THE AUTOLING SYSTEM$^{1,2}$

by

Sheldon Klein
William Fabens
Robert G. Herriot
William J. Katke
Michael A. Kuppin
Alicia E. Towster

Technical Report #43
September 1968
ABSTRACT

The AUTOLING system represents an attempt to replace the human linguist with a machine in the process of linguistic fieldwork with an informant. To the extent that the attempt succeeds, the analytic and heuristic methodology of live linguists can be considered formalized.

The current system consists of three as yet unjoined components: a morphological analyzer, a program for learning context-free phrase structure grammar, and a program for learning monolingual and bilingual transformations. All programs are written in ALGOL and operational on the Burroughs B-5500 computer. The capabilities of the system are illustrated with examples of its treatment of selected problems in English, Latin, Roglai, Indonesian, Thai, Chinese and German.

1 A portion of this paper was presented at the 42nd Annual Meeting of the Linguistic Society of America, December 1967, Chicago, under the title "Testing an Automated Linguistic Fieldworker".

2 This research has been supported in part by the Wisconsin Alumni Research Foundation and the National Science Foundation.
TABLE OF CONTENTS

1.0 Introduction 1
1.1 On Discovery Procedures, Algorithmic and Otherwise 1
1.2 The Philosophy Behind AUTOLING 2
2.0 Morphological Analyser 4
3.0 Phrase Structure Heuristic Learning Program 6
3.1 Rule Testing Heuristics 9
  3.1.1 Substitution and Informant Queries 9
  3.1.2 Parsing Illegal Sentences and Recycling 12
3.2 Test Problems 14
  3.2.1 Embedding Illustration: Artificial Language 16
  3.2.2 English I 17
  3.2.3 English II 19
  3.2.4 Latin: Koutsoudas #26 23
  3.2.5 Roglai: Koutsoudas #58 28
  3.2.6 Indonesian: Koutsoudas #43 and #9 Combined 35
  3.2.7 Thai 46
  3.2.8 Mandarin Chinese 56
3.3 Planned Improvements in the Phrase Structure Learning Program 65
4.0 Transformation Learning Program 67
4.1 Bottom-to-Top Transformation Learning 67
4.2 Top-to-Bottom Transformation Learning 69
  4.2.1 Program Logic 69
  4.2.2 Features of the Program 75
  4.2.3 Learning and Application of an Active-Passive Transformation 77
  4.2.4 Learning a Bilingual Transformation 80
5.0 Proving the Linguist Superfluous 82
References 84
1.0 Introduction.

The AUTOLING system* is, in conception, an on-line computer program that replaces a human linguist in analytic interaction with a live informant. The discovery procedures are heuristic rather than algorithmic; an algorithm is here defined as a method that guarantees success; a heuristic as an analytic approach that may work, but does not necessarily guarantee success in all cases.

1.1 On Discovery Procedures, Algorithmic and Otherwise.


In 1959 [18], Solomonoff suggested an automatic method for phrase structure language grammar discovery that, while adding little to the methodology of linguistic analysis, introduced the problem to a computing audience and suggested the use of an informant.

Distributributional analysis grammar discovery procedures alone will not work for two reasons. The first is practical: the combinatorial computations required in the analysis of say a million words of running text would yield intermediate data exceeding the memory capacity of any known computer—and the computation time of course would be commensurately prohibitive. More formally, E. Shamir demonstrated in 1962 [16] that it is impossible to obtain a discovery method for context free phrase structure grammars using natural language text alone as input. (He also suggested that the use of an informant would not alter the problem.)

*Written in ALGOL for the Burroughs B-5500 Computer.
Paul Garvin, in 1961 [2] suggested the use of heuristics in guiding the choice of distributional tests. His methodology still used only edited texts as a data base.

E. Mark Gold, in 1964 [4], indicates a formal demonstration that a discovery algorithm for learning context free grammar is possible if interaction with an informant is permitted.

Garvin discussed the use of heuristics in automatic informant work in 1965 [2].*

Work on the AUTOLING system has been described in 1967 [10, 11].

Other programs which involve some sort of language learning include those of Knowlton 1962 [12], Uhr 1964 [20] McConlogue & Simmons 1965 [14], and Siklossy 1968 [17].

1.2 The Philosophy Behind AUTOLING

The basic goal of AUTOLING research is to replace the linguist rather than aid him. To the extent that this goal is attained (in the form of a computer program system), the analytic methodology of live linguists can be considered formalized.

Human linguists obtain grammars without performing massive distributional analysis through the use of heuristics which permit them to make testable hypotheses that, when verified, eliminate the need for great masses of distributional analyses.

The ideal AUTOLING system would incorporate all the heuristics that a good linguist uses in fieldwork, and be capable of undertaking the analytic process via interaction with a live informant.

*At a RAND Corporation Linguistics Colloquium in December 1964, S. Klein suggested to Paul Garvin, the invited speaker, that he might use the concept of heuristic problem solving and game playing in application to automated informant work.
However, we have not attempted to deal with phonology in the project. The intended informant for the system would be bilingual in English and some other language, and be capable transcribing his non-English language in phonemic notation on a teletype keyboard.

The ideal AUTOLING system would, initially, ask the informant a battery of prestored questions ('How do you say...?') designed to elicit material for morphological analysis. At a somewhat later stage of interaction a phrase structure learning component would attempt to learn a phrase structure grammar, accepting as inputs the informant's responses rewritten morphophonemicall. A continuous interaction between the phrase structure learning component and the morphological program would take place. Ideally, the program would learn monolingual transformations, and also bilingual ones so that the program might generate its own query list in English. That is, a sentence is generated to test a hypothesized rule, then translated into English; the system would then output a 'How do you say' message followed by the English translation of the test sentence; the informant's reply is then matched against this prediction.

This paper, however, deals with the current reality of the AUTOLING system. At the moment it consists of three disconnected components: a morphological analyzer that is not informant interactive; an informant-interactive, context-free phrase structure learning program (the most developed component) which does test rules, but with a 'Can you say' message followed by the test production in the language under analysis rather than its English translation; and an informant-interactive program that learns monolingual and bilingual transformations. Also, in the actual systems, no query lists are used. The infor-
ment implicitly provides his own at certain times by feeding in pertinent data.

The reader will please note that all approaches that might involve a human linguist giving the machine advice have been avoided because of the primary goal of the research -- the replacement of the linguist and the concomitant formalization of fieldwork methodology.

There is of course a key secondary goal that demands the same kind of approach: the AUTOLING system logic will be incorporated in the learning component of Klein's computer simulation of Historical Change in Language system [8, 9]. The exact function of AUTOLING in this system is described in a paper presented at the 10th International Congress of Linguists in 1967 [10].

2.0 Morphological Analyzer

This program is essentially the work of Alicia E. Towster. Although a great deal of work has gone into it, it is relatively undeveloped from the point of view of integration into the rest of the AUTOLING system. A number of early versions have been moderately successful in analyzing problems taken from Nida [15], Koutsoudas [13], but have revealed basic problems that have led to major revisions.

One of the key problems is that of control of the gloss metalanguage. Inherent in the analytic procedure is a parallel comparison of strings in the language and their glosses. If the glosses are left in English, the program is limited to making morphological cuts that match the English gloss units. Human analysts draw upon a larger semantic data base and reinterpret the glosses as the problem demands. To incorporate this reinterpretation in a
program demands a formalization of the gloss reinterpretation process. This in turn makes mandatory the incorporation of a universal semantic list (not necessarily the semantic distinctive features of Katz and Fodor [7] or Chomsky [1]) with rules for rewriting English glosses in terms of those features, even though many might not be semantic distinctive features in English.

If the claim for 'universality' of such semantic features disturbs some linguists, they might view them as simple an inventory of all (hopefully) possible semantic units any language in the world might assign to a given English gloss.

The determination of such a list of semantic features is, of course, a momentous task. Accordingly, only a small listing of such features suitable for handling perhaps five or ten assorted text book problems have been determined.

The decision to rewrite glosses in terms of semantic features creates another major analytic problem for a program: the determination of which such features are actually distinctive; or in other terms, what bundles of semantic primitives are to be treated as units in a given language, and how they are distributed. The problem can be made clear from a 'trivial' English example:

<table>
<thead>
<tr>
<th>Form</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>I eat</td>
<td>1st person, singular, human, animate, eat, indicative, present tense.</td>
</tr>
<tr>
<td>you eat</td>
<td>2nd person, singular, human, animate, eat, indicative, present tense.</td>
</tr>
</tbody>
</table>
Clearly, 'I' and 'you' are uniquely associated with the features '1st person' and '2nd person'. But what of the remainders?

In an effort to handle the problem a distinction between primary and secondary glosses is noted. 'I' and 'you' are assigned '1st person' and '2nd person', respectively, as primary glosses. The remaining features with each input are assigned to both members of each cut but labelled as secondary glosses.

Later analysis determines the final assignment of the features, and their relabelling as primary.

Other problems that arise in analysis are those of over-cutting. Occasionally what should emerge as a single morpheme is split into two components. Proper analysis of the semantic features associated with these over-cut units at a later stage should provide enough data for recombination, but the pertinent heuristics have not yet been programmed.

Additional analytic heuristics which are in a state of flux pertain to input-sequence sensitivity. Currently, the inputs are analyzed in block units containing a fixed number of forms and glosses. The selection of new analytic blocks for intermediate stages of analysis as a function of what cuts have occurred in the preliminary stages involve heuristics that are also yet to be programmed. Heuristics for grouping allomorphs, and determining morphophonemic rules also await implementation.

3.0 Phrase Structure Heuristic Learning Program

This system, which learns unordered, context-free phrase structure rules is not yet connected to the Morphological Analysis program. It accepts
as inputs sentences written with spaces between morphological units. A multi-
path parser yields all possible parses of each input. If no complete parses are
obtained, the top nodes of the incomplete parses are ordered according to the
number of uncombined nodes remaining in each parse.

These incomplete parse top nodes serve as inputs to the basic heuristic 1.

Heuristic 1

X Y Z \[ \longrightarrow \] *S_{1} \rightarrow X Y S

[where \( S_{1} \) stands for rule: \( *S_{1} \) indicates sentence rule]

That is, coin as a rule the closure of the parse. In the case that no rules of
the provisional grammar apply to the input string, the string itself is treated
as a parse, and heuristic 1 yields:

\[ *S_{1} \rightarrow \text{morpheme}_{1} \text{ morpheme}_{2} \ldots \text{ morpheme}_{n} \]

This of course is what happens to the first input to the system. The remaining
learning heuristics cover the various cases in which units in identical environ-
ments are assigned to classes.

Heuristic 2 states that two single morphemes in identical environments
are assigned to the same class.

Heuristic 2

\[ S_{1} \rightarrow X m_{1} Y \] \[ S_{3} \rightarrow m_{1} \]
\[ S_{2} \rightarrow X m_{2} Y \] \[ S_{3} \rightarrow m_{2} \]

where \( X \) and \( Y \) are strings consisting of terminal, non-terminal or a mixed
combination of units, and \( m_{1} \) and \( m_{2} \) are single morphemes, and either
X or Y may be empty but not both.

Heuristic 2 also applies in the case that \( m_1 \) and \( m_2 \) are strings of one or more non-terminals.

Heuristic 3 states that if a terminal element (morpheme) and a single non-terminal element occur in identical environments, the morpheme is added to the non-terminal class.

**Heuristic 3.**

\[
\begin{align*}
S_1 & \rightarrow X \quad S_1 \quad Y & S_1 & \rightarrow m_1 \\
S_2 & \rightarrow X \quad m_j \quad Y & \Rightarrow & S_1 & \rightarrow m_j \\
S_1 & \rightarrow m_i & & S_1 & \rightarrow X \quad S_1 \quad Y
\end{align*}
\]

where \( X \) and \( Y \) are strings of terminal, non-terminal or mixed elements, and where \( X \) or \( Z \) may be empty but not both, and where \( m_i \) and \( m_j \) are morphemes.

There are also what might be termed negative heuristics, designed to prevent premature ad hoc rule coining. Namely, if one of the strings occurring in an environment identical with some other string should contain one morpheme plus anything else, no new rules are coined (other than to add the tree top to the rule list via heuristic 1). Also, no new rules are coined in the case of frame overlap, e.g.

\[
\begin{align*}
S_1 & \rightarrow X \quad Y & S_1 & \rightarrow X \quad A \quad B \quad Y \\
S_2 & \rightarrow X \quad A \quad Y & \text{or} & S_2 & \rightarrow X \quad A \quad A \quad B \quad Y
\end{align*}
\]
where $X$ and $Y$ are defined as above, and $A$ and $B$ are strings of non-terminals.

The last pattern matching, phrase structure learning heuristic coins recursive rules;

**Heuristic 4.**

\[
\begin{align*}
S_1 & \rightarrow X \ A \ Y \\
S_2 & \rightarrow X \ A \ A \ Y \\
S_3 & \rightarrow A \\
\implies S_3 & \rightarrow S_3 \ A \\
S_1 & \rightarrow X \ S_3 \ Y
\end{align*}
\]

where $X$ and $Y$ are defined as before and $A$ is a string of one or more non-terminals. Actually, this heuristic is not really needed for the coining of recursive rules. The other heuristics are capable of learning indirect recursion in a variety of ways.

3.1 Rule Testing Heuristics

The remaining heuristics pertain to the testing of the validity of the rules coined by heuristics 2, 3, & 4. Except for heuristic 5, they are not numbered, and are best described in terms of the flow of the program.

3.1.1 Substitution and Informant Queries

The validity of each rule coined by heuristics 2-4 are tested via substitution, and test sentence generation. Each time a new class is coined, the system inspects each rule in the grammar for tokens of elements of the new class. When such a rule is found, the system tentatively makes the substitution of the new class name for the token of its element, and
performs a test generation in the following manner:

Only full sentences are offered to the informant for acceptance or rejection. Accordingly, if the substitution is made in an unstarred rule (non-sentence), the system climbs upwards through the hierarchy of rules to the first starred rule it can find that might yield the test rule in a generation path. The program then generates randomly downwards but with a forced choice of the rule in which the substitution was made.

The test sentence is outputted on the teletype with the query:

CAN YOU SAY:

If the informant types:

YES

the substitution is made.

If the reply is NO, no substitution is made, and the test sentence is added to the illegal list. If a given rule contains two strings in the domain of the substitution, each is tried individually and serially, and if any succeeds, the next substitution is in the revised string.

If substitution should yield duplicate rules, one of the duplicates is deleted and all reference to the deleted rule are replaced by references to the remaining one.

The system also avoids the creation by substitution of unexitable node loops involving rules of the type:

\[
S_i \rightarrow S_j \\
\vdots \\
S_j \rightarrow S_i
\]
In the case that all substitutions of a newly posited rule fail, the heuristic is assumed to have failed, and the next one is tried. If all heuristics 2–4 fail for a given top node, then the remaining top nodes are each subjected to the same heuristics and testing. In the event of all failures, the original top node is added as a rule via heuristic 1. There remains one special rule coining heuristics governing class splitting applicable to morphemes. It is treated in this section because its application also involves rule testing.

Suppose the grammar contains:

\[
\begin{align*}
S_1 & \rightarrow m_1 \\
S_1 & \rightarrow m_2 \\
S_2 & \rightarrow X \ S_1 \ Y \\
S_3 & \rightarrow W \ S_1 \ Z \\
\end{align*}
\]

where \(X, Y, W, \) & \(Z\) are strings of terminals or non-terminals or both.

Suppose that, via heuristic 2, the rule

\[
S_1 \rightarrow m_3
\]

is posited, and that the substitution test succeeds in \(S_2\), but fails in \(S_3\). In that case the following rules are coined

\[
\begin{align*}
S_1 & \rightarrow m_1 \\
S_1 & \rightarrow m_2 \\
S_4 & \rightarrow S_1 \\
S_4 & \rightarrow m_3 \\
S_2 & \rightarrow X \ S_4 \ B \\
S_3 & \rightarrow W \ S_1 \ Z \\
\end{align*}
\]

and the rule \(S_1 \rightarrow m_3\) is thrown out.
3.1.2. Parsing Illegal Sentences and Recycling.

Parsing

Each test sentence rejected by the informant is added to a list of 'illegals'. All illegals determined during a single testing cycle are treated as belonging to the same frame. A frame is the interaction and processing that takes place after an input sentence from the informant. All the rule substitutions, test sentences and informant 'yes' and 'no' responses arising from the informant input sentence are treated as belonging to the same testing frame.

As each rule is coined or updated, the parsing component of the program attempts to parse each illegal sentence of the current frame. If an illegal sentence parses, then the postulated new rule or substitution is disregarded.

After each five informant inputs, the system attempts to parse all of its illegals, from all previous processing frames. Because the testing is inherently incomplete, bad rules may have slipped into the grammar. If at this time any of the illegal sentences is successfully parsed, the system enters a recycle mode, which is described below. (At one stage of development the program parsed all illegals from all frames before coining a new rule. This procedure proved too costly in processing time and the current methodology was used instead.) It also happens that some illegals are the product of posited spurious recursive rules, and may accordingly be very long as well as illegal. Such sentences if parsed to completion required as much as 5 or even 10 minutes of computer time. After determining that 99% of parsable sentences were parsed in 1 minute or less, we made the system
assume a sentence is unparseable if the completed parse is not found in less than one minute. This limitation is for the parsing of illegals only; no limitation exists on parsing time for new inputs from the informant.

Recycling

As indicated earlier, an attempt is made to parse all illegal sentences from all previous frames after every 5 informant sentence inputs. If an illegal is parsed, then some bad rule has slipped past the tests. Accordingly the system destroys its entire grammar, but saves the list of past input sentences and the list of illegal sentences. The analysis then begins anew, but with the following differences. The system first processes the recorded input list instead of immediately soliciting new inputs from the informant. The input list is reordered slightly as a rough attempt to overcome the sequence sensitivity of the learning process. Specifically, the last 5 informant sentence inputs before the recycle are put at the head of the list and processed first.

Also, the illegal sentence that caused the recycle is made a permanent member of the current frame illegal set, and an attempt is made to parse it during each rule coining step in the recycle mode. If there have been other recycles, the responsible illegals are also made permanent members of the 'current frame illegal set'.

There may be recycles within recycles with a limit of depth 3. If this limit is reached the system gives up on the whole analysis. No limit exists for unnested recycles.
3.2 Test Problems

In the following copies of actual teletype output, the only human inputs are:

1. An input sentence following a program generated NEXT.
2. *TYPE in response to a computer generated NEXT which is a request by the human for a listing of the current grammar.
3. YES and NO in response to program requests, CAN YOU SAY THIS followed by a program generated test sentence.

Other messages the system can accept are:

5. A *SAVE followed by a single digit makes the system periodically save the total state of the learning process on disc file at periodic intervals. At each saving point the system outputs a message, indicated the name of the saved file in terms of the initial digit after the first *SAVE message, plus a sequence number.
6. A *RESTART message followed by a file number reinitializes the program to the state stored under that file name. This feature permits the continued analysis, over a long period of time, of up to 9 different languages. It also permits partial backup restarts in case of accidents or erroneous inputs.

Some of the problems were taken from Koutsoudas' test, Writing Transformational Grammars: An Introduction [13]. The sentences were entered as inputs in the order they occurred in each problem. Often the program generated
as a test case a problem sentence not yet entered. In this case the informant usually omitted the repetition. Each time the computer grammar was adequate for parsing an input, it outputed the message:

PARSED OK

Often, the informant's responses to the machine's test utterances did not come from a full knowledge of the language. The informant behaved consistently, but may have said YES and NO to hypothetical constructions that may have deserved different replies.
3.2.1 Embedding Illustration: Artificial Language.

BEGIN AUTOLING
NEXT: \( \overline{X \ Y} \)

NEXT: \( \overline{*TYPE} \)
\[ *S_1 := X \ Y \]

\[ \overline{X \ X \ Y \ Y} \]

NEXT: \( \overline{*TYPE} \)
\[ *S_1 := X \ Y \]
\[ *S_2 := X \ S_1 \ Y \]

\[ \overline{X \ X \ X \ Y \ Y \ Y} \]
CAN YOU SAY:
\[ X \ X \ X \ X \ X \ X \ Y \ Y \ Y \ Y \ Y \ Y \]

YES →

NEXT: \( \overline{*TYPE} \)
\[ *S_1 := X \ Y \]
\[ *S_2 := X \ S_3 \ Y \]
\[ S_3 := S_2 \]
\[ S_3 := S_1 \]

\[ \overline{X \ X \ X \ X \ Y \ Y \ Y \ Y} \]
-PARSED OK-

NEXT:
\[ \overline{X \ X \ X \ X \ X \ X \ X \ X \ X \ Y \ Y \ Y \ Y \ Y \ Y \ Y \ Y \ Y \ Y} \]
-PARSED OK-
3.2.2 English I

BEGIN AUTOLING
NEXT:
I WANT HIM TO GO ←

NEXT: I NEED HIM TO GO ←

NEXT: **TYPE
*S1 := I S2 HIM TO GO
S2 := NEED
S2 := WANT

I WANT HER TO GO ←
CAN YOU SAY:
   I NEED HER TO GO
YES ←

NEXT: **TYPE ←
*S1 := I S2 S3 TO GO
S2 := NEED
S2 := WANT
S3 := HER
S3 := HIM

YOU WANT HER TO GO ←
CAN YOU SAY:
   YOU NEED HER TO GO
YES ←

NEXT: **TYPE ←
*S1 := S4 S2 S3 TO GO
S2 := NEED
S2 := WANT
S3 := HER
S3 := HIM
S4 := YOU
S4 := I

YOU WANT HIM TO RUN ←
CAN YOU SAY:
   YOU NEED HIM TO RUN
YES ←
**PARSING ILLEGALS**
NEXT: HE WANT S HER TO GO←
CAN YOU SAY:
    HE WANT S HIM TO RUN
YES←

NEXT: *TYPE←
*S1 := S4 S2 S3 TO S5
S2 := NEED
S2 := WANT
S3 := HER
S3 := HIM
S4 := YOU
S4 := I
S5 := RUN
S5 := GO
*S6 := HE S2 S S3 TO S5
SHE WANT S HIM TO RUN←
CAN YOU SAY:
    SHE WANT S HER TO GO
YES←

NEXT: *TYPE←
*S1 := S4 S2 S3 TO S5
S2 := NEED
S2 := WANT
S3 := HER
S3 := HIM
S4 := YOU
S4 := I
S5 := RUN
S5 := GO
*S6 := S7 S2 S S3 TO S5
S7 := SHE
S7 := HE
SHE WANT S HER TO GO←
-PARSED OK-
BEGIN AUTOLING
NEXT:

EAT THE CAT ←

NEXT: EAT A CAT ←

NEXT: EAT THE CAT S ←
CAN YOU SAY:
    EAT A CAT S
NO ←

NEXT: *TYPE ←
*S1 := EAT S2 CAT
S2 := A
S2 := THE
*S3 := EAT THE CAT S

EAT A DOG ←
CAN YOU SAY:
    EAT THE DOG
YES ←
CAN YOU SAY:
    EAT THE DOG S
YES ←

NEXT: *TYPE ←
*S1 := EAT S2 S4
S2 := A
S2 := THE
*S3 := EAT THE S4 S
S4 := DOG
S4 := CAT

KNOW THE DOG ←
CAN YOU SAY:
    KNOW A DOG
YES ←
CAN YOU SAY:
    KNOW THE CAT S
YES ←
*PARSING ILLEGALS*
NEXT: *TYPE ←
*S1 := S5 S2 S4
S2 := A
S2 := THE
*S3 := S5 THE S4 S
S4 := DOG
S4 := CAT
S5 := KNOW
S5 := EAT

EAT THESE CAT S ←
CAN YOU SAY:
    EAT THESE DOG S
YES ←

NEXT: *TYPE ←
*S1 := S5 S2 S4
S2 := A
S2 := THE
*S3 := S5 S6 S4 S
S4 := DOG
S4 := CAT
S5 := KNOW
S5 := EAT
S6 := THESE
S6 := THE

EAT SOME CAT S ←

NEXT: EAT SOME CAT ←

NEXT: *TYPE ←
*S1 := S5 S2 S4
S2 := SOME
S2 := A
S2 := THE
*S3 := S5 S6 S4 S
S4 := DOG
S4 := CAT
S5 := KNOW
S5 := EAT
S6 := SOME
S6 := THESE
S6 := THE

EAT THOSE CAT S ←
NEXT: I EAT A CAT←

CAN YOU SAY:
 I KNOW A CAT

YES←

*PARSING ILLEGALS*

NEXT: *TYPE←
*S1 := S5 S2 S4
S2 := SOME
S2 := A
S2 := THE
*S3 := S5 S6 S4 S
S4 := DOG
S4 := CAT
S5 := KNOW
S5 := EAT
S6 := THOSE
S6 := SOME
S6 := THESE
S6 := THE
*S7 := I S1

YOU EAT A CAT←

CAN YOU SAY:
 YOU KNOW A DOG

YES←

NEXT: *TYPE←
*S1 := S5 S2 S4
S2 := SOME
S2 := A
S2 := THE
*S3 := S5 S6 S4 S
S4 := DOG
S4 := CAT
S5 := KNOW
S5 := EAT
S6 := THOSE
S6 := SOME
S6 := THESE
*S7 := S8 S1
S8 := YOU
S8 := I
THE DOG S KNOW A CAT

CAN YOU SAY:
THOSE CAT S KNOW SOME DOG
YES

NEXT:  *TYPE
*S1 := S5 S2 S4
S2 := SOME
S2 := A
S2 := THE
*S3 := S5 S6 S4 S
S4 := DOG
S4 := CAT
S5 := KNOW
S5 := EAT
S6 := THOSE
S6 := SOME
S6 := THESE
S6 := THE
*S7 := S8 S1
S8 := YOU
S8 := I
*S9 := S6 S4 S S1
3.2.4 Latin: Koutsoudas #26

PROBLEM 26: LATIN

1. puer virum videt  
   The boy sees the man.
2. vir puerum videt  
   The man sees the boy.
3. puer virum defendit  
   The boy defends the man.
4. vir puero nocet  
   The man harms the boy.
5. puer viro nocet  
   The boy harms the man.
6. puer viro subvenit  
   The boy helps the man.
7. puer viri meminit  
   The boy remembers the man.
8. puer viro meminit  
   The boy remembers the man.
BEGIN AUTOLING

NEXT:
PUER VIR UM VIDE T←

NEXT:
VIR PUER UM VIDE T←

NEXT:
PUER VIR UM DEFENDI T←
CAN YOU SAY:
VIR PUER UM DEFENDI T
YES←

NEXT: *TYPE←
*S1 := PUER VIR UM S3 T
*S2 := VIR PUER UM S3 T
S3 := DEFENDI
S3 := VIDE

VIR PUER O NOCE T←

NEXT: PUER VIR O NOCE T←
*PARSING ILLEGALS*

NEXT: PUER VIR O SUBVENI T←
CAN YOU SAY:
VIR PUER O SUBVENI T
YES←

NEXT: *TYPE←
*S1 := PUER VIR UM S3 T
*S2 := VIR PUER UM S3 T
S3 := DEFENDI
S3 := VIDE
*S4 := VIR PUER O S6 T
*S5 := PUER VIR O S6 T
S6 := SUBVENI
S6 := NOCE

PUER VIR I MEMINI T←

NEXT: PUER VIR O MEMINI T←
CAN YOU SAY:
VIR PUER O MEMINI T
YES←
CAN YOU SAY:
PUER VIR I SUBVENI T
NO←
NEXT: *TYPE ←
*S1 := Puer Vir Um S3 T
*S2 := Vir Puer Um S3 T
S3 := Defendi
S3 := Vide
*S4 := Vir Puer O S6 T
*S5 := Puer Vir O S6 T
S6 := Memini
S6 := Subveni
S6 := Noce
*S7 := Puer Vir I Memini T

[The sentence inputs that follow have been added to the problem]

VIR Puer Um Vide T ←
-Parsed OK-

NEXT: VIR Puer I Memini T ←
Can you say:
  Vir Puer I Noce T
NO ←
*Parsing Illegals*

NEXT: *TYPE ←
*S1 := Puer Vir Um S3 T
*S2 := Vir Puer Um S3 T
S3 := Defendi
S3 := Vide
*S4 := Vir Puer O S6 T
*S5 := Puer Vir O S6 T
S6 := Memini
S6 := Subveni
S6 := Noce
*S7 := Puer Vir I Memini T
*S12 := Vir Puer I Memini T
Puer Vir O Noce T ←
-Parsed OK-

NEXT: Puer Oppid Um Vide T ←
Can you say:
  Puer Oppid Um Defendi T
YES ←
Can you say:
  Oppid Puer O Subveni T
YES ←
Can you say:
  Oppid Puer I Memini T
YES ←
CAN YOU SAY:
   OPPID Puer UM VIDE T
YES ←
CAN YOU SAY:
   Puer OPPID O MEMINI T
YES ←
CAN YOU SAY:
   Puer OPPID I MEMINI T
YES ←

NEXT: *TYPE ←
*S1 := Puer S13 UM S3 T
*S2 := S13 Puer UM S3 T
S3 := DEFENDI
S3 := VIDE
*S4 := S13 Puer O S6 T
*S5 := Puer S13 O S6 T
S6 := MEMINI
S6 := SUBVENI
S6 := NOCE
*S7 := Puer S13 I MEMINI T
*S12 := S13 Puer I MEMINI T
S13 := OPPID
S13 := VIR

VIR OPPID UM VIDE T ←
CAN YOU SAY:
   OPPID OPPID UM VIDE T
YES ←
CAN YOU SAY:
   Puer Puer I MEMINI T
YES ←
CAN YOU SAY:
   Puer Puer O MEMINI T
YES ←
CAN YOU SAY:
   Puer Puer UM VIDE T
YES ←
CAN YOU SAY:
   Puer Puer O NOCE T
YES ←
CAN YOU SAY:
   OPPID OPPID O SUBVENI T
YES ←
CAN YOU SAY:
   Puer Puer UM DEFENDI T
YES ←
PROBLEM 58: ROGLAI

1. ama naw
   Father went.

2. ama naw tubray
   Father went yesterday.

3. aday ?bak bu
   The child ate rice.

4. ama naw juday aday ?bak bu
   Father went after the child ate rice.

5. aday ?bak bu tubray
   The child ate rice yesterday.

6. aday ?bak bu juday ama naw
   The child ate rice after the father went.

7. tubray ama naw
   Yesterday father went.

8. juday aday ?bak bu ama naw
   After the child ate rice, the father went.

9. tubray aday ?bak bu
   Yesterday the child ate rice.

10. juday ama naw aday ?bak bu
    After father went, the child ate rice.

11. aday naw musu-p
    The child went early in the morning.

12. ama ?bak ika-t
    Father ate fish.

13. aday naw juday ama ?bak ika-t
    The child went after the father ate fish.

14. aday naw juma ama ?bak ika-t
    The child went before father ate fish.

15. musu-p aday naw
    Early in the morning the child went.

16. juday ama ?bak ika-t aday naw
    After the father ate fish, the child went.

17. juma ama ?bak ika-t aday naw
    Before the father ate fish, the child went.

18. ama naw juma aday ?bak bu
    Father went before the child ate rice.

19. juma aday ?bak bu ama naw
    Before the child ate rice, the father went.

20. aday ?bak ika-t tubray
    The child ate fish yesterday.

21. tubray aday ?bak ika-t
    Yesterday the child ate fish.

22. juday ama naw aday ?bak ika-t
    After father went, the child ate fish.

23. juma ama naw aday ?bak ika-t
    Before father went, the child ate fish.
BEGIN AUTOLING
NEXT: AMA NAW ←

NEXT: ADEY NAW ← [Added input]

NEXT: AMA NAW TUBREY ←
CAN YOU SAY:
  ADEY NAW TUBREY
YES ←

NEXT: ADEY QBEK BU ←
CAN YOU SAY:
  AMA QBEK BU
YES ←

NEXT: AMA NAW JUDEY ADEY QBEK BU ←
CAN YOU SAY:
  ADEY NAW JUDEY AMA QBEK BU
YES ←
*PARSING ILLEGALS*

NEXT: ADEY QBEK BU TUBREY ←
CAN YOU SAY:
  AMA QBEK BU TUBREY
YES ←
CAN YOU SAY:
  AMA QBEK BU JUDEY AMA QBEK BU
YES ←
CAN YOU SAY:
  AMA NAW JUDEY ADEY NAW
YES ←

NEXT: *TYPE ←
*S1 := S2 NAW
S2 := ADEY
S2 := AMA
*S3 := S6 TUBREY
*S4 := S2 QBEK BU
*S5 := S6 JUDEY S6
S6 := S4
S6 := S1

ADEY QBEK BU JUDEY AMA NAW ←
-PARSED OK-
NEXT:  TUBREY AMA NAW ←
CAN YOU SAY:
TUBREY ADEY QBEK BU
YES ←

NEXT:  JUDEY ADEY QBEK BU AMA NAW ←
CAN YOU SAY:
JUDEY AMA QBEK BU ADEY QBEK BU
YES ←

NEXT:  JUDEY AMA NAW ADEY QBEK BU ←
-PARSED OK-
*PARSING ILLEGALS*

NEXT:  ADEY NAW MUSUUP ←
CAN YOU SAY:
ADEY QBEK BU MUSUUP
YES ←
CAN YOU SAY:
MUSUUP ADEY QBEK BU
YES ←

NEXT:  *TYPE ←
*S1 := S2 NAW
S2 := ADEY
S2 := AMA
*S3 := S6 S9
*S4 := S2 QBEK BU
*S5 := S6 JUDEY S6
S6 := S4
S6 := S1
*S7 := S9 S6
*S8 := JUDEY S6 S6
S9 := MUSUUP
S9 := TUBREY

AMA QBEK IKAAT ←
CAN YOU SAY:
ADEY QBEK IKAAT
YES ←
NEXT: *TYPE ←
*S1 := S2 NAW
S2 := ADEY
S2 := AMA
*S3 := S6 S9
*S4 := S2 QBEK S10
*S5 := S6 JUDEY S6
S6 := S4
S6 := S1
*S7 := S9 S6
*S8 := JUDEY S6 S6
S9 := MUSUUP
S9 := TUBREY
S10 := IKAAT
S10 := BU

ADEY NAW JUDEY AMA QBEK IKAAT ←
-PARSED OK-

NEXT: ADEY NAW JUMA AMA QBEK IKAAT ←
CAN YOU SAY:
    ADEY NAW JUMA AMA NAW
YES ←
CAN YOU SAY
    JUMA AMA NAW AMA NAW
YES ←
CAN YOU SAY:
   AMA NAW JUDEY AMA NAW
YES ←
CAN YOU SAY:
    JUDEY ADEY NAW ADEY NAW
YES ←
CAN YOU SAY:
    TUBREY AMA QBEK IKAAT
YES ←
CAN YOU SAY:
    TUBREY ADEY NAW
YES ←
CAN YOU SAY:
    AMA QBEK IKAAT TUBREY
YES ←
CAN YOU SAY:
    AMA QBEK BU JUMA AMA NAW
YES ←
CAN YOU SAY:
    MUSUUP ADEY NAW ADEY QBEK IKAAT AMA QBEK BU
NO ←
CAN YOU SAY:
   ADEY QBEK BU JUMA AMA QBEK IKAAT AMA NAW ADEY NAW AMA QBEK BU
NO ←

NEXT: *TYPE ←
  *S1 := S2  NAW
  S2 := ADEY
  S2 := AMA
  *S3 := S5
  *S4 := S2 QBEK S10
  *S5 := S6  S12
  S6 := S4
  S6 := S1
  *S7 := S8
  *S8 := S12  S6
  S9 := MUSUUP
  S9 := TUBREY
  S10 := IKAAT
  S10 := BU
  S11 := JUMA
  S11 := JUDEY
  S12 := S11  S6
  S12 := S9

MUSUUP ADEY NAW ←
-PARSED OK-
*PARSING ILLEGALS*

NEXT:  JUDEY AMA QBEK IKAAT ADEY NAW ←
-PARSED OK-

NEXT:  JUMA AMA QBEK IKAAT ADEY NAW ←
-PARSED OK-

NEXT:  AMA NAW JUMA ADEY QBEK BU ←
-PARSED OK-

NEXT:  JUMA ADEY QBEK BU AMA NAW ←
-PARSED OK-

NEXT:  ADEY QBEK IKAAT TUBREY←
-PARSED OK-

NEXT:  TUBREY ADEY QBEK IKAAT ←
-PARSED OK-

NEXT:  JUDEY AMA NAW ADEY QBEK IKAAT ←
-PARSED OK-
As indicated earlier, fewer heuristics operate on strings containing individual morphemes. Accordingly, to add more members to the verb classes, the cover morphemes VERBTRAN and VERBINTRAN were introduced into the inputs with the following result:

AMA VERBINTRAN ←
CAN YOU SAY:
    ADEY VERBINTRAN
YES ←

NEXT: *TYPE ←
*S1 := S2 S15
S2 := ADEY
S2 := AMA
*S3 := S5
*S4 := S2 QBEK S10
*S5 := S6 S12
S6 := S4
S6 := S1
*S7 := S8
*S8 := S12 S6
S9 := MUSUUP
S9 := TUBREY
S10 := IKAAAT
S10 := BU
S11 := JUMA
S11 := JUDEY
S12 := S11 S6
S12 := S9
S10 := IKAAT 
S10 := BU 
S11 := JUMA 
S11 := JUDEY 
S12 := S11 S6 
S12 := S9 
S15 := VERBINTRAN 
S15 := NAW 

ADEY VERBTRAN BU ←
CAN YOU SAY:
   ADEY VERBTRAN IKAAT
YES ←
CAN YOU SAY:
   AMA VERBTRAN BU
YES ←
CAN YOU SAY:
   AMA VERBTRAN IKAAT
YES ←
CAN YOU SAY:
   ADEY VERBINTRAN JUMA AMA NAW
YES ←

NEXT: *TYPE ←
*S1 := S4
S2 := ADEY
S2 := AMA
*S3 := S5
*S4 := S2 S17
*S5 := S6 S12
S6 := S1
*S7 := S8
*S8 := S12 S6
S9 := MUSUUP
S9 := TUBREY
S10 := IKAAT 
S10 := BU 
S11 := JUMA 
S11 := JUDEY 
S12 := S11 S6 
S12 := S9 
S15 := VERBINTRAN 
S15 := NAW 
S16 := VERBTRAN 
S16 := QBEEK 
S17 := S16 S10 
S17 := S15
Some additional sentences were added to the corpus to link the two problems.

**Problem 43: Indonesian**

1. guru itu makan soman̄ka
   teacher the eat watermelon
   *The teacher is eating watermelon.*

2. guru itu rupãna makan soman̄ka
   *The teacher is apparently eating watermelon.*

3. rupãna guru itu makan soman̄ka
   *The teacher is apparently eating watermelon.*

4. guru itu makan soman̄ka rupãna
   *The teacher is apparently eating watermelon.*

5. guru itu disini kamarin makan soman̄ka
   teacher the here yesterday eat watermelon
   *The teacher ate watermelon here yesterday.*

6. disini kamarin guru'itu makan soman̄ka
   *The teacher ate watermelon here yesterday.*

7. guru itu makan soman̄ka disini kamarin
   *The teacher ate watermelon here yesterday.*

8. makan soman̄ka guru itu
   *The teacher is eating watermelon.*

9. rupãna makan soman̄ka guru itu
    *Apparenty the teacher is eating watermelon.*

10. makan soman̄ka guru itu rupãna
    *The teacher is eating watermelon apparently.*

11. makan soman̄ka rupãna guru itu
    *The teacher apparently is eating watermelon.*

12. disini kamarin makan soman̄ka guru itu
    *The teacher ate watermelon here yesterday.*

13. makan soman̄ka guru itu disini kamarin
    *The teacher ate watermelon here yesterday.*

14. makan soman̄ka disini kamarin guru itu
    *The teacher ate watermelon here yesterday.*
9. INDONESIAN

1. oraŋ itu makan kacan kamarin
   The man ate peanuts yesterday.
2. oraŋ itu kamarin makan kacan
   The man ate peanuts yesterday.
3. kamarin oraŋ itu makan kacan
   The man ate peanuts yesterday.
4. kamarin makan kacan oraŋ itu
   The man ate peanuts yesterday.
5. makan kacaŋ kamarin oraŋ itu
   The child ate peanuts yesterday.
6. makan kacaŋ oraŋ itu kamarin
   The child bought peanuts yesterday.
7. ana? itu makan kacan kamarin
   The teacher bought peanuts yesterday.
8. ana? itu kamarin mamboli kacaŋ
   The teacher bought cookies yesterday.
9. kamarin guru itu mamboli kacaŋ
   The man bought cookies yesterday.
10. kamarin mamboli kuwe guru itu
    The child ate peanuts yesterday.
11. mamboli kuwe oraŋ itu kamarin
    When did the man eat cookies?
12. makan kacaŋ kamarin ana? itu
    When did the child buy cookies?
13. kapan oraŋ itu makan kuwe
    When did the teacher eat peanuts?
14. kapan mamboli kuwe ana? itu
    When did the man buy peanuts?
15. guru itu kapan makan kacaŋ
    When did the child buy cookies?
16. mamboli kacaŋ kapan oraŋ itu
    When did the teacher eat cookies?
17. mamboli kuwe ana? itu kapan
    Who ate peanuts yesterday?
18. guru itu makan kuwe kapan
    Who ate cookies yesterday?
19. siapa makan kacaŋ kamarin
    Who bought cookies yesterday?
20. siapa kamarin makan kuwe
    Who bought peanuts yesterday?
21. kamarin siapa mamboli kuwe
    Who ate peanuts yesterday?
22. yaŋ kamarin mamboli kacaŋ siapa
    Who bought cookies yesterday?
23. yaŋ makan kacaŋ siapa kamarin
    Who ate peanuts yesterday?
24. yaŋ mamboli kuwe kamarin siapa
    Who bought cookies yesterday?
BEGIN AUTOLING
NEXT:  GURU ITU MAKA_N SEMANGKA ←

NEXT:  ORANG ITU MAKA_N SEMANGKA ←  [added to problem]

NEXT:  ANAQ ITU MAKA_N SEMANGKA ←  [added to problem]

NEXT:  GURU ITU MEMBELI SEMANGKA ←  [added to problem]
CAN YOU SAY:
  ANAQ ITU MEMBELI SEMANGKA
YES ←
NEXT:  GURU ITU MEMBELI KUWE ←  [added to problem]
CAN YOU SAY:
  ORANG ITU MAKA_N KUWE
YES ←
*PARSING ILLEGALS*

NEXT:  GURU ITU MAKA_N KACANG ←  [added to problem]

NEXT:  *TYPE ←
*S1 := S2 ITU S3 S4
S2 := ANAQ
S2 := ORANG
S2 := GURU
S3 := MEMBELI
S3 := MAKA_N
S4 := KACANG
S4 := KUWE
S4 := SEMANGKA

GURU ITU RUPANYA MAKA_N SEMANGKA ←

CAN YOU SAY:
  GURU ITU RUPANYA MEMBELI KACANG
YES ←

NEXT:  RUPANYA GURU ITU MAKA_N SEMANGKA ←
CAN YOU SAY:
  RUPANYA ANAQ ITU MEMBELI SEMANGKA
YES ←

NEXT:  GURU ITU MAKA_N SEMANGKA RUPANYA ←
CAN YOU SAY:
  GURU ITU MEMBELI SEMANGKA RUPANYA
YES ←
NEXT: GURU ITU DISINI KAMARIN MAKAN SEMANGKA ←
CAN YOU SAY:
   ORANG ITU DISINI KAMARIN MEMBELI KUWE
YES ←
*PARSING ILLEGALS*

NEXT: DISINI KAMARIN GURU ITU MAKAN SEMANGKA ←
CAN YOU SAY:
   DISINI KAMARIN ANAQ ITU MAKAN KUWE
YES ←

NEXT: GURU ITU MAKAN SEMANGKA DISINI KAMARIN ←
CAN YOU SAY:
   GURU ITU MAKAN KUWE DISINI KAMARIN
YES ←

NEXT: *TYPE ←
   *S1 := S2 ITU S3 S4
   S2 := ANAQ
   S2 := ORANG
   S2 := GURU
   S3 := MEMBELI
   S3 := MAKAN
   S4 := KACANG
   S4 := SEMANGKA
   *S5 := S2 ITU RUPANYA S3 S4
   *S6 := RUPANYA S1
   *S7 := S1 RUPANYA
   *S8 := S2 ITU DISINI KAMARIN S3 S4
   *S9 := DISINI KAMARIN S1
   *S10 := S1 DISINI KAMARIN

MAKAN SEMANGKA GURU ITU ←
CAN YOU SAY:
   MEMBELI KUWE ORANG ITU
YES ←

NEXT: RUPANYA MAKAN SEMANGKA GURU ITU ←
CAN YOU SAY:
   RUPANYA MAKAN KUWE ORANG ITU
YES ←
CAN YOU SAY:
   MAKAN KUWE ANAQ ITU DISINI KAMARIN
YES ←
CAN YOU SAY:
   MAKAN KACANG GURU ITU RUPANYA
YES ←
CAN YOU SAY:
   DISINI KAMARIN MEMBELI KUWE ORANG ITU
YES ←
NEXT: *TYPE ←
*S1 := S2 ITU S3 S4
S2 := ANAQ
S2 := ORANG
S2 := GURU
S3 := MEMBELI
S3 := MAKAN
S4 := KACANG
S4 := KUWE
S4 := SEMANGKA
*S5 := S2 ITU RUPANYA S3 S4
*S6 := RUPANYA S12
*S7 := S12 RUPANYA
*S8 := S2 ITU DISINI KAMARIN S3 S4
*S9 := DISINI KAMARIN S12
*S10 := S12 DISINI KAMARIN
*S11 := S3 S4 S2 ITU
S12 := S11
S12 := S1

MAKAN SEMANGKA GURU ITU RUPANYA ←
-PARSED OK-
*PARSING ILLEGALS*

NEXT: MAKAN SEMANGKA RUPANYA GURU ITU ←
CAN YOU SAY:
    MEMBELI KUWE RUPANYA GURU ITU
YES ←

NEXT: DISINI KAMARIN MAKAN SEMANGKA GURU ITU ←
-PARSED OK-

NEXT: MAKAN SEMANGKA GURU ITU DISINI KAMARIN ←
-PARSED OK-

NEXT: MAKAN SEMANGKA DISINI KAMARIN GURU ITU ←
CAN YOU SAY:
    MEMBELI KUWE DISINI KAMARIN ANAQ ITU
YES ←
NEXT: *TYPE ←
*S1 := S2 ITU S3 S4
S2 := ANAQ
S2 := ORANG
S2 := GURU
S3 := MEMBELI
S3 := MAKAN
S4 := KACANG
S4 := KUWE
S4 := SEMANGKA
*S5 := S2 ITU RUPANYA S3 S4
*S6 := RUPANYA S12
*S7 := S12 RUPANYA
*S8 := S2 ITU DISINI KAMARIN S3 S4
*S9 := DISINI KAMARIN S12
*S10 := S12 DISINI KAMARIN
*S11 := S3 S4 S2 ITU
S12 := S11
S12 := S1
*S13 := S3 S4 RUPANYA S2 ITU
*S14 := S3 S4 DISINI KAMARIN S2 ITU

ORANG ITU MAKAN KACANG KAMARIN ←
CAN YOU SAY:
    MAKAN KACANG ORANG ITU KAMARIN

YES ←
CAN YOU SAY:
    KAMARIN ANAQ ITU MAKAN KUWE

YES ←
CAN YOU SAY:
    ORANG ITU KAMARIN MEMBELI KUWE

YES ←
CAN YOU SAY:
    MAKAN SEMANGKA RUPANYA ORANG ITU

YES ←
CAN YOU SAY:
    ANAQ ITU DISINI RUPANYA MEMBELI SEMANGKA

NO ←
CAN YOU SAY:
    DISINI RUPANYA ANAQ ITU MAKAN SEMANGKA

NO ←
CAN YOU SAY:
    MAKAN SEMANGKA ORANG ITU DISINI RUPANYA

NO ←
CAN YOU SAY:
    MAKAN KACANG DISINI RUPANYA ANAQ ITU

NO ←
PARSING ILLEGALS*
NEXT: *TYPE ←
*S1 := S2 ITU S3 S4
S2 := ANAQ
S2 := ORANG
S2 := GURU
S3 := MEMBELI
S3 := MAKAN
S4 := KUWE
S4 := SEMANGKA
*S5 := S2 ITU S15 S3 S4
*S6 := S15 S12
*S7 := S12 S15
*S8 := S2 ITU DISINI KAMARIN S3 S4
*S9 := DISINI KAMARIN S12
*S10 := S12 DISINI KAMARIN
*S11 := S3 S4 S2 ITU
S12 := S11
S12 := S1
*S13 := S3 S4 S15 S2 ITU
*S14 := S3 S4 DISINI KAMARIN S2 ITU
S15 := KAMARIN
S15 := RUPANYA

ORANG ITU KAMARIN MAKAN KACANG ←
-Parsed OK-

NEXT: KAMARIN ORANG ITU MAKAN KACANG ←
-Parsed OK-

NEXT: KAMARIN MAKAN KACANG ORANG ITU ←
-Parsed OK-

NEXT: MAKAN KACANG KAMARIN ORANG ITU ←
-Parsed OK-

NEXT: MAKAN KACANG ORANG ITU KAMARIN ←
-Parsed OK-

NEXT: ANAQ ITU MAKAN KACANG KAMARIN ←
-Parsed OK-

NEXT: ANAQ ITU KAMARIN MEMBELI KACANG ←
-Parsed OK-

NEXT: KAMARIN GURU ITU MEMBELI KACANG ←
-Parsed OK-
NEXT: KAMARIN MEMBELI KUWE GURU ITU ←
- PARSED OK -

NEXT: MEMBELI KUWE ORANG ITU KAMARIN ←
- PARSED OK -

NEXT: MAKAN KACANG KAMARIN ANAQ ITU ←
- PARSED OK -

NEXT: KAPAN ORANG ITU MAKAN KUWE ←
CAN YOU SAY:
   MEMBELI SEMANGKA KAPAN GURU ITU
YES ←
CAN YOU SAY:
   GURU ITU KAPAN MEMBELI KACANG
YES ←
CAN YOU SAY:
   MAKAN KACANG GURU ITU KAPAN
YES ←

NEXT: *TYPE ←
*S1 := S2 ITU S3 S4
S2 := ANAQ
S2 := ORANG
S2 := GURU
S3 := MEMBELI
S3 := MAKAN
S4 := KACANG
S4 := KUWE
S4 := SEMANGKA
*S5 := S2 ITU S15 S3 S4
*S6 := S15 S12
*S7 := S12 S15
*S8 := S2 ITU DISINI KAMARIN S3 S4
*S9 := DISINI KAMARIN S12
*S10 := S12 DISINI KAMARIN
*S11 := S3 S4 S2 ITU
S12 := S11
S12 := S1
*S13 := S3 S4 S15 S2 ITU
*S14 := S3 S4 DISINI KAMARIN S2 ITU
S15 := KAPAN
S15 := KAMARIN
S15 := RUPANYA

KAPAN MEMBELI KUWE ANAQ ITU ←
- PARSED OK -
NEXT: GURU ITU KAPAN MAKAN KACANG ←
-PARSED OK-

NEXT: MEMBELI KACANG KAPAN ORANGE ITU ←
-PARSED OK-

NEXT: MEMBELI KUWE ANAQ ITU KAPAN ←
-PARSED OK-
*PARSING ILLEGALS*

NEXT: GURU ITU MAKAN KUWE KAPAN ←
-PARSED OK-

NEXT: SIAPA MAKAN KACANG KAMARIN ←
CAN YOU SAY:
   SIAPA MAKAN KACANG KAPAN
NO ←

NEXT: SIAPA KAMARIN MAKAN KUWE ←
CAN YOU SAY:
   SIAPA RUPANYA MAKAN SEMANGKA
NO ←

NEXT: *TYPE ←
*S1 := S2 ITU S3 S4
S2 := ANAQ
S2 := ORANG
S2 := GURU
S3 := MEMBELI
S3 := MAKAN
S4 := KACANG
S4 := KUWE
S4 := SEMANGKA
*S5 := S2 ITU S15 S3 S4
*S6 := S15 S12
*S7 := S12 S15
*S8 := S2 ITU DISINI KAMARIN S3 S4
*S9 := DISINI KAMARIN S12
*S10 := S12 DISINI KAMARIN
*S11 := S3 S4 S2 ITU
S12 := S11
S12 := S1
*S13 := S3 S4 S15 S2 ITU
*S14 := S3 S4 DISINI KAMARIN S2 ITU
S15 := KAPAN
S15 := KAMARIN
S15 := RUPANYA
*S16 := SIAPA MAKAN KACANG KAMARIN
*S17 := SIAPA KAMARIN MAKAN KUWE
KAMARIN SIAPA MEMBELI KUWE ←
CAN YOU SAY:
   KAPAN SIAPA MEMBELI SEMANGKA
NO ←

NEXT: YANG KAMARIN MEMBELI KACANG SIAPA ←
CAN YOU SAY:
   YANG RUPANYA MAKAN KUWE SIAPA
NO ←

NEXT: YANG MAKAN KACANG SIAPA KAMARIN ←
CAN YOU SAY:
   YANG MAKAN SEMANGKA SIAPA KAPAN
NO ←
*PARSIN ILLEGALS*

NEXT: YANG MEMBELI KUWE KAMARIN SIAPA ←
CAN YOU SAY:
   YANG MEMBELI KACANG RUPANYA SIAPA
NO ←

NEXT: *TYPE ←
*S1 := S2 ITU S3 S4
S2 := ANAQ
S2 := ORANG
S2 := GURU
S3 := MEMBELI
S3 := MAKAN
S4 := KACANG
S4 := KUWE
S4 := SEMANGKA
*S5 := S2 ITU S15 S3 S4
*S6 := S15 S12
*S7 := S12 S15
*S8 := S2 ITU DISINI KAMARIN S3 S4
*S9 := DISINI KAMARIN S12
*S10 := S12 DISINI KAMARIN
*S11 := S3 S4 S2 ITU
S12 := S11
S12 := S1
*S13 := S3 S4 S15 S2 ITU
*S14 := S3 S4 DISINI KAMARIN S2 ITU
S15 := KAPAN
S15 := KAMARIN
S15 := RUPANYA
*S16 := SIAPA MAKAN KACANG KAMARIN
*S17 := SIAPA KAMARIN MAKAN KUWE
*S18 := KAMARIN SIAPA MEMBELI KUWE
*S19 := YANG KAMARIN MEMBELI KACANG SIAPA
*S20 := YANG MAKAN KACANG SIAPA KAMARIN
*S21 := YANG MEMBELI KUWE KAMARIN SIAPA
SIAPA MEMBELI KACANG KAMARIN ← [added to the problem]
CAN YOU SAY:
    SIAPA MEMBELI SEMANGKA KAMARIN
YES ←

NEXT: SIAPA MAkan KUWE KAMARIN ← [added to the problem]
- PARSED OK -

NEXT: SIAPA MAkan KACANG KAMARIN ← [added to the problem]
- PARSED OK -

NEXT: *TYPE ←
*S1 := S2 ITU S3 S4
S2 := ANAQ
S2 := ORANG
S2 := GURU
S3 := MEMBELI
S3 := MAkan
S4 := KACANG
S4 := KUWE
S4 := SEMANGKA
*S5 := S2 ITU S15 S3 S4
*S6 := S15 S12
*S7 := S12 S15
*S8 := S2 ITU DISINI KAMARIN S3 S4
*S9 := DISINI KAMARIN S12
*S10 := S12 DISINI KAMARIN
*S11 := S3 S4 S2 ITU
S12 := S11
S12 := S1
*S13 := S3 S4 S15 S2 ITU
*S14 := S3 S4 DISINI KAMARIN S2 ITU
S15 := KAPAN
S15 := KAMARIN
S15 := RUPANYA
*S16 := SIAPA MAkan KACANG KAMARIN
*S17 := SIAPA KAMARIN MAkan KUWE
*S18 := KAMARIN SIAPA MEMBELI KUWE
*S19 := YANG KAMARIN MEMBELI KACANG SIAPA
*S20 := YANG MAkan KACANG SIAPA KAMARIN
*S21 := YANG MEMBELI KUWE KAMARIN SIAPA
*S22 := SIAPA S3 S4 S15
3.2.7 Thai

Problem constructed by Peter Lee of the University of Wisconsin Linguistics Department, who also acted as informant. The tones are not indicated as they are not pertinent to this particular problem. Glosses of the sentences have been added to the original teletype printout.

BEGIN AUTOLING

NEXT: KHUN CHOOP PHAK←
("You like vegetables")

NEXT: KHUN CHOOP BURII←
("You like cigarettes")

NEXT: *TYPE←
*S1 := KHUN CHOOP S2
S2 := BURII
S2 := PHAK

KHUN CHOOP KHAJ←
("You like eggs")

NEXT: *TYPE←
*S1 := KHUN CHOOP S2
S2 := KHAJ
S2 := BURII
S2 := PHAK

KHUN HIW KHAJ←
("You are hungry for eggs")

CAN YOU SAY:
   KHUN HIW BURII

NO←

NEXT: *TYPE←
*S1 := KHUN CHOOP S2
S2 := KHAJ
S2 := BURII
S2 := PHAK
*S5 := KHUN HIW KHAJ

KHUN HIW KHAAW←
("You are hungry for rice")

*PARSING ILLEGALS*
NEXT: *TYPE←
*S1 := KHUN CHOOP S2
S2 := KHAJ
S2 := BURII
S2 := PHAK
*S5 := KHUN HIW S6
S6 := KHAAW
S6 := KHAJ
*S10 := KHUN HIW S6

KHUN HIW PHAK←
("You are hungry for vegetables")

NEXT: *TYPE←
*S1 := KHUN CHOOP S2
S2 := KHAJ
S2 := BURII
S2 := PHAK
*S5 := S10
S6 := PHAK
S6 := KHAAW
S6 := KHAJ
*S10 := KHUN HIW S6

KHUN HIW NAAM
("You are thirsty.")

NEXT: KHUN HIW NAAM MAJ←
CAN YOU SAY:

KHUN HIW KHAJ MAJ

YES←

NEXT: *TYPE←
*S1 := KHUN CHOOP S2
S2 := KHAJ
S2 := BURII
S2 := PHAK
*S5 := S10
S6 := NAAM
S6 := PHAK
S6 := KHAAW
S6 := KHAJ
*S10 := KHUN HIW S6
*S11 := S5 MAJ

KHUN CHOOP KHAJ MAJ←
("Do you like eggs?")
CAN YOU SAY:

KHUN CHOOP BURII MAJ
("Do you like cigarettes?")

YES←
NEXT:  *TYPE ←
     *S1 := KHUN CHOOP S2
     S2 := KHAJ
     S2 := BURII
     S2 := PHAK
     *S5 := S10
     S6 := NAAM
     S6 := PHAK
     S6 := KHAAW
     S6 := KHAJ
     *S10 := KHUN KIWi S6
     *S11 := S12 MAJ
     S12 := S1
     S12 := S5

KHUN CHOOP PHAK MAJ -
-PARSED OK-
*PARSING ILLEGALS*

NEXT:  KHAW MII NAAM MAJ ←
CAN YOU SAY:
       KHAW MII PHAK MAJ
YES ←

("Do you like vegetables?")

NEXT:  PHOM MII PHAK ←
CAN YOU SAY:
       PHOM MII KHAJ
YES ←

("Does he have water?")

("Does he have vegetables?")

("I have vegetables.")

("I have eggs.")

NEXT:  *TYPE ←
     *S1 := KHUN CHOOP S2
     S2 := KHAJ
     S2 := BURII
     S2 := PHAK
     *S5 := S10
     S6 := NAAM
     S6 := PHAK
     S6 := KHAAW
     S6 := KHAJ
     *S10 := KHUN HIWi S6
     *S11 := S12 MAJ
     S12 := S1
     S12 := S5
     *S13 := KHAW MII S6 MAJ
     *S14 := PHOM MII S6

KHAW CHOOP KHAJ ←
CAN YOU SAY:
       KHAW CHOOP PHAK
YES ←

("He likes eggs.")

("He likes vegetables.")
CAN YOU SAY:  
    KHAW HIW PHAK  
("He is hungry for vegetables.")

CAN YOU SAY:  
    KHUN MII KHAAW MAJ  
("Do you have rice?")

YES ←

NEXT:  TYPE ←
*S1 := S15 CHOOP S2
S2 := KHAJ
S2 := BURII
S2 := PHAK
*S5 := S10
S6 := NAAM
S6 := PHAK
S6 := KHAAW
S6 := KHAJ
*S10 := S15 HIW S6
*S11 := S12 MAJ
S12 := S1
S12 := S5
*S13 := S15 MII S6 MAJ
*S14 := PHOM MII S6
S15 := KHAW
S15 := KHUN

KHUN MII KHAJ ←
CAN YOU SAY:  
    KHUN MII NAAM  
("You have water.")

YES ←

CAN YOU SAY:  
    PHOM CHOOP KHAJ  
("I like egg(s).")

YES ←

CAN YOU SAY:  
    PHOM MII KHAAW MAJ  
("Do I have rice?")

YES ←

CAN YOU SAY:  
    PHOM HIW KHAJ  
("I'm hungry for egg(s).")

YES ←

CAN YOU SAY:  
    KHAW MII KHAAW MAJ  
("Does he have rice?")

YES ←

CAN YOU SAY:  
    KHAW MII PHAK  
("He has vegetable(s).")

YES ←

CAN YOU SAY:  
    PHOM HIW·NAAM  
("I am thirsty.")

YES ←

CAN YOU SAY:  
    PHOM HIW KHAAW  
("I'm hungry (for rice.)")

YES ←
CAN YOU SAY:
   PHOM MII NAAM MAJ
YES ←
CAN YOU SAY:
   PHOM MII KHAJ MAJ
YES ←
CAN YOU SAY:
   KHUN MII KHAJ MAJ
YES ←
CAN YOU SAY:
   KHAW CHOO P KHAJ MAJ
YES ←

NEXT: *TYPE ←
*S1 := S15 CHOO P S2
S2 := KHAJ
S2 := BURII
S2 := PHAK
*S5 := S10
S6 := NAAM
S6 := PHAK
S6 := KHAAW
S6 := KHAJ
*S10 := S14
*S11 := S13
S12 := S1
S12 := S5
*S13 := S17 MAJ
*S14 := S15 S16 S6
S15 := PHOM
S15 := KHAW
S15 := KHUN
S16 := MII
S16 := HIW
S17 := S10
S17 := S12

PHOM KIN KHAJ KHRAP ←
CAN YOU SAY:
   PHOM KIN KHAAW KHRAP
YES ←
*PARSING ILLEGALS*

NEXT: DICHAN KIN KHAAW KHA ←
CAN YOU SAY:
   DICHAN KIN KHAJ KHA
YES ←

("Do I have water?")

("Do I have egg(s)?")

("Do you have egg(s)?")

("Does he like egg(s)?")

Morphological Note:

PHOM = 'I' (masc. speaker)
DICCHAN = 'I' (fem. speaker)
KHRAP = 'sir or ma'am' (masc. speaker)
KHA = 'sir or ma'am' (fem. speaker)

("I eat egg(s), sir or ma'am.")

("I eat rice, sir or ma'am.")

("I eat rice, sir or ma'am.")

("I eat egg(s), sir or ma'am.")
NEXT: *TYPE ←
*S1 := S15 CHOOP S2
S2 := KHAJ
S2 := BURII
S2 := PHAK
*S5 := S10
S6 := NAAM
S6 := PHAK
S6 := KHAAW
S6 := KHAJ
*S10 := S14
*S11 := S13
S12 := S1
S12 := S5
*S13 := S17 MAJ
*S14 := S15 S16 S6
S15 := PHOM
S15 := KHAW
S15 := KHUN
S16 := MII
S16 := HIW
S17 := S10
S17 := S12
*S18 := S15 KIN S6 KHRAP
*S19 := DICCHAN KIN S6 KHA

DICCHAN KAMLANG KIN KHAJ KHA ←
("I am eating egg(s) sir or ma'am.")
CAN YOU SAY:
DICCHAN KAMLANG KIN KHAAW KHA
("I am eating rice  sir or ma'am.")
YES ←

NEXT: *TYPE ←
*S1 := S15 CHOOP S2
S2 := KHAJ
S2 := BURII
S2 := PHAK
*S5 := S10
S6 := NAAM
S6 := PHAK
S6 := KHAAW
S6 := KHAJ
*S10 := S14
*S11 := S13
S12 := S1
S12 := S5
*S13 := S17 MAJ
*S14 := S15 S16 S6
S15 := PHOM
S15 := KHAW
S15 := KHUN
S16 := MII
S16 := HIW
S17 := S10
S17 := S12
*S18 := S15 KIN S6 KHRAP
*S19 := DICHAN KIN S6 KHA
*S20 := DICHAN KAMLANG KIN S6 KHA

KHAW KIN PHAK ➔
CAN YOU SAY:
   KHAW HIW PHAK KHRAP
   CAN YOU SAY:
   DICHAN KIN NAAM KHA
   CAN YOU SAY:
   DICHAN KAMLANG HIW NAAM KHA
   CAN YOU SAY:

("He eats vegetable(s).")
("He is hungry for vegetable(s)
   sir or mām.")
("I am drinking, sir or mām.")
("I am (and have been for some
time) thirsty, sir or mām.")

NEXT: *TYPE ➔
*S1 := S15 CHOOP S2
S2 := KHAJ
S2 := BURII
S2 := PHAK
*S5 := S10
S6 := NAAM
S6 := PHAK
S6 := KHAAW
S6 := KHAJ
*S10 := S14
*S11 := S13
S12 := S1
S12 := S5
*S13 := S17 MAJ
*S14 := S15 S16 S6
S15 := PHOM
S15 := KHAW
S16 := KHUN
S16 := KIN
S16 := MII
S16 := HIW
S17 := S10
S17 := S12
*S18 := S15 S16 S6 KHRAP
*S19 := DICHAN S16 S6 KHA
*S20 := DICHAN KAMLANG S16 S6 KHA
DICHTAN CA HIW NAAM KHA ←
CAN YOU SAY:
   DICHTAN CA KIN PHAK KHA
YES ←

NEXT: NAKRIAN CA KIN KLUAJ ←
CAN YOU SAY:
   NAKRIAN KAMLANG HIW KLUAJ
YES ←
*PARSING ILLEGALS*

NEXT: NAKRIAN KAMLANG KIN KLUAJ ←
-PARSED OK-

NEXT: NAKRIAN KAMLANG RIAN NANGSYY ←
CAN YOU SAY:
   NAKRIAN CA RIAN NANGSYY
YES ←

NEXT: *TYPE ←
*S1 := S15 CHOOP S2
 S2 := KHAJ
 S2 := BURII
 S2 := PHAK
*S5 := S10
 S6 := NAAM
 S6 := PHAK
 S6 := KHAAW
 S6 := KHAJ
*S10 := S14
*S11 := S13
 S12 := S1
 S12 := S5
*S13 := S17 MAJ
*S14 := S15 S16 S6
 S15 := PHOM
 S15 := KHAW
 S15 := KHUN
 S16 := KIN
 S16 := MII
 S16 := HIW
 S17 := S10
 S17 := S12
*S18 := S15 S16 S6 KHRAP
*S19 := DICHTAN S16 S6 KHA
*S20 := DICHTAN S21 S16 S6 KHA
 S21 := CA
 S21 := KAMLANG
*S22 := NAKRIAN S21 S16 KLUAJ
*S23 := NAKRIAN S21 RIAN NANGSYY

("I will be thirsty, sir.")
("I will eat vegetable(s), sir.")
("Students will eat banana(s)."
("The students are and have been hungry for bananas.
((The) students are eating bananas.)
((The) students are studying (books).)
((The) students will study (books).)
KHUUU KAMLANG SOON NAKRIAN ←  
CAN YOU SAY: 
KHUUU CA SOON NAKRIAN 
YES ←  

NEXT: KHUUU KAMLANG SOON NANGSYY ←  
CAN YOU SAY: 
KHUUU CA SOON NANGSYY 
YES ←  
CAN YOU SAY: 
NANGSYY CA RIAN NANGSYY 
NO ←  
CAN YOU SAY: 
NANGSYY KAMLANG KIN KLUAJ 
NO ←  
CAN YOU SAY: 
NAKRIAN KAMLANG RIAN NAKRIAN 
NO ←  

NEXT: *TYPE ←  
*S1 := S15 CHOOP S2 
S2 := KHAJ 
*S2 := BURII 
S2 := PHAK 
*S5 := S10 
S6 := NAAM 
*S6 := PHAK 
S6 := KHAAW 
S6 := KHAJ 
*S10 := S14 
*S11 := S13 
S12 := S1 
S12 := S5 
*S13 := S17 MAJ 
*S14 := S15 S16 S6 
S15 := PHOM 
S15 := KHAW 
S15 := KHUN 
S16 := KIN 
S16 := MII 
S16 := HIW 
S17 := S10 
S17 := S12 
*S18 := S15 S16 S6 KHRAP 
*S19 := DICHAN S16 S6 KHA 
*S20 := DICHAN S21 S16 S6 KHA 
S21 := CA 
S21 := KAMLANG
*S22 := NAKRIAN S21 S16 KLUAJ
*S23 := NAKRIAN S21 RIAN NANGSYY
*S24 := KHRUU S21 SOON S25
S25 := NANGSYY
S25 := NAKRIAN

KHRUU JAJ MAAK ←
("The teacher is very big.")

NEXT: KHRUU KHAW KAW MAAK ←
("His teacher is very old.")

CAN YOU SAY:
KHRUU PHOM KAW MAAK
("My teacher is very old.")

YES ←
*PARSING ILLEGALS*

NEXT: *TYPE ←
*S1 := S15 CHOOP S2
S2 := KHAJ
S2 := BURII
S2 := PHAK
*S5 := S10
S6 := NAAM
S6 := PHAK
S6 := KHAAW
S6 := KHAJ
*S10 := S14
*S11 := S13
S12 := S1
S12 := S5
*S13 := S17 MAJ
*S14 := S15 S16 S6
S15 := PHOM
S15 := KHAW
S15 := KHUN
S16 := KIN
S16 := MII
S16 := HIW
S17 := S10
S17 := S12
*S18 := S15 S16 S6 KHRAP
*S19 := DICCHAN S16 S6 KHA
*S20 := DICCHAN S21 S16 S6 KHA
S21 := CA
S21 := KAMLANG
*S22 := NAKRIAN S21 S16 KLUAJ
*S23 := NAKRIAN S21 RIAN NANGSYY
*S24 := KHRUU S21 SOON S25
S25 := NANGSYY
S25 := NAKRIAN
*S26 := KHRUU JAJ MAAK
*S27 := KHRUU S15 KAW MAAK
3.2.8 Mandarin Chinese

Problem constructed by Margaret A. Naeser of the University of Wisconsin Linguistic Department. Ai Chen Ting of the University of Wisconsin East Asian Languages and Literature Department served as informant. Transcription is in the Yale romanization system. Tones are not indicated as they were not pertinent to this particular problem (no minimal pairs used). Glosses have been added to the original output.

*SYSTEM REINITIALIZED*

NEXT: WO SHWO HWA ←
("I speak words."")

NEXT: TA SHWO HWA ←
("He speaks words."")

NEXT: NI SHWO HWA ←
("You speak words."")

NEXT: WO MEN SHWO HWA ←
CAN YOU SAY:
NI MEN SHWO HWA
("I plural speak words."")

YES ←
("You plural speak words."")

NEXT: WO YOU SHU ←
CAN YOU SAY:
TA YOU SHU
("I have books."")

YES ←
("He has books."")

*PARSING ILLEGALS*

NEXT: *TYPE ←
*S1 := S2 SHWO HWA
S2 := NI
S2 := TA
S2 := WO
*S3 := S2 MEN SHWO HWA
*S4 := S2 YOU SHU
WO YOU JUNG GWO SHU ←
CAN YOU SAY:
TA YOU JUNG GWO SHU
("I have China Land books."")
("Chinese"
("He has China Land books."")
NEXT:  WO YOU FA GWO SHU ←
CAN YOU SAY:
   TA YOU FA GWO SHU
YES ←

NEXT:  WO SHWO FA GWO HWA ←
CAN YOU SAY:
   NI SHWO FA GWO HWA
YES ←

NEXT:  WO YOU YING GWO SHU ←
CAN YOU SAY:
   NI SHWO YING GWO HWA
YES ←

NEXT:  WO YOU DE GWO SHU ←
CAN YOU SAY:
   NI SHWO DE GWO HWA
YES ←

*PARSING ILLEGALS*

NEXT:  WO KAN SHU ←
CAN YOU SAY:
   TA KAN SHU
YES ←
CAN YOU SAY:
   TA KAN YING GWO SHU
YES ←

NEXT:  TA CHANG GER ←
CAN YOU SAY:
   NI CHANG GER
YES ←

NEXT:  *TYPE ←*
*S1 := S2 SHWO HWA
S2 := NI
S2 := TA
S2 := WO
*S3 := S2 MEN SHWO HWA
*S4 := S2 S8 SHU
*S5 := S2 S8 S6 GWO SHU
S6 := DE
S6 := YING
S6 := FA
S6 := JUNG
*S7 := S2 SHWO S6 GWO HWA
S8 := KAN
S8 := YOU
*S9 := S2 CHANG GER

("I have French Land books.")
("He has French Land books.")
("I speak French Land words.")
("You speak French.")

("I have English Land books.")
("You speak English Land words.")
("I have German Land books.")
("You speak German Land words.")

("I read books.")
("He reads books.")
("He reads English Land books.")
("He sings songs.")
("You sing songs.")
TA CHANGE JUNG GWO GER ←
CAN YOU SAY:
    NI CHANG JUNG GWO GER
YES ←

NEXT: WO CHR FAN ←
CAN YOU SAY:
    TA CHR FAN
YES ←

NEXT: TA CHR JUNG GWO FAN ←
CAN YOU SAY:
    TA CHR YING GWO FAN
YES ←
*PARSING ILLEGALS*

NEXT: TA DZWO FAN ←
CAN YOU SAY:
    WO DZWO FAN
YES ←
CAN YOU SAY:
    WO DZWO YING GWO FAN
YES ←

NEXT: TA CHR TSAI ←
CAN YOU SAY:
    TA DZWO TSAI
YES ←
CAN YOU SAY:
    NI CHR JUNG GWO TSAI
YES ←

NEXT: *TYPE ←
*S1 := S2 SHWO HWA
S2 := NI
S2 := TA
S2 := WO
*S3 := S2 MEN SHWO HWA
*S4 := S2 S8 SHU
*S5 := S2 S8 S6 GWO SHU
S6 := DE
S6 := YING
S6 := FA
S6 := JUNG
*S7 := S2 SHWO S6 GWO HWA
S8 := KAN
S8 := YOU
*S9 := S2 CHANG GER

("He sings China Land songs.")
Chinese
("You sing Chinese songs.")

("I eat food.")

("He eats food.")

("He eats China Land food.")
Chinese
("He eats England Land food.")
English

("He prepares food.")

("I prepare food.")

("I prepare England Land food.")
English

("He eats vegetables.")

("He prepares vegetables.")

("You eat Chinese vegetables.")
WO HE CHA ←
CAN YOU SAY:
   TA HE CHA
YES ←
("I drink tea.")
("He drinks tea.")

NEXT: WO HE JYOU ←
CAN YOU SAY:
   NI HE JYOU
YES ←
("I drink wine.")
("You drink wine.")

NEXT: WO HE PIYOU ←
*PARSING ILLEGALS*

NEXT: WO HE FA GWO PIYOU ←
CAN YOU SAY:
   NI HE FA GWO PIYOU
YES ←
("I drink beer.")
("You drink French beer.")

NEXT: SHEI HE FA GWO PIYOU ←
CAN YOU SAY:
   SHEI SHWO HWA
YES ←
("Who drinks French beer?")
("Who speaks words?")
CAN YOU SAY:
   SHEI HE JYOU
("Who drinks wine?")

*TYPE ←
*S1 := S2 SHWO HWA
S2 := SHEI
S2 := NI
S2 := TA
S2 := WO
*S3 := S2 MEN SHWO HWA
*S4 := S2 S8 SHU
*S5 := S2 S8 S6 GWO SHU
S6 := DE
S6 := YING
S6 := FA
S6 := FA
S6 := JUNG
*S7 := S2 SHWO S6 GWO HWA
S8 := KAN
S8 := YOU
*S9 := S2 CHANG GER
*S10 := S2 CHANG S6 GWO GER
*S11 := S2 S13 S14
*S12 := S2 S13 S6 GWO S14
S13 := DZWO
S13 := CHR
S14 := TSAI
S14 := FAN
*S15 := S2 HE S16
S16 := PIJYOU
S16 := JYOU
S16 := CHA
*S17 := S2 HE S6 GWO S16
YES ←
CAN YOU SAY:
    SHEI DZWO FAN
    ("Who prepares food?")
YES ←
CAN YOU SAY:
    SHEI DZWO JUNG GWO FAN
    ("Who prepares Chinese food?")
YES ←
CAN YOU SAY:
    SHEI CHANG DE GWO GER
    ("Who sings German songs?")
YES ←
CAN YOU SAY:
    SHEI CHANG GER
    ("Who sings songs?")
YES ←
CAN YOU SAY:
    SHEI YOU SHU
    ("Who has books?")
YES ←
CAN YOU SAY:
    SHEI SHWO FA GWO HWA
    ("Who speaks French words?")
YES ←
CAN YOU SAY:
    SHEI KAN DE GWO SHU
    ("Who reads German books?")
YES ←
CAN YOU SAY:
    SHEI MEN SHWO HWA
NO ←
NEXT: *TYPE ←
*S1 := S18 SHWO HWA
*S2 := NI
*S2 := TA
*S2 := WO
*S3 := S2 MEN SHWO HWA
*S4 := S18 S8 SHU
*S5 := S18 S8 S6 GWO SHU
*S6 := DE
*S6 := YING
*S6 := FA
*S6 := JUNG
*S7 := S18 SHWO S6 GWO HWA
*S8 := KAN
*S8 := YOU
*S9 := S18 CHANG GER
*S10 := S18 CHANG S6 GWO GER
*S11 := S18 S13 S14
*S12 := S18 S13 S6 GWO S14
*S13 := DZWO
*S13 := CHR
*S14 := TSAL
*S14 := FAN
*S15 := S18 HE S16
*S16 := PIYOU
*S16 := JYOU
*S16 := CHA
*S17 := S18 HE S6 GWO S16
*S18 := S2
*S18 := SHEI

WO YAU SHWO HWA ←
CAN YOU SAY:
    TA YAU SHWO HWA
YES ←

("I want to speak words."
  speak
("He wants to speak words."
  speak

NEXT: WO SYIHWAN SHWO HWA ←
CAN YOU SAY:
    SHEI SYIHWAN SHWO HWA
YES ←

("I desire to speak words."
  speak
("Who desires to speak words."
  speak

("I can speak.")

NEXT: WO KEYI SHWO HWA ←
*PARSING ILLEGALS*

NEXT: WO DEI SHWO HWA ←

("I must speak words."
  speak
NEXT: WO DEI SHWO JUNG GWO HWA ←
CAN YOU SAY:
WO SYIHWAN SHWO FA GWO HWA
YES ←
CAN YOU SAY:
WO SYIHWAN KAN SHU
YES ←
CAN YOU SAY:
WO YAU CHANG GER
YES ←
CAN YOU SAY:
TA SYIHWAN CHANG DE GWO GER
YES ←
CAN YOU SAY:
TA KEYI DZWO FA GWO FAN
YES ←
CAN YOU SAY:
NI DZWO TSAI
YES ←
CAN YOU SAY:
NI SYIHWAN HE CHA
YES ←
CAN YOU SAY:
NI HE FA GWO CHA
YES ←
CAN YOU SAY:
NI DEI SHWO HWA
YES ←
CAN YOU SAY:
TA DEI KEYI SHWO HWA
NO ←
CAN YOU SAY:
NI YAU YOU JUNG GWO SHU
YES ←
CAN YOU SAY:
NI SYIHWAN SHWO HWA
YES ←
CAN YOU SAY:
NI SYIHWAN KAN SHU
YES ←
CAN YOU SAY:
SHEI KEYI CHANG GER
YES ←
CAN YOU SAY:
SHEI CHANG JUNG GWO GER
YES ←
CAN YOU SAY:
TA CHR FA GWO FAN
YES ←

("I must speak Chinese words.")
Chinese
("I desire to speak French.")
("I desire to read books.")
("I want to sing songs.")
("He desires to sing German songs.")
("He can prepare French food.")
("You prepare vegetables.")
("You desire to drink tea.")
("You drink French tea.")
("You must speak.")
("You want to have Chinese books.")
("You desire to speak.")
("You desire to read books.")
("Who can sing songs?")
("Who can sing Chinese songs?")
("He eats French food.")
CAN YOU SAY:
WO KEYI DZWO TSAI
YES ←
CAN YOU SAY:
SHEI HE CHA
("I can prepare vegetables.")
YES ←
CAN YOU SAY:
TA DEI HE YING GWO CHA
("Who drinks tea?")
YES ←
CAN YOU SAY:
TA DEI SHWO HWA
("He must drink English tea.")
YES ←
CAN YOU SAY:
NI YAU SHWO HWA
("He must speak.")
YES ←
CAN YOU SAY:
NI YAU SHWO HWA
("You want to speak.")
YES ←
CAN YOU SAY:
TA YAU YOU FA GWO SHU
("He wants to have French books.")
YES ←
CAN YOU SAY:
SHEI SHWO YING GWO HWA
("Who speaks English?")
YES ←

NEXT: WO MEN DEI SHWO HWA ←
("We must speak.")
CAN YOU SAY:
SHEI MEN SYIHWAN SHWO HWA
NO ←

NEXT:
WO MEN DEI DAU JER LAI SHWO HWA ←
("We must come here to speak.")
CAN YOU SAY:
WO KEYI MEN DEI DAU JER LAI SHWO HWA
NO ←

NEXT: SHEI DEI DAU JER LAI SHWO HWA ←
("Who must come here to speak?")
CAN YOU SAY:
WO YAU DAU JER LAI SHWO HWA
("I want to come here to speak.")
YES ←
*PARSING ILLEGALS*

NEXT: WO YAU DAU JYA LAI SHWO HWA ←
("I want to come home to speak.")
CAN YOU SAY:
SHEI YAU DAU JYA LAI SHWO HWA
("Who wants to come home to speak.")
YES ←

NEXT: WO YAU DAU SYWESYAU LAI SHWO HWA ←
("I want to come to school to speak.")
CAN YOU SAY:
NI SYIHWAN YAU DAU SYWESYAU LAI SHWO HWA
NO ←
CAN YOU SAY:
NI KEYI YAU DAU SYWESYAU LAI SHWO HWA
NO ←
CAN YOU SAY:  
TA YAU DAU SYWESYAU LAI SHWO HWA  
YES ←  
("He wants to come to school to speak."")

NEXT: WO YAU DAU FANGDZ LAI SHWO HWA ←  
("I want to come to the house to speak."")

NEXT: *TYPE ←
*S1 := S22 SHWO HWA
S2 := NI
S2 := TA
S2 := WO
*S3 := S2 MEN SHWO HWA
*S4 := S22 S8 SHU
*S5 := S22 S8 S6 GWO SHU
S6 := DE
S6 := YING
S6 := FA
S6 := JUNG
*S7 := S22 SHWO S6 GWO HWA
S8 := KAN
S8 := YOU
*S9 := S22 CHANG GER
*S10 := S22 CHANG S6 GWO GER
*S11 := S22 S13 S14
*S12 := S22 S13 S6 GWO S14
S13 := DZWO
S13 := CHR
S14 := TSAI
S14 := FAN
*S15 := S22 HE S16
S16 := PIYOU
S16 := JYOU
S16 := CHA
*S17 := S22 HE S6 GWO S16
S18 := S2
S18 := SHEI
*S19 := S1
S20 := DEI
S20 := KEYI
S20 := SYIHWAN
S20 := YAU
S21 := S2 S20
S21 := S18
S22 := S21
S22 := S18 S20
*S23 := WO MEN DEI SHWO HWA
*S24 := WO MEN DEI DAU JER LAI SHWO HWA
*S25 := S22 DAU JER LAI SHWO HWA
*S26 := S22 YAU DAU S29 LAI SHWO HWA
S29 := FANGDZ
S29 := SYWESYAU
S29 := TWA
3.3 Planned Improvements in the Phrase Structure Learning Program.

The system is very sensitive to the order in which inputs are presented. At the moment the class splitting heuristic 5 does not apply under certain input sequence circumstances. For example, given the inputs:

the girl is tall
a girl is tall,

yielding the grammar:

\[ *S_1 \rightarrow S_2 \text{ girl is tall} \]
\[ S_2 \rightarrow \text{ the} \]
\[ S_2 \rightarrow \text{ a} \]

If the next input is:

the girl s are tall

The system will add the rule:

\[ *S_3 \rightarrow S_2 \text{ girl s are tall} \]

**without** checking to see if

a girl s are tall

is legal. This test would have been made if the last input had been the first input. As a result the system maintains an illegal rule which may not be corrected for a very long time, if ever. If an 'a' should occur with a plural noun in a later test for another rule and be rejected by the informant, the system will merely reject the rule currently under test. In such circumstance 3 different things may happen. The system may recycle and correct the error in a later learning run; the system may recycle recursively to a depth 3 and quit; or, more frequently, learn a very complicated grammar which is capable of parsing all the inputs from the informant, but which, from a generative point of view, still contains illegal rules.
ADDENDUM to page 65

The sensitivity problem described on page 65 is now corrected via random checking. The error might still occur on a random basis if $S_2$ has more than two members.

The improvement is reflected in the examples of Section 3, e.g. page 19.
This flaw can probably be corrected by the following heuristic procedure which will be added to the system: if a given top node string derived from an informant input sentence contains any nonterminals that are classes of morphemes, generate test sentences through the top node string selecting the other members of each morpheme class, and apply the class splitting heuristic 5, each time the informant rejects a test case.

Another method for obtaining a cleaner grammar would be to treat the right half of each rule in which a valid substitution has been made as an informant input, and subject it to the input rule coining heuristics. This improvement will be attempted, but the refining may come only at periodic intervals rather than after every rule change because of computation time problems.

The major improvement of the system will come from converting it to a context sensitive phrase structure learning system. The data structures already have appropriate links for associating context with individuals rules. Such an improvement will also require the construction of a context sensitive multi-path parser.

The heuristics for context sensitive learning will be supplemental to, and on the pattern of those for context free learning. Basically, if a context free rule is to be rejected on the basis of an informant's rejection of a text sentence, the system will attempt to reformulate the rule with a context restriction.

The formulation of the context restrictions themselves will initially be rather specific, but may grow in generality of statement via application of the heuristics already used in learning the context free rules.
Delete second paragraph of page 66.

Improvement already in system and reflected in examples of Section 3.
4.0 Transformation Learning Program.

Our work on a transformation learning program yielded two learning methods, bottom-to-top and top-to-bottom. The top-to-bottom method, which is entirely the work of William Fabens, was the one actually implemented. The bottom-to-top method, however, lends itself more readily to rule modification via informant interaction, and will be implemented and used in future versions of the system. Both methods require as input first a P-marker, (that is, an input sentence with a tree structure derived either from parsing or generation); and second, the sentence (without a tree) into which the first sentence is to be transformed. In each method the sentences may be in different languages. Both methods yield learning of bilingual transformations. The learning of monolingual transformations is a special case.

4.1 Bottom-to-Top Learning.

This method yields learning of the least general case first, and gradually increases the level of generality acceptable to an informant. Consider an input P-Marker

```
S
 /\   /
NP1  NP1
 /\   /\   /
D1 N1 Vsg1 Sg1 D2 N2
     |    |    |
The man eat s a fish
```

and a 'target' sentence (the desired transform)

```
a fish is eaten by the man
```
Increasingly complex transformations are coined by climbing up the trees of both sentences: in the case of the first, the given P-marker; in the case of the second, the implicit local tree structures existing in common with the first.

Accordingly, the lowest level transformation that could be coined is:

$$T_1: \text{the man eats a fish} \Rightarrow \text{a fish is eaten by a man}.$$ 

A somewhat higher level transformation would incorporate what is common to both inputs one level above the terminal string level. e.g.

$$T_2: \text{D}_1 \text{ N}_1 \text{ V}_1 \text{ Sg}_1 \text{ D}_2 \text{ N}_2 \Rightarrow \text{D}_2 \text{ N}_2 \text{ is V}_1 \text{ en by } \text{D}_1 \text{ N}_1$$

The assignment of higher level nodes to the elements on the right was determined by their existence in the P-marker of the first sentence. Test sentences generated via this transformation are offered an informant. Should he reject any, the level of generality of the transformation is decreased. For example, should this transformation not work if a different member of class $V_1$ is used, the transformation would be reformulated:

$$T_3: \text{D}_1 \text{ N}_1 \text{ eat Sg}_1 \text{ D}_2 \text{ N}_2 \Rightarrow \text{D}_2 \text{ N}_2 \text{ is eaten by } \text{D}_1 \text{ N}_1$$

And a special verb class containing 'eat' might be formulated at a later stage. Assuming, however that $T_2$ is accepted, the program would search for additional higher level common units, e.g.

```
NP_1
  D_1 N_1 V_1 Sg_1 D_2 N_2 \Rightarrow NP_2
  |   |   |   |   |   |
the man eats a fish  a fish is eaten by the man
```
This suggests the transformation:

\[ T_4: \quad NP_1 \ V_1 \ Sg_1 \ NP_2 \quad \Longrightarrow \quad NP_2 \text{ is } V_1 \text{ en by } NP_1 \]

There would, of course, be intermediate stages before \( T_4 \) is obtained, e.g. the coining of:

\[ NP_1 \ V_1 \ Sg_1 \ D_2 \ N_2 \quad \Longrightarrow \quad D_2 \ N_2 \text{ is } V_1 \text{ en by } NP_1 \]

If the informant accepts the test cases for \( T_4 \), it is accepted as the most general transformation to be learned, as no more common elements can be found between the explicit tree of the source and the implicit tree of the transform string. Should the informant reject test sentences derived from \( T_4 \), the system would again retreat in level of generality.

The transformation might also be subject to change and updating because of changes in the nature of the phrase structure grammar.

4.2 Top-to-Bottom Transformation Learning.

As indicated, this program is implemented and working. (Some bugs still exist in it, but these did not seriously interfere with the test examples presented here.) As indicated the logic and program are due to William Fabens.

4.2.1 Program Logic.

The output of the learning process is a list of ordered transformations, which operate from the top of the tree downwards. This block of transformations is also restated by the program (through substitution of terms) as a single transformation identical with the type derived by the bottom-to-top learning method.
Let us consider the learning process with the same example as in section 4.1. The input P-marker is "the man eat s a fish" plus its attendant tree structure, and again the transform: "a fish is eat en by the man". The system asks the translation of each morpheme in the input sequence. The correct reply is the equivalent morpheme in the transform string. (This is a substitute for dictionary lookup in the general case of bilingual transformation learning.) If there are no equivalents, the informant replies 'NONE'. If two morphemes are identical in the transform, the informant subscripts his replies to identify the relative positions of each like morpheme. (In the case of bilingual learning, this is in no sense word for word translation. Rather it is a method for locating equivalent phrase units).

Each morpheme in the transform sentence is placed at the bottom of its own push down stack. Above each morpheme in its stack is a list, in sequence, of the tree nodes on the path leading from it to the S node in the original input P-marker. Thus

\[
\begin{align*}
\text{a} & \quad \text{D}_2 \quad \text{NP}_2 \quad \text{VP}_1 \\
\text{fish} & \quad \text{N}_2 \quad \text{NP}_2 \quad \text{VP}_1 \\
\text{is} & \quad \text{V}_1 \quad \text{Vsg}_1 \quad \text{VP}_1 \\
\text{en} & \quad \text{from} \\
\text{by} & \quad \text{the} \quad \text{D}_1 \quad \text{NP}_1 \\
\text{man} & \quad \text{N}_1 \quad \text{NP}_1 \\
\end{align*}
\]
Next the tops of the stacks are scanned. If the same node occurs discontinuously on the top level, it is added to the tops of the intervening stacks. Accordingly, $VP_1$ is added to the top of the stack above 'is':

\[
\begin{align*}
\text{a} & \quad D_2 & \quad NP_2 & \quad VP_1 \\
\text{fish} & \quad N_2 & \quad NP_2 & \quad VP_1 \\
\text{is} & \quad VP_1 \\
\text{eat} & \quad V & \quad Vsg_1 & \quad VP_1 \\
\end{align*}
\]

If one or more morphemes are at the top of an adjacent stack (immediate adjacency in a forward linear scan) the node at the top of the stack of the last sequence of like top nodes is added to the tops of the morpheme stacks. In this case the result is that $VP_1$ is now added to the stacks containing 'en' and 'by'. (The ordering is admittedly arbitrary: one may ask why $NP_1$ was not added instead of $VP_1$). The result is:

\[
\begin{align*}
\text{a} & \quad D_2 & \quad NP_2 & \quad VP_1 \\
\text{fish} & \quad N_2 & \quad NP_2 & \quad VP_1 \\
\text{is} & \quad VP_1 \\
\text{eat} & \quad V_1 & \quad Vsg_1 & \quad VP_1 \\
\text{en} & \quad VP_1 \\
\text{by} & \quad VP_1 \\
\text{the} & \quad D_1 & \quad NP_1 \\
\text{man} & \quad N_1 & \quad NP_1 \\
\end{align*}
\]
The system then tabulates all possible tree branch transformations starting at the top. The source tree yields the left hand formulation of each transformation. The left half of the first of the series of ordered transformations is:

$$S(NP_1 \quad VP_1) \implies$$

which may be interpreted as 'NP_1 and VP_1' dominated by an S node. An S node was implicitly at the top of the transform pushdown stacks.

Accordingly, the right half of the transformation is formulated as S dominated whatever is currently at the top of the pushdown stacks, where adjacent strings of like nodes are treated as a single node:

$$T_a: \quad S(NP_1 \quad VP_1) \implies S(VP_1 \quad NP_1)$$

For the next step, the top nodes of the pushdown stacks above the transform tree are deleted and the node redistribution process described above is repeated except that morphemes at the top of the stacks no longer receive adjacent nodes. In this case no new nodes are added to any of the stacks.

```
a \quad D_2 \quad NP_2
fish \quad N_2 \quad NP_2
is
eat \quad V_1 \quad Vsg_1
en
by
the \quad D_1
man \quad N_1
```
The strings of VP₁'s and NP₁'s have been removed from the tops of the pushdown stacks. The right half of the next transformation is formed from what was under the VP₁ nodes:

\[ \Longrightarrow \text{VP}_1 (\text{NP}_2 \text{ is Vsg}_1 \text{ en by}) \]

The left half is derived from what is under the VP₁ node in the input tree; yielding:

\[ T_b: \text{VP}_1 (\text{Vsg}_1 \text{ NP}_2) \Longrightarrow \text{VP}_1 (\text{NP}_2 \text{ is Vsg}_1 \text{ en by}) \]

Similarly, what is beneath the deleted NP₁ in the source tree and the stack form the next transformation:

\[ \text{NP}_1 (D_1 \text{ N}_1) \Longrightarrow \text{NP}_1 (D_1 \text{ N}_1) \]

which is an identity transformation (a consequence of the method). Identity transformations, although calculated, are suppressed in the teletype output. Again the program strips the top nodes from the push down stacks:

\[ a \quad D_2 \]
\[ \text{fish} \quad N_2 \]
\[ (\text{is}) \quad \text{(removed from the stack)} \]
\[ \text{eat} \quad V_1 \]
\[ (\text{en}) \quad \text{(removed from the stack)} \]
\[ (\text{by}) \quad \text{(removed from the stack)} \]
\[ \text{the} \]
\[ \text{man} \]
The removed \( \text{NP}_2 \) yields an identity transformation which is disregarded:

\[
\text{NP}_2 \ (D_2 \ N_2) \implies \text{NP}_2 \ (D_2 \ N_2)
\]

However the removal of the \( \text{Vsg}_1 \) unit yields a non-trivial transformation:

\[
\text{T}_c: \quad \text{Vsg}_1 \ (V \ Sg) \implies \text{Vsg}_1 (V)
\]

The deleted \( D_1 \) and \( N_1 \) nodes again yield trivial identity transformations:

\[
D_1 \ (\text{the}) \implies D_1 \ (\text{the})
\]

\[
N_1 \ (\text{man}) \implies N_1 \ (\text{man})
\]

Repeating the process, the identity transformations

\[
D_2 (a) \implies D_2 (a)
\]

\[
N_2 \ (\text{fish}) \implies N_2 \ (\text{fish})
\]

are also coined.

Repeated substitutions in the battery of ordered transformations yield a rule identical to that derived by the bottom-to-top learning method:

\[
\text{T}_a: \quad S(\text{NP}_1 \ V\text{P}_1) \implies S(\text{VP}_1 \ \text{NP}_1)
\]

yields via substitution of the terms in \( \text{T}_b \) for \( \text{VP}_1 \):

\[
S(\text{NP}_1 \ \text{Vsg}_1 \ \text{NP}_2) \implies S(\text{NP}_2 \ \text{is} \ \text{Vsg}_1 \ \text{en} \ \text{by} \ \text{NP}_1)
\]

followed by substitution for \( \text{Vsg}_1 \) from transformation \( \text{T}_c \):

\[
S(\text{NP}_1 \ V_1 \ Sg_1 \ \text{NP}_2) \implies S(\text{NP}_2 \ \text{is} \ V_1 \ \text{en} \ \text{by} \ \text{NP}_1)
\]

which is identical to the transformation \( \text{T}_4 \) derived in section 4.1.
4.2.2 Features of the Program.

The system begins its input of the starting P-marker by outputting a message:

\[
\text{SENT :=}
\]

The human replies with the nodes deriving from \text{SENT}. The nodes must be bracketed by quotes and separated by commas. For example the human might reply:

"NP" , "VP"

followed by an arrow. The system will then query the expansion of each node, descending the left most branch of the tree first. E.g. the next query would then be:

. NP :=

The dot indicates the relative depth in the tree. If the last element was a terminal, the arrow key alone is punched. At the end of the input of a tree the system asks:

DO YOU WANT ANOTHER PARSE?

If the answer is YES, (in case of error) the program disregards the last inputted tree and offers one the chance of another tree input. If the answer is NO, the system outputs the message:

PLEASE TRANSLATE

followed by the terminal elements of the tree in sequential order. The human must then input the desired transform string, with each morpheme in quotes and separated by commas. The system will then ask for morpheme equivalents. If two or more morphemes are identical in either string the system assigns or interprets subscripts where appropriate.
After the transformation is learned, the system prints the rules, both in the ordered form and in the expanded, single expression notation. The program also uses '+' to indicate two units descended from the same node, and '-' to indicate more remote ancestry. The program then outputs a message:

WANT TO REAPPLY LAST TRANSFORMATION?

A YES answer means that the next time a tree exhibiting nodes in the domain of the transformation is inputted, the program will apply the transformation.

If the new P-marker is more complex then that of the learning input (but with nodes still in the domain of the transformation), the system will still correctly apply the rule (except in case of extreme complexity because of an uncorrected bug in the program).

Section 4.2.2 lists a copy of the teletype output with keypunch errors and extraneous system messages excluded) for learning of an active-passive transformation, and its application to a more complex sentence with different vocabulary. Section 4.2.3 lists a similar teletype transcript for the learning of an English to German bilingual transformation. (The print out of the dominating node on the right half of each transformation is suppressed on the teletype output.)
4.2.3 Learning and Application of an Active-Passive Transformation.

SENT := "NP", "VP" ←

. NP := "D", "NP" ←

. . D := ←

. . NP := ←

. VP := ←

DO YOU WANT ANOTHER PARSE? YES ←

SENT := "D" ←

. D := ←

DO YOU WANT ANOTHER PARSE? YES ←

SENT := "NP", "VP" ←

. NP := "JOHN" ←

. . JOHN := ←

. VP := "VP", "NP" ←

. . VP := "VSTMA", "SGA" ←

. . . VSTMA := "WANT" ←

. . . . WANT := ←

. . . SGA := "S" ←

. . . . S := ←

. . NP := "MARY" ←

. . . MARY := ←

DO YOU WANT ANOTHER PARSE? NO ←

PLEASE TRANSLATE

JOHN WANT S MARY

"MARY", "IS", "WANT", "ED", "BY", "JOHN" ←
C( JOHN ) := "JOHN" ←
C( WANT ) := "WANT" ←
C( S ) := "NONE" ←
C( MARY ) := "MARY" ←

SENT ( NP , VP ) ::= VP NP

VP ( VP(1) , NP(1) ) ::= NP(1) IS VP(1) ED BY

VP(1) ( VSTMA , SGA ) ::= VSTMA

OR: NP - VSTMA + SGA - NP(1) - ::= NP(1) + IS - VSTMA - ED + BY - NP

TRANSLATION IS...
MARY IS WANT ED BY JOHN

WANT TO REAPPLY LAST TRANSFORMATION? YES ←

SENT ::= "NP" , "VP" ←

. NP ::= "D" , "NP" ←

.. D ::= "THE" ←

... THE ::= ←

.. NP ::= "NP" , "MOD" ←

... NP ::= "MAN" ←

... MAN ::= ←

... MOD ::= "PREP" , "NP" ←

... PREP ::= "IN" ←

.... IN ::= ←

.... NP ::= "D" , "NP" ←

.... D ::= "THE" ←

..... THE ::= ←

..... NP ::= "PARK" ←

..... PARK ::= ←
\[ \begin{align*}
\text{VP} &: \text{"VP", "VP"} \\
\text{VSTMA} &: \text{"VSTMA", "SGA"} \\
\text{KISS} &: \text{"KISS"} \\
\text{SGA} &: \text{"ES"} \\
\text{ES} &: \text{""} \\
\text{NP} &: \text{"D", "NP"} \\
\text{D} &: \text{"A"} \\
\text{A} &: \text{""} \\
\text{NP} &: \text{"ADJ", "NP"} \\
\text{ADJ} &: \text{"PRETTY"} \\
\text{PRETTY} &: \text{""} \\
\text{NP} &: \text{"GIRL"} \\
\text{GIRL} &: \text{""} \\
\text{DO YOU WANT ANOTHER PARSE? NO} &: \text{""} \\
\text{T( A )} &: \text{"A"} \\
\text{T( PRETTY )} &: \text{"PRETTY"} \\
\text{T( GIRL )} &: \text{"GIRL"} \\
\text{T( KISS )} &: \text{"KISS"} \\
\text{T( THE )} &: \text{"THE"} \\
\text{T( MAN )} &: \text{"MAN"} \\
\text{T( IN )} &: \text{"IN"} \\
\text{T( THE(1) )} &: \text{"THE"} \\
\text{T( PARK )} &: \text{"PARK"} \\
\text{TRANSLATION IS...} \\
\text{A PRETTY GIRL IS KISSED BY THE MAN IN THE PARK} \end{align*} \]
4.2.4 Learning a Bilingual Transformation.

SENT :=
  NP :=
   PN :=
    HE :=
  VP :=
   VP :=
   IS :=
  NP :=
   NP :=
   D :=
    A :=
   NP :=
   MAN :=
    RC := "RP",
   RP :=
   WHO :=
    VP :=
     V :=
      V :=
       LIVE :=
        SG :=
        S :=
        MOD :=
        PREP :=
IN :=

NP :=

D :=

THE :=

NP :=

CITY :=

DO YOU WANT ANOTHER PARSE?

PLEASE TRANSLATE

HE IS A MAN WHO LIVE S IN THE CITY

C( HE ) :=

C( IS ) :=

C( A ) :=

C( MAN ) :=

C( WHO ) :=

C( LIVE ) :=

C( S ) :=

C( IN ) :=

C( THE ) :=

C( CITY ) :=

SENT ( NP, VP ) := NP VP

VP ( VP(1), NP(1) ) := VP(1) NP(1)

NP(1) ( NP(2), RC ) := NP(2)

NP(2) ( D, NP(3) ) := D RC -ENDER NP(3)

RC ( RP, VP(2) ) := VP(2)

VP(2) ( V, MOD ) := MOD V
MOD ( PREP, NP(4) ) := PREP NP(4)

NP(4) ( D(1) . NP(5) ) := D(1) NP(5)

V ( V(1) , SG ) := V(1)

OR: NP - VP(1) - D + NP(3) - RP - V(1) + SG - PREP - D(1) + NP(5) -
:= NP - VP(1) - D - PREP - D(1) + NP(5) - V(1) - ENDER + NP(3) -

[A compounded error: the collapsed transformation used the identity trans-
formations in the expansion. With corrections, it should read]

NP - VP(1) - D + NP(3) - RP - V(1) + SG - MOD :
NP - VP(1) - D - MOD - V(1) - ENDER + NP(3)

5.0 Proving the Linguist Superfluous.

In terms of speed in producing a grammar, the phrase structure learning
component of the AUTOLING system seems to have an advantage over the
human linguist. For example, the Indonesian problem of Section 3.2.6
required about 45 minutes of the 'informant's' time at the teletype, including the usage
of less than 4 minutes of computer central processor time (in a time sharing
environment involving relatively light demands by other users). AUTOLING's
time advantage over a human analyst appears to increase with the size of the
corpus, but precise tests have not been carried out.

With respect to completeness and quality, the existing AUTOLING
system is not yet ready to replace human linguists. But gradual improve-
ments are inevitable, and eventually the role of the human fieldworker may be
challenged seriously, particularly if the state of the art ever permits the
incorporation of adequate vocal communication between informant and com-
puter.

The proof of the adequacy of the machine as linguist might eventually be
demonstrated through a variant of the Turing test for artificial intelligence [19].
Let 5 or 10 human linguists each spend a set amount of time individually
working with the same informant. Let the machine linguist do the same
Then let the grammars produced by all participants be presented, anonymously, to another group of linguists who must attempt to spot the machine's grammar. If the machine linguist is not determined as such with statistically significant frequency, one may assume it is at least as good a fieldworker as the weakest human analyst in the test group.

While such a success might make 'data collecting' linguists superfluous, it should free most for work in linguistic theory. Of course by that time the computer will have become an essential tool of the theorist, not just for data collection and analysis, but as a means of checking the implications of theoretical formulations and models.
REFERENCES


