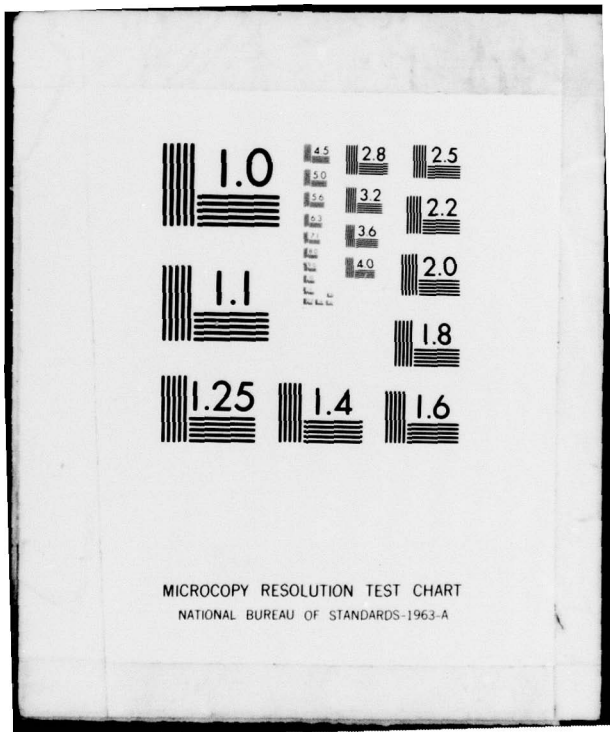
AD-A080 100 OHIO STATE UNIV RESEARCH FOUNDATION COLUMBUS F/G 9/2
TRANSLATION OF PHRASE STRUCTURED PROGRAMMING LANGUAGES.(U)
DEC 79 W BUTTELMANN AFOSR-75-2811
UNCLASSIFIED OSURF-760242/784151 AFOSR-TR-80-0088 NL

END
DATE FILMED
3 - 80
DDC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
6.3
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

RF Project 760242/784151
Final Report

the ohio state university
research foundation
1314 kinnear road
columbus, ohio
43212

TRANSLATION OF PHRASE STRUCTURED PROGRAMMING LANGUAGES

William Buttelmann
Department of Computer and Information Science

For the Period
May 1, 1975 - June 30, 1979

U. S. AIR FORCE
Air Force Office of Scientific Research
Bolling AFB, D.C. 22209

Grant No. AFOSR-75-2811

December, 1979

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER (19) (18) AFOSR-TR-80-0088	2. GOVT ACCESSION NO. / 3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) (6) TRANSLATION OF PHRASE STRUCTURED PROGRAMMING LANGUAGES.	5. TYPE OF REPORT & PERIOD COVERED: Final 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) (10) William/Buttelmann	8. CONTRACT OR GRANT NUMBER(s) (15) ✓AFOSR-75-2811
9. PERFORMING ORGANIZATION NAME AND ADDRESS Ohio State University Research Foundation, 1314 Kinnear Road Columbus, OH 43212	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS (16) 61102F (17) 2304/A2
11. CONTROLLING OFFICE NAME AND ADDRESS (11) Air Force Office of Scientific Research/NM Bolling AFB, Washington, D. C. 20332	12. REPORT DATE December 1979 13. NUMBER OF PAGES 13
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) (9) Final rept. 1 May 75-30 Jun 79	15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report) (12) 15

Approved for public release; distribution unlimited.

(14) OSURF-764242/784151

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

formal semantics, formal translation, formal language theory, syntax-directed translation, semantic-directed translation, phrase structure grammars, phrase structure languages, complexity of translation

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

An integrated formal theory of phrase structure linguistic descriptions has been developed, based on classical phrase structure linguistic theory and on a new formal theory of phrase structure semantics developed by this research project. The theory provides for both context-sensitive syntax and semantic structure, interpretations, the meaning of certain linguistic entities such as morphemes, phrases, sentences, and statements, and specifies the nature of and relationship between syntactic and semantic ambiguity.

The theoretical model of syntax and semantics was developed in order to study

OVER

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED 267 360 LB

20. Abstract cont.

the computational aspects of semantic interpreters and language translators. Based on the model, a theory of interpretation and translation was constructed, and a number of results in these two areas have been developed. ~~We show~~ IT WAS SHOWN how linguistic properties can be treated either as context-sensitive or context-free and either as semantic or syntactic information. This theory establishes the basis for further research into the computational complexity of semantic interpreters and especially into the problem of minimizing the complexity of interpreters by proper trade-offs between syntax and semantics and/or "context-sensitive"-ness and "context-free"-ness. We also develop several methods of translation based on syntax-driven or semantic-driven strategies and give results relating the different translation methods in terms of computational power and complexity. Finally, we study the problem of automatically generating a translator from the formal specifications of the syntax and semantics of two languages. Some early results in this area are given.

(12)

Translation of Phrase Structured Programming Languages
Final Technical Report
to the
U.S. Air Force Office of Scientific Research
Grant AFOSR 75-2811

by

William Buttelmann

Department of Computer & Information Science
The Ohio State University
Columbus, Ohio 43210

0. Introduction.

This report gives a summary of all the results of the research performed under the project supported by Grant AFOSR 75-2811. It is important to keep in mind that the research project supported by this grant began before the grant support commenced and is still an ongoing project. This report concentrates on the results obtained during the period of the grant.

Most of the technical results have been published in the journal articles and Ph.D. dissertations given in the list of publications at the end of this report, and the technical details of proofs and algorithms are given in those publications, copies of which have been forwarded to the AFOSR at the time of submission for publication. Therefore this report will state definitions and results in an informal manner, in order to present the results in a form more understandable by a reader not familiar with the formalism of research in this field. Instead of repeating the technical definitions and theorems in this report, we shall concentrate on the implications and importance of the results.

The principal investigator and staff of this project wish to take this opportunity to express their gratitude to the United States Air Force for supporting this research, and particularly for the helpful and encouraging cooperation of Lt. Colonel George W. McKemie, under whose supervision this project was conducted.

The report is divided into three areas:

1) Formal theory of semantics, language definitions, and phrase structure languages with semantics.

2) Formal theory of language translation.

3) Open problems for further research.

1. Formal theory of semantics, language definitions, and language.

Formal language theory has been an acknowledged field of theoretical computer science since at least 1954. And significant contributions have been made in the theory of syntax. But there have been no general theoretical models of semantics and of the relation between syntax and semantics. Numerous concrete models of particular semantics have been developed (Refs. 1-8), but none are abstract enough to match the generality of our formal models of syntax. On the other hand, a general theory of translation requires a comprehensive theory of both syntax and semantics. Consequently, one of the major efforts in this project was to develop a general theory of semantics for phrase structure grammars. Because it is so foundational to the rest of the research in this project, the semantic model has evolved during the entire project, and it is still being developed and refined. However, certain basic definitions for semantics which were developed during the period of funding by the grant appear to be fundamental, and have led to the beginnings of a general, cohesive theory of linguistic description and language translation. The following outlines the basic theoretical developments and results in this area.

The chief contributions of this portion of the research project have been to give precise formal definitions for the basic concepts in semantics, and to force a rethinking of certain established concepts in the area of formal language theory--in particular, the concepts of "phrase", "sentence", "syntactic contextual dependence", "ambiguity", and "language". We will explain these points in more detail later in this section of the report.

The following is an outline of the developments and results in this area:

1) We have developed the first formal general theory of semantics for phrase structure languages.

2) The theory states precisely:

- what meaning is,
- how meaning is specified in a linguistic description,
- just what the relationship is between syntax and semantics,
- how meaning is computed as a function of both syntactic and semantic information.

3) The theory includes a theory of semantic context and context-sensitive semantic functions which shows:

- how syntactic and semantic context differ,
- how synactic and semantic context may be interdependent,
- how both syntactic and semantic contextual information affect the meaning of a phrase or sentence.

4) The theory provides precise definitions of syntactic and semantic

structure and syntactic and semantic ambiguity which shows:

- how syntactic and semantic ambiguity are distinct but related properties of language,

- how a partial knowledge of one kind (syntax or semantics) may be used to resolve ambiguities of the other kind--a process extremely important in reducing the processing time of language understanding and translation systems.

5) Given the formal definition of phrase-structure semantics, we specify a formal definition for language definitions (or in the parlance of linguistics, of "linguistic descriptions"). A phrase structure language definition is a phrase structure grammar together with a phrase structure semantics. A language definition may have an unrestricted, context-sensitive, context-free, or regular syntax and a context-sensitive or context free semantics.

6-) The formal definition of "language definition" gives rise to precise definitions for phrases, sentences, languages, and the set of meanings of a phrase or sentence. The definitions of phrase, sentence, and language differ from the standard definitions in classical formal language theory because our language definition system now has a semantic component. Phrases and sentences are no longer just the frontiers of certain syntax structures, but the meaningful frontiers of these structures. In addition to the usual syntactic criteria, a string of symbols must satisfy certain semantic criteria in order to be a phrase or sentence. A language is not just a set of sentences generated by a grammar, but the set of sentences (meaningful sentences of the grammar) together with their meanings assigned by the semantics. In other words, the language of a linguistic description is the set of ordered pairs, (s,m), where s is a sentence of the grammar and m is one of the meanings of s assigned by the semantics.

7) The hierarchy of grammars and the two different kinds of semantics — context-free and context-sensitive — give rise to a hierarchy of linguistic descriptions ranging from the most general (unrestricted syntax and context-sensitive semantics) to the most restricted (regular syntax and context-free semantics). A number of fundamental issues in theoretical linguistics arise when we consider in just what since this apparent hierarchy of linguistic descriptions is a true hierarchy. Some of these issues are further outlined below.

8) The formal definition of language definitions gives rise to a number of ways in which language definitions are related. To begin with, there are the two classical relations on grammars suggested by Chomsky: weak equivalence and strong equivalence. However a number of other relations suggests themselves given that we have a richer model of language definition. The following seem relevant at the present time. Suppose D and D' are language definitions.

- Weak equivalence. D and D' are weakly equivalent iff they define the same language.

- Strong equivalence. D and D' are strongly equivalent iff they

define the same set of linguistic structures -- i.e., the same set of syntactic structures and the same set of semantics structures for each syntactic structure.

- Tree equivalence. D and D' are tree-equivalent iff they define the same set of skeleton syntax trees. By skeleton syntax trees we mean the following: In phrase structure linguistics, every syntactic structure is either a tree or a tree augmented with certain context-sensitive structure. Thus, the underlying phrase structure is always tree-like. If one starts with an arbitrary syntactic structure and eliminates all the context-sensitive structural information, one is left with a data structure which is a tree that shows the underlying phrasal relationships of the original sentence. We call this underlying phrase structure the skeleton tree of the original syntactic structure. Thus, tree-equivalent language definitions define the same phrase structural relationships, but may do so by means of different context sensitive mechanisms. Since the basic phrase structures of a language seem intuitively to be more fundamental than the context restrictions, tree equivalence seems to be an important relation on language definitions.

- Strong tree equivalence. D and D' are strongly tree-equivalent iff they are tree-equivalent and assign the same semantic structures to each skeleton tree. Language definitions which are strongly tree-equivalent define the same phrase relationships and assign the same meanings to tree-equivalent sentences.

Obviously, other relations are definable as well. But these four have turned out to be useful in the research so far. A little thought will show that strong equivalence refines strong tree-equivalence, which refines both tree-equivalence and weak equivalence. However, neither weak equivalence nor tree-equivalence refine each other, in general.

9) Now we can say more about the hierarchy of language definitions. Certain types of language definitions have turned out to be important, so we have given them special names. In particular, a language definition is:

- general if it has an unrestricted syntax and context-sensitive semantics,

- context-free if both its syntax and semantics are context-free, and

- regular if its syntax is regular and its semantics is context-free.

A major (and counter-intuitive) result is the following:

Theorem: Every language definable by a general language definition is also definable by a regular language definition.

This theorem means that, in terms of language definition power, the hierarchy of linguistic descriptions is not a true hierarchy. Any language that can be defined by any phrase structure linguistic description can be defined by the simplest (or weakest) linguistic description!

10) A similar but more important result is the following:

> Theorem: If D is a general language definition, there is a context-free language definition D' with the property that D and D' are tree-equivalent and weakly equivalent.

This theorem seems to make an important statement in linguistic theory: that any language definable by a general phrase structure language definition (having a context-sensitive or unrestricted syntax and a context-sensitive semantics) can also be defined with a context-free syntax and context-free semantics! And the context-free definition will not alter the basic phrase - structure of the language! This result gives us the important fact that, in terms of language definition power, context-sensitiveness is not necessary! It is always possible to define a language without context-sensitive rules without altering its basic phrase structure! Such a result is extremely counter-intuitive. In fact, the proof of the theorem is constructive and shows just how to convert context-sensitive information to context-free information.

This result says, in effect, that in deciding whether a particular property of a language should be treated as a context-free or context-sensitive property, or whether it should be treated as a syntactic or semantic property, the criteria to be used do not deal with definitional or generational power, but must refer to other aspects of linguistic or computational issues. This point leads naturally to a number of open questions about the nature of language and linguistic descriptions, which are mentioned in Section 3 of this report.

2. Formal theory of language translation.

There have been several attempts at a formal characterization of language translation — notably sequential transducers and gsm mappings (Ref. 9), g sm maps (Ref. 10) and the study of syntax-directed translations in Refs. 11-13.

However, none of these models incorporate any concepts of semantics, and therefore fail to get at the essential requirement of a translation — that it be meaning-preserving.

Given our formal model of linguistic definition with explicit syntax and semantics, a formal theory of translation developed itself naturally. The chief results are outlined here:

1) The first task was to formulate a sensible definition of "translation". This is straightforward only when the source and target languages are unambiguous and each sentence has a unique meaning. For

ambiguous languages and languages with many sentences sharing common meanings, several definitions suggest themselves. The following is a catalog of some. Let L be a source language and L' a target language.

- An ordinary translation (or just "translation" for short) from L to L' is a function on L to L' which assigns to each member (s,m) of L the set of all members (s',m') of L' where m = m'.

- A full translation from L to L' is a function which assigns to each member (s, m) of L the set of all members (s', m') of L' where m = m'.

Note that in our sense of translation, we are given both the source sentence and its meaning. Intuitively, it doesn't make sense to consider translating if you can't figure out the meaning of what you are trying to translate.

2) If one considers translation as a process or procedure, two fundamentally different strategies suggest themselves at once:

- semantic-driven translation - a procedure which, given a source sentence s and its meaning m uses the semantic rules of the source and target languages to produce one (or more) sentences in the target language having the same meaning.

- syntax-driven translation - a procedure which, given a source sentence s and its meaning m uses the syntax rules of the source and target languages to produce one (or more) sentences in the target languages having the same meaning.

3) One of the most important results of our research is to show the existence of a particular class of translators (which we call "table translators") which have the following properties (our results on table translators were published in Refs. 14-16):

- They are defined on languages defined by context-free language definitions.

- No semantic computation is performed during the translation.

This is a very important feature, since semantic functions can be arbitrarily complex. In fact it is this feature of tree translators which enables the next property.

- They perform ordinary translation by first parsing the input sentence and then translating the parse tree. Thus, we can divide their computation time into:

  parsing time + actual translation time.

- The actual translation time is a linear function of the length of the input sentence! The coefficients of the function depend only on the syntax of the source and target languages, not on their semantics.

This property makes table translators fast as translators go, since any sequential translator must read the input, a process which alone takes linear time. In fact, it is hard to conceive of a translator operating in a sequential manner which can run faster. Since context free languages can be parsed in cubic time or less, then, if the input sentence is syntactically and semantically valid, translation can occur in no worse than cubic time!

- The class of languages for which tree translators exist is the entire class of phrase structure languages for which any computable translation exists!

The last two properties of tree translators make their discovery rather remarkable. We have found a translation method which can be used on any phrase structure language pair for which a computable translation exists, and which runs in linear space and in the time it takes to parse plus an increment which is a linear function of the input!

4) It could appear at first reading that the discovery of the table translator "solves" in some sense the language translation problem. But, of course, it doesn't. This development really points out more exactly what the translation problem is. In the first place, the fact that actual translation time is linear is a little misleading, since the coefficients of the linear function depend on the number of entries in the syntax table driving the translator, and this can be huge. The real translation problem is not performing the translation, but finding the translation -- that is to say, the right set of rules for driving the translation.

Our project has studied this problem in some detail, and it is yet today one of the chief efforts in the project. During the period of this grant, we have approached this problem from different perspectives:

i) Integrate the strategies of semantic-driven and syntax-driven translation.

ii) Study the problem of generating table translators from given linguistic descriptions of source and target languages.

These efforts are discussed in (5) and (6) below:

5) Refs. 17 and 18 report the results of a study integrating semantic-driven and syntax-driven translation strategies. A hierarchy of translators is defined with increasing translation power and which exhibit improved computation time for certain classes of languages. Some very important issues were uncovered during this study which need further development -- in particular:

- The concept of "semantic deviance", its role in the translation process and the effect of its presence on the computational complexity of translation.

- The interplay of semantic and syntax considerations during the translation process, and how they can be best balanced to

reduce translation time.

Our papers report some results in these areas, but more work needs to be done.

6) The research in the area of producing translators from linguistic descriptions has not yet been reported in the literature because we have as yet only a few results. The results we have are given here. Suppose D and D' are two linguistic descriptions, and we wish to produce a syntax-driven translator from the language of D to the language of D', driven by the grammar rules of D and D':

- Such a translator (using the given linguistic definitions D and D') may not exist, even though there may be other definitions for the same languages for which such a translator does exist.
- It is undecidable, in general, whether such a translator exists.
- It is undecidable, in general, whether any computable translation from the source to the target language exists.
- If a computable translation does exist, then a syntax-driven translator (in fact, a table translator) exists for (and is defined by) some pair of definitions D" and D"' for the language.
- There is, in general, no effective way to construct from D and D' the pair of definitions D" and D"' which define the table translator.
- We have an algorithm which, given a pair of definitions, will produce the table translator defined by the definitions on their languages, provided such a translator exists.

The development of this algorithm appears to be an important contribution of this research, and has led to several further areas of study discussed in Section 3 of this report.

3. Open problems.

1) The implications of the theorem given in item (10) of Section 1 are important in linguistic theory, language design, and will have practical implications for the efficiency of language processes. The fact that the very model of language which established a formally distinct syntax and semantics leads to the proof that there are not theoretical reasons in the model for making any particular linguistic property either syntactic or semantic -- this fact -- has important philosophical implications for linguistic theory. Just what are the right criteria for deciding whether a given property should be treated syntactically or semantically? We know that, for a given language, some definitions seem to be "better" in some sense than others. Why? Parsimony? Computational complexity of parsers

and/or interpretors? Complexity of the definition itself? There is much research to be done in this area.

2) A related question deals with a process we call evaluation. Now that we have a formal model of language which explicitly incorporates semantics, we can talk about the process of evaluating the meaning (or meanings) of a sentence. Evaluation is the function which computes for a given sentence s all its meanings. Given the semantic context, an evaluation gives all the meanings assigned by the semantics to s within that context. In our model, evaluation is the analogy of parsing in classical formal language theory, and we need a formal study of evaluation similar to the study of parsing. Of particular practical interest is the question of how to coordinate the use of syntactic and semantic information during the evaluation process in order to minimize overall evaluation time by reducing backtracking and resolving ambiguity. We need to study the question, When and how should syntactic/semantic information be used to resolve ambiguities encountered during the evaluation process?

3) Item (5) of Section 2 describes our results in a study of translation srategies integrating semantic and syntactic-driven strategies. The results so far are interesting and satisfying, but more work needs to be done. Our studies so far have identified a hierarchy of techniques which vary in the classes of languages they will translate. But the important issue is to understand the nature of the trade-off between the use of syntactic and semantic information during translation so as to minimize the time and space cost of translation. It is desirable to find semantic and syntax-directed translation schemes for a significant class of languages (say, the context-free or even the context-sensitive languages) which run in polynomial time with small coefficients and in polynomial space.

4) The immediate open problems in the area of translator generation is the following:

i) What is a good upper bound on the problem of generating a table translator from source and target language definitions?

ii) Identify significant and useful classes of languages for which table translators running in polynomial time and space can be generated in reasonable time.

A host of further problems suggest themselves, as a result of this research. It is the author's opinion that, as a result of this grant, fundamental issues in the theoretical nature of syntax and semantics have been identified and defined in such a way that we can now study in a careful and precise way the nature and interrelationships of syntax and semantics, the nature of meaning and semantic evaluation of phrases and sentences within a context, the nature of ambiguity and techniques for its resolution, the theoretical nature of translation and the rules of syntactic and semantic knowledge in the translation process, and the complexity and generality of the translator generation process. This project has established a foundation on which much future research in semantics and translation will depend.

References

(1) Tarski, A. (1936). Der Wahrheitsbergriff in den formalasierten Sprachen. In Studia Philosophica, I, pp. 261-304. Originally published in 1933.

(2) Knuth, D. E. (1968). Semantics of context-free languages. In Mathematical Systems Theory, 2, pp. 127-146.

(3) Benson, D. B. (1970). Syntax and Semantics: a categorial view. In Information and Control, 16, pp 738-773.

(4) Katz, J., and Fodor, J. A. (1964). The structure of a semantic theory. In Language, 39, pp. 170-210. Reprinted in Fodor and Katz (eds.) The Structure of Language. Englewood Cliffs, New Jersey: Prentice-Hall, pp. 479-518.

(5) Lucas, P., Lauer P., and Stigleitner, H., (1968), Method and notation for the formal definition of programming languages, IBM Leb. Tech. Report, TR25.087, Vienna.

(6) Scott, D., and Strachey, C., (1971), Towards a mathematical semantics for computer languages, in "Proceedings of the Symposium on Computers and Automata," Polytechnic Institute of Brooklyn.

(7) Tennet, R. D., (1976), The Denotational Semantics of Programming Languages, CACM 19, 8 (Aug. 1976), pp. 437-453.

(8) Goguen, J. A., Thatcher, J. W., Wagner, E. G., and Wright, J. B., (1977), Initial Algebra Semantics and Continuous Algebras, JACM 24, 1 (Jan. 1977), pp. 68-95.

(9) Ginsberg, Seymour, (1966), "The Mathematical Theory of Context-Free Languages," McGraw-Hill, New York.

(10) Thatcher, James W., (1970), Generalized sequential machine maps, J. Comput. Syst. Sci., 4, pp. 339-367.

(11) Aho, A. V., and Ullman, J. D., (1969), Properties of syntax-directed translations, J. Comput. Syst. Sci 3, pp. 319-334.

(12) Aho, A. V., and Ullman, J. D., (1969), Syntax-directed translations and the pushdown assembler, J. Comput. Syst. Sci. 3, pp. 37-56.

(13) Aho, A. V., and Ullman, J. D., (1971), Translations on a context-free grammar, Inform. Contr. 19, pp 439-475.

(14) Krishnaswamy , R., (1976). "Methodology and generation of language translators", Ph.D. Dissertation, The Ohio State University, Columbus, Ohio.

(15) Krishnaswamy , R., and Buttelmann, H. William, (1978), Formal methodology of translation I. Semantic preserving translations, Inform.

Contr. 39, 3, 247-271.

(16) Krishnaswamy , R., and Buttelmann, H. William, (1978), Formal methodology of translation II. Capabilities and limitations, Inform. Contr. 39, 3, 272-283.

(17) Pyster, Arthur, (1975), "Formal Translation of Phrase-Structure Languages," Ph.D. Dissertation, The Ohio State University, Columbus, Ohio.

(18) Pyster, Arthur, and Buttelmann, H. William, (1978), Semantic-syntax-directed translations, Inform. Contr. 36, 3, 320-361.

Publications

Pyster, Arthur, (1975), "Formal Translation of Phrase-Structure Languages," Ph.D. Dissertation, The Ohio State University, Columbus, Ohio.

Krishnaswamy , R., (1976), "Methodology and generation of language translators", Ph.D. Dissertation, The Ohio State University, Columbus, Ohio.

Pyster, Arthur, and Buttelmann, H. William, (1978), Semantic-syntax-directed translations, Inform. Contr. 36, 3, 320-361.

Krishnaswamy , R., and Buttelmann, H. William, (1978), Formal methodology of translation I. Semantic preserving translations, Inform. Contr. 39, 3, 247-271.

Krishnaswamy , R., and Buttelmann, H. William, (1978), Formal methodology of translation II. Capabilities and limitations, Inform. Contr. 39, 3, 272-283.

Dickey, Frederick Joseph, (1979), "Translations Between Programming Languages," Ph.D. Dissertation, The Ohio State University, Columbus, Ohio.