Garden-Pathing in a Statistical Dependency Parser

Marisa Ferrara Boston and John T. Hale
Department of Linguistics and Languages
Michigan State University
East Lansing, MI 48824
{mferrara, jthale}@msu.edu

Abstract

This study differentiates between probability models that lead to garden-pathing and those that fail to do so in an incremental dependency parser. Models that take into account intermediate parser-states and part-of-speech pairs correctly reflect human preferences in three well-known cases: Main Verb vs. Reduced Relative ambiguities, Prepositional Phrase Attachment and Subject-Object ambiguities. Dependency width and direction features were not crucial in these examples, but may ultimately be helpful in accounting for other human sentence processing data. The results support computational proposals about human processing that prioritize stack memory and part-of-speech information over surface distance and dependency direction.

1 Introduction

Garden-pathing is a kind of temporary ambiguity in natural language that is widely thought to reflect fundamental properties of the human sentence processing mechanism. Example 1 shows a garden path sentence that is typically difficult for humans to parse.

(1) The horse **raced** past the barn fell.

The difficulty in interpreting the sentence in Figure 1 arises when the verb "raced" is mis-analyzed as the main verb of the sentence (the correct analysis defines "raced past the barn" as a reduced relative modifying the noun "horse"). The phenomenon can be characterized as arising from: a) fallible heuristic strategies such as the Canonical Sentoid Strategy (CSS) (Bever 1970); b) tree-structural heuristics like Minimal Attachment and Late Closure (Frazier 1979); c) lexical preferences (Ford, Bresnan, & Kaplan 1982; Macdonald, Perlmutter & Seidenberg 1994); or d) a pressure to assign semantic roles (Pritchett 1988; Gibson 1991). This paper presents an account of garden-pathing based on the features that guide a statistical dependency parser. It uses *k*-best search to implement Frazier's (1979) idea that garden-pathing is due to mistaken pruning of the correct analysis. However, the *k*-best approach generalizes this idea by loosening the requirement that only one parse is maintained at any given time. Using the garden path example from Example 1, if the parser-action that analyzes "horse" to be the subject of the verb "raced" has a low probability and ranks fourth or fifth, it will not be chosen as a possible analysis by a *k=3* parser and the parser will not garden path. If, however, the transition does rank in the top three, given the probabilities determined by the feature, the parser will garden path as a human would.

We follow Tesnière (1959) and Hayes (1964) in describing sentence structure in terms of word-to-word connections called *dependencies*. Figure 1 depicts an English sentence where the *head* word "bought" has links to its *dependents* "John" and "book".

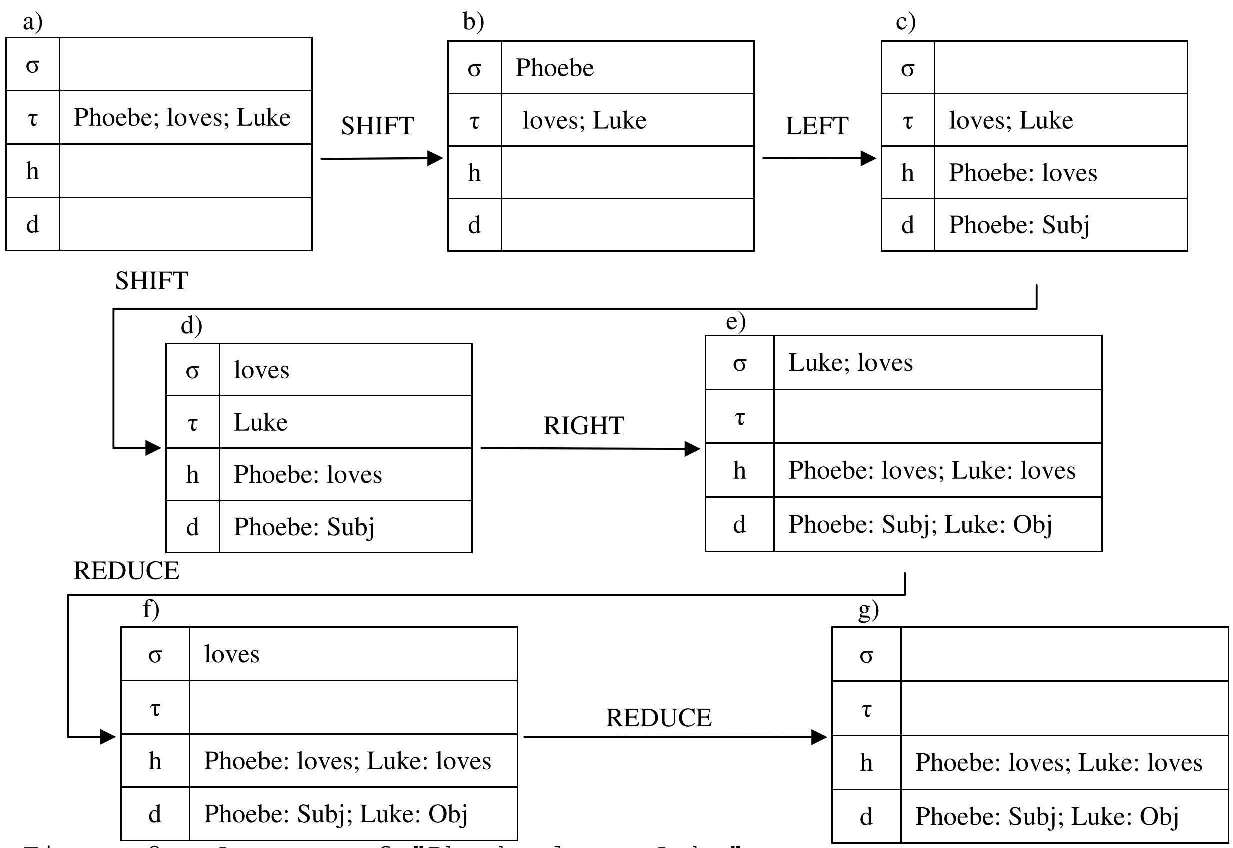


Figure 2: A parse of "Phoebe loves Luke".

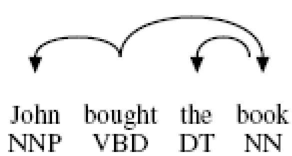


Figure 1: A DG sentence.[1]

We apply four kinds of statistical features (Figure 4) in a non-deterministic incremental dependency parser, examining each one's usefulness for targeting garden-path analyses that ensnare human readers in three well-studied cases (Sections 3.1-3.3). The results support models of human sentence processing that attend more to parser-state and part-of-speech pair information than surface distance or dependency direction. Before proceeding to the results, Section 2 sketches the overall methodology.

2 Methodology

Nivre's (2004a) incremental dependency parser assigns a dependency graph to each initial substring of a well-formed sentence. It does this by keeping track of a parser configuration that aggregates four data structures, listed in definition (2).

(2) Nivre-defined parser configuration (2006).

1. **σ** : A stack of already-parsed non-reduced words.
2. **τ**: An ordered input list of unparsed words.
3. **h**: A function from dependent words to head words.
4. **d**: A function from dependent words to arc types.

As with other pushdown automata, Nivre's parser defines stack-manipulating operations, or transitions, between configurations, defined in (3).

[1] All dependency graphs were output using Matthias Kromann's DGGraph tool (2002).

Feature	Description
Configuration	The probability of a transition *T* is the probability of the transition given the current configuration (conditioned by the top three elements in σ and the first element in τ).
Part-of-Speech (POS) Pair	The probability of an arc *R* from $word_i$ to $word_j$ is the probability that POS($word_i$) heads POS($word_j$)
Surface Distance	The probability of arc *R* from $word_i$ to $word_j$ is the probability of the number of words between POS($word_i$) and POS($word_j$).
Directionality	The probability of an arc *R* from $word_i$ to $word_j$ with direction *d* (left or right) is the probability that an arc from POS($word_i$) to POS($word_i$) is of type *d*.

Figure 4: Feature descriptions.

(3) Nivre-defined transition types (2006).

1. **LEFT-ARC**: A left arc is drawn from the current word being parsed *i* to the first word *j* of σ, making *j* the head of *i*; *i* is popped off σ.
2. **RIGHT-ARC**: A right arc is drawn from the first word *j* on σ to the current word being parsed *i,* making *i* the head of *j*; *j* is pushed onto σ.
3. **SHIFT**: Shifts the current word being parsed *j* onto σ without drawing any arcs.
4. **REDUCE**: Pops σ (applies only if the top word has a head).

Figure 2 exemplifies the changes to a Nivre parser's configuration in the course of left-to-right dependency analysis of the sentence in Figure 3. In the initial configuration (2a) the stack σ is empty, the input list τ contains the full input string, and the h and d functions are empty. Because Left-arc, Right-arc, and Reduce are not viable options, the parser must shift the first word "Phoebe" onto the stack, leading to the configuration in 2b. A Left-arc from "loves" to "Phoebe" pops "Phoebe" off the stack and adds head information to h (the head of "Phoebe" is "loves") and arc-type information to d (the arc from "loves" to "Phoebe" could be labeled with the Subject function). Similarly, once the parser is in the state shown in 2d, a Right-arc transition leads to configuration 2e, which now defines the Object function of the sentence. Finally, when no inputs are left, the parser uses the Reduce action to pop the stack until the parser is in the final configuration, with σ and τ empty (2g). What remains is the information in h and d, which is used to draw the dependency analysis for the sentence shown in Figure 3.

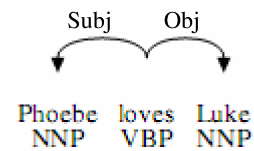


Figure 3: A dependency analysis of "Phoebe loves Luke".

In our implementation, transitions are chosen by stochastic models, or features. These features are informed by converted sentences from the full Wall Street Journal corpus of the Penn Treebank. We converted the sentences with Yamada's (2004) Ptb-conv 3.0 tool, which applies Collins' (1999) heuristic head-table to induce dependency graphs from the Penn Treebank's phrase-structure markup. We chose Ptb-conv over other dependency conversion tools because its outputs best matched certain linguistic analyses we desired. For example, Collins' SBAR rule interprets relative pronouns to be heads of relative clauses and dependents of the modified noun, an analysis in keeping with Gibson's Dependency Locality Theory.

We follow Nivre (2004b) in using a generative probability model to rank parser actions. A *k*-best search algorithm explores the top *k=3* parser configurations according to the ranking established by the probabilities. Figure 4 specifies the model's features, similar to those defined by Hall & Novák (2005). Some features take into account the parts of speech to be connected by potential dependency arcs, while others explicitly model the width or direction of a potential arc.

We compared the sequence of maximum-likelihood parser actions with human performance on three well-known ambiguity resolution problems in an effort to relate feature distributions to human sentence processing heuristics.

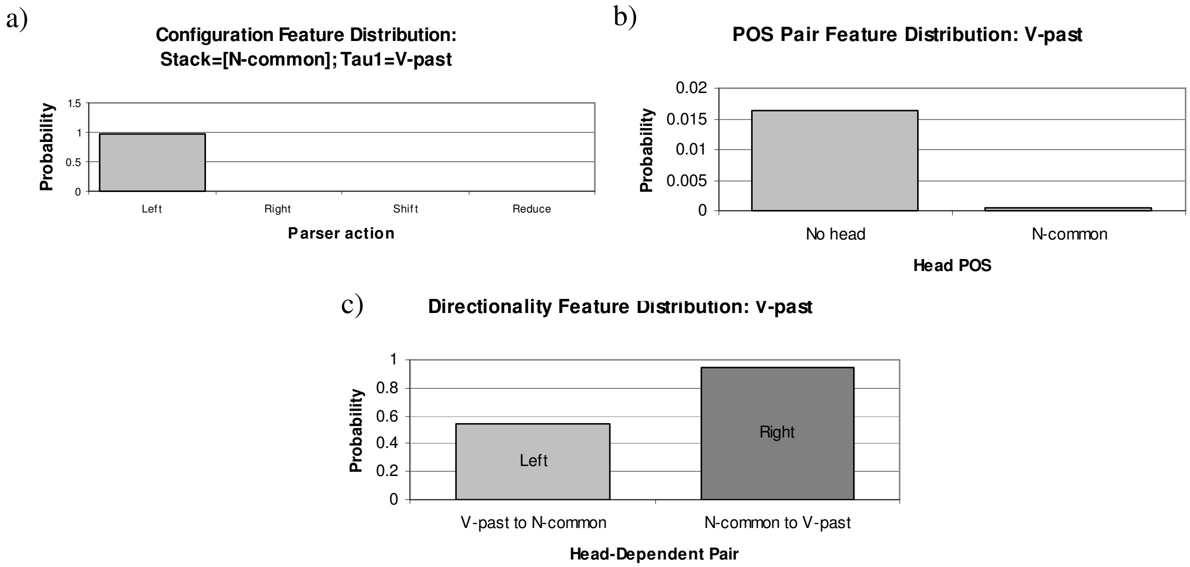


Figure 7: Feature distribution diagrams for MV vs. RR readings.

3 Results

Some features allow the dependency parser to avoid garden-pathing on ambiguous structures while others more closely match the human data. The results are sub-divided by the specific garden-pathing examples, beginning with the Main Verb versus Reduced-Relative reading.

3.1 Main Verb vs. Reduced-Relative Reading

The first sentences tested were those that exhibit Main Verb versus Reduced-Relative ambiguities, like the standard garden path example given in Figure 1. Figure 5 shows the development of a garden path in the sentence "The doctor sent for the patient arrived."

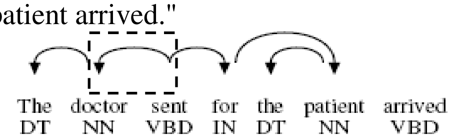


Figure 5: The main verb interpretation.

The box in Figure 5 highlights the arc responsible for the human misinterpretation of the third word "sent" as the main verb of the sentence rather than as the head of a reduced-relative depending on the noun "doctor". This latter situation is depicted in Figure 6.

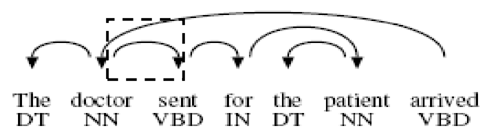


Figure 6: The reduced-relative interpretation.

A dependency parser guided by the Configuration or POS Pair features misinterprets the sentence just as a human would. Figure 7a shows that when a noun ("doctor") is solely on the stack and a verb ("sent") is the next input word, there is a high probability of a Left-arc action, making "sent" the head of "doctor". Figure 7b shows that POS Pair feature favors a Shift transition over the alternative Right-arc transition that links the verb "sent" to the noun "doctor". The Directionality feature councils against the action that leads the parser up the garden-path: Figure 7c illustrates how, on the probability model estimated from the Treebank, a Right-arc from a nominal head "doctor" to a dependent verb "sent" is more probable than the reverse action—a Left-arc from a verbal head "sent" to a dependent noun "doctor"—which leads to a garden path. The Surface Distance feature is inconsequential because it can not differentiate between the two analyses.

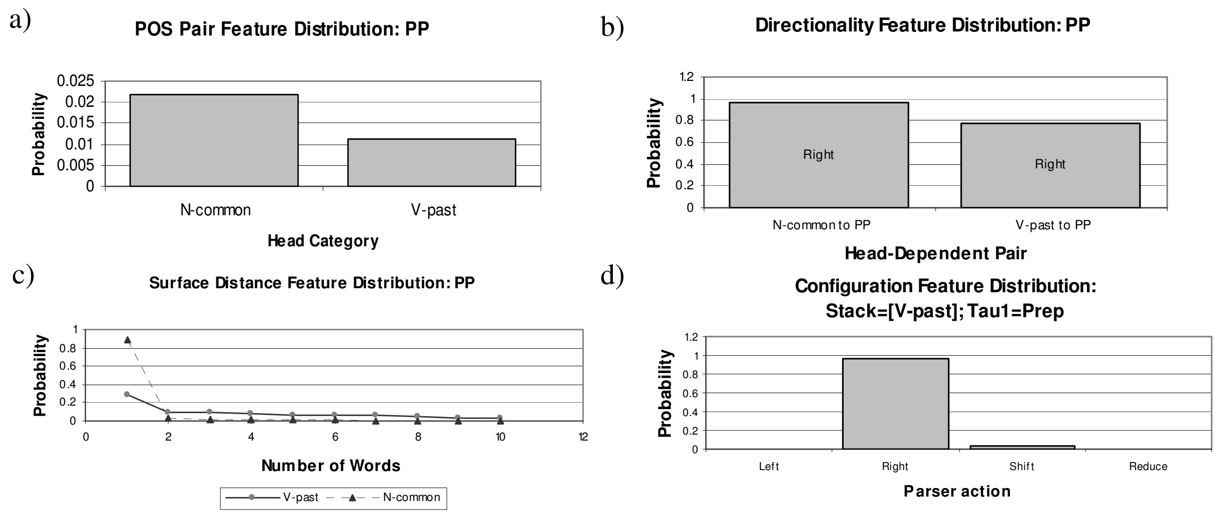
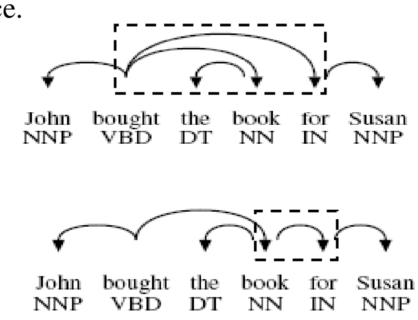


Figure 9: Feature distribution diagrams for PP-Attachment.

Collectively, the results in Figure 7 indicate that the Configuration and POS Pair features reflect human preferences better than the Directionality feature does on its own. The former two features together implement Bever's (1970) CSS. A model prioritizing Configuration and POS Pair over Directionality is thus consistent with the human preference in the Main Verb versus Reduced-Relative ambiguity.

3.2 Prepositional-Phrase Attachment

Ambiguity resolution is also required for Prepositional Phrase (PP) Attachment. Figures 8a and 8b show alternative attachment sites for the PP, which give rise to different readings of the sentence.



a)

b)

Figure 8: Alternative readings for PP-Attachment

Figure 8a shows high attachment, or the benefactive reading of the sentence, where the book is intended for Susan. 8b depicts low attachment, where the PP attaches to the noun. The POS Pair feature favors the low attachment reading (Figure 9a), as do the Directionality (Figure 9b) and Surface Distance (Figure 9c) features. However, Configuration allows the parser to find the high attachment reading (Figure 9d). In the Nivre parser, this happens because the noun "book" (and its dependent determiner "the") has already been removed from the stack when the attachment site for "for" is being considered (cf. Shieber 1983, Perreira 1985) The only alternative is for the preposition "for" to attach to the verb "bought". Models of the human sentence processor where ambiguity resolution is guided more strongly by the kind of hierarchical information in the Configuration feature are consistent with Frazier's (1979) Minimal Attachment.

Ford, Bresnan, & Kaplan (1982) note that PP-Attachment is associated with lexical preferences as in Example 4.

(4)
a. The woman **wanted** the dress on the rack.
b. The woman **positioned** the dress on the rack

The human sentence processor favors low PP-Attachment with "want" whereas it prefers high PP-Attachment with "position". In the Penn Treebank annotation both verbs receive the same POS. However, dependency parsers guided by the POS Pair feature would be able to differentiate between "want"- and "position"-type verbs if given an enriched tag set. Such an approach constitutes one way to realize Ford, Bresnan, & Kaplan's (1982) suggestion that their lexical distinctions be grounded in distributional frequencies.

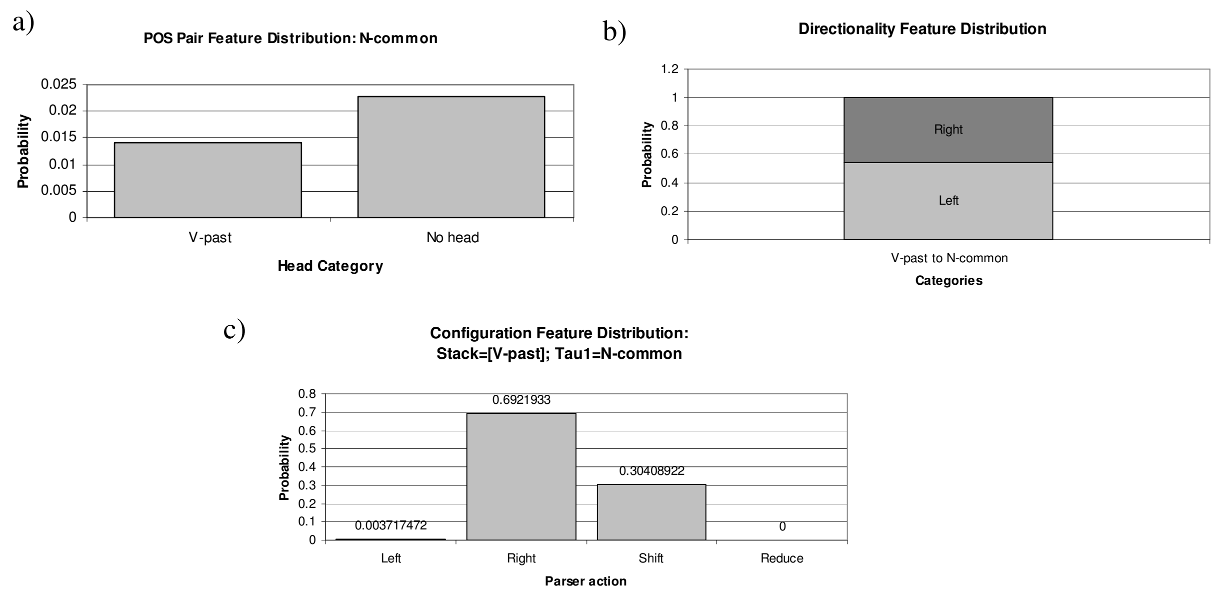
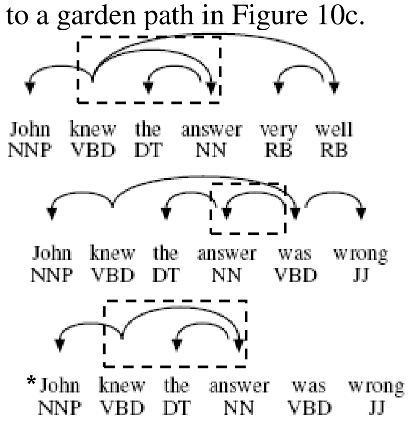


Figure 11: Feature distribution diagrams for Subject-Object ambiguities.

3.3 Subject-Object Ambiguities

A Subject-Object ambiguity arises when a parser assumes a noun to be the direct object (DO) of a verb (Figure 10a), unaware that further input may reveal a second verb to which the noun can attach as a subject (S) (Figure 10b). The human preference is for the reading in Figure 10a, which leads to a garden path in Figure 10c.



a)

b)

c)

Figure 10: Subject-Object ambiguity.

Both the POS Pair and the Directionality features favor the S reading of the sentence, where the noun "answer" is not immediately attached as an object of the verb "knew". The POS Pair feature rates a Shift transition higher than a Right-arc transition (Figure 11a), and Directionality slightly prefers a Left-arc from a verbal head "was" to a noun "answer". But, the Configuration feature (11c) is able to produce the correct human garden path by choosing a Right-arc transition to attach the noun "answer" to the verb "knew".

The Configuration feature prefers the action that leads to the garden path DO reading over alternatives leading to the S reading. This particular type of ambiguity is not as difficult for the human processor in the sense that both readings of the sentence (10a and 10b) are acceptable (Kimball 1973; Ferreira & Henderson 1990). Therefore, models of the human processor that weight these features more closely would best reflect this situation.

3.4 Analysis

The feature set in Section 2 is able to model human garden-pathing in MV vs. RR, PP-Attachment, and Subject-Object ambiguities. In all three cases, the Configuration feature agrees with the human preference while the other features combine to demonstrate the alternative grammatical readings. Figure 12 shows a hierarchy of the features based on our results.

Configuration >> POS Pair >> {Surface Distance, Directionality}

Figure 12: Feature ranking.

The above hierarchy reflects the fact that Configuration will always choose the human garden path. POS Pair is able to model garden-pathing in the MV vs. RR sentences, and is able to guide the parser in easy recovery of the alternative readings for PP-Attachment and Subject-Object ambiguities. The final two features, Surface Distance and Directionality, choose the alternative readings when applicable. This hierarchy constitutes a distributional basis for human parsing preferences.

4 Conclusions and Future Work

Our results reveal that an incremental dependency parser trained on the Wall Street Journal corpus is able to model human garden-pathing and ambiguity resolution. An adequate model might combine the features in Figure 4 such that Configuration and POS Pair probabilities guide initial parsing preferences, whereas POS Pair, Directionality and Surface Distance probabilities figure only in the recovery of alternative analyses. The key difference between our work and other studies with implemented stochastic models of human sentence processing, such as Hale 2001 and Park & Brew 2006, is that our parser specifies which grammatically-interpretable action is being performed at each time step. This allows the garden-pathing intuition to be formalized as fixed-beam search. Our methodology also differs from the previous studies in its reliance on dependency structures, which are independently well-motivated in psycholinguistics (Gibson 2000) and allow for a greater domain of locality compared to finite state approaches. The specific findings from this study are directly applicable in more expressive formalisms that have a dependency interpretation, such as the class of mildly context-sensitive grammars (Joshi et al. 1991).

This study naturally leads to future work that considers these features in relation to psycholinguistic theories such as Gibson's Dependency Locality Theory (2000). We aim to reveal a more fine-tuned account of human parser heuristics by testing additional psycholinguistic constructions against a parser informed by multi-feature models.

Acknowledgements

This work was supported by an IGERT training grant from the National Science Foundation.

References

Bever, T.G. 1970. The Cognitive Basis for Linguistic Structures. In R. Hayes (ed.), *Cognition and Language Development*. New York: Wiley & Sons: 277-360.

Collins, M. 1999. *Head-Driven Statistical Models for Natural Language Processing*. Ph.D. Dissertation, University of Pennsylvania.

Ferreira, F. and John Henderson. 1990. Use of Verb Information in Syntactic Parsing: Evidence from eye movements and word-by-word self-paced reading. *Journal of Experimental Psychology: Learning, Memory and Language*. 30: 725-745.

Ford, M., Joan Bresnan, and Ronald Kaplan. 1982. A Competence-Based Theory of Syntactic Closure. In Joan Bresnan (ed.) *The Mental Representation of Grammar*. Cambridge, MA: The MIT Press: 727-796.

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

Gibson, E. 1991. *A Computational Theory of Human Linguistic Processing: Memory limitations and processing breakdown*. Ph.D. Dissertation, Carnegie Mellon University.

Gibson, E. 2000. The Dependency Locality Theory: A distance-based theory of linguistic complexity. In A. Marantz et al. (eds.) *Image, Language, Brain*. Cambridge, MA: MIT Press: 95-126.

Hale, J. 2001. A Probabilistic Earley Parser as a Psycholinguistic Model. *Proceedings of NAACL 2001*: 1-8.

Hall, K. and Václav Novák. 2005. Corrective Modeling for Non-Projective Dependency Parsing. *Proceedings of the 9th International Workshop on Parsing Technologies (IWPT '05)*: 42-52.

Hayes, D.G. 1964. Dependency Theory: A formalism and some observations. *Language*, 40: 511-525.

Joshi, A.K., K. Vijay-Shankar, and David Weir. 1991. The Convergence of Mildly Context-Sensitive Grammar Formalisms. In P. Sells et al. (eds.) Processing of Linguistic Structures. Cambridge, MA: MIT Press: 31-81.

Kimball, J. 1973. Seven Principles of Surface Structure Parsing in Natural Language. *Cognition*, 2: 15-47.

Kromann, M. 2002. DGGraph Dependency Graph Drawing Tool. Available online at http://www.id.cbs.dk/~mtk/DGgraph/.

Lewis, R. 1993. *An Architecturally-based Theory of Human Sentence Comprehension*. Ph.D. Dissertation, Carnegie Mellon University.

Macdonald, M., Neal Perlmutter, and Mark Seidenberg. 1994. Lexical Nature of Syntactic Ambiguity Resolution. *Psychological Review* 101(4): 676-703.

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

Nivre, J. 2004b. Inductive Dependency Parsing. MSI Report 04070. Växjö University: School of Mathematics and Systems Engineering.

Nivre, J. 2006. *Inductive Dependency Parsing*. Springer.

Park, J. and Chris Brew. 2006. A Finite-State Model of Human Sentence Processing. *Proceedings of Coling/ACL 2006*: 49-56.

Perreira, F. 1985. A New Characterization of Attachment Preference. In D. Dowty et al. (eds.) *Natural Language Parsing: Psychological, computational, and theoretical perspectives*. Cambridge: Cambridge University Press: 307-319.

Pritchett, B. 1988. The Grammatical Basis of Language Processing. *Language* 64(3): 539-576.

Shieber, S. 1983. Sentence Disambiguation by a Shift-Reduce Parsing Technique. *Proceedings of the Eighth International Conference on Artificial Intelligence*: 699-703.

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

Yamada, H. 2004. Ptb-conv 3.0: Dependency generator tool. Available online at http://www.jaist.ac.jp/~h-yamada/.