

kansas  
working  
papers  
in  
linguistics

---

volume 3  
1978

Anthony Staiano and Feryal Yavaş  
Editors

Funded by Student Activity Fees  
University of Kansas  
Lawrence, Kansas  
1978

P  
J  
.K36  
v. 3  
1978

## Introduction

This third volume of the Kansas Working Papers in Linguistics covers a diversity of topics which range from general linguistic theory to child language. To provide coherency, we have, therefore, grouped the papers into a number of major sections as reflected in the Table of Contents. What follows is our attempt to capture the major point of each paper, organized according to those sections.

The first paper is Ken Miner's "On the Notion 'Restricted Linguistic Theory': Toward Error Free Data in Linguistics." Miner maintains that linguistic theories must be more firmly grounded on secure data bases. He contends that the attempt to construct theories based on limited data from a few languages leads to serious errors. Rather than seeking to construct general theories, Miner advocates that we should limit ourselves to "restricted theories" which may be confined to one language family.

The Phonetics-Phonology section contains four very different papers. Geoff Gathercole's research demonstrates that instrumental evidence can play a crucial role in phonological analysis. His instrumental research on strong and weak stops in Kansas Potawatomi clearly indicates that the underlying contrast between these series is preserved even in final positions, not neutralized as heretofore supposed. In addition, the paper provides evidence for the interaction between stress and the syntactic structure of Potawatomi.

Mehmet Yavas' paper on the implications of borrowing for Turkish phonology provides a modus operandi for the analysis of languages which have lexicons replete with loan words. In the case of Turkish, previous analyses, though recognizing the importance of loan words, have neglected to incorporate them into their descriptions. Drawing evidence from borrowing, Yavas proposes that current treatments of vowel and consonant harmony should be drastically revised: consonant harmony plays the pivotal role in determining the vowel choice, not conversely. By so analyzing Turkish, he is able to account for a wide range of data unaccounted for by treatments which assume the primacy of vowel harmony.

Robert Rankin's study of Quapaw as a dying language supports the evidence from child language acquisition, aphasia, and comparative linguistics that there exists a universal hierarchy of sound-type complexity. As Quapaw functioned less and less as a native language, principled changes occurred in its phonology: the types of series lost and the order in which they were lost were determined by their relative complexity, with the most marked being lost first.

Code-mixing is the topic of Maria Dobozy's paper. Taking a letter written by a bilingual American-Hungarian as her data, Dobozy describes the phonological rules that are operating in such a code-mixing, with special emphasis on vowel harmony. She demonstrates that vowel harmony is an important process in the system and plays a central role in the rendition of English words by such speakers.

The first paper in the Syntax-Semantics section is Gerald Denning's, "Meaning and Placement of Spanish Adjectives." Denning attempts to clarify the problems of the differences in the meaning and treatment

of restrictive adjectives in three dialects of Spanish. He argues that a strict generative semantic approach will not handle the data and suggests an analysis within the framework of pragmatics.

Virginia Gathercole provides a cross-linguistic study of the use of the deictic verbs "come" and "go." She formulates the uses of "come" and "go" in eleven languages by extending Talmy's (1975) model for verbs of motion to include a presuppositional component. Gathercole divides the contexts in which "come" and "go" are used into (a) immediate deixis and (b) extended deixis. Her goal is to characterize the use of deictic verbs of motion in the eleven languages studied by a limited number of assertional and presuppositional components and thus suggest a possible universal framework for such verbs.

Whereas Denning and Gathercole focus on language related issues, Juan Abugattas takes a more general, philosophical approach in his discussion of speech acts. He claims that previous speech act analyses used the sentence as the basic unit. Abugattas believes, however, that we must go beyond the sentence: "social reality" dictates that we categorize sets of sentences into speech acts, which he calls "complex acts."

Kurt Godden's paper, "Problems in Machine Translation Between Thai and English Using Montague Grammar," brings us to a specific language oriented concern: how to mechanically translate sentences, in particular those containing restrictive relative clauses, from one language to the other. He enumerates the problems related to such a task and proposes a solution involving meaning postulates and context within a Montague framework.

Historical and Comparative Linguistics is represented by Karen Booker's "On the Origin of Number Marking in Muskogean." Booker reconstructs two proto-Muskogean number markers, one dualizer and one pluralizer which were first used with intransitive verbs of location and then generalized to locative transitives. Later these markers spread to intransitive non-locatives. Booker maintains that the highly complex suppletive verb system of Muskogean arose when these markers lost their original meaning.

Three papers, Esther (Etti) Dromi's analysis of the acquisition of locative prepositions by Hebrew children, Gregory Simpson's study of children's categorization processes, and John More's review of relative clause research, constitute the Child Language Acquisition section of the working papers. Dromi's study, which is one of the few published works in the acquisition of Hebrew, compares the order of acquisition of Hebrew locatives with Brown's (1973) order for English and also with Slobin's (1973) universals. Among her findings, Hebrew al ("on") is acquired later than English on. Her findings for Hebrew locatives are particularly interesting in that they allow a comparison of the acquisition of prefixes with that of full prepositions. Her conclusions point to the pivotal role that morphological complexity plays in the order of acquisition of locatives in Hebrew.

Gregory Simpson's major concern has to do with the process by which children form conceptual categories. He argues, on the basis of experimental data, that overextensions should not be taken as evidence

for category formation. His data suggest a distinction between concept formation and object naming, a distinction not made in previous studies. "Function," what objects can do or what can be done to them, determines how that object is conceptualized, but an object's perceptual properties may determine the name given to it. Therefore, "the child may know that two objects don't really belong together, but gives them the same name until he has more evidence."

The acquisition of relative clauses has been a topic of great interest among psycholinguists. John More presents a valuable critical review of the recent literature with special emphasis on the debate between Dan Slobin (1971), Amy Sheldon (1974), Michael Smith (1975), Tavakolian (1977), and deVilliers et al. (1976). The Minimal Distance Principle, the Noun-Verb-Noun Strategy, the Parallel Function Hypothesis, and Slobin's operating principles are compared, along with the formulations of deVilliers and Tavakolian.

Five major topic areas are represented in this third volume of the Kansas Working Papers in Linguistics. Each paper in its own way is a contribution to linguistic scholarship: some provide evidence in new areas of inquiry, others bring new evidence to bear on old questions, while still others suggest future courses of research.

Anthony Staiano and Feryal Yavas

Editors

## CONTENTS

Volume 3, 1978

Linguistic Theory

- On the Notion "Restricted Linguistic Theory": Toward  
Error Free Data in Linguistics.  
Kenneth L. Miner 1

Phonetics-Phonology

- Instrumental Phonetic Studies and Linguistic Analysis:  
The Case of Kansas Potawatomi  
Geoff Gathercole 20
- Borrowing and Its Implications for Turkish Phonology  
Mehmet Yavaş 34
- The Unmarking of Quapaw Phonology: A Study of  
Language Death.  
Robert L. Rankin 45
- Vowel Harmony and Code-Mixing: A Description of  
Phonemic Substitution in an American-Hungarian Text.  
Maria Dobozy 53

Syntax-Semantics

- Meaning and Placement of Spanish Adjectives.  
Gerald L. Denning 68
- Towards a Universal for Deictic Verbs of Motion  
Virginia C. Gathercole 72
- Speech Acts, Functions and Presuppositions  
Juan Abugattas A. 89
- Problems in Machine Translation Between Thai and English  
Using Montague Grammar  
Kurt Godden 94

Historical-Comparative Linguistics

- On the Origin of Number Marking in Muskogean  
Karen M. Booker 100

Language Acquisition

- The Use of Locative Prepositions by Hebrew Speaking  
Children 2:0 - 3:0 Years Old  
Esther (Etti) Dromi 113
- Static and Dynamic Properties as Bases for Children's  
Categorization  
Greg Simpson 131
- The Relative Clause in Child Language: A Review  
John Blake More 143

PROBLEMS IN MACHINE TRANSLATION BETWEEN THAI AND  
ENGLISH USING MONTAGUE GRAMMAR

Kurt Godden

Introduction

Joyce Friedman is currently doing research on Montague grammar and she is writing, along with co-researchers, computer programs that can parse English sentences, translate these parses into LISP<sup>1</sup> analogues of intensional logic, translate these LISP expressions back into English parses, and interpret or assign meanings to these LISP expressions by reference to a machine model analogous to Montague's possible worlds, time points, and entities. Assuming the success of Friedman's programs, it should be possible then to use them to translate one natural language into another if one has Montague-based grammars for similar fragments of both languages and the appropriate computer programs.

In this paper, I would like to discuss some of the problems that such a translation would have to solve. Two genetically unrelated languages, English and Thai,<sup>2</sup> will be used for the examples for several reasons. First, using unrelated languages, one would expect to meet more difficult problems due to different syntactic constructions than if related languages were used. Therefore, if a machine translation is shown to be possible for unrelated languages, then there should be a greater range of application since it should be a somewhat easier task to do the same for closely related languages. Second, certain constructions in English and Thai are similar enough to illustrate the relative ease with which a Montague theory can handle the translation. Third, these same constructions are at the same time different enough in certain respects to present difficult problems which, if adequately treated by the theory, reveal the theory's power and utility. Such a syntactic construction in the two languages is the restrictive relative clause (RC). While many of the syntactic RC types are extremely similar in terms of word order and rule operation, and the semantic representations in both languages serve to identify uniquely the referent of the head noun, different kinds of relative clauses in one language may be the same kind of RC in the other or else a relative clause of one type in one language must be expressed as an RC of a different type in the other language. These differences are problems not only for the syntax of a Montague grammar, but also for the semantic component.

In the following sections I describe the kinds of RC's in the two languages, how a Montague-type grammar describes them, and then I discuss specific problems that a machine would encounter in translating RC's.

### Brief Comparison of English and Thai Relative Clauses

In Godden (1978), I extend Montague's (1974) rules to account for sentences like those in (1).

- 1 a. the boy <sup>{that}</sup><sub>who</sub> went to the market  
 b. the books <sup>{that}</sup><sub>which</sub> he writes  
 c. the house where you live  
 d. the boy whose father is tall  
 e. the girl who hit the boy who looked at her

Let us look at the corresponding Thai sentences when we try to translate (1), preserving the same relative clause types.

- 2 a. dèk phuučaj thii paj talaat  
 child male that go market  
 the boy that went to the market  
 b. nangs++ thii khaw khian  
 book that he write  
 the book(s) that he writes, or  
 the book(s) that he wrote  
 c. baan thii khun juu  
 house that you live  
 the house where you live  
 d. \*dèk phuučaj thii phoo khooj khaw suuj  
 child male that father poss. he be-tall  
 the boy whose father is tall (or, closer to the Thai, the  
 boy such that the father of him is tall)  
 e. dèk phuujiij thii tii dèk phuučaj thii mooj thae  
 child female that hit child male that look-at her  
 the girl who hit the boy who looked at her

Note first the similarities of the syntactic forms between Thai and English. In both languages, the first noun in the embedded sentence that is co-referential to the head noun is replaced by a relative pronoun which is moved to the beginning of the embedded sentence. Where the embedded noun is the subject, this movement takes place vacuously. It can be seen that the basic form and functioning of the syntactic rules for Thai will be the same as those for English. Since the semantic rules in Montague grammar change each syntactic operation into a semantic form, the semantic representations for the most part would also be the same for both languages.

### Problems in Translating Relative Clauses

One obvious difference between the Thai and English RC's is in the choice of relative pronouns. While English has several, Thai has only one<sup>3</sup>. This presents no problems as far as the English relative pronouns who(m), which, and that are concerned since they are entirely syntactic in nature; that is, the choice of one or the other depends only on gram-



mathematical gender. There is no semantic difference in the intensional logic expression. However, looking at the relative clauses of place expressed with where in English, we encounter difficulties.

For English, the relative pronoun where is introduced by an adverbial variable, whereas who(m), which, and that are from noun phrase variables. Therefore, for where, the semantics is different from the others. In Thai, if we had only one relative pronoun, there would be no difference in its semantics between reference to places and reference to things. If a machine were to translate from Thai to English then, the Thai expression for the house where you live could come out in English as \*the house which you live. The problem is revealed even more clearly in (3) and (4).

3. rōŋ thii khun dāj jōk wāj pēn rōŋ thii jaaw  
trench that you aspect dig asp. be trench that be-long

4 a. The trench that you dug is the trench that is long.

b. The trench where you dug is the trench that is long.

Either of (4a) or (4b) is a possible translation of (3). In (4a) we could have a situation where the speaker is saying that someone dug a trench and that trench is long. (4b) could describe a very different situation where a trench was already dug (by anyone or anything) and some person was inside that trench digging something, perhaps another smaller trench, and that larger trench is long. Both that and where RC's serve to limit the reference of the head noun from a set of many possible objects to one specific object. However, as (3) and (4) show, the actual state of the world can be quite different between the two types of RC's. In translating (3) to English, if thii is translated always as a noun phrase, there is no way to represent the meaning of (4b) and therefore no way to derive (4b).

The obvious solution is to regard thii as one of two homophonous, but syntactically and semantically disjoint, relative pronouns. One thii would correspond to English who(m), which, and that and the other thii to where. This would assure the correct translation from English to Thai. This analysis of thii is required not only to translate it into English, but it is necessary for any descriptive grammar of Thai in order to represent the two different meanings of (3) in intensional logic.

A far more difficult problem to solve is revealed in (1d) and (2d). Thai has no construction analogous to English possessive RC's. (1d) would have to be translated as (5).

5. dək phuučaj thii mii phōō suŋ  
child male that have father be-tall  
the boy who has a tall father

(5) is still an RC, but of a very different type and one that expresses a different embedded proposition. Of course, when interpreted semantically (1d) and (5) would refer to the same individual, but how can the translation be done with a machine that uses intensional logic, not its interpretation, as the level of mediation between the two languages? Notice that the problem only arises when translating from English to Thai. When translating (5) into English the machine would simply produce the English gloss given for (5). However, for (1d) we have an in-

tensional logic expression that is roughly shown in (6)<sup>4</sup>.

6.  $\lambda x_n (\text{boy}'(x_n) \wedge \text{his}'(\hat{\text{father}})'(\hat{\text{be-tall}})')$

There is no RC construction in Thai that could give this expression. Therefore, our machine could not give us a Thai translation using RC's.

One solution, which is quite undesirable, would be to write another computer program that would take us from the interpretation model into expressions of intensional logic. Such a program would have to take representations of entities, worlds, and time points and produce intensional logic expressions which could be regarded as true or false and which represent relationships among the entities, worlds, and times. This would include building expressions of relationships that do not exist. In short, a program like this could produce as output an indefinite number of expressions representing any number of real or hypothetical relationships. Restricting a program like this would be virtually an impossible task.

A different approach to the problem would be to take (6) and derive any Thai expression with an equivalent meaning. If our goal is to convey in one language what was expressed in another, then why restrict ourselves to the same syntactic constructions in this translation?

Taking this approach, we see that (6) represents a conjunction. Thus, we could take the English sentence John is the boy whose father is tall, come to a logical expression which contains (6), then arrive at a conjunction in Thai,

จอน เปน เด็ก ภูมู๋จาย ล้า ฝอจ๊ว หนอจ๊ว หน้าว สูนุ  
 John be child male and father poss. he be-tall  
 John is the boy and his father is tall.

For this example, it appears that we can do this. However, this is merely an accident. If (6) is used as the subject of the verb instead of the object, the anomaly reveals itself.

7. The boy whose father is tall likes Mary.

8. \*เด็ก ภูมู๋จาย ล้า ฝอจ๊ว หนอจ๊ว หน้าว สูนุ จ๊จ๊ป มารีอิ  
 child female and father poss. he be-tall like Mary  
 \*The boy and his father is tall likes Mary.

If (7) were allowed to be translated into (8), an ungrammatical sentence would result. The reason is because in Montague's intensional logic common nouns are represented as predicates. In (6), boy'( $x_n$ ) is an open sentence. However, if it could be converted to English (it<sup>n</sup> cannot), it would be expressed as a common noun, which happens to be conjoined to a sentence. But natural languages can not conjoin common nouns and sentences. This is guaranteed by Schachter's (1976) Coordinate Constituent Constraint.

How, then, could our machine translate English possessive RC's into Thai? This answer can be found within Montague's theory itself. Montague lists nine "meaning postulates" which restrict the possible interpretations of intensional logic to those natural for English. One of these meaning postulates assures "the natural definition of seek as try to find," (p.264). This meaning postulate states that x's seeking some-

thing is logically equivalent to  $x$ 's trying to find something. We can allow for a similar meaning postulate to solve our problem. This meaning postulate will simply state that our interpretations (for either language) are restricted to those in which an expression  $x$  has an  $ab$  (where  $x$  and  $b$  are entities and  $a$  is an attribute) is logically equivalent to an expression  $x$ 's  $b$  is  $a$ . In this way, we can take (1d) into an intensional logic expression containing (6). Then since (6) contains an expression of the form  $x$ 's  $b$  is  $a$ , viz.  $his'$  ( $\wedge$ father') ( $\wedge$ be-tall')<sup>5</sup>, where  $his'$  refers to  $boy'(x_n)$ , our meaning postulate will allow the machine to change this to the logically equivalent form  $x$  has an  $ab$ , roughly  $\lambda x_n (boy'(x_n) \wedge x_n(\wedge$ have'( $\wedge$ tall( $\wedge$ father')))). This last expression is then converted easily to (5), the desired result.

### Non-RC-Related Problems

The section above discussed problems related specifically to translating RC's for Thai and English. In the course of the discussion, the reader may have noticed some other possible difficulties our task would have to meet. For example, consider (2b). Here we can see a problem that arises naturally out of the fact that English is more of an inflecting language than Thai.

In translating a Thai sentence to English, how does our machine know what tense to use in the English sentence and what number to associate with the subject and verb? The Thai RC in (2b) can have its embedded sentence express either the present or the past tense and its head noun either singular or plural. Of course, this depends on the context and results in no ambiguity for native speakers who have access to this context. It follows then that our machine must also have access to the context in order to translate correctly.

In Oh and Godden (1978), a new theory of grammar is presented which associates a linguistic expression with two things. The first is its "basic" meaning, its intensional logic form in a Montague theory; and the second is a set of contexts in which the linguistic expression may be used. A context is viewed as a set of propositions which must be interpreted just as the "basic" proposition of the sentence. Applying this to our present problem, Friedman's program that converts sentence parsings into logical forms would be modified to change parsings into logical forms with associated contexts. In translating Thai to English, the user of the system would supply the Thai sentence along with tense and number specified in the accompanying context. The other program that converts logical expressions to natural language would also have to be modified to take two inputs, the converted logical expression and the context, also in the form of intensional logic expressions.

It may appear superficially that inputting the system with a sentence and also with tense and number is an ad hoc solution. However, the tense and number are just part of what may be many other statements, in particular presuppositions and deictic categories, which naturally account for and explain a wide range of phenomena concerning language use and meaning.

One more point should be mentioned about (1) and (2). Comparing (1a) and (2a), we see that the English contains a preposition which has

no counterpart in Thai. Since this preposition has a reflex in the semantic component, we need to provide for it in Thai. Again, a meaning postulate will be able to handle the problem. This meaning postulate will state that something of the form (approximately)  $\text{go}'(\wedge \text{to}'(x))$  is logically equivalent to  $\text{go}'(x)$ . In going from Thai to English we would obtain the latter. Then, not finding a possible English phrase to express  $\text{go}'(x)$ , the system could refer to the meaning postulate, convert the expression into the form acceptable for English, and the results would be as desired.

### Conclusion

Of course, not all the problems that machine translation would encounter have been discussed. However, I hope that some of the easy and the not so easy problems for a specific translation task have been pointed out, and solutions suggested. I believe that no difficulties presented here are insurmountable when natural extensions of the theoretical approach and modifications of the programs used are incorporated.

### Footnotes

1. "LISP" is an acronym for List Processing Language and refers to a computer language used widely in the field of Artificial Intelligence.
2. Three native speakers of standard Thai were used as informants.
3. This statement is true only within a certain style of speech, casual-polite speech. If one speaks in a very formal manner, the number of relative pronouns increases. The casual-polite style is discussed here since it is the more common mode of expression in everyday Thai.
4. The translation of the is ignored for ease of exposition; whose is treated as a constant; and be-tall is taken as an intransitive verb phrase.
5. Now we need to regard be-tall as the translation in abbreviated form of be plus an adjective.

### References

- Friedman, Joyce. 1977. Computer Programs for Montague Grammar. Unpublished.
- Godden, Kurt. 1978. Me and my RC: what's good enough for Montague ain't good enough for me. Proceedings of the 1977 Mid-America Linguistics Conference. Columbia: University of Missouri.
- Montague, Richard. 1974. The proper treatment of quantification in ordinary English. Formal Philosophy, ed. by R. Thomason. London: Yale University Press.
- Oh, Choon-Kyu and Kurt Godden. 1978. Presuppositional Grammar. Syntax and Semantics: Presupposition, vol. 11, ed. by C-K. Oh and D. A. Dinneen. New York: Academic Press.
- Schachter, Paul. 1977. Constraints on coordination. Language 53.