Modeling garden paths in a statistical dependency parser: Chinese, German, English

Marisa Ferrara Boston, Zhong Chen, and John T. Hale

Department of Linguistics and Languages, Michigan State University

Introduction

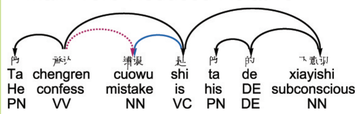
This study differentiates between probability models that lead to garden-pathing and those that fail to do so in an incremental dependency parser. We use Dependency Grammar (Tesnière 1959) to describe sentence structure in terms of word-to-word connections called dependencies.

We apply two sets of statistical features, state-based and non state-based, and examine each one's usefulness for targeting garden-path analyses that ensnare human readers in three languages.

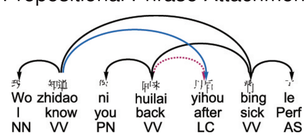
Chinese, German, and English garden paths...

These sentences are ambiguous between two interpretations: a human-preferred locally-correct analysis, indicated by the dashed arc, and a globally-correct interpretation, indicated by a solid blue arc.

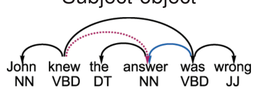
Main Verb-Reduced Relative



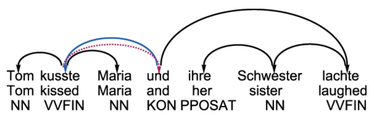
Prepositional Phrase Attachment



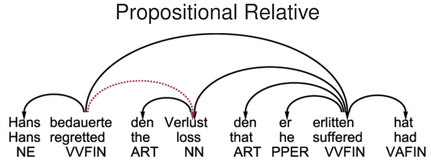
Subject-object



Conjunction



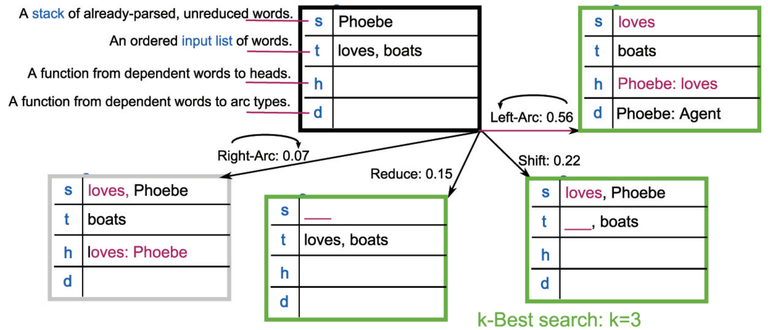
Propositional Relative



The data were aggregated from a variety of Chinese (Hsiao and Gibson 2003), English (Bever 1970), and German (Sailer 2004; Agricola 1968) psycholinguistic and linguistic studies.

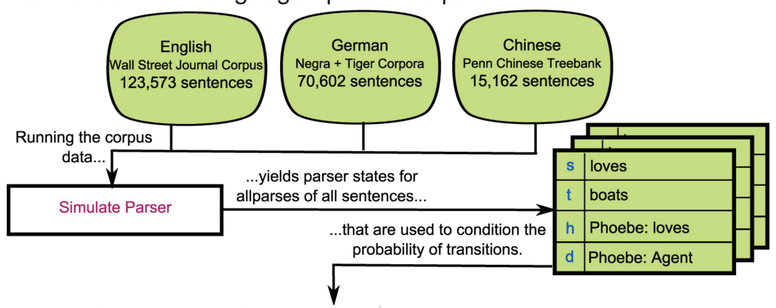
...are run through a parser...

Our parser is built to the specifications of Nivre 2004 with an added k-best search to implement garden-pathing (Frazier 1979). States encapsulate incremental analyses, and four possible actions can be taken to transition between states.



...that is informed by state-based and non state-based features...

The features, or statistical models, are trained on converted sentences from language-specific corpora.



	State-based features		Non state-based features
Stack3	Top 3 stack elements and input word.	Distance	Surface distance between top stack element and input word.
Stack1	Top stack element and input word.	Next	Next input word.
Top	Top stack element.	Position	Position of input word.

State-based features take into account internal parser states, while non state-based features take into account string information.

...state-based features result in human-like performance

Features that counsel for the human-preferred action are marked with a ✔, while those that counsel for the globally-correct action are marked with a ✗.

	State-based features			Non state-based features		
Sentence Type	Stack3	Stack1	Top	Distance	Next	Position
Chinese						
Main Verb-Reduced Relative		✔	✔		✗	✗
Subject-Object	✔	✔	✔	✗	✗	✗
Prepositional Phrase Attachment	✔	✔			✔	✗
German						
Propositional Relative Clause	✔	✔	✔	✔	✔	✔
Conjunction	✗	✔	✗	✗	✔	
Prepositional Phrase Attachment		✔	✔		✗	✗
English						
Main Verb-Reduced Relative	✔	✔	✔	✗	✗	
Subject-Object	✔	✔	✔	✗		✗
Prepositional Phrase Attachment	✔	✔		✗	✗	✗
Total	6	9	6	1	3	1

This leads to a feature hierarchy that defines a distributional basis for human parsing preferences.

Stack1 ≫ Stack3 ≫ Top ≫ Next ≫ Distance ≫ Position

Conclusion

The results reveal that garden-pathing models are best implemented by parsers that attend more to parser-state information than non state-based information.

References

Agricola, E. 1968. Syntaktische Mehrdeutigkeit (Polysyntaktizität) bei der Analyse des Deutschen und des Englischen. Schriften zur Phonetik, Sprachwissenschaft und Kommunikationsforschung. Akademie Verlag, Berlin.

Bever, T.G. 1970. The cognitive basis for linguistic structures. In R. Hayes (ed.), Cognition and Language Development. New York: Wiley and Sons. 277-360.

Frazier, L. 1979. On Comprehending Sentences: Syntactic parsing strategies. Ph.D. Dissertation, University of Connecticut.

Hsiao, F. and Gibson E. 2003. Processing relative clauses in Chinese. Cognition 90: 3-27.

Nivre, J. 2004. Incrementality in Deterministic Dependency Parsing. Proceedings of the Workshop on Incremental Parsing (ACL): 50-57.

Sailer, M. 2004. Propositional relative clauses in German. In Stefan Müller (ed.), Proceedings of the 11th International Conference on HPSG. Stanford:CSLI. 223-243.

Tesnière, L. 1959. Élements de syntaxe structurale. Editions Klincksiek.

{mferrara,chenzho3,jthale}@msu.edu

http://www.msu.edu/~mferrara/