \text{Prcps} \rangle$ s.t.
 - ▶ $\langle \text{Types}, \sqsubseteq \rangle$ wrt. $\text{Attr} = \{\text{CAT}, \text{SUBCAT}, \text{PHON}, \text{SEM}, \text{SLASH}, \text{QSTORE}\} \cup (\text{LiceeFeat} \cap \text{LicorFeat})$,
 - ▶ and Prcps contains three schemas, e.g. Merge, Covert Move and Overt Move.

Motivation

Comparison and computational support.



Patrick Farrell.

Functional shift as category underspecification.
English Language and Linguistics, 5(1):109–130, 2001.



Mark Gold.

Language identification in the limit.
Information and Control, 10:447–474, 1967.



Makoto Kanazawa.

Learnable classes of categorial grammars.
CSLI Publications, Stanford, California, 1998.



Martin Pickering and Steven Frisson.

Processing ambiguous verbs.
Journal of Experimental Psychology, 27(2):556–573, 2001.



Martina Penke, Ulrike Janssen, and Sonja Eisenbeiss.

Psycholinguistic evidence for the underspecification of morphosyntactic features.
Brain and Language, 90:423–433, 2004.



Anders Søgaard and Petter Haugereid.

A tractable typed feature structure grammar for Mainland Scandinavian.
Nordic Journal of Linguistics, 30(1), 2007.



Edward Stabler.

Derivational minimalism.

In Christian Retore, editor, *Logical Aspects of Computational Linguistics*, volume 1328 of *Lecture Notes in Artificial Intelligence*, pages 68–95. Springer, Berlin, Germany, 1997.