Abstract

Olney (2007) presents an unsupervised grammar induction model that uses semantic similarity to induce syntactic structure. A key element of this model is the operational definition of syntactic heads as being semantically substitutable for their phrases. This paper describes the history of this operational definition for heads and tests its validity with respect to four computational implementations. The paper concludes with implications for these results on the operational definition of heads proposed by theoretical linguists as well as the model presented by Olney (2007).

Semantic Heads for Grammar Induction

Andrew M. Olney
Institute for Intelligent Systems
University of Memphis
Memphis, TN 38152
aolney@memphis.edu

August 6, 2007

1 Introduction

Previous work on unsupervised grammar induction [1, 3, 11] has made use of the distributional hypothesis [9], which characterizes words by their contexts. Under the distributional hypothesis, a phrase that occurs in the same environments as a single word, i.e. has a similar distribution, is likely to be have the same syntactic function as the single word. Hierarchical descriptions of sentences can then be built by attaching such phrases to a higher level node (represented by the single substituting word) until only a single root node remains [1].

Olney (2007) recently proposed a semantically-oriented model based on the distributional hypothesis. This model is distinguished from previous models in that it does not make use of part of speech tags or nonterminal nodes, yet it still manages to significantly outperform a right branching baseline. A key element of this model is the operational definition of syntactic heads as being semantically substitutable for their dependents, as discussed in the theoretical linguistics literature [17, 10, 4]. This paper explores using semantic substitutability to determine syntactic heads and presents perhaps the first computational evaluation of this notion.

2 Semantic Heads

Heads are theoretically-motivated linguistic elements with a primary role in syntactic description. In X-bar theory, a phrase contains a single head which determines the syntactic type of the phrase [2]. In dependency grammars, heads have a similar role, except that syntactic relations exists solely between words [16]. The importance of heads across these theories suggests that the proper identification of heads is of central importance to a language learner.

Despite the wide use of heads and head-like notions in syntactic theory, there has been much debate as to precisely what a head is [17, 10, 4]. However, some agreement exists on a loose semantic definition of head: X is a head of X+Y if X describes *a kind of* thing described by X+Y [17, 10, 4]. Beyond this initial agreement, differences emerge. Zwicky (1985) equates *a kind of* with semantic arguments. He argues that "green car" describes a kind of car, rather than a kind of green. This example is endocentric, since the head, "car" appears within the phrase "green car." Hudson (1987) makes the opposite claim, that *a kind of* refers to semantic functors, e.g. "on the desk" refers to a kind of location rather than a kind of desk. This example is exocentric, since the meaning "location" is not directly attributable to any particular word in "on the desk," but to the phrase as a whole. Croft (1996) unites these two perspectives by noting that semantic functors are heads for government relationships and semantic arguments are heads for modification relationships.

Given the theoretical importance *a kind of* has in identifying heads, it is worthwhile exploring machine learning methods that can acquire this kind of knowledge. In particular, Latent Semantic Analysis (LSA) is a vector space method capable of measuring the similarity between words and collections of words [5, 6, 12]. LSA has been shown to closely approximate vocabulary acquisition in children [12], grade essays as reliably as experts in English composition [7], and understand student contributions in tutorial dialogue [8, 14]. These results are particularly impressive considering that LSA creates its knowledge representation without human intervention.

3 Methodology

We present four methods for identifying heads using LSA. The basic methodology is to create an LSA space and to compare the semantic similarity of a dependency pair's elements to the whole. For example, "green" and "car" would both be compared with "green car." Using the Penn Treebank [13], heads found using these four methods are compared with manually identified heads.

The four methods presented use this basic methodology along the dimensions +/- order and +/- endocentric. The ordered methods use unigrams and bigrams as basic elements, inherently preserving word order. The minus endocentric, or exocentric, methods do not compare a dependency element to the whole, but rather to the nearest unigram neighbor of the whole. For example "in bed" may have a nearest unigram neighbor, "sleepy," which is more similar to "bed" than to "in." Furthermore, the construction of the LSA spaces varied in terms of local and global context. Global context represents the traditional LSA calculation, in which $cell_{ij}$ denotes the number of times $term_i$ appeared in $document_j$. In local context, $cell_{ij}$ is the number of times $term_j$ occurred before the target $term_i$, and the value of $cell_{i(j+n)}$ is the number of times $term_j$ occurred after the target $term_i$, where n is the number of terms in the corpus. Both local and global spaces were constructed using both unigrams and bigrams as terms to preserve word order.

4 Results & Discussion

Results in Table 1 show that only the ordered methods were significantly better than chance, and that unordered methods were significantly worse (p = .05). There was no significant difference between endocentric and exocentric methods (p=.05). These results suggest that LSA is capturing "*a kind of*"-like information on a more abstract level than endocentric and exocentric, which would make LSA similarity closer to the loose semantic definition of head described in the literature [17, 10, 4]. However, the low overall discriminability of LSA, 57% in the best case, further suggests that semantic similarity is not the only factor in determining headhood. It appears likely that there is another element to determining headness that is missing from the discussion amongst theoretical linguists.

These results have similar significance to the model proposed by Olney (2007). It is somewhat surprising that this model can outperform a right branching baseline even though the method of determining headhood has a weak discriminability of 57%. It seems likely therefore that an improvement in the ability to determine heads will be a major source of improvement in this model.

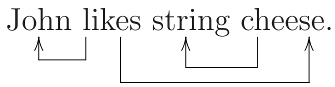

Figure 1: A Dependency Graph

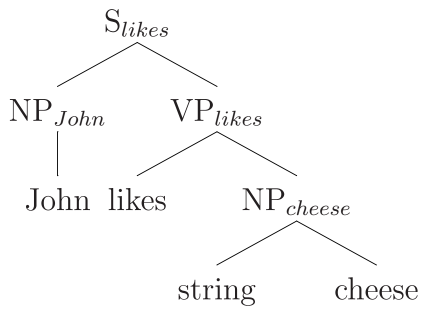

Figure 2: A Lexicalized Tree

5 Results

Table 1: Head Discrimination Results for WSJ10

Method Ordered/Endocentric	Local Context Percentage Correct	Global Context Percentage Correct
-/-	42.3%	41.7%
-/+	42.3%	41.6%
+/-	56.8%	49.8%
+/+	57.3%	48.3%

References

[1] Eric Brill and Mitchell Marcus. Automatically acquiring phrase structure using distributional analysis. In *Speech and Natural Language: Proceedings of a Workshop Held at Harriman, New York*, pages 155–160, Philadelphia, February 23-26 1992. Association for Computational Linguistics.

[2] Noam Chomsky. Remarks on nominalization. In R. Jacobs and P. S. Rosenbaum, editors, *Readings in English Transformational Grammar*. Ginn and Co., Waltham, Massachusetts, 1970.

[3] Alexander Clark. Unsupervised induction of stochastic context-free grammars using distributional clustering. In Walter Daelemans and Rémi Zajac, editors, *Proceedings of the ACL 2001 Workshop on Computational Natural Language Learning*, pages 105–112, Philadelphia, July 6-7 2001. Association for Computational Linguistics.

[4] William Croft. What's a head? In Laurie Zaring and Johan Rooryck, editors, *Phrase structure and the lexicon*, pages 35–75. Kluwer, Dordrecht, 1996.

[5] Scott C. Deerwester, Susan T. Dumais, George W. Furnas, Thomas K. Landauer, and Richard A. Harshman. Indexing by latent semantic analysis. *Journal of the American Society of Information Science*, 41(6):391–407, 1990.

[6] Susan Dumais. Improving the retrieval of information from external sources. *Behavior Research Methods, Instruments and Computers*, 23(2):229–236, 1991.

[7] P.W. Foltz, S. Gilliam, and S. Kendall. Supporting Content-Based Feedback in On-Line Writing Evaluation with LSA. *Interactive Learning Environments*, 8(2):111–127, 2000.

[8] A.C. Graesser, P. Wiemer-Hastings, K. Wiemer-Hastings, D. Harter, T.R.G. Tutoring Research Group, and N. Person. Using latent semantic analysis to evaluate the contributions of students in AutoTutor. *Interactive Learning Environments*, 8(2):129–147, 2000.

[9] Zellig Harris. Distributional structure. *Word*, 10:140–162, 1954.

[10] Richard A. Hudson. Zwicky on heads. *Journal of Linguistics*, 23:109–132, 1987.

[11] Dan Klein and Christopher D. Manning. A generative constituent-context model for improved grammar induction. In *Proceedings of the 40th Annual Meeting of the ACL*, pages 128–135, Philadelphia, July 7-12 2002. Association for Computational Linguistics.

[12] Thomas K. Landauer and Susan T. Dumais. A solution to plato's problem: The latent semantic analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review*, 104:211–240, 1997.

[13] Mitchell P. Marcus, Mary A. Marcinkiewicz, and Beatrice Santorini. Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, 19(2):313–330, 1993.

[14] B. A. Olde, D. Franceschetti, A. Karnavat, and A. C. Graesser. The right stuff: Do you need to sanitize your corpus when using Latent Semantic Analysis? In *Proceedings of the 24th Annual Meeting of the Cognitive Science Society*, pages 708–713, Mahwah, NJ, 2002. Erlbaum.

[15] A. M. Olney. Latent semantic grammar induction: Context, projectivity, and prior distributions. In *Proceedings of TextGraphs-2: Graph-Based Algorithms for Natural Language Processing*, pages 45–52, Rochester, NY, April 2007. Association for Computational Linguistics.

[16] Lucien Tesniere. *Elements de syntaxe structurale*. Klincksieck, Paris, 1959.

[17] Arnold M. Zwicky. Heads. *Journal of Linguistics*, 21:1–29, 1985.