PHRASE STRUCTURE GRAMMARS AND NATURAL LANGUAGES

Gerald Gazdar

Cognitive Studies Programme
University of Sussex
Brighton BN1 9QN

ABSTRACT

During most of the last two decades, computational linguists and AI researchers working on natural language have assumed that phrase structure grammars, despite their computational tractability, were unsatisfactory devices for expressing the syntax of natural languages, however, during the same period, they have come to realize that transformational grammars, whatever their linguistic merits, are computationally intractable as they stand. The assumption, unchallenged for many years, that PSG's were inadequate for natural languages is based on arguments originally advanced by transformational linguists in the late 1950»s and early 1960's. but recent work has shown that none of those arguments were valid. The present paper draws on that work to argue that (i) there is no reason, at the present time, to think that natural languages are not context-free languages, (ii) there are good reasons to think that the notations needed to capture significant syntactic generalizations will characterize phrase structure grammars or some minor generalization of them, and (iii) there are good reasons for believing that such grammars, and the monostratal representations they induce, provide the necessary basis for the semantic interpretation of natural languages. If these arguments are valid, then the prospects for a fruitful interaction between theoretical linguistics and AI are much brighter than they would otherwise be.

I INTRODUCTION

Consider the following quotations:

> As already mentioned, a context-free phrase structure grammar is not sufficient to describe or analyze the whole range of syntactic constructions which occur in natural language texts (cf. Chomsky 1957). Even if one disregards the theoretical linguist's demand for satisfactory descriptions of the syntactic structure of sentences, there are strong reasons to design a more powerful parser than the one described above.
> Cases in point are the phenomenon of agreement within noun phrases, the correspondence in the verb phrase between the form of the main verb and the type of the auxiliary, and subject verb agreement. The argument structure of predicates, that is, their various types of objects and complements, represents another kind of context-sensitivity in natural language. ... The best strategy seems to be to take care of the particular types of context-sensitivity recognized by linguistic theory by means of special procedures which act as a superstructure of the algorithm for context-free analysis. ... In addition to the above-mentioned drawbacks a context-free phrase structure grammar has difficulty handling word order variation in a natural way.
> (welin 1979: 62-63)

> One significant -use of the general context-free methods is as part of a system of processing natural languages such as English. We are not suggesting that there is a context-free grammar for English. It is probably more appropriate to view the grammar/parser as a convenient control structure for directing the analysis of the input string. The overall analysis is motivated by a linguistic model which is not context free, but which can frequently make use of structures determined by the context-free grammar.
> (Graham, Harrison & Ruzzo 1980: 415-116)

These two passages have a number of things in common, of which three are relevant here. Firstly, the issue of whether natural languages (NL's) are context-free languages (CFL's) and are susceptible to analysis by context-free phrase structure grammars (CF-PSG's) is one the authors take to be relevant to parsing. Secondly, both passages assume that this issue has already been resolved, and resolved in the negative. Thirdly, I did not have to look for them, I merely bumped into them, as it were, in the course of recent reading. However, I am sure that if I had been in the business of finding passages with this kind of flavour in the parsing literature of the past 20 years, then I could have found dozens, probably hundreds.

The purpose of the present paper is simply to draw the attention of computational linguists to the fact that the issue of the status of NL's with respects to the CFL's and CF-PSG's is not resolved, and to the fact that all the published arguments seeking to establish that NL's are not CFL's, or that CF-PSG's are not adequate for the analysis of NL's, are completely without force. Of course, this does not entail that NL's are CFL's or that CF-PSG's constitute the appropriate formal theory of NL grammars. But it does have as a consequence that computational linguists should not just give up on CF-PSG's on the grounds that theoretical linguistics has demonstrated their inadequacy. No such demonstration exists.

In assessing whether some formal theory of grammar is an adequate theory for NL's, at least the following three criteria are relevant, and have been historically. (i) Does it permit NL's qua sets of strings to be generated? (ii) Does it permit significant generalizations to be expressed? (iii) Does it support a semantics, that is, does it provide a basis on which meanings can be assigned to NL expressions in a satisfactory manner?

In the remainder of this paper, I shall consider these three criteria in turn with reference to the adequacy of CF-PSG's as grammars for NL's. The issues are large, and space is limited, so my discussion will take the form, for the most part, of annotated references to the literature where the various issues are properly dealt with.

II <u>GENERATING NATURAL LANGUAGE STRING SETS</u>

The belief that CF-PSG's cannot cope with syntactic concord and long-distance dependencies, and hence that NL's are not CFL's, but, say, properly context-sensitive languages, is well entrenched. One textbook goes so far as to assert that 'the grammatical phenomenon of Subject Predicate Agreement is sufficient to guarantee the accuracy of [the statement that] English is not a CF-PSG language' (Grinder & Elgin 1973: 59). The phenomenon guarantees no such thing, of course. Nor is the character of the problem changed when agreement is manifested across unbounded distances in strings (pace Bach 1974: 77, Bresnan 1978: 38). Indeed, finite state languages can exhibit such dependencies (see Pullum & Gazdar, 1982).

The introductory texts and similar expository works in the field of generative grammar offer nothing that could be taken seriously as an argument that NL's are not CFL's. However, five putatively non-specious arguments to this effect are to be found in the more technical literature. These are based on the following phenomena:

(a) English comparative clauses
(Chomsky 1963: 378-9),
(b) the decimal expansion of pi
(Elster 1978: 43-44),
(c) 'respectively'
(Bar-Hillel & Shamir 1960: 96,
Langendoen 1977: 4-5),
(d) Dutch subordinate clauses
(Huybregts 1976),
(e) Mohawk noun incorporation
(Postal 1964).

Pullum & Gazdar (1982) show that (a) is based on a false empirical claim and a false claim about formal languages, (b) has no bearing on English or any other natural language since it depends on a confusion between grammar and arithmetic, (c) is based on a false empirical claim, and the facts, such as they are, are relevant to semantics rather than syntax in any case, (d) provides no basis for any string set argument [1], and (e) Postal crucially failed to take account of one class of permissible incorporations - once these are recognized, the formal basis of his argument collapses.

Thus, Pullum & Gazdar (1982) demonstrate that every published argument purporting to show that one or another NL is not a CFL is invalid, either formally, or empirically, or both. Whether any NL, construed as a string set, falls outside the class of CFLs remains an open question, just as it was twenty five years ago.

III <u>CAPTURING SIGNIFICANT GENERALIZATIONS</u>

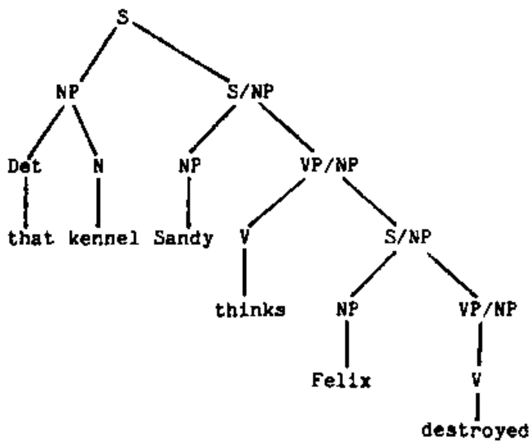
Argumentation purporting to show that CF-PSG's will miss significant generalizations about some NL phenomenon has been woefully inadequate. Typically it consists simply of providing or alluding to some CF-PSG which obviously misses the generalization in question. But, clearly, nothing whatever follows from such an exhibition. <u>Any</u> framework capable of handling some phenomenon at all will typically make available indefinitely many ugly analyses of the phenomenon. But this fact is neither surprising nor interesting. What is surprising, and rather disturbing, is that arguments of this kind (beginning, classically, in chapter 5 of Chomsky 1957) have been taken so seriously for so long.

Capturing significant generalizations is largely a matter of notation. But CF-PSG's, taken as a class of mathematical objects, have properties which are theirs independently of the notations that might be used to define them. Thus they determine a certain set of string sets, they determine a certain set of tree sets, they stand in particular equivalence relations, and so on. An analogy from logic is pertinent here: the truth function material implication Just is material implication whether you notate it with an arrow, or a hook, or the third letter of the alphabet, and whether you use prefix, infix, or postfix positioning of the symbol.

Over its 25 year history, transformational grammar developed a whole armoury of linguistically useful notations, and many of these can Just as well be used in characterizing CF-PSG's. Three such notational devices merit individual mention: (a) complex symbols, (b) rule schemata, and (c) mappings from one set of rules into another (metarules).

Harman (1963) deserves the credit for first seeing the potential of PSG's incorporating complex symbols. The use of a finite set of complex symbols, in place of the traditional finite set of monadic symbols, leaves the mathematical properties of CF-PSG's unchanged. Every CF-PSG employing complex symbols generates a tree set that is isomorphic to the tree set generated by some CF-PSG not employing complex symbols. However, complex symbol systems including features, "X-bar conventions", and "slash categories", etc., allow many significant syntactic generalizations to be captured rather straightforwardly.

For example, Gazdar & Pullum (1981) show how the use of complex symbols for subcategorization in a CF-PSG can capture generalizations that had had to be stipulated in the standard transformational account employing context-sensitive lexical insertion. And in Gazdar (1981a, 1981b) and Sag (1982a, 1982b) complex symbols called "slash categories" are shown to be able to capture the generalizations underlying the class of unbounded dependency constructions in English (e.g. relative clauses, wh-questions, topicalization, etc.). Figure 1 illustrates with a topicalization example:



Interestingly, such an analysis of unbounded dependencies is able to capture a generalization about their interaction with coordination that was never satisfactorily captured in transformational grammar (see Gazdar, Pullum, Sag & Wasow 1982).

Rule schemata allow generalizations to be captured by collapsing sets of rules with some common property into a single statement. In a CF-PSG, one can capture the familiar generalization that only like-constituents conjoin with a schema along the following lines:

```
α -> α and α
where α is any category.
```

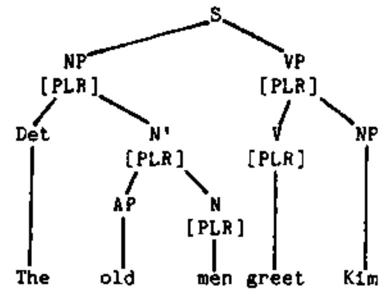
The generalization that this captures was not captured in classical TG: part of it was expressed in the base rules, and the rest was intendedly expressed in the formulation of a transformation called "Coordination Reduction".

Another example of the power of rule schemata (applied to complex symbols) involves agreement phenomena:

```
S -> NP[$] VP[$]
```

where $ ranges over permissible combinations of agreement features.

This schema, taken together with a widely assumed, and putatively univeral, convention of feature transfer (the Head Feature Convention, see Gazdar 1982) suffices to capture all the straightforward facts about subject verb agreement in English. Figure 2 illustrates:



Indeed, as Borsley (1983) and Stucky (1983, appendix) demonstrate, far more subtle and complex agreement phenomena succumb to analysis by phrase structure grammars equipped with features and rule schemata.

Like a rule schema, a metarule is a grammar characterization device (i.e. a clause in the definition of the grammar). But unlike a rule schema, it is one which enables one to define one set of rules in terms of another set, antecedently given. Generalizations which would be lost if the two sets of rules were merely listed are captured by the metarule [2].

For example, suppose that our grammar contains, inter alia, the following set of rules expanding VP:

```
VP -> V[1] NP
VP -> V[2] NP NP
VP -> V[3] NP PP
VP -> V[4] NP VP
VP -> V[5] NP S
VP -> V[6] NP NP S
```

Then we can augment the grammar by means of the following metarule:

```
VP -> V NP W        ==>

VP[PAS] -> V W (PP[by])
```

This says that for every rule in the grammar which expands VP as a verb followed by an NP possibly followed by other stuff (we don't care what), there is also to be a rule expanding a passive VP as the verb followed by the other stuff (if there was any) followed optionally by an agentive PP. This metarule will thus add the following rules to our grammar:

```
VP[PAS] -> V[1] (PP[by])
VP[PAS] -> V[2] NP (PP[by])
VP[PAS] -> V[3] PP (PP[by])
VP[PAS] -> V[4] VP (PP[by])
VP[PAS] -> V[5] S (PP[by])
VP[PAS] -> V[6] NP S (PP[by])
```

These rules will now allow the grammar to generate passive sentences directly.

Note that transformations were mappings from sets of structures to sets of structures, whereas metarules are mappings from sets of rules to sets of rules. Gazdar & Sag (1981) and Gunji (1983) show how metarules can capture reflexive pronoun generalizations in the definitions of PSG's for English and Japanese, respectively, and Gazdar, Pullum & Sag (1982) use metarules to provide accurate and nonredundant analyses for "Subject-auxiliary inversion", adverb placement, and VP ellipsis in a grammar of the English auxiliary system.

So-called "free word order" languages have sometimes been alleged to pose a problem in principle for the generalization-capturing powers of CF-PSG's. That they do not is amply demonstrated in horrocks (1963), Pullum (1982) and Stucky (1983) [3].

There is, thus, at the present time, no reason whatsoever to think that the goal of capturing linguistically significant generalizations is in any way inconsistent with the use of CF-PSG's.

IV_ SUPPORT OF A_ SEMANTICS

In asking whether some theory of syntax "supports a semantics", we are asking whether there exists some semantic theory which will interpret the structures provided by the syntax in a manner consistent with our judgments concerning ambiguity, synonymy, entailment, etc. Within linguistics, the key semantic development of the 1970's was the appearance of a model theory for natural languages originating in the work of Montague (see Montague 1974; Dowty et al. 1981). Before Montague, linguists had been disposed either to do their semantics in the syntax, a practice that reached its apogee in the work of the Generative Semanticists, or not at all, as in Chomsky's oeuvres. But the sophisticated machinery (including lambda abstraction, meaning postulates, higher order quantification, intension and extension operators, etc.) which Montague had made available meant that semantics could now be done as such. This had all kinds of implications for syntax.

Here are some examples, (i) Heny (1970) and Cooper (1975) showed that quantifier scope ambiguities could be handled entirely in the semantics, without the need for quantifier movement in the syntax. More recently, Cooper (1983) has provided a general semantic treatment of pronouns and quantifier binding which directly interprets a phrase structure syntax, (ii) Dowty (1978) showed that the semantic properties of dative, passive, unapeoified object, "Raising", and "Equi" constructions could all be provided for directly, without corresponding syntactic operations that moved or deleted NP's. (iii) Also in 1978, three sets of authors independently proposed closely related cross-categorial semantic theories for coordination (Cooper 1979; Gazdar 1980a; Keenan & Faltz 1978). This work, which is further advanced in Partee & Rooth (1983), completely undercuts all semantic motivation for "Coordination Reduction" transformations or equivalent operations in the syntax, (iv) McCloskey (1979) was able to show that the deep structure/surface structure distinction was irrelevant to giving a semantics for relative clauses and wh-questions: structures of either kind provide a suitable locus for semantic interpretation, (v) Klein (1980, 1981a, 1981b, 1982) has demonstrated that the meaning of all the various comparative constructions found in English can be derived directly from their surface syntactic forms as given by a CF-PSG along the lines of Gazdar (1980b). (vi) Wasow, Sag, & Nunberg (1982) argue that the proper treatment of idioms is one which semantically interprets them compositionally. Their analysis vitiates all the arguments for syntactic movement rules that depend on the claim that sentences like tabs were kept on every suspect cannot be compositionally interpreted. In a similar vein, Sag (1982c) and Sag 4 Klein (1982) have shown that the syntactic distribution of the English "dummy subjects" it and there (as in such sentences as *it* appears to be obvious that we lost, and there seems to have been a lion in the park) follow from an appropriate semantic theory of such expressions, without any need for raising operations in the syntax. (vii) And, in related work, Klein & Sag (1962) develop a semantic theory for phrase structure grammars for natural languages which largely eliminates the need to stipulate semantic translation rules. Instead, these rules follow from a function which takes as arguments (1) the semantic type of a lexical item, and (2) the phrase structure rule responsible for introducing the lexical item.

Thus there is every reason to believe that CF-PSG's can support appropriate semantic theories for NL's at least as well as multilevel syntactic frameworks [4, 5], Indeed, there is an ad hominem argument which suggests that they may be preferable. It is noteworthy that almost all the linguists currently working out the implications of Montague's semantic legacy (e.g. Bach, Cooper, Dahl, Dowty, Karttunen, Keenan, Klein, Partee, Peters, and Sag, to name but ten) have gravitated towards essentially monostratal syntactic theories although this concreteness is not obviously presaged in Montague's own, rather abstract, approach to syntax.

1 CONCLUSIONS

In brief, I have tried to make plausible the following three claims, (i) There is no reason, at the present time, to think that NL's are not CFL's. (ii) There are good reasons for thinking that the notations that we need to capture significant syntactic generalizations will characterize CF-PSG's, or some minor generalization of them, such as indexed grammars, (iii) There are very good reasons for believing that such grammars, and the monostratal representations they induce, provide the necessary basis for the semantic interpretation of NL's. And that, concomitantly, there is no semantic motivation for syntactic operations that move, delete, permute, copy or substitute constituents.

The relevance of these claims, which emerge from current theoretical linguistics, to the computational modelling of natural language processing is something which is argued for or assumed in an increasing body of recent work [6].

ACKNOWLEDGEMENTS

This research was supported by grant HR 7829/1 from the SSRC (UK). I am grateful to Aravind Joshi, Ewan Klein, Geoff Pullum, Ivan Sag, Henry Thompson, and two anonymous referees for some helpful suggestions. An ancestor of this paper was presented at an SERC workshop on natural language parsing organized by Karen Sparck-Jones and Yorick Wilks and held at the University of Essex in April 1982.

FOOTNOTES

1. Whether or not a CF-PSG can generate the correct structural descriptions for the relevant set of Dutch sentences is an intriguing open question. _If the grammatical set of trees is exactly the set that Bresnan, Kaplan, Peters & Zaenen (1982) assume it to be, then, as they show, the answer is no. But see, now, Pullum (1983) and Thompson (1983).

2. The idea of using one grammar to generate another originates in computer science with the work of van Wijngaarden (1969) who used the technique to give a perspicuous syntax for ALGOL68. A good introduction to his work can be found in Cleaveland & Uzgalis (1975). Janssen (1980) employs a van Wijngaarden-style two-level grammar to define a generalization of Montague's PTQ syntax. Recent work in computational linguistics that employs metarules includes Gawron et al. (1982), Kay (1983), Konolige (1980), Robinson (1980), Schubert and Pelletier (1982), Shieber, Stucky, Uszkoreit and Robinson (1983), and Thompson (1982).

3. The interesting parsing issues that "free word order" languages give rise to are addressed in Shieber (1983) and Uszkoreit (1983).

4. This is not to suggest, of course, that all significant semantic problems are solved by the shift to surface structure syntax. One currently interesting puzzle concerns how one binds pronouns that appear in "dislocated" constituents (Cooper 1983; Dahl 1981, 1982; Engdahl 1982a, 1982b). Another concerns the possibility of multiple wh-type dependencies in Scandinavian languages, and the variable-binding issue that this gives rise to (see Engdahl 1980; Maling & Zaenen 1982). One promising strategy for getting a solution to both these problems entails grounding the semantics on an indexed grammar (Aho 1968; Hopcroft & Ullman 1979, pp. 389-390) rather than on a CF-PSG. Indexed grammars are similar to CF-PSG's which employ complex symbols, except that there is no finite limit on the number of distinct complex symbols that can be used. This generalization of CF-PSG is potentially relevant to the issue mentioned in footnote 1, above, and to the nesting of equative and comparative clauses (see Klein 1981b). Cf., also, the tree adjunction defined MCSL's of Joshi (1982) (the MCSL's properly include the CFL's but are properly included by the indexed languages).

5. I assume that this conclusion would carry over to non-Montague approaches to semantics. However, I have restricted myself here to the Montague paradigm since it is by far and away the most detailed and extensive framework for NL semantics available at the present time. That Montague's ideas are compatible with a computational orientation to language is evidenced or argued in a large body of recent work: e.g., Bronnenberg et al. (1960), Friedman (1978, 1981), Friedman, Moran & Warren (1978), Friedman & Warren (1978), Fuchi (1981), Gunji & Sondheimer (1980), Hobbs & Rosenschein (1978), Indurkhya (1981), Ishimoto (1962), Janssen (1976, 1977, 1983), Landsbergen (1981, 1982), Matsumoto (1981, 1982), Moran (1980), Nishida et al. (1981), Nishida & Doshita (1962, 1983), Root (1981), Saheki (1982), Sawamura (1961), Sondheimer & Gunji (1978), and Warren (1979). See also Gunji (1981) for a description of the programming language EIL - Extended Intensional Lisp.

6. E.g., Bear & Karttunen (1979), Ejerhed (1980), Fodor (1982a, 1982b), Gawron et al. (1982), Joshi (1983), Joshi & Levy (1982), Kay (1963), Konolige (1980), Pullum (1983), Pulman (1983), Robinson (1980, 1982), Rosenschein & Shieber (1982), Ross (1981, 1982), Schubert (1982), Schubert & Pelletier (1982), Shieber (1983a, 1963b), Thompson (1981, 1982, 1983), and Uszkoreit (1983). And see Berwick & Weinberg (1982) for a rather extended metatheoretical disquisition.

REFERENCES

Aho, Alfred (1968) Indexed grammars - an extension of context-free grammars. Journal of the ACM 15, 647-671.

Bach, Emmon (1974) Syntactic Theory. New York: Holt, Rinehart & Winston.

Bar-Hillel, Y., and E. Shamir (1960) Finite state languages: formal representations and adequacy problems. Reprinted in Y. Bar-Hillel (1964) Language and information. Reading, Ma.: Addison-Wesley, 87-98.

Bear, John and Lauri Karttunen (1979) PSG: a simple phrase structure parser. Texas Linguistic Forum 15, 1-46.

Berwick, Robert and Amy Weinberg (1982) Parsing efficiency, computational complexity, and the evaluation of grammatical theories. Linguistic Inquiry 13, 165-191.

Borsley, Robert (1983) A Welsh agreement process and the status of VP and S. In Gerald Gazdar, Ewan Klein, and Geoffrey Pullum (eds.) Order, concord and constituency. Dordrecht: Foris, 57-74.

Bresnan, Joan (1978) A realistic transformational grammar. In M. Halle, J.W. Bresnan, and G.A. Miller (eds.) Linguistic theory and psychological reality. Cambridge, Ma.: MIT Press.

Bresnan, Joan, Ronald Kaplan, Stanley Peters and Annie Zaenen (1982) Cross-serial dependencies in Dutch. Linguistic Inquiry 13, 613-635.

Bronnenberg, W.J.H.C., H.C. Bunt, S.P.J. Landsbergen, R.J.H. Sena, W.J. Schoenmakers, and E.P.C. van Utteren (1980) The question-answering system PHLIQA 1. In L. Bole (ed.) Natural language question-answering systems. Munich: Carl Hanser Verlag.

Chomsky, Noam (1957) Syntactic structures. The Hague: Mouton.

Chomsky, Noam (1963) Formal properties of grammars. In R.D. Luce, R.R. Bush, and E. Galanter (eds.) Handbook of mathematical psychology, Volume II New York Wiley.

Cleaveland, J. and R. Uzgalis (1975) Grammars for programming languages: what every programmer should know about grammar. New York: Elsevier.

Cooper, Robin (1975) Montague's semantic theory and transformational syntax. PhD dissertation, University of Massachusetts at Amherst.

Cooper, Robin (1979) Model theory for a fragment of English. Unpublished manuscript, University of Wisconsin at Madison.

Cooper, Robin (1983) Quantification and syntactic theory. Dordrecht: Reidel.

Dahl, Osten (1981) Bound pronouns in dislocated constituents - the case for contextually interpreted variables. Mimeo, University of Stockholm.

Dahl, Osten (1982) Converting in bound pronouns: the problem and its solution. Mimeo, University of Stockholm.

Dowty, David (1978) Governed transformations as lexical rules in a Montague Grammar. Linguistic Inquiry 9, 393-426.

Dowty, David, Robert Wall, and Stanley Peters (1981) Introduction to Montague semantics. Dordrecht: Reidel.

EJerhed, Eva (1980) A context-free phrase-structure parser for English auxiliaries. Paper presented to the Fourth Groningen Round Table Conference on Auxiliaries, Groningen, Holland, July 1980.

Elster, J. (1978) Logic and society: contradictions and possible worlds. New York: Wiley.

Engdahl, Elisabet (1980) The syntax and semantics of questions in Swedish. PhD dissertation, University of Massachusetts at Amherst.

Engdahl, Elisabet (1982a) Constituent questions, topicalization, and surface structure interpretation. In D. Flickinger, M. Macken, and N. Wiegand (eds.) Proceedings from the First Annual West Coast Conference on Formal Linguistics. Stanford, 256-267.

Engdahl, Elisabet (1982b) A note on the use of lambda-conversion in generalized phrase structure grammars. Linguistics and Philosophy 4, 505-515.

Fodor, Janet (1982a) Phrase structure parsing and the island constraints. Linguistics and Philosophy, to appear.

Fodor, Janet (1982b) How can grammars help parsers? Paper presented at the Conference on Syntactic Theory and How People Parse Sentences, Ohio State University, May 14-15, 1982.

Friedman, Joyce (1978) Computational and theoretical studies in Montague Grammar at the University of Michigan. SISTM Quarterly 1, 62-66.

Friedman, Joyce (1981) Expressing logical formulas in natural language. In J. Groenendijk, T. Janssen, and M. Stokhof (eds.) Formal methods in the study of language. Mathematical Centre Tracts, Amsterdam, 113-130.

Friedman, Joyce, D. Moran, and D. Warren (1978) An interpretation system for Montague Grammar. American Journal of Computational Linguistics microfiche 74, 23-96.

Friedman, Joyce, and David Warren (1978) A parsing method for Montague Grammars. Linguistics and Philosophy 2, 347-372.

Fuchi, Kasuhiro (1981) Natural language and its formal representation: a case study of translation in Montague style from a programmer's point of view. Paper presented to the First Colloquium on Montague Grammar and Related Topics, Kyoto, February 1981.

Gawron, Jean Mark, Jonathan King, John Lamping, Egon Loebner, Anne Paulson, Geoffrey Pullum, Ivan Sag & Thomas Wasow (1982) The GPSG linguistics system. In Proceedings of the 20th Annual Meeting of the" Association for Computational Linguistics, 74-81. Menlo Park: ACL. Also distributed as Hewlett Packard Computer Science Technical Note CSL-82-5.

Gazdar, Gerald (1980a) A cross-categorial semantics for coordination. Linguistics and Philosophy 3, 407-409.

Gazdar, Gerald (1980b) A phrase structure syntax for comparative clauses. In T. Hoekstra, H.v.d. Hulst, and M. Moortgat (eds.) Lexical grammar. Foris Publications, Dordrecht, 165-179.

Gazdar, Gerald (1981a) Unbounded dependencies and coordinate structure. Linguistic Inquiry 12, 155-184.

Gazdar, Gerald (1981b) On syntactic categories. Philosophical Transactions (Series *B) of* the Royal Society 295, 267-283.

Gazdar, Gerald (1982) Phrase structure grammar. In Pauline Jacobson and Geoffrey Pullum (eds.) The nature of syntactic representation. D. Reidel, Dordrecht, 131-186.

Gazdar, Gerald, and Geoffrey Pullum (1981) Subcategorization, constituent order and the notion "head". In M. Moortgat, H.v.d. Hulst and T. Hoekstra (eds.) The scope of lexical rules. Foris Publications, Dordrecht, 107-123.

Gazdar, Gerald, and Ivan Sag (1981) Passive and reflexives in phrase structure grammar. In J. Groenendijk, T. Janssen, and M. Stokhof (eds.) Formal methods in the study of language. Mathematical Centre Tracts, Amsterdam, 131-152.

Gazdar, Gerald, Geoffrey Pullum, and Ivan Sag (1982) Auxiliaries and related phenomena in a restrictive theory of grammar. Language 58, 591-638.

Gazdar, Gerald, Geoffrey Pullum, Ivan Sag, and Tom Wasow (1982) Coordination and transformational grammar. Linguistic Inquiry 13, 663-676.

Graham, Susan, Michael Harrison, and Walter Ruzzo (1980) An improved context-free recognizer. ACM Transactions on Programming Languages and Systems 2,415-462.

Grinder, John, and Suzette Elgin (1973) Guide to transformational grammar. New York! Holt, Rinehart and Winston.

Gunji, Takao (1981) Toward a computational theory of pragmatics - discourse, presupposition, and, implicature. PhD dissertation, Ohio State University.

Gunji, Takao (1983) Generalized phrase structure grammar and Japanese reflexivization. Linguistics and Philosophy 6, 115-156.

Gunji, Takao, and Norman Sondheimer (1980) The mutual relevance of model-theoretic semantics and artificial intelligence. SMIL: Journal of Linguistic Calculus 3, 5-42.

Harman, Gilbert (1963) Generative grammars without transformation rules: a defense of phrase structure. Language 39, 597-616.

Heny, Frank (1970) Semantic operations on base structures. PhD dissertation, UCLA.

Hobbs, Jerry, and Stanley Rosenschein (1978) Making computational sense of Montague's intensional logic. Artificial Intelligence 9, 287-306.

Hopcroft, John, and Jeffrey Ullman (1979) Introduction to automata theory, languages, and computation. Reading, Ma.: Addison-Wesley.

Horrocks, Geoffrey (1983) The order of constituents in Modern Greek. In Gerald Gazdar, Ewan Klein, and Geoffrey Pullum (eds.) Order, concord and constituency. Dordrecht: Foris, 95-111.

Huybregts, M.A.C. (1976) Overlapping dependencies in Dutch. Utrecht Working Papers in Linguistics 1, 24-65.

Indurkhya, B. (1981) Sentence analysis programs based on Montague grammar. M.E.E. thesis, Netherlands Universities Foundation for International Cooperation.

Ishimoto, I. (1982) A Lesniewskian version of Montague grammar. In J. Horecky (ed.) COLING 82. Amsterdam: North Holland, 139-144.

Janssen, Theo (1976) A' computer program for Montague Grammar: theoretical aspects and proofs for the reduction rules. Amsterdam Papers in Formal Grammar I, 154-176.

Janssen, Theo (1977) Simulation of a Montague Grammar. Annals of System Research 6, 127-140.

Janssen, Theo (1980) On problems concerning the quantification rules in Montague grammar. In C. Rohrer (ed.) Time, tense, and quantifiers. Tubingen: Max Niemeyer, 113—134-

Janssen, Theo (1983) Foundations and applications of Montague Grammar. Amsterdam: Mathematisch Centrum.

Joshi, Aravind (1982) How much context-sensitivity is required, if any, for assigning reasonable structural descriptions? Paper presented at the Conference on Syntactic Theory and How People Parse Sentences, Ohio State University, May 14-15, 1982.

Joshi, Aravind (1983) Factoring recursion and dependencies: an aspect of tree-adjoining grammars (TAG) and a comparison of some formal properties of TAGs, GPSGs, PLGs, and LFGs. In Proceedings of the 21st Annual Meeting of the Association for Computational Linguistics. Menlo Park: ACL"

Joshi, Aravind, and Leon Levy (1982) Phrase structure trees bear more fruit than you would have thought. American Journal of Computational Linguistics 8, 1-11.

Kay, Martin (1983) When meta-rules are not meta-rules. In K. Sparck-Jones and Y. Wilks (eds.) Automatic natural language parsing. Chichester: Ellis Horwood.

Keenan, Edward, and Leonard Faltz (1978) Logical types for natural language. UCLA Occasional Papers in Linguistics _3.

Klein, Ewan (1980) A semantics for positive and comparative adjectives. Linguistics and Philosophy 4, 1-45.

Klein, Ewan (1981a) The interpretation of adjectival, nominal, and adverbial comparatives. In J. Groenendijk, T. Janssen, and M. Stokhof (eds.) Formal methods in the study of language. Mathematical Centre Tracts, Amsterdam, 381-398.

Klein, Ewan (1981b) The syntax and semantics of nominal comparatives In M. Moneglia (ed.) Atti de seminario su tempo e verbale strutture quantificate in forma loglca. Presso l'Accademia della Crusca, Florence, 223-253.

Klein, Ewan (1982) The interpretation of adjectival comparatives. Journal of Linguistics 18, 113-136.

Klein, Ewan, and Ivan Sag (1982) Semantic type and control. In M. Barlow, D. Flickinger & I.A. Sag (eds.) Developments in Generalized Phrase Structure Grammar; Stanford Working Papers in Grammatical Theory, Volume 2. Bloomington: Indiana University Linguistics Club, 1-25. Also: to appear in Linguistics and Philosophy.

Konolige, Kurt (1980) Capturing linguistic generalizations with metarules in an annotated phrase-structure grammar. In Proceedings of the 18th Annual Meeting of the Association for Computational Linguistics, 43-48. Menlo Park: ACL

Landsbergen, Jan (1981) Adaptation of Montague grammar to the requirements of parsing. In J. Groenendijk, T. Janssen, and M. Stokhof (eds.) Formal methods in the study of language. Mathematical Centre Tracts, Amsterdam, 399-419.

Landsbergen, Jan (1982) Machine translation based on logically isomorphic Montague grammars. In J. Horecky (ed.) COLING 82. Amsterdam: North Holland, 175-182.

Langendoen, D. Terence (1977) On the inadequacy of type-2 and type-3 grammars for human languages. In P.J. Hopper (ed.) Studies in descriptive and historical linguistics. Amsterdam: John Benjamin, 159-171.

Maling, Joan, and Annie Zaenen (1982) A phrase structure account of Scandinavian extraction phenomena. In Pauline Jacobson and Geoffrey Pullum (eds.) The nature of syntactic representation. D. Reidel, Dordrecht, 229-282.

Matsumoto, Yuji (1981) Software implementation of Montague grammar and related problems. In Shogo Iguchi (ed.) Formal approaches to natural language: Proceedings of the First Colloquium on Montague Grammar and Related Topics. Kyoto: Kyoto Working Group of Montague Grammar, 148-158.

Matsumoto, Yuji (1982) A Montague grammar of Japanese with special regard to meaning adjustment. Paper presented to the Second Colloquium on Montague Grammar and Related Topics, Kyoto, March 1982.

McCloskey, James (1979) Transformational syntax and model theoretic semantics. Dordrecht: Reidel.

Montague, Richard (1974) Formal philosophy. New Haven: Yale University Press.

Moran, Douglas B. (1980) Dynamic partial models. PhD dissertation, University of Michigan.

Nishida, Toyo-Aki, Masaki Kiyono, and Shuji Doshita (1981) An English-Japanese machine translation system based on formal semantics of natural language. In Shogo Iguchi (ed.) Formal approaches to natural language: Proceedings of the First* Colloquium on Montague Grammar and Related Topics. Kyoto: Kyoto Working Group of Montague Grammar, 104-147.

Nishida, Toyo-Aki, and Shuji Doshita (1982) An English-Japanese machine translation system based on formal semantics of natural language - a progress report. In J. Horecky (ed.) COLING 82. Amsterdam: North Holland, 277-282.

Nishida, Toyo-Aki and Shuji Doshita (1983) An application of Montague Grammar to English-Japanese machine translation. In Proceedings, Conference on Applied NaturaT Language Processing, 1-3 February 1983, Santa Monica, California. ~~Menlo Park: ACL.

Partee, Barbara, and Mats Rooth (1983) Generalized conjunction and type ambiguity. To appear in R. Bauerle, C. Schwarze and A.v. Stechow (eds.) Meaning, use and interpretation of language. Berlin: de Gruyter.

Postal, Paul (1964) Limitations of phrase structure grammars. In J.A. Fodor & J.J. Katz (eds.) The structure of language: readings in the philosophy of language. Englewood Cliffs: Prentice-Hall, 137-151.

Pullum, Geoffrey (1982) Free word order and phrase structure rules. In James Pustejovsky and Peter Sells (eds.) Proceedings of the Twelfth Annual Meeting of the North Eastern Linguistic Society. Graduate Linguistics Student Association, University of Massachusetts, Amherst, Mass., 209-220.

Pullum, Geoffrey (1983) Context-freeness and the computer processing of human languages. In Proceedings of the 21st Annual Meeting of the Association for ComputationaT Linguistics. Menlo Park: ACL.

Pullum, Geoffrey and Gerald Gazdar (1982) Natural languages and context free languages. Linguistics and Philosophy 4, 471-504.

Pulman, Stephen (1983) Generalised phrase structure grammar, Earley's algorithm, and the minimisation of recursion. In K. Sparck-Jones and ¥. Wilks (eds.) Automatic natural language parsing. Chichester: Ellis Horwood.

Robinson, Jane (1980) Computational aspects of the use of metarules in formal grammars. Research Proposal No. ECU 80-126, S.R.I. International.

Robinson, Jane (1982) DIAGRAM: a grammar for dialogs. Communications of the ACM 25, 27-47.

Root, Rebecca (1981) SMX: a program for translating English into Montague's intensional logic. Unpublished manuscript, University of Texas at Austin.

Rosenschein, S.J., and S.M. Shieber (1982) Translating English into logical form. In Proceedings of the 20th Annual Meeting of the Association for Computational Linguistics. Menlo Park: ACL——

Ross, Kenneth (1981) Parsing English phrase structure. PhD dissertation, University of Massachusetts at Amherst.

Ross, Kenneth (1982) An improved left-corner parsing algorithm. In J. Horecky (ed.) COLING 82. Amsterdam: North Holland, 333-338.

Sag, Ivan (1982a) Coordination, extraction, and generalized phrase structure. Linguistic Inquiry 13, 329-336.

Sag, Ivan (1982b) On parasitic gaps. In D. Flickinger, M. Macken, and N. Wiegand (eds.) Proceedings from the First Annual West Coast Conference on Formal Linguistics. Stanford, 35-46. Also to appear in Linguistics and Philosophy.

Sag, Ivan (1982c) A semantic theory of "NP-movement" dependencies. In Pauline Jacobson and Geoffrey Pullum (eds.) The nature of syntactic representation. D. Reidel, Dordrecht, 427-466.

Sag, Ivan, and Ewan Klein (1982) The syntax and semantics of English expletive pronoun constructions. In M. Barlow, D. Flickinger & I.A. Sag (eds.) Developments in Generalized Phrase Structure Grammar: Stanford Working Papers in Grammatical Theory, Volume 2. Bloomington: Indiana University Linguistics Club, 92-136

Saheki, Motoji (1982) A software program for a language like natural language. Paper presented to the Second Colloquium on Montague Grammar and Related Topics, Kyoto, March 1982.

Sawamura, Hajime (1981) Intensional logic as a basis of algorithmic logic. Paper presented to the First Colloquium on Montague Grammar and Related Topics, Kyoto, February 1981.

Schubert, Lenhart (1982) An approach to the syntax and semantics of affixes in 'conventionalized' phrase structure grammar. In Proceedings of the 4th Biennial Conference of the Canadian Society for Computational Studies of Intelligence, T59-195.

Schubert, Lenhart, and Jeffry Pelletier (1982) From English to logic: context-free computation of "conventional" logical translations. American Journal of Computational Linguistics 8, 27-44.

Shieber, Stuart (1983a) Direct parsing of ID/LP grammars. MS, SRI International.

Shieber, Stuart (1983b) Sentence disambiguation by a shift-reduce parsing technique. In Proceedings of the 21st Annual Meeting of the Association for Computational Linguistics. Menlo Park: ACL.

Shieber, Stuart, Susan Stucky, Hans Uszkoreit, and Jane Robinson (1983) Formal constraints on metarules. In Proceedings of the 21st Annual Meeting of the" Association for Computational Linguistics. Menlo Park: ACL.

Sondheimer, Norman, and Takao Gunji (1978) Applying model-theoretic semantics to natural language understanding: representation and question-answering. In Proceedings of the Seventh International Conference on Computational Linguistics, Bergen.

Stucky, Susan (1983) Verb phrase constituency and linear order in Makua. In Gerald Gazdar, Ewan Klein, and Geoffrey Pullum (eds.) Order, concord and constituency. Dordrecht: Foris, 75-94.

Thompson, Henry (1981) Chart parsing and rule schemata in PSG. In Proceedings of the 19th Annual Meeting of the" Association for Computational Linguistics, 167-172. Menlo Park: ACL.

Thompson, Henry (1982) Handling metarules in a parser for GPSG. Edinburgh D.A.I. Research Paper No. 175. Also: In M "Barlow, D. Flickinger & I.A. Sag (eds.) Developments in Generalized Phrase Structure Grammar: Stanford Working Papers in Grammatical Theory, Volume 2. Bloomington: Indiana University Linguistics Club, 26-37.

Thompson, Henry (1983) Crossed serial dependencies: a low-power parseable extension to GPSG. In Proceedings of the 21st Annual Meeting of the Association for Computational Linguistics. Menlo Park: ACL.

Uszkoreit, Hans (1983) A framework for processing partially free word order. In Proceedings of the 21st Annual Meeting of the Association for Computational Linguistics. Menlo Park: ACL.

van Wijngaarden, A. (1969) Report on the algorithmic language ALGOL68. Numerische Mathematik 14, 79-218.

Warren, David (1979) Syntax and semantics in parsing: an application to Montague Grammar. PhD dissertation, University of Michigan.

Wasow, Tom, Ivan Sag, and Geoffrey Nunberg (1982) Idioms: an interim report. In Preprints of the Plenary Session Papers, The XIIIth International Congress of Linguists, August 29 - September 4, 1982. TokyoTrm:

Welin, Carl Wilhelm (1979) Studies in computational text comprehension. Stockholm: Institute of Linguistics.

Formal Theories of Language Acquisition: Practical and Theoretical Perspectives

Daniel N. Osherson
Center for Cognitive Science
M I T
Cambridge, MA 02139 USA

Michael Stob
Department of Mathematics
Calvin College
Grand Rapids, MI 49506 USA

Scott Weinstein
Department of Philosophy
University of Pennsylvania
Philadelphia, PA 19104 USA

ABSTRACT

Learning Theory is the study of systems that implement functions from evidential states to theories. The theoretical framework developed in the theory makes possible the comparison of classes of algorithms which embody distinct learning strategies along a variety of dimensions. Such comparisons yield valuable information to those concerned with inference problems in Cognitive Science and Artificial Intelligence. The present paper employs the framework of Learning Theory to study the design specifications of inductive systems which are of interest in the domain of language acquisition.

Section 0: Introduction

Learning Theory is tho investigation of systems that implement functions from evidential states to theories. Of central concern is the characterization of conditions under which such functions stabilize to accurate theories of a given environment. Within the theory, the informal notions of "evidence," "theory," "stabilization," "accuracy," and "environment" are replaced by precise definitions. Alternative formulations of these concepts yield alternative models within the theory. The vigorous development of Learning Theory began with a celebrated paper by Gold (1967). Angluin & Smith (1982) provide a valuable review of formal results.

Learning Theory is motivated by both scientific and technological concerns. Scientifically, the theory has proved useful in the analysis of human learning, particularly, language acquisition (see Osherson, Stob & Weinstein, forthcoming, for a review of issues). Technologically, the theory helps specify what is learnable 1n principle, and may thus guide the construction of practical systems of inductive inference.

Learning Theory yields potentially valuable insights about problems of inductive inference in the context of Cognitive Science and Artificial Intelligence. The theory provides the framework for systematic comparison of various learning algorithms. Such comparisons are particularly useful in determining the relative strength of classes of algorithms which embody distinct abstract learning strategies, in assessing their resource requirements, and in predicting their behavior in various environments. When combined with empirical studies of language acquisition, Learning Theory may provide constraints on the character of the learning strategies implemented by children, and reflect in turn on the character of the class of languages which may be acquired. Such studies in Cognitive Science may be of importance to system builders in Artificial Intelligence. They suggest that the search for ideal learning strategies' is not well motivated. Rather, by focussing on learners who embody different "styles" of learning, and by investigating their properties, the theory allows a comparison of the optimality of distinct approaches to learning along a multitude of dimensions. In addition, through the analysis of classes of algorithms that embody distinct learning strategies, this theoretical framework may provide a useful complement to studies of *ad hoc* systems built to perform inductive inference in problem domains of limited scope.

The present paper reviews some of our recent work on practical inference and relates it to problems in language acquisition. We consider design specifications for Inductive systems relevant to (a) the speed of Inference, (b) the simplicity of inferred theories, (c) the likelihood of inferential success, and (d) the resilience of such systems in environments subject to informational imperfection. Attention is restricted to learning paradigms in which only "positive information" is available about the language or data-set to be inferred; direct information about nonmembership is not offored to the learner; Angluin & Smith (1982) survey results relevant to learning paradigms in which both positive and negative information is assumed available.

Our exposition is organized as follows. The next section provides definitions and construals at the heart of contemporary learning theory. Section 2 exhibits theorems proper to the topics (a) - (d) listed above. Proofs of these theorems can be found in Osherson, Stob & Weinstein (1982, 1983d). In Section 3 we consider the relation between results reviewed here and language acquisition by children.

Section 1: The Gold model

1.1 Sequences, languages, texts

N is the set of natural numbers. We take the notions *finite sequence* (in N) and *infinite sequence* (in N) to be basic. The set of all finite sequences is denoted: SEQ. For $n \in N$, and infinite sequence, t: rng(t) is the set of numbers appearing in t; t_n is the nth member of t; and $\bar{t}_n$ is the finite sequence of length n in t.

Let $\varphi_0, \varphi_1, \ldots, \varphi_i, \ldots$ be a fixed list of all partial recursive functions of one variable, and assume the list to be *acceptable* in the sense of Rogers (1967, Ch. 2). For $i \in N$, let W_i = domain φ_i, the recursively enumerable subset of N with *index* i. *Languages* are identified with nonempty members of $\{W_i \mid i \in N\}$. The collection of all languages is denoted: $\mathcal{RE}$. For $L \in \mathcal{RE}$, $i \in N$, if $L = W_i$, then i is said to be for L.

A *text for* $L \in \mathcal{RE}$, is any infinite sequence such that rng(t) = L. The class of all texts for L is denoted: T_L. Given a collection, $\mathcal{L}$, of languages, "$T_{\mathcal{L}}$" denotes $\bigcup_{L \in \mathcal{L}} T_L$, the class of all texts for languages in $\mathcal{L}$.

1.2 Learning functions

Let G be a fixed, computable isomorphism between SEQ and N. A *learning function* is any function from N into M; such a function will be thought of as operating on members of SEQ (via G), yielding indices for recursively enumerable sets. Learning functions may be total or partial, recursive or nonrecursive. The (partial) recursive learning functions are just $\Phi, \Phi 1, \ldots,$ $\Phi. \ldots$ The class of all learning functions is denoted: F. The class of all recursive learning functions (partial or total) is denoted: F^{rec}.

1.3 Convergence, identification

Let $f \in \mathcal{F}$, $t \in T_{\mathcal{RE}}$, $i \in N$. We say that f *converges to* i *on* t just in case (a) $f(\sigma)$ is defined for all σ in t, and (b) for all but finitely many $n \in N$, $f(\bar{t}_n) = i$. Intuitively, f converges to i on t just in case (a) f never becomes "stuck" in examining ever longer finite sequences in t, and (b) f eventually conjectures i, and never departs from it thereafter.

Let $f \in \mathcal{F}$, $t \in T_{\mathcal{RE}}$. f is said to *identify* t just in case f converges to an index, i, on t such that W_i = rng(t). f *identifies* $L \in \mathcal{RE}$ just in case f identifies every $t \in T_L$. f *identifies* $\mathcal{L} \subseteq \mathcal{RE}$ just in case f identifies every $t \in T_{\mathcal{L}}$; in this case $\mathcal{L}$ is said to be *identifiable.*

Gold's Theorem (1967): Let $\mathcal{L}$ include all finite languages and any infinite language. Then $\mathcal{L}$ is not identifiable.

Section 2: Practical learning

2.1 Efficient inference

Useful learning must not take too much time. This vague admonition can be resolved into two demands: (i) the learner must not examine too many inputs before settling for good on a correct hypothesis, and (11) the learner must not spend too long examining each input. Learners satisfying (i) will be called "text-efficient;" learners satisfying

(ii) will be called "time-efficient;" learners satisfying both (i) and (ii) will be called "efficient." In this section these requirements are formulated precisely and examined for their impact on identifiability.

2.1.1 Text-efficiency

Following Gold (1967, Section 10), we define the partial functional CONV:$\mathcal{F} \times T_{\mathcal{RE}} \rightarrow N$ as follows. For all $f \in \mathcal{F}$, $t \in T_{\mathcal{RE}}$.

$$\text{CONV}(f,t) = \mu n[(\forall m > n)(f(\bar{t}_m) = f(\bar{t}_{m-1}))].$$

CONV is defined on $f \in \mathcal{F}$, $t \in T_{\mathcal{RE}}$ if f converges on t, in which case CONV(f,t) is the length of the smallest $\sigma \in$ SEQ in t such that f's last revised conjecture in t is made on σ.

Now let $F \in \mathcal{F}$, $\mathcal{L} \subseteq \mathcal{RE}$. We say that f identifies $\mathcal{L}$ *text efficiently* just in case f identifies $\mathcal{L}$, and for all $g \in \mathcal{F}$ that identify $\mathcal{L}$,

(*) if $(\exists t \in T_{\mathcal{L}})(\text{CONV}(g,t) < \text{CONV}(f,t))$, then $(\exists s \in T_{\mathcal{L}})(\text{CONV}(f,s) < \text{CONV}(g,s))$.

f is said to identify $\mathcal{L}$ *text efficiently with respect to* $\mathcal{F}^{rec}$ just in case f identifies $\mathcal{L}$, and for all $g \in \mathcal{F}^{rec}$ that identify $\mathcal{L}$, (*) holds.

Intuitively, f identifies $\mathcal{L}$ text-efficiently [with respect to $\mathcal{F}^{rec}$] just in case no other [recursive] learning function that identifies $\mathcal{L}$ is strictly faster than f in terms of convergence delay. This notion of text efficiency yields:

Proposition 1: A collection, $\mathcal{L}$, of languages is identifiable if and only if some $f \in \mathcal{F}$ identifies $\mathcal{L}$ text efficiently.

Proposition 1 shows that text efficiency is not a restrictive design feature relative to the class of all learning functions. In contrast, the next proposition shows that text efficiency is restrictive relative to the class of recursive learning functions.

Proposition 2: There is some $\mathcal{L} \subseteq \mathcal{RE}$ such that (i) some $f \in \mathcal{F}^{rec}$ identifies $\mathcal{L}$, but (ii) no $f \in \mathcal{F}^{rec}$ identifies $\mathcal{L}$ text efficiently with respect to $\mathcal{F}^{rec}$.

2.1.2 Time efficiency

A learner is time efficient if it reacts quickly to new inputs. We formalize this notion only for the case of recursive learning functions. A computational complexity measure in the sense of Blum (1967a) is imposed upon our acceptable ordering of partial recursive functions. The measure justifies reference to the number of "steps" required for φ_i to halt on j $(i,j \in N)$. For $i,s,k \in N$, we let $\varphi_{i,s}(k)$ be the output (if any) after s steps in the computation of $\varphi_i(k)$; "$|\varphi_i(k)|$" denotes $\mu s[\varphi_{i,s}(k)$ halts].

Now let $\mathcal{L} \subseteq \mathcal{RE}$, let $i \in N$, and let $h: N \rightarrow N$ be a total recursive function. We say that φ_i identifies $\mathcal{L}$ *h-time efficiently* just in case (i) φ_i identifies $\mathcal{L}$, and (ii) for every $t \in T_{\mathcal{L}}$, there is a $k \in N$ such that for all $j > k$:

$$|\varphi_i(\bar{t}_{j+1})| \leq |\varphi_i(\bar{t}_j)| + h(t_{j+1}).$$

Intuitively, φ_i identifies $\mathcal{L}$ h-time efficiently if for every $t \in T_{\mathcal{L}}$, φ_i eventually takes no more than $h(t_{n+1})$ additional steps to respond to $\bar{t}_{n+1}$ than to respond to $\bar{t}_n$: that is, except for a constant, the growth in φ_i's response time to ever longer initial segments of t is bounded by h. h-time efficiency turns out not to restrict the classes of languages identifiable by recursive learning function. This is the burden of the next proposition.

Proposition 3: There is a recursive function, h, such that for all $\mathcal{L} \subseteq \mathcal{RE}$, some $f \in \mathcal{F}^{rec}$ identifies $\mathcal{L}$ if and only if some $\varphi_i \in \mathcal{F}^{rec}$ identifies $\mathcal{L}$ h-time efficiently.

Indeed, any h such that $h(x) \geq x$ almost everywhere can be chosen in Proposition 3.

2.1.3 Efficiency

Let *I* C RE>6, and let h be a total recursive function. Φ_i is said to identify L *h-efficientlyy* just in case (i) Φi identifies L text efficiently with respect to F^{rec} (in particular, *Φi* identifies L*)*, and (ii) Φj is h-time efficient for L. h-efficiency, that is. combines the virtues of text efficiency (with respect to F^{rec}) and h-time efficiency. The next proposition shows that, for any h, h-efficiency is more restrictive than text efficiency as a design feature of recursive learning functions.

Proposition 4: For every total recursive function, h, there is a collection, L*,* of languages such that (i) some f € F^{rec} identifies L text efficiently, but (ii) no Φ € Frec identifies L h-efficiently.

Proposition 5: For some total recursive function, h, there is L C RE such that (i) some f € F^{rec} identifies L text efficiently, (ii) some Φ_i identifies *L* h-time efficiently, but (iii) no Φ_j identifies L h-efficiently.

2.2 Simple conjectures

To be useful, a learner should not only converge rapidly to a correct theory of its environment, it should also converge to a relatively simple theory: excessively complex theories, even if true, are of little practical use. To study the impact of such simplicity constraints on learnability, a total recursive size measure, S:N -♦ N, is now imposed on our acceptable ordering of partial recursive functions. Intuitively, S may be conceived as mapping indices to sizes, S(i) being the length of the program for *Φ1* corresponding to index i. The measure is governed by the following two axioms, due to Blum (1967b).

Axiom 1: For all i € N, there are only finitely many j € N such that S(j) ■ i.

Axiom 2: The predicate "j € $S^{-1}(1)$," for i_fj € N, is decidable.

Define the function MS:°RE -> N as follows. For L € RE, RE(L) uJ[(3k)(W_k = L & S(k) = j]. Thus, MS(L) is the size of the smallest program that accepts L. Concern about simple conjectures may take the following form. Let g be a total recursive function, let f € F, and let L C RE. f is said to identify L *g-s1mply* just in case f identifies *L,* and for all t € T_L, f converges on t to an index, j, such that S(j) < g(MS(rng(t)). To exemplify, let g be \x.2x. Then f identifies *L* g-simply just in case f identifies L*,* and for all L € L and t € T_L, f converges on t to an index of size no greater than twice MS(L) (the size of the smallest program that accepts L * rng(t)).

Text efficiency and g-simplicity are more restrictive design features of recursive learning functions than either is alone. This is the content of the next proposition.

Proposition 6: There is *L* C RE such that (i) some f € F^{rec} identifies L text efficiently, (ii) for any total recursive function, g, such that g(x) > x for all x € N, some f € F^{rec} identifies L g-simply, but (iii) for every f € F^{rec}, and every total recursive function, h, if f identifies *L* text efficiently with respect to F^{rec}, then f does not identify L h-simply.

2.3 Learning in likely environments'[1]

In some environments each potential element of a language is associated with a fixed probability of occurrence, invariant through time. Such environments may be thought of as infinite sequences of stochastically independent events, the probability of a given element, e, appearing in the n+lst position being independent of the contents of positions 0 through n.

To study such environments, each L € RE is associated with a probability measure* m, on N such that for all x € N, x € L if and only 1f $m_L(\{x\})$ > 0. (Recall that every L € RE 1s nonempty; see Section 1.1.) Next, we impose on ^{TRE} its Baire topology, RE; that 1s, for each *a* € SEQ, we take B_g ■ {t € ^{T}6bSI^a is in T to basic open set of F RE. For each L € R£, we define the (unique) complete probability measure, M_L, on $\mathcal{T}_{\mathcal{RE}}$ by stipulating that for all $\sigma \in$ SEQ, $M_L(B_\sigma) = \Pi_{j < lh(\sigma)} m_L(\sigma_j)$. We now assume the existence of a fixed collection, $\{M_L | L \in \mathcal{RE}\}$ of measures on correponding members of $\{T_L | L \in \mathcal{RE}\}$. Intuitively, for measurable $S \subseteq T_L$, $M_L(S)$ is the probability that an arbitrarily selected text for L is drawn from S.

The following facts are easy to establish. For all $L, L' \in \mathcal{RE}$,

(I) if $L \neq L'$, then $M_L(T_{L'}) = 0$;

(II) for $f \in \mathcal{F}$, $M_L(\{t \in T_L | f$ identifies $t\})$ is defined.

In the stochastic context just discussed, the Gold definition of language identification seems needlessly restrictive. Rather than requiring identification of every text for a given language, L, it seems enough to require identification of any subset of T_L of sufficient probability. We are thus led to the following definition. Let $f \in \mathcal{F}$, $L \in \mathcal{RE}$. f is said to *measure-one identify* L just in case $M_L(\{t \in T_L | f$ identifies $t\}) = 1$. f *measure-one identifies* $\mathcal{L} \subseteq \mathcal{RE}$ just in case f measure-one identifies every $L \in \mathcal{L}$; in this case, $\mathcal{L}$ is said to be *measure-one identifiable*. The definition of measure-one identifiability is inspired by Wexler & Culicover (1980, Ch. 3).

Measure-one identification of a language differs from ordinary identification only by a set of measure zero. The next proposition reveals the significance of this small difference.

Proposition 7: $\mathcal{RE}$ is measure-one identifiable.

Let $\mathcal{L} = \{L_i | i \in N\}$ be an indexed collection of languages, and let $\{p_i | i \in N\}$ be the corresponding measures on them. $\mathcal{L}$ is said to be *uniformly measured* just in case the predicates "$x \in L_y$" and "$p_x(\{y\}) = z$" are decidable (the decidability of the latter predicate actually implies that of the former). Minor modifications in the proof of Proposition 7 yield the following.

Proposition 8: Let $\mathcal{L}$ be a uniformly measured collection of languages. Then, some $f \in \mathcal{F}^{rec}$ measure-one identifies $\mathcal{L}$.

Thus, in contrast to Gold's Theorem (Section 1.3, above), any uniformly measured collection of languages consisting of all finite sets and any infinite set is measure-one identifiable.

2.4 Imperfect environments

2.4.1 Noisy texts

A *noisy text* for a language, L, is any text for a language of the form $L \cup D$, where D is a finite set. Thus, a noisy text for a language, L, can be pictured as a text for L into which any number of intrusions from a finite set have been inserted. Since the empty set is finite, texts for L count as noisy texts for L. We say that a learning function, f, *identifies a language*, L, *on noisy text* just in case f converges to an index for L on every noisy text for L. A learning function, f, *identifies a collection*, $\mathcal{L}$, *of languages on noisy text* just in case f identifies every language in $\mathcal{L}$ on noisy text.

It is clear that noisy text renders impossible the identification of the collection of all finite languages. The following proposition provides a less obvious example of the disruptive effects of such environments for recursive learning functions.

Proposition 9: There is a collection, $\mathcal{L}$, of languages such that (a) every language in $\mathcal{L}$ is infinite and disjoint from every other language in $\mathcal{L}$, (b) some recursive learning function identifies $\mathcal{L}$, and (c) no recursive learning function identifies $\mathcal{L}$ on noisy text.

2.4.2 Incomplete texts

An *incomplete text* for a language, L, is defined to be a text for L-D, where D is any finite set. An incomplete text for a language, L, can be pictured as a text for L from which all occurrences of a given finite set of sentences have been removed. Texts for L count as incomplete texts for L. We say that a learning function, f. *identifies a language,* L, *on incomplete* text just in case f converges to an index for L on every incomplete text for L. Identifiability of collections of languages on incomplete text is defined straightforwardly.

Proposition 10: There is a collection, L, of languages such that (a) every language in L is infinite and disjoint from every other language in *L,* (b) some recursive learning function identifies 1, and (c) no recursive learning function identifies *L* on incomplete text.

2.4.3 Noisy and incomplete text compared

Let $\mathcal{C}_{\mathcal{N}}$ be the family of all collections, *L*, of languages such that *L* can be identified (by arbitrary learning function) on noisy text. Define $\mathcal{C}_{\mathcal{I}}$ similarly with respect to incomplete text. We have:

Proposition 11: **$\mathcal{C}_{\mathcal{N}}$ is a proper subset of $\mathcal{C}_{\mathcal{I}}$.**

2.4.4 Finite-difference identification on imperfect text

Let L, L' $\in \mathfrak{RE}$. L is said to be a *finite variant* of L' just in case (L - L') $\cup$ (L - L') is finite. A learning function, f, is said to *finite-difference identify a language,* L, on *noisy* text just in case for every noisy text, t, for L, f converges on t to an index for a finite-variant of L; f *finite-difference identifies* a collection, *L*, *of* languages on noisy text just in case for every L$\in$*L*, f finite-difference identifies L on noisy text. Finite-difference identification on incomplete text is defined similarly.

Given the margin of error tolerated in finite-difference identification, one might doubt that imperfection restricts this kind of learning. It is thus natural to conjecture:

L is finite-difference identifiable [by recursive learning function] if and only if 1 is finite-difference identifiable on noisy text [by recursive learning function]; and

L is finite-difference identifiable [by recursive learning function] if and only if L is finite-difference identifiable on incomplete text [by recursive learning function].

The next two propositions show that both versions of both conjectures are false.

Proposition 12: There is a collection, L, of languages such that some recursive learning function identifies L, but no learning function (recursive or not) finite-difference identifies *L* on noisy text.

Proposition 13: There 1s a collection, JL, of languages such that some recursive learning function identifies L, but no learning function (recursive or not) finite-difference identifies L on incomplete text.

Section 3: Language acquisition and formal models of Inference

The circumstances of normal language acquisition by children appear to share a fundamental feature with the inference paradigms discussed above. Infants apparently have no direct access to the nonsentences (so labeled) of the target language. This assertion rests on evidence that children are seldom corrected for ungrammatical utterances *per* se, nor do they communicate more successfully with grammatical than with ungrammatical sentences (Brown & Hanlon, 1970). In short, children learn their language on texts.

As a consequence of this shared environmental feature, the propositions adduced above are relevant to the study of language acquisition. Thus: (a) Propositions 1 - 4 reveal some of the consequences of efficient learning by children; if language acquisition proceeds efficiently in the senses earlier defined, the class of learnable languages is narrower on this account, (b) If the class of natural languages Is infinite, then a corollary of results in Section 2.2 shows that infinitely many grammars conjectured by children are wildly oversized (by any reasonable measure of size), (c) If each sentence in the target language can be associated with a lower bound on the probability of Its occurrence in the child's linguistic environment, then Proposition 7 shows that the set of languages that can be learned with certainty is very large. And (d) Propositions 9 and 10 reveal the surprising consequences for language acquisition of even mild imperfections in children's linguistic environments. These kinds of connections between Formal Learning Theory and language acquisition by children have become increasingly central to linguistic theory and developmental psycholinguistics (see Wexler & Culicover, 1980; Osherson, Stob & Weinstein 1983a).

The results reviewed in this paper represent only one of several perspectives on language acquisition offered by Formal Learning Theory. Thus, in addition to efficiency, soveral other learning strategies plausibly attributed to children have been investigated from the learning theoretic point of view (Osherson, Stob & Weinstein 1983b). These include such response tendencies as (a) restriction to hypotheses compatible with available data, (b) gradual shifts in hypotheses rather than large leaps, (c) perseveration on conjectures that predict the available linguistic data, (d) restriction to grammars with nontrivial "recursive" rules, and (e) exclusion of long-past data in hypothesis selection. Additionally, criteria of successful acquisition less stringent than identification have been formulated and studied in the context of contemporary linguistic theory and language acquisition. These issues are discussed in Osherson & Weinstein (1983c).

REFERENCES

Angluin, D. & Smith, C.H., A survey of inductive inference: Theory and methods, Yale University Technical Report 250, October 1982.

Blum, M., On the size of machines, Information & Control. 11(3). 1967a. pp. 257-285.

Blum, M., A machine-Independent theory of the complexity of the recursive functions, JACM, 14(2). 1967b, pp. 322-336.

Gold, E. M.. Language Identification In the limit, Information & Control, 10. 1967, 447-474.

Osherson, f). . Stob, M. & Weinstein. S., Environments for learning, manuscript, 1982.

Osherson, D., Stob, M. & Weinstein, S., Analysis of a language acquisition paradigm, manuscript, 1983a.

Osherson, D., Stob, M. & Weinstein, S., Learning strategies, Information A Control, 1983b.

Osherson, D. & Weinstein, S., Criteria of learning, Information & Control, 1983c.

Osherson, D. , Stob, M. & Weinstein, S., Three topics in the theory of practical learning, manuscript, 1983d.

Osherson, D. , Stob, M. & Weinstein, S., Introduction to formal learning theory, (forthcoming).

Rogers, H. , Theory of recursive functions and effective computabil1ty, McGraw-Hill. 1967.

Wexler, K. & Culicover, P., Formal principles of language acquisition, MIT Press, 1980.