GRAMMARS AS USER MODELS

Ethel Schuster
Department of Computer and Information Science
University of Pennsylvania
Philadelphia, PA 10104

ABSTRACT*

This paper describes how grammars of native languages can be used is tutoring systems to facilitate instruction in a second language. The grammar functions as a user model, which enables the system to customise its responses by addressing problems that may be due to interference from the native language. Its correction strategy is based upoa comparison of the native language model with a model of the target language. The problems and solutions presented in this paper are related to the more general question of how modelling previous knowledge facilitates instruction in a new skill.

1. INTRODUCTION

People learning a second laaguage sometimes kave problems that caa be attributed to interference of their native language. This interference shows up in systematic errors in both speech and written text. Our concern is with <u>syntax</u> and syntactic interference. We waat to capture a student's knowledge of his/her native grammar that leads to this interference and use it to assist the student in learning that of the second laaguage.

Much research in AI has been devoted to the development of question answering systems, expert systems, and tutorial systems. Part of this research has involved enabling these systems to produce more "natural," "cooperative" responses, to recognize misconceptions and to provide explanations. One approach to improved interaction has been to incorporate a user model, e.g. |1, 5, 7). The claim we make in this paper is that grammars can serve as user models. We address the question of how correspondences between the grammars of two languages can provide an account of <u>syntactic</u> errors made by native speakers of one language attempting to learn a second one. This account can then be used in correcting the student. As aa example of this, we describe a computer tutoring system, VF* which focusses on the acquisition by non-English speakers of English constructions formed from a verb plus particle or verb plus prepositional phrase.

It has often been claimed that people rely heavily on their previous knowledge when learning a new skill [6, 8). While this previous knowledge caa sometimes help their learning, it caa oftea hinder it as well: Halast and Moraa (2), for example, describe a case ia which people reasoa incorrectly by analogy from a previous skill. Ia learning a second language, one's native language may be a source of both correct and incorrect analogies. Thus VF* includes a grammar of the native language in order to detect and correct errors due to incorrect analogy. This grammar also enables Vp[s] to explain the differences and similarities ia the verbal constructs of the two languages in focus.

2. OUR APPROACH TO SECOND LANGUAGE

The fundamental claim of one theory of second laaguage acquisition, the contrastive analysis theory (CA) |3j, is that when people are learning a secoad language L_2, the patterns of the language to be learned are matched with those of the native language L_1. Those that match are fairly easy to learn while those that do not, become more difficult and result in errors.

The work presented here has been developed along the lines of CA theory. Since interference errors reflect differences between the grammar of the student's native language and that of the target laaguage, they are oftea predictable from a contrastive analysis of the two grammars. This work focusses on one problem that appears on the syntactic level of L_2 acquisition, that is, the use of the complex construction verb plus preposition/particle, ia the English of non-native speakers.**

Consider the following interaction between a tutorial system for English and a student:

```
Tutor:  TRANSLATE  THE  FOLLOWING  SENTENCE:
        E1 penso'  en viajar    a  Europa ahora.
Student: He thought in travelling to Europe nov.
Tutor:   Incorrect. The correct answer ia
        'He thought of travelling to Europe now.' or
        'He thought about travelling to Europe now'.
        In English, you can use the verb <to think>
        aa in the following cases:
        - 'There is little opportunity to <think out>
          what the long-tern solution nay be'
        - 'With the beginning of the new day we
          had to <think> things <over> again'
        - 'She haa to <think up> a catchy nana for this
          system'
```

The system here provides the correct answer but it fails to recognise that the user has translated directly from Spanish.*** It leaves students to determine the relationship between the use of certain structures or patterns ia the new domain (or laaguage) by themselves. Here, the tutorial system could have provided a more adequate explanation by informing the user why the error occurred.

**Note that we are dealing only with <u>second</u> laaguage acquisitioa, aot third or fourth. Here the learner generalises only from his/her native language.

***By 'direct translation' we mean the most common translation used for a certain word, that is, the first one found when looking up that word ia a dictionary. This is different from the 'corresponding translation' which is the actual translatios of the words and which varies in the context of the sentence. For example, the direct translation of <en> is <in> while the corresponding translation of <en> when used with the verb <pensar>, <to think>, is <of> or <about>.

That is, it could have added:

> *Note that <en> is often translated as <in> in English, but here it is not.*

In order to provide such an explanation, a tutorial system must have some model of the user's knowledge of his/her native language, and of what features might cause interference. This is the point of departure of this work: to see what role learners' familiarity with one language (Spanish) plays in their learning a new language (English).

2.1. User Model in **VP***

Instead of having individual user models for each user, **VP*** has a canonical user model describing the grammar of the native language, in this case Spanish. One may argue that there are several dialects of Spanish and that the grammar represented in the system does not *exactly* correspond to the one any particular user has. We have assumed a grammar of standard Spanish for this system because the issues it addresses do not focus on idiomatic expressions but on specific language structures which do not show as much idiolectal variation.

When the student's previous knowledge of a verb-particle or verb-prepositional phrase construction can be seen to interfere with his/her learning new information, **VP*** has been designed to recognize the error and correct it. Consider the following:

VP*:   TRANSLATE THE FOLLOWING SENTENCE
         Moris penso   en comprar un carro.
STUDENT: Moris thought in buying a car.
VP*;   You used the incorrect preposition <in>.
         In English you can use <think of> or <think
         about> in this sentence. *Note that
         the direct translation of
         <think of> --<pensar de>~- does not exist in
         Spanish.* In English you can also use
         <think up> (an excuse, invent); <think over>
         (review); <think out> (consider, examine).

Providing this helpful response requires the system to maintain a model of the way verbs and prepositions are paired in both the student's native language, Spanish, and in English. For example, <pensar en> in Spanish translates directly to the English <think in> but corresponds to <think of> or <think about>. Therefore, a Spanish sentence that has <pensar en> translates to English as <think of> or <think about>. The same goes for <sonar con> which directly translates to <dream with> but corresponds to the English <dream of> or <dream about>. Another class of errors come from the fact that a verb in Spanish corresponds to a verb with particle in English, as in <recoger> corresponds to <pick up>; <escoger> to <pick out> and so on. Thus the student may either forget to include the particle, use a incorrect particle or insert the correct particle in the wrong place.

The user model in **VP*** anticipates the possibility that the student may overlook the differences and use a one-to-one correspondence mechanism. If the student does use this kind of one-to-one correspondence, the user model enables **VP*** to recognize the error and deal with it in a satisfactory manner.

3. THE VP* SYSTEM

VP* presents students with a sequence of Spanish sentences to translate into English. After presenting a sentence, the system reads in the student's translation and extracts its verb and preposition, assuming that the translation is grammatical. If the student misspells a word or provides an answer that the system cannot understand, **VP*** gives him/her another chance to answer. If the English parse fails due to an incorrect or missing preposition, the system looks at the verb phrase to try to extract the verb (and the preposition, in the case of an incorrect one). In either case, VP* then looks into the user model to obtain information about the Spanish verb. From there, it may hypothesize the source of the error. If interference from Spanish could be the source of error, the system provides the correct answer along with information that indicates the error is due to this interference. If the system cannot identify the error as one caused by interference, it provides the student with the correct answer and proceeds to the next exercise.

3.1. Implementation

VP* has been implemented in UNH Prolog version 1.3. Prolog was chosen as an implementation vehicle because it provides good pattern matching mechanisms and parsing techniques. VP* uses two grammars — one grammar for Spanish and one for English. Each grammar includes a dictionary and information about the verbs in each language and their associated prepositions. The system also has a table of direct translations of verbs and prepositions. In addition, a parse tree is stored for the correct answers to each exercise. These parse trees are matched against the user's•••• answers in order to locate any errors in those answers.

VP* presents the student with Spanish sentences to translate into English. For each exercise the system accepts as input the student's English translation, which it attempts to parse using the English grammar. If this attempt fails, then an attempt is made to parse the sentence using Spanish grammar rules in order to provide a parse tree for the system to analyze. If this latter attempt also fails (i.e. the translation contains unknown words or misspellings) then **VP*** allows the student to try again.

If a parse is obtained (using the English grammar) the system compares it with the stored parse tree for the sentence in order to determine if it is indeed the appropriate translation. If the translation that the student provides can be parsed but it is not the translation of the given sentence, the system allows the student to try again. If the parse has succeeded and corresponds to the stored parse tree, **VP*** acknowledges the correct translation and offers additional information about the verb and/or preposition in English or Spanish. For example, if the student types:

VP*:   TRANSLATE THE FOLLOWING SENTENCE:
         Yo soñé con el viaje.
STUDENT: I dreaatd of the trip.

the system provides an answer such as:

> Correct I
> Note that the direct translation of <soñar con>
> --<dreaa with>—does not exist in English.
> In English you can also use <dream about>
> in this sentence.

If the parse does not correspond to the stored parse but it is correct by Spanish rules, the system matches the verb phrase from the failed parse against the previously stored correct parse to figure out the differences between the correct stored parse and that of the student's. From this comparison, it finds the wrong prepositions and/or missing particles. If it finds an incorrect preposition with a correct verb, it looks up the direct Spanish translation of the incorrect preposition. For example, if in the previous example the

---

student typed:

>   I dreamed with the trip.

the system obtains the incorrect pair < dream with>, looks it up in its table of direct translations nnd finds it to be <sonar con>. **VP** proceed to its user model nnd finds this pnir in it. Therefore it is able to explain the incorrect preposition. It generates an error message indicating the error as being due to interference from Spanish. It then goes to the English grammar nnd looks np the correct preposition(s) that go with the verb nnd provides the student with more information, e.g. other prepositions thai can be used and so on. For example, **VP** provides the student with information about other preposition(s) that may be used with a verb:

>   In English you can une <dream of> or
>   <drean about> in thin sentence.

4, P O S S I B L E   E X T E N S I O N S

While **VP** has been designed to provide corrective responses in tutoring the use of verb and prepositions in English to Spanish speakers, the method presented here of using grammars as user models can be extended to other language features as well. For example, in English, attributive adjectives typically precede the nouns they modify, while in Spanish, they typically follow them.***** Hence it is common for native Spanish speakers to say in English:

- tht house small
- a ear green

VP can be extended to deal with this type of syntactic interference. A second example concerns English question patterns for Spanish speakers. Since Spanish marks subject agreement on verbs, non-emphatic subjects may be deleted. English, however, requires an overt subject which standardly precedes the verb. Now, English and Spanish have a similar question pattern as in the following:

EVGLISH	SPANISH
It he a student?	Es un estudiants?
	(It a student?)

In English, the auxiliary is moved to the start of the sentence and the subject is required. In Spanish, a subject pronoun is not required, there are restrictions on the verbs that use this pattern, and usually both the auxiliary and verb are moved to the beginning of the sentence. Thus Spanish speakers learning English may have the following problems: They may not place *be* before *he,* may omit the subject pronoun, or may use the same pattern with verbs other than *be* and *have.* **VP** may deal with these interference problems given the information about Spanish question patterns in its user model.

Most importantly, we believe we can generalise this approach. That is, instead of having to hand build special mechanisms for each case, we can derive predictable errors from the grammars themselves, along with a general mechanism for taking the particular difference lending to the current interference error and generating an appropriate response. This is the most challenging direction for research.

*****There is a small, closed set of Spanish adjectives (e.g. gran, burn) which may also precede the noun, although their use entails some change in semantic referent [4, p. 17).

5 . C O N C L U S I O N S

Earlier work has shown that people often rely on their knowledge of one domain when learning a new similar domain. A similar situation seems to be the case for language learning where students learning a second language use much of the knowledge they have of their native language.

VP has addressed the question of how we can represent the knowledge of certain aspects of a language in a computer system and use this knowledge to provide the student of a second language with information tailored to his/her knowledge. In particular, it has focused on the problem of how correspondence of grammars of two languages can provide a sufficient basis for explaining the possible origin of grammatical errors made by native speakers of one language when learning a second one.

R E F E R E N C E S

[1| Brown, J.S. and R. R. Burton. "Diagnostic Models for Procedural Bugs in Basic Mathematical Skills." *Cognitive Science* £(1078), 155-192.

[2] Halasi, F. and T.P. Moran. Analogy Considered Harmful. Proceedings CHI-82, CH1-82, Gaithersburg, MD., March , 1082.

|3| Lado, R. *Linguistic Across Cultures.* University of Michigan Press, Ann Arbor, MI, 1057.

(4) Poplack, S. Syntactic Structure and Social Function of Code-switching. In *Latino Language and Communicative Behavior,* Duran, R., Ed.,Ablex, New Jersey, 1081, pp. 160-184.

|5| Rich, E. "User Modeling via Stereotypes." *Cognitive Science* 5(1070), 320-354.

[6] Rumelhart, D.E. and D.A.Norman. Analogical Processes in Larning. In *Cognitive Skills and their Acquisition,* J.R. Anderson, Ed.,Erlbaum, Hillsdale, New Jersey, 1081, pp. 335-360.

(7) Wallis, J.W. and E.H.Shortliffe. Explanatory Power for Medical Expert Systems. Stanford University, Stanford, California, December, 1081.

[8] Winston, P.H. Learning and Reasoning by Analogy: The Details. AI Memo 520, MIT, Cambridge, Massachusetts, May, 1080.