PHRASE STRUCTURE TREES BEAR MORE FRUIT THAN YOU WOULD HAVE THOUGHT*

Aravind K. Joshi
Department of Computer and Information Science
The Moore School/D2
University of Pennsylvania
Philadelphia, PA 19104

and

Leon S. Levy
Bell Telephone Laboratories
Whippany, NJ 07981

EXTENDED ABSTRACT**

There is renewed interest in examining the descriptive as well as generative power of phrase structure grammars. The primary motivation has come from the recent investigations in alternatives to transformational grammars [e.g., 1, 2, 3, 4]. We will present several results and ideas related to phrase structure trees which have significant relevance to computational linguistics.

We want to accomplish several objectives in this paper.

1. We will give a brief survey of some recent results and approaches by various investigators including, of course, our own work, indicating their interrelationships.

Here we will review the work related to the notion of node admissibility starting with Chomsky, followed by the work by McCawley, Peters and Ritchie, Joshi and Levy, and more recent work of Gazdar.

We will also discuss other amendments to context-free grammars which increase the descriptive power but not the generative power. In particular, we will discuss the notion of categories with holes as recently introduced by Gazdar [3]. There is an interesting history behind this notion. Sager's parser explicitly exploits such a convention and, in fact, uses it to do some coordinate structure computation. We suspect that some other parsers have this feature also, perhaps implicitly. We will discuss this matter, which obviously is of great interest to computational linguists.

2. Our work on local constraints on structural descriptions, [5, 6], which is computationally relevant both to linguistics and programming language theory, has attracted some attention recently; however, the demonstration of these results has remained somewhat inaccessible to many due to the technicalities of the tree automata theory. Recently, we have found a way of providing an intuitive explanation of these results in terms of <u>interacting</u> finite state machines (of the usual kind). Besides providing an intuitive and a more transparent explanation of our results, this approach is computationally more interesting and allows us to formulate an interesting question: How large a variable set (i.e., the set of nonterminals) is required for a phrase structure grammar or how much information does a nonterminal encode? We will present this new approach.

3. We will present some new results which extend the "power" of local constraints without affecting the character of earlier results. In particular, we will show that local constraints can include, besides the proper analysis (PA) predicates and domination (δ) predicates, more complex predicates of the following form.

(1) (PRED N_1 N_2 ... N_n)

where N_1, N_2, ... N_n are nonterminals mentioned in the PA and/or δ constraint of the rule in which (1) appears and PRED is a predicate which, roughly speaking, checks for certain domination or left-of (or right-of) relationships among its arguments. Two examples of interest are as follows.

(2) (CCOMMAND A B C)

CCOMMAND holds if B immediately dominates A and B dominates C, not necessarily immediately. Usually the B node is an S node.

(3) (LEFTMOSTSISTER A B)

LEFTMOSTSISTER holds if A is the leftmost sister of B.

We will show that introduction of predicates of the type (1) do not change the character of our result on local constraints. This extension of our earlier work has relevance to the formulation of some long distance rules without transformations (as well as without the use of the categories with holes as suggested by Gazdar). We will discuss some of the processing as well as linguistic relevance of these results.

4. We will try to compare (at least along two dimensions) the local constraint approach to that of Gazdar's (specifically his use of categories with holes) and to that of Peters' use of linked nodes (as presented orally at Stanford recently).

The dimensions for comparison would be (a) economy of representation, (b) proliferation of categories, by and large semantically vacuous, and (c) computational relevance of (a) and (b) above.

5. Compositional semantics [8] is usually context-free, i.e., if nodes B and C are immediate descendants of node A, then the semantics of A is a composition (defined appropriately) of the semantics of B and semantics of C. Semantics of A depends only on nodes B and C and not on any other part of the structural description in which A may appear. Our method of local constraints (and to some extent Peters' use of linked nodes) opens the possibility of defining the semantics of A not only in terms of the semantics of B and C, but also in terms of some parts of the structural description in which A appears. In this sense, the semantics will be context-sensitive. We have achieved some success with this approach to the semantics of programming languages. We will discuss some of our preliminary ideas for extending this approach to natural language, in particular, in specifying scopes for variable binding.

6. While developing our theory of local constrains and some other related work, we have discovered that it is possible to characterize structural descriptions (for phrase structure grammars) entirely in terms of trees without any labels, i.e., trees which capture the grouping structure without the syntactic categories (which is the same as the constituent structure without the node labels [7]. This is a surprising result. This result

* This work was partially supported by NSF grant MCS79-08401.

** Full paper will be available at the time of the meeting.

provides a way of determining how much "structure" nonterminals (syntactic categories) encode and therefore clearly, it has computational significance. Moreover, to the extent that the claim that natural languages are context-free is valid, this result has significant relevance to learnability theories, because our result suggests that it might be possible to "infer" a phrase structure grammar from just the grouping structure of the input (i.e., just the phrase boundaries). Further, the set of structural descriptions without labels are directly related to the structural descriptions of a context-free grammar; hence, we may be able to specify "natural" syntactic categories.

In summary, we will present a selection of mathematical results which have significant relevance to many aspects of computational linguistics.

SELECTED REFERENCES

[1] Bresnan, J.W., "Evidence for an unbounded theory of transformations," Linguistic Analysis, Vol. 2, 1976.

[2] Gazdar, G.J. M., "Phrase structure grammar," to appear in The Nature of Syntactic Representation, (eds. P. Jacobson and G.K. Pullum), 1979.

[3] Gazdar, G.J.M., "English as a context-free language," unpublished ms., 1978.

[4] Gazdar, G.J.M., "Unbounded dependencies and coordinate structure," unpublished ms. 1979.

[5] Joshi, A.K. and Levy, L.S., "Local transformations," SIAM Journal of Computing, 1977.

[6] Joshi, A.K., Levy, L.S., and Yueh, K., "Local constraints in the syntax and semantics of programming languages," to appear in Journal of Theoretical Computer Science, 1980.

[7] Levy, L.S. and Joshi, A.K., "Skeletal structural descriptions," Information and Control, Nov. 1978.

[8] Knuth, D.E., "Semantics of context-free languages," Mathematical Systems Theory, 1968.

[9] Sager, N., "Syntactic analysis of natural languages," in Advances in Computers (eds. M. Alt and M. Rubinoff), vol. 8, Academic Press, New York, 1967.