ONE ANALYZER FOR THREE LANGUAGES

by

Robert Wilensky and Michael Morgan

Memorandum No. UCB/ERL M81/67

14 September 1981

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

TITLE PAGE

ONE ANALYZER FOR THREE LANGUAGES

by

Robert Wilensky and Michael Morgan

Memorandum No. UCB/ERL M81/67

14 September 1981

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

ONE ANALYZER FOR THREE LANGUAGES

by

Robert Wilensky and Michael Morgan

Memorandum No. UCB/ERL M81/67

14 September 1981

One Analyzer for Three Languages*

Robert Wilensky and Michael Morgan

Computer Science Division

Department of EECS

University of California, Berkeley

Berkeley, California 94720

ABSTRACT

We have been developing a model of natural language use. The primary point of this model is to provide a general framework in which it is possible to conveniently express knowledge about the meaning of a language's utterances. Programs for understanding and producing natural language based upon this model work by interpreting a knowledge base of such facts. Thus improving this knowledge base directly expands the language processing abilities of the system without any substantial programming effort.

We have recently tested our model by taking the understanding component of our system, called PHRAN (PHRasal ANalyzer), which was originally written to understand English text, and supplying it with knowledge bases for Spanish and Chinese.

While we had no theoretical predispositions about what components of this system should be transferable to other languages, we found that we were able to encode the relevant knowledge about these languages using our existing representational scheme. Once this was done, PHRAN was able to understand both these languages to a substantial extent without any modification to its control structure. Moreover, we encountered no problems in describing or processing these languages that would suggest serious difficulties in extending the system's knowledge bases beyond their current capacity.

1. Introduction

We have been developing a model of natural language use. The primary point of this model is to provide a general framework in which it is possible to

*Research sponsored in part by the Office of Naval Research under contract N00014-80-C-0732.

conveniently express knowledge about the meaning of a language's utterances. Programs for understanding and producing natural language based upon this model work by interpreting a knowledge base of such facts. Thus improving this knowledge base directly expands the language processing abilities of the system without any substantial programming effort.

We have recently tested our model by taking the understanding component of our system, called PHRAN (PHRasal ANalyzer), which was originally written to understand English text, and supplying it with knowledge bases for Spanish and Chinese.

While we had no theoretical predispositions about what components of this system should be transferable to other languages, we found that we were able to encode the relevant knowledge about these languages using our existing representational scheme. Once this was done, PHRAN was able to understand both these languages to a substantial extent without any modification to its control structure. Moreover, we encountered no problems in describing or processing these languages that would suggest serious difficulties in extending the system's knowledge bases beyond their current capacity.

In this report, we first briefly outline the assumptions of the model and describe the systems built in accordance with them. The rest of the report is concerned with the details of the Chinese and Spanish versions of PHRAN and of the representation we used to encode knowledge about the meaning of utterances in these languages.

2. Assumptions of the Model

The primary motivation our model was the realization that natural language users possess a large body of knowledge about what the utterances of their language mean. Thus constructing comprehensive language processing systems

would entail solving problems of how to represent, index, and manipulate language knowledge so as to produce a system that would be robust, modular, extensible, and easy to modify. In short, we would have all the problems usually associated with the phrase "knowledge engineering."

Moreover, these problems would be particularly pronounced in the case of language processing. The knowledge basis would not only be large, but in a sense, open ended. Furthermore, linguistic theories are notoriously complex and short-lived, and thus any particular formulation of the knowledge we chose would probably be subject to substantial revision. Finally, we are concerned with using this knowledge for more than one task, as we wish to use it both for understanding and producing language utterances.

The following sections describe these motivations in more detail.

2.1. The Importance of Non-generative Language

Language users know a great number of facts about their language. They know the meanings of a large number of words, and know the rules for relating these meanings to the occurrence of those words in an utterance. Moreover, they know the significance of a set of meaningful linguistic units that are not necessarily understood in terms of their components. We call all such units "phrases". Included in this set are idioms, canned phrases, lexical collocations, cliches, structural formulas, and other non-generative language structures. For example, the language user needs to know the particular fact that "out of the blue" means unexpectedly; that "by and large" means more or less; that "<person1> bear <person2> a <sentiment>" is a way of expressing a continued sentiment of one person toward another (as in "John bears Mary a grudge"); and that "<person1> give <person2> <object>" describes a transfer of possession. Our conjecture is that such units constitute a very considerable fraction of the language knowledge needed by an intelligent language processor.

In most theories of language, non-generative forms are usually considered to be theoretically uninteresting entities, or irritating special cases that violate the aesthetics of one's theory. However, if such structures do play a central role in language use, then most language processing is actually the application of these special case rules.

This is precisely the point of view we take. That is, while our model allows for the more traditional, very general word-to-meaning mappings, these mappings play no priviledged role. Both generative and non-generative knowledge is represented and applied uniformly - the only difference is in the degree of abstractness of the knowledge encoded.

Once this view is taken, both language analysis and language production become kinds of data base management problems. The knowledge about the meanings of phrases of different shapes and abstractness is the large data base. The problem is to represent this knowledge so it can be applied uniformly, and so it can be accessed correctly and efficiently for various language processing tasks.

It should be mentioned that we do not view our acceptance of a theory based on special cases as an abandonment of the hope of finding scientifically interesting generalizations to make about language. In fact, we believe there are principles of language use that can be derived from our approach. They are just not the principles one normally associates with language structure. Rather they are general principles of the application of this language knowledge. Interestingly, they are instances of more general principles that are also applicable to knowledge application that have no relation to language per se. Some of these principles are discussed in Wilensky and Arens (1980).

2.2. Sharable Knowledge Base

Language analysis and language production are of course very different problems. In language analysis, the task is to identify the meanings of incoming utterances; in production, the goal is to choose a language form the best encodes one's idea and intentions.

However, in spite of these differring natures, it is reasonable to ask what knowledge these tasks share in common. As the language the user speaks and understands is more or less the same, it would not seem unreasonable that knowledge he uses to encode an idea in a sentence and knowledge he uses to understand the meaning of that very same sentence should somehow be related.

In our model, it is assumed that the knowledge used for analysis and for production is by and large the same. That is, there is only one data base of knowledge about the meaning of a language's forms. This knowledge base may be indexed and therefore accessed differently for different tasks, thus accounting for some of the asymmetries between the analysis and production. But the language knowledge used by both tasks is the same knowledge represented the same way.

There are a number of reasons for believing that this assumption may be true for human language processors, although in the end, this would seem to be an empirical question. For example, people do not generally use words that they cannot understand, a possibility if their understanding and production knowledge were uncoupled. Also, it is certainly possible to talk about the meaning and use of a word or phrase independently of whether one is understanding or saying it. In fact, in our common understanding, separate analysis and production definitions for words, etc., are not recognized as acceptable.

However, the knowledge engineering reasons for this decision are more compelling. By having the knowledge of the two components be a shared data

base, only one form of representation is needed. Moreover, the addition of new knowledge to this data base extends the capabilities of both systems simultaneously. One need only assert a piece of knowledge about the meaning of a phrase to the data base, and system will be able to understand that phrase when it occurs, as well as be able to use that phrase to express an idea for which the phrase is appropriate. As this requirement forces knowledge to be represented declaratively, the other benefits of such representations are enjoyed as well.

2.3. Benefits of Declarative Representations

If language knowledge is to consist of one large data base used both for analysis and production, then it is imperative that this knowledge be stored in a highly declarative format. Only in this manner can the same knowledge be used for two quite different tasks.

Structuring the knowledge in this fashion entails several traditional knowledge engineering advantages. For example, in this format, knowledge about the language is kept separate from the processing strategies that apply this knowledge to the understanding and production tasks. Thus adding new knowledge requires only adding new assertions to the data base, not writing and debugging new code.

In addition, other knowledge besides the meaning of a phrase can be easily associated with such declaractive representations. For example, the context in which a certain phrase is appropriate may be stored together with the meaning of that phrase; the analyzer can use such knowledge to help infer the context, the production mechanism to decide whether or not to use that phrase in a particular situation.

Such additional information would be more difficult to introduce into a system that did not have phrasal knowledge stored as objects, i. e., that wasn't

phrasally oriented and didn't use declarative representations.

3. PHRAN and PHRED

Initially, our model was used as the basis for a natural language analysis program called PHRAN (PHRasal ANalyzer) written by Yigal Arens (Wilensky and Arens, 1980). PHRAN reads English sentences and produces representations from them that encode their meaning. Subsequently, the model was used as the basis for a language production mechanism, PHRED (PHRal English Diction), developed by Steven Upstill. PHRED takes meaning representations as input and expresses them in English sentences.

Both PHRAN and PHRED share a common data base of language knowledge. This data base contains declarative representations about what the phrase of the English language mean. This knowledge is stored in the form of *pattern-concept* pairs. A pattern is a phrasal construct of varying degrees of specificity. For example, it may be an exact literal string, such as "so's your old man"; it may be a pattern of limited flexibility such as "<nationality> restaurant", or "<person> <kick> the bucket"; or it may be a very general phrase such as "<person> <give> <person> <object>".

The concept part of a pattern-concept pair is a conceptual template that represents the meaning of the associated phrase. The conceptual template is a piece of meaning representation with possible references to pieces of the associated phrasal pattern. The meaning representation we use is a variant of Conceptual Dependency (Schank, 1975). Together, these pairs associate different forms of utterances with their meanings. For example, associated with the phrasal pattern "<nationality> restaurant" is the conceptual template denoting a restaurant that serves <nationality> type food; associated with the phrasal pattern "<person1> <give> <person2> <object>" is the conceptual template that denotes a transfer of possession by <person1> of <object> to <person2> from

<person1>.

PHRAN understands by reading the input text and trying to find the phrasal patterns that apply to it. As it reads more of the text it may eliminate some possible patterns and suggest new ones. At some point it may recognize the completion of one or more patterns in the text. It may then have to chose among possible conflicting patterns. Finally, the conceptual template associated with the desired pattern is used to generate the structure denoting the meaning of the utterance.

PHRED produces sentences that encode an idea by examining the same knowledge base. However, PHRED starts with a meaning representation it wishes to express, and tries to find conceptual templates that match it. If it finds more than one such template, it may have to chose between them. Then the phrasal pattern associated with the chosen conceptual template will be used to express the idea. Since these patterns may have variable pieces that relate to variable pieces of the conceptual template, PHRED must now find a way of expressing each subpart. The knowledge base of pattern-concept pairs is again consulted to find a way to do so; this knowledge is then used in a manner described by the initial pattern to form the appropriate mode of expression.

PHRAN and PHRED serve as the front and back end to various natural language processing systems. In general, PHRAN and PHRED perform that part of language processing that requires detailed knowledge of the specific language involved; the other components of the system perform reasoning based on more general, non-linguistic world knowledge. For example, PAM (Plan Applier Mechanism) is a story understanding program being developed at Berkeley that can make inferences about the goals and plans of a story's characters. PAM also knows about the relevant saliency of the story components it encounters, so it can distinguish the *points* of a story from the story's less interesting parts.

PAM uses PHRAN to read the initial sentences of the story and produce representations of their meaning. After it has read the story, made the necessary inferences and recognized the story points, PAM uses PHRED to create a summary of that story in English by generating just those parts of the story representation that constitute the story points. The following example is meant to demonstrate some of PHRAN and PHRED's language processing capabilities:

Input to PHRAN:

JOHN GRADUATED COLLEGE. JOHN LOOKED FOR A JOB. THE XENON CORPORATION GAVE JOHN A JOB. JOHN WAS WELL LIKED BY THE XENON CORPORATION. JOHN WAS PROMOTED TO AN IMPORTANT POSITION BY THE XENON CORPORATION.

JOHN GOT INTO AN ARGUMENT WITH JOHN'S BOSS. JOHN'S BOSS GAVE JOHN'S JOB TO JOHN'S ASSISTANT. JOHN COULDN'T FIND A JOB. JOHN COULDN'T MAKE A PAYMENT ON HIS CAR AND HAD TO GIVE UP HIS CAR. JOHN ALSO COULDN'T MAKE A PAYMENT ON HIS HOUSE, AND HAD TO SELL HIS HOUSE, AND MOVE TO A SMALL APARTMENT.

JOHN SAW A HIT AND RUN ACCIDENT. THE MAN WAS HURT. JOHN DIALED 911 THE MAN'S LIFE WAS SAVED. THE MAN WAS EXTREMELY WEALTHY, AND REWARDED JOHN WITH A MILLION DOLLARS. JOHN WAS OVERJOYED. JOHN BOUGHT A HUGE MANSION AND AN EXPENSIVE CAR, AND LIVED HAPPLY EVER AFTER.

After processsing by PAM...

Summary generated by PHRED:

JOHN WORKED FOR THE XENON CORPORATION.
THE XENON CORPORATION FIRED JOHN.
JOHN COULD NOT PAY FOR HIS HOUSE AND HIS CAR.
JOHN WAS BROKE.
A MAN GAVE JOHN SOME MONEY.
JOHN WAS RICH.
JOHN GOT A NEW CAR AND A NEW HOUSE.

In addition, PHRAN has been able to understander utterances of considerably greater complexity. For example, the following are examples of sentences PHRAN can understand, taken from Newsweek (March, 1980):

> Oilmen are encouraged by the amount of natural gas discovered in the Baltimore Canyon, an undersea trough about 100 miles off the New Jersey coast.

Tenneco, one of 39 companies engaged in drilling in the area, thinks its leased tract contains a marketable supply of gas.

The following are also examples of sentences PHRAN can process, selected to represent the variety of constructs it can handle:

Mary wanted to talk to the man who brought her son home.

The young man was told to drive quickly over to Berkeley.

John has gotten into another argument with his boss.

The man rewarded Bill with a million dollars for saving his life.

The book marker wanted by your mother is in the red box.

Willa's best friend is a bum who lives in Madison square.

If John gives Bill the big apple then Bill won't be hungry.

The school bus was driven by Mary's friend to The Big Apple.

John has kicked the bucket.

John kicked the red bucket.

The bucket was kicked by John.

The old French man's brother picked the book up.

If Mary brings John we will go to a Chinese restaurant.

Willa gives me a headache.

As the knowledge base shared by PHRAN and PHRED contains phrasal patterns of various levels of abstraction, each mechanism handles generative as well as non-generative constructs with a single processing strategy. Again, since their knowledge base is in fact shared, adding a single item to both data bases makes that item usable both to the language understanding and language production components.

PHRAN and PHRED are both written in UCILISP and run on a KL-10 at Berke-

ley. Their knowledge base contains over three hundred patterns. PHRAN averages about 1-5 cpu-sec to analyze a sentence; PHRED takes about the same time to generate one from an underlying concept.

No attempt has been made to optimize these programs, which are both currently uncompiled code.

3.1. PHRAN and PHRED in a model of language

As the discussion above indicates, PHRAN and PHRED are viewed as components of an integrated system that possesses other knowledge application abilities. While we talk about PHRAN and PHRED as separate components, they are not meant to be autonomous. Both PHRAN and PHRED rely on interactions with memory and inference modules that are not specificly language oriented. For example, in determining whether a phrase matches some component of a pattern, PHRAN may need to know if one of the meanings associated with that pattern has a certain feature, e. g., whether it represents an animate or inanimate object. The answers to such questions require retrieving or possibly deriving an answer that has nothing to do with language per se, and hence is outside of PHRAN's domain of expertise. The answer is in fact obtained by consulting a separate memory component, another part of the system within which PHRAN is running.

Thus PHRAN and PHRED only represent that part of our understanding mechanism that is language specific. This may be thought of as the very beginning and the very tail end of our language use apparatus. While we do not view these as autonomous components, we do view them as separable from the other components into which they are integrated. This view is in contrast to Schank, Lebowitz, and Birnbaum (1980), who emphasize the importance of integrated language understanding, but do not view the language specific component as having an independent status.

More traditional views of language have tended to break up language use into structural, meaning, and use components, i. e., syntax, semantics, and pragmatics. Our model, as well as Schank, Lebowitz, and Birnbaum's is not based on this particular distinction; in our case, syntactic, semantic and pragmatic knowledge may all be entertwined both functionally and structurally within a single pattern.

However, unlike Schank, Lebowitz, and Birnbaum, we wish to preserve a level of language processing that is distinct from other sorts of knowledge application. Namely, this is the level of application of language specific knowledge as embodied by the PHRAN-PHRED knowledge-base. Our work indicates to us that this knowledge is applied in a fashion quite different from general world knowledge, and therefore that it is worth preserving a distinction here. For example, the kind of processing needed to understand the relation between sentences of a text appears unrelated to that needed to understand the way in which words of a sentence join together to produce a meaning. These ideas are elaborated by Arens (1981).

The importance of this distinction is two-fold. First, it indicates that while language understanding is highly integrated with non-linguistic processing, that a separable level of language-specific processing is still isolatable. Secondly, since these other processes are viewed as being different in nature from PHRAN and PHRED, then we cannot view these programs as models for the entire process. That is, an integrated understander will contain components that are not designed along the same lines as PHRAN; a general production mechanism will contain parts that do not resemble the control structure of PHRED.

4. Using PHRAN for Other Languages

Patterns were developed for PHRAN to parse Chinese and Spanish sentences without requiring any change in PHRAN's mechanism. These patterns handle

grammatical constructions, idiomatic expressions, and, in the case of Chinese, recognize new words from patterns of one-syllable terms. No changes were made to the actual PHRAN code to tailor it to either of the languages. Moreover, the entire time spent on doing this encoding was about 6 half-time graduate student months by a student previously unfamiliar with PHRAN; in fact, it turned out not to be necessary to learn the details of how PHRAN works in order to complete this project.

Although these patterns were by and large modeled after those of English PHRAN, the patterns developed for Spanish and Chinese give PHRAN a better idea of these languages' syntax than the English patterns presently do of English syntax. The English patterns can be divided into basically three classes: patterns dealing with noun-phrases, patterns to handle adverbs, and patterns to handle the full sentence. A noun-phrase, for example, can include an article and adjectives. PHRAN recognizes an article/adjective/noun combination, outputs the appropriate conceptual information about how the adjective describes the noun, and turns the phrase into a single noun-phrase term.

Adverbs are handled specially because they can appear in many places in a phrase. Thus, when PHRAN sees an adverb, it removes it from the phrase. Then, after PHRAN gets a conceptualization for the phrase minus the adverb, PHRAN adjusts the conceptualization to account for the adverb.

The full sentence patterns divide English sentences into two types: active and passive. With the active case, for example, English PHRAN indexes these sentence patterns under the phrase (NOUN-PHRASE VERB), called the "target." Then, after matching the target, that is, after seeing a noun-phrase followed by a verb, PHRAN considers all the active voice patterns associated with the particular verb.

Spanish and Chinese PHRAN also use these three categories. With noun-

phrases, Spanish differs slightly because adjectives can come after the noun, but otherwise they are basically the same as in English. In Chinese, noun-phrases can be quite complicated, possibly just consisting of an article or modifier and dropping the noun altogether.

The big difference between Spanish and Chinese PHRAN and English PHRAN, however, lies in the adjective and sentence categories. First, English PHRAN does not make any special allowance for phrases that provide supplementary information, usually in the form of prepositional phrases. It handles a phrase like "to Mary" as an optional part of a particular verb. So to handle a sentence like "John sold the book to Mary," the verb "sell" would have associated with it the pattern [(PERSON) (ROOT SELL) (PHYSOB) ([(TO (PERSON))])], where ([(TO (PERSON))]) is an optional part of the pattern. In Spanish, these phrases can come in many places in the sentence. Even in English these phrases can come elsewhere in the sentence, as in "To you I give this book."

One possible solution, following in the spirit of English PHRAN's optional parts, would be to place these options throughout the pattern, in all the places they may appear. The problem with this approach is that frequently, many different options can apply. Consider the sentence "John sold the book to Mary in the park at noon for twenty dollars." The optional part approach would require including all four of the above options in the pattern.

The approach taken here classifies these supplementary phrases as "adverbs," that is, the patterns use the PHRAN adverb mechanism to independently remove these phrases from consideration in the sentence, with each phrase later modifying the resulting conceptualization. With this approach, each phrase need only have one pattern identifying it as an "adverb" instead of being included as an optional part to all verbs to which it may apply, and it can come anywhere in the sentence.

Adverbs can have restrictions, too, so that the pattern [TO (LOCATION)] modifies only change-of-location conceptualizations, whereas the pattern [AT (TIME)] can apply to any action. When a new pattern is suggested and there is an outstanding adverb, PHRAN checks the pattern to see if it meets the restriction of the adverb. If it meets the restriction and the pattern matches, PHRAN will modify the representation associated with the pattern appropriately.

Another difference between Chinese and Spanish PHRAN and English PHRAN lies in the categorization of sentence types. Whereas English PHRAN uses only the active/passive voice distinction, Spanish PHRAN and Chinese PHRAN classify sentences by their subject, verb, and object structure. These categories are subject/verb, subject/verb/direct-object, subject/verb/indirect-object/, subject/verb/indirect-object/direct-object, and subject/verb/phrase (the last category corresponding to a sentence containing an embedded clauses, as in "It seems that he is a good person"). Chinese PHRAN combines the second and third categories.

These distinctions are especially important for Spanish. Some pronouns in Spanish can be both indirect and direct objects, and an indirect pronoun is never marked, as opposed to English where the preposition "to" frequently precedes the indirect object. By using this categorization, when PHRAN sees the sentence "Juan me patea" ("John me kicks," that is, "John kicks me"), the sentence matches the general target pattern (NOUN-PHRASE DIR VERB). The entry associated with this target pattern says to look at the subject/verb/direct-object category of the verb, "patear". Since "me" can be either a direct or an indirect pronoun, the sentence also matches the target pattern (NOUN-PHRASE INDIR VERB), directing PHRAN to look under the subject/verb/indirect-object category of the verb "patear," The verb "patear" can only take a direct object, however, so there will be no entry under its subject/verb/indirect-object

category. PHRAN then uses the direct-object representation, disambiguating the pronoun.

Indexing patterns by verbs and indirect and direct objects involves a certain amount of syntactic parsing, enough to fit a phrase to one of the sentence categories. However, PHRAN does not perform a complete syntactic parse on a sentence. PHRAN does not distinguish, for instance, among different types of supplementary phrases, treating "here" and "at the house" as the same type of entity, even though syntactically "here" is an adverb and "at the house" is a prepositional phrase.

The sentence categorization also works well when using rewriting rules. Often a basic pattern must be rewritten to handle certain situations. The above Spanish sentence demonstrates that when using a pronoun, the object comes before the verb instead of after it. So PHRAN indexes an entry by the pattern (NOUN-PHRASE DIR VERB), that is, PHRAN activates the entry after seeing a noun-phrase, followed by a direct object and a verb. Since the basic direct-object sentence type places the object after the verb, the entry applies the rewriting rule [1 (AND 3 (DIR)) 2] to the direct-object pattern for "patear", [(PERSON) (ROOT PATEAR) (OR (PERSON) (PHYSOB))], producing the pattern [(PERSON) (AND (OR (PERSON) (PHYSOB)) (DIR)) (ROOT PATEAR)]. This rewritten pattern places the object before the verb and requires that it be a direct-object pronoun.

With this categorization, each category has a fixed length, with each member of the pattern having a fixed meaning; in the direct-object case above, the category has length 3, where the first item is the subject, the second is the verb, and the third is the direct object. So when using a rewriting rule with the subject/verb/direct-object category, one always knows that the pattern being rewritten will have three items in it, with 1 corresponding to the subject, 2 to

the verb, and 3 to the direct object.

It is probably the case that the different flavor to the rules between English PHRAN and Spanish and Chinese PHRAN does not reflect important differences in these languages. More likely, it suggests that our English patterns could be improved by structuring them similarly. However, we have not yet applied this structuring to the English patterns to find out exactly how conducive they will be to this formulation. It may be the case that while some of the same structuring applies, it will be less useful in English. For example, consider the use of the word "in" with the verb "arrive" to mean about the same as "to", as in "John arrived in New York." If we tried to handle prepositional phrases with the adverb mechanism a la Spanish and Chinese PHRAN, the system would not find "John arrived to New York" awkward, and would probably need an explicit reference to "in" in the "arrive" pattern in order to come up with the appropriate meaning of "arrive in." Thus, at least in this respect, much of the current structure of PHRAN's English patterns are likely to be preserved.

The following sections illustrate some pattern constructions in Chinese and Spanish. The examples are actual PHRAN input/output, edited to include an English translation after the input.

5. Chinese PHRAN

PHRAN understands the Pin-Yin romanization of the Chinese language, the official romanization of the People's Republic of China which is now coming into common use. The number which appears at the end of a word indicates which of the four tones is used to pronounce the word.

5.1. Noun-phrases in Chinese

Chinese noun-phrases can be quite complicated. PHRAN gives them the general form SUBSTANTIVE, a Chinese grammatical term. A SUBSTANTIVE has

the following basic structure: [DEMONSTRATIVE] [NUMBER] [MEASURE-WORD] [MODIFIER-LIST] [NOUN]. A demonstrative is either "zhei4" or "zhe4," meaning "this," or nei4" or "na4," meaning "that." Number is self-explanatory. Measure words are used because every noun in Chinese must be "measured." Sometimes English has a corresponding measure word, as in "three cups sugar." Usually, however, English has no equivalent. The Chinese measure word frequently has no actual meaning, unlike "cup" which has a real English and Chinese meaning. Instead, it is simply a required part of the syntax, so PHRAN recognizes measure words but ignores them in constructing the semantic definition. Finally, a modifier-list is a string of one or more modifiers, which tend to be adjective-like and whose structure will be discussed later.

Besides the full form shown above, each of the following subsets also constitutes a SUBSTANTIVE: [DEMONSTRATIVE], [DEMONSTRATIVE] [MEASURE-WORD], [NUMBER] [MEASURE-WORD], [DEMONSTRATIVE] [NUMBER] [MEASURE-WORD], plus all the previous forms followed by any one of [MODIFIER-LIST], [NOUN], or [MODIFIER-LIST] [NOUN], plus any of [MODIFIER-LIST], [NOUN], or [MODIFIER-LIST] [NOUN] by themselves. Modifiers may also sometimes come before the demonstrative. The following examples depict some instances of these SUBSTANTIVE forms:

```
Input:    NEI4 GE4
~         (That [measure])
~ Uses pattern [(P-O-S DEMONSTRATIVE) (P-O-S MEASURE)]
~  to create a term with form SUBSTANTIVE, giving it the
~  value SPEC14

Output:
SPEC14


Input:    SAN1  BEN3
~         (Three [measure])
~ Uses pattern [(P-O-S NUMBER) (P-O-S MEASURE)]
~  to create a term with form SUBSTANTIVE, giving
~  it the value SPEC15 and outputing the information
~  that the object is a group of 3 things.
```

Output:
(GROUP (OBJECT SPEC15) (DEGREE 3))
SPEC15

Input: HEN3 HAO3 DE1
~ (Very good)
~ Uses pattern:
~ [(P-O-S DEGREE) (P-O-S ADJECTIVE) DE1]
~ to create a term with from SUBSTANTIVE,
~ giving the term the value SPEC2 and outputing
~ the information that the object is "good."

Output:
(GOOD (OBJECT SPEC2) (DEGREE VERY))
SPEC2

Input: NEI4 LIANG3 GE4 FEI1 CHANG2 CHANG2 DE1
~ (That two extremely long
~ ie. those two extremely long ones)
~ Uses pattern
~ [(P-O-S DEMONSTRATIVE) (P-O-S NUMBER) (P-O-S MEASURE)]
~ to handle "NEI4 LIANG3 GE4," creating a term of
~ form SUBSTANTIVE and speech part SELECTOR.
~ It gives the term the value SPEC17 and outputs the
~ numerical information that the object is a group
~ of 2 things.
~ Then it uses the pattern
~ [(P-O-S SELECTOR) (P-O-S DEGREE) (P-O-S ADJECTIVE) DE1]
~ and creates a term with speech part NOUN-PHRASE
~ and form SUBSTANTIVE, retaining the value SPEC17 and
~ outputting the information that the object is "long."
Output:

(GROUP (OBJECT SPEC17) (DEGREE 2))
(LONG (OBJECT SPEC17) (DEGREE EXCEPTIONALLY))

SPEC17

Input: YI2 GE4 HAO3 REN2
~ (A good person)
~ Uses pattern [(P-O-S NUMBER) (P-O-S MEASURE)] to
~ change "yi2 ge4" into a term of form SUBSTANTIVE
~ and part of speech SELECTOR and assigns some value
~ to the object. Since the number is 1 (from a), no
~ group information is output.
~ Then it uses the pattern
~ [(P-O-S SELECTOR) (P-O-S ADJECTIVE) (P-O-S NOUN)]
~ to create a term with part of speech NOUN-PHRASE
~ and with form SUBSTANTIVE. It outputs the adjectival
~ information that the object is good and creates the
~ name "person6" to refer to it.

Output:
((MAN (OBJECT PERSON6)) (PERSON (OBJECT PERSON6)))

(GOOD (OBJECT PERSON6))

PERSON6

As shown in the explanation of the examples, when PHRAN sees a pattern of words that qualifies as a Chinese substantive, PHRAN collapses these words to a single term with form SUBSTANTIVE. PHRAN has a separate pattern for each substantive possibility which does not contain a substantive embedded within it. If the pattern is not a noun, PHRAN creates a term which has form SUBSTANTIVE and part of speech SELECTOR, as with the first example "Nei4 ge4." When first creating a selector, PHRAN also generates a name for the noun of the form SPECn to use to refer to in case the noun is omitted. Then, if other words follow which can combine with the SELECTOR to give a substantive, PHRAN creates a new SUBSTANTIVE term and makes it either a NOUN-PHRASE or a SELECTOR, depending on whether the pattern contains a noun or not. If the term is a selector, there is the possibility that the noun-phrase may contain still more of the following words. Once the term is a noun-phrase, however, the term is complete. Since the noun-phrase need not contain a noun, sentence patterns never check whether a term is a noun-phrase, but rather whether it is a substantive.

A serious conceptual problem arises with this approach that PHRAN does not presently handle. Some modifiers have different conceptual meanings depending on the particular noun. The term "wo3 de1" (my), for instance, usually signifies ownership, as in "wo3 de1 shu1" (my book). The phrase "wo3 de1 tou2" (my head), however, means that the head is a part of the person's body, not that he owns the head. PHRAN must process "wo3 de1" without knowing what the noun is, so it has no way to decide between the two representations.

This problem is inherent in the fact that Chinese noun-phrases can omit the

noun, the reason that forces PHRAN to pre-process the modifiers. It has a counterpart in English in that, while English always requires a noun or pronoun, the word "one" can be used, providing absolutely no information with which to select the appropriate representation. The sentence "Wo3 de1 hen3 hao3," meaning "Mine is good," provides no information to distinguish between the different representations entailed by "my." In this case, the same problem occurs in English since "mine" does not specify the particular noun either, referring to a previously specified item. In both languages, PHRAN would require contextual information to pick the correct representation. Such a contextual mechanism is now under development for PHRAN, and will hopefully alleviate this problem for all these languages.

5.2. Recognizing compound words

Besides grammatical structure considerations, PHRAN's pattern-based mechanism also works well in constructing new words. All words in Chinese are one syllable long. There exist a finite number of these one syllable words and PHRAN can recognize all of them, though not necessarily knowing all the meanings. But words can be grouped together to create a new term. The word "XIAN1," meaning "first," and "SHENG1," meaning "born," together become "XIAN1 SHENG1," meaning "teacher." The word "XUE2" means "study," so "XUE2 SHENG1" is "student." Peking is "BEI3 JING1," from "BEI3," meaning "north," and "JING1," meaning "capital."

Multi-word "words" are not restricted to nouns, either. The word "FEI1," a word with a strong negative connotation, together with the word "CHANG2," meaning "often," become the adverb of degree "FEI1 CHANG2," meaning extremely. The two words "XI3" and "HUAN1" together become the verb "XI3 HUAN1," which means "to like."

PHRAN's pattern mechanism makes it easy to specify a multi-word "word" as a pattern and create the new term. The entry for "Bei3 jing1" looks like

```
(INDEX-UNDER-PATTERN (BEI3 JING1)
 [(NIL
   [BEI3 (*)]
   [P-O-S 'NOUN
    FORM 'SUBSTANTIVE
    DESCRIPTION '(PEKING CITY PLACE LOCATION)
    PREDS '(CITY LOCATION)
    CD-FORM 'PEKING])])
```

It just uses the two words "bei3" and "jing1" as the target index. Once PHRAN sees these two words appear consecutively, it activates this entry, producing the pattern [BEI3 (*)]. The BEI3 says just match the word BEI3 and the lone star is an automatic match. No more checking is required because the index mechanism would not activate this entry had it not just read the expression "bei3 jing1." The other information in the entry is just the same as that for a normal one-word entry. PHRAN uses the DESCRIPTION field so it can refer to the noun in patterns by one of its descriptors, such as "location." The PREDS field contains the predicates associated with the noun.

Chinese also uses whole words as "suffixes." The word "GUO2," for example, meaning literally "kingdom," when preceded by various "country name" words signifies the Chinese word for that country. "YING1" sounds similar to England, so "YING1 GUO2" is England. "FA3" approximates the sound of France, so "FA3 GUO2" is France. The name for China, "ZHONG1 GUO2," has a more meaningful base, literally meaning "middle kingdom." As above the entry for such words contains normal one-word noun information:

```
(INDEX-UNDER-PATTERN (NATHEAD GUO2)
 [(NIL
   [(NATHEAD) (*)]
   [P-O-S 'NOUN
    FORM 'SUBSTANTIVE
    DESCRIPTION '(NATION COUNTRY PLACE LOCATION ABSTOB)
    ADJ (VALUE 1 ADJ)
    PREDS '(NATION LOCATION)
```

CD-FORM (VALUE 1 NATION)])])

"Nathead" appears as a descriptor for all word entries that can name a country, such as "Ying1," "Fa3," and "Zhong1" mentioned above. The NATION field of these entries contains the name of the nation, such as "England." The ADJ field contains the adjectival form of the name, as with "English."

A place, such as a country or city, followed by the word "REN2," meaning "person," means a person from that place. So we get

Input: MEI3 GUO2 REN2
~ (America person)
~ Uses the pattern [(PLACE) REN2].
Output:
((PERSON (OBJECT PERSON1)) (MAN (OBJECT PERSON1)))

(ORIGIN (OBJECT PERSON1) (LOCATION USA))

PERSON1

A country followed by "HUA4," meaning "speech," is one way of signifying that country's language. Thus "ZHONG1 GUO2 HUA4" is Chinese and "YING1 GUO2 HUA4" is English. The pattern [(COUNTRY) HUA4] handles this case. A different method precedes the word "WEN2," meaning vaguely "literary," by the same "country name" type of word that precedes "GUO2" to form the language name. This method uses the pattern [(NATHEAD) WEN2]. So "ZHONG1 GUO2" is China and "ZHONG1 WEN2" is Chinese. Both the methods use the adjectival form associated with the country for their representations.

In those cases where the country name does not incorporate "GUO2," the whole name usually precedes "WEN2." Thus one has

Input: XI1 BAN1 YA2
~ (Spain)
Output:
((LOCATION (OBJECT SPAIN)) (NATION (OBJECT SPAIN)))

SPAIN

and

```
Input:     XI1 BAN1 YA2 WEN2
~          (Spanish )
~ Uses the pattern [(NATHEAD) WEN2].
~ Since the whole country name must appear before the WEN2
~   to construct the language name, the entry for Spain,
~   [XI1 BAN1 YA2], incorporates NATHEAD as one of its
~   descriptors.
Output:
((LANGUAGE (OBJECT SPANISH)))

SPANISH
```

5.3. Idioms and special phrases

PHRAN can also handle special phrases in Chinese. Normally all modifiers are followed by the particle "DE1." Some often-used phrases or one-word adjectives can omit the DE1. A personal pronoun, for example, can immediately precede the word "PENG2 YOU3," meaning "friend," as in

```
Input:     WO3 PENG2 YOU3
~          (I  friend
~     ie. my friend)
~ Uses the pattern [(PERSON) (FRIEND)]
Output:
((PERSON (OBJECT PERSON2))
(FRIEND (OBJECT PERSON2))
(REFERENT (OBJECT EG01))
(PERSON (OBJECT EG01)))

(RELATION (ACTOR EG01) (FRIEND PERSON2))

PERSON2
```

A language word can immediately precede a word indicating some type of literature. So "DE2 WEN2," meaning "German," and "ZA2 ZHI4," meaning "magazine," become

```
Input:     DE2 WEN2 ZA2 ZHI4
~          (German  magazine)
~ Uses the pattern [(LANGUAGE) (LITERATURE)]
Output:
((PHYSOB (OBJECT MAGAZINE1))
(LITERATURE (OBJECT MAGAZINE1))
(MAGAZINE (OBJECT MAGAZINE1))
(LANGUAGE (OBJECT GERMAN)))
```

```
(LITERATURE (OBJECT MAGAZINE1) (TYPE GERMAN))

MAGAZINE1
```

Using the same pattern, "FA3 WEN2," meaning French, and "BA04," meaning "newspaper," become

```
Input:     FA3 WEN2 BA04
~          (French  newspaper)
Output:
((PHYSOB (OBJECT NEWSPAPER1))
(LITERATURE (OBJECT NEWSPAPER1))
(NEWSPAPER (OBJECT NEWSPAPER1))
(LANGUAGE (OBJECT FRENCH)))

(LITERATURE (OBJECT NEWSPAPER1) (TYPE FRENCH))

NEWSPAPER1
```

A special phrase in Chinese as well as English is a place followed by the word for "restaurant," as in the English "Chinese restaurant." The meaning is a place that serves Chinese-type food as opposed to a restaurant in China or a restaurant from China. So PHRAN produces

```
Input:     ZHONG1 GUO2 FAN4 DIAN3
~          (China       restaurant)
~ Uses the pattern [(PLACE) (RESTAURANT)]
Output:
(RESTAURANT (OBJECT RESTAURANT1) (TYPE CHINESE))

RESTAURANT1
```

5.4. Chinese PHRAN's basic sentence categories

Chinese PHRAN uses four basic sentence types: SIMPLE, SINGLE, DOUBLE, and SENTOBJ. A SIMPLE sentence describes some action with no object involved. The verb "XUE2 XI2" (study), for example, has the information

```
((SIMPLE)
   [(FORM SUBSTANTIVE) (NAME XUE2XI2)]
   [CONCEPT '($STUDY (TENSE ?TENSE) (MOOD ?MOOD) (MODE ?MODE)
              (ACTOR ?SUBJECT))
    OLD-CD (VALUE 2 CD)
    CD (OR (TERM'S OLD-CD) (TERM'S CONCEPT))])
```

associated with it. The pattern says there should be an item with form SUBSTANTIVE followed by an item with name "xue2xi2." Verbs are identified by the NAME field. The associated concept specifies that the conceptual information associated with "xue2 xi2" is the STUDY plan. The actor is the subject. Tense, mood, and mode are optional fields. The tense is almost never specified in Chinese. The mood is the type of question, such as interrogative, and does not appear for normal declarative sentences. The mode field indicates the manner of an action, such as a PTRANS with MODE "fast" to indicate running. The field OLD-CD appears because combining verbs (see later) creates a new verb term with a conceptual form associated with it. This conceptualization would incorporate the present term's CONCEPT and would be the OLD-CD. The actual triggering pattern for a normal SIMPLE sentence is

```
(INDEX-UNDER-PATTERN (SUBSTANTIVE VERB)
(GET (FROM-END 1 NAME))
[1 (* AND 2 (SEP NEXT))]
(CD-FORM '?CD
MODE (VALUE 2 MODE)
SUBJECT (VALUE 1)
TENSE (VALUE 2 TENSE)
DO (SAVE-PREDICATES (TERM'S SUBJECT) (VALUE 1 PREDS))
P-O-S 'SENTENCE) SIMPLE)
```

Upon seeing a SUBSTANTIVE followed by a VERB, the mechanism gets the pattern associated with the particular verb name under the SIMPLE category. Since this pattern matches the normal SIMPLE sentence, the "rewriting" is just the original order, taking the first item, (FORM SUBSTANTIVE), followed by the second item, (NAME XUE2XI2), to give [(FORM SUBSTANTIVE) (* AND (NAME XUE2XI2) (SEP NEXT))]. The rewriting line, [1 (* AND 2 (SEP NEXT))], accomplishes this straightforward transformation. The asterisk simply indicates to PHRAN the current position in the pattern. The (SEP NEXT) clause specifies that the next word should be a separator of some kind, usually the end of the sentence. This clause prevents PHRAN from seeing a

sentence with an object and mistakenly grabbing just the SUBJECT and VERB and calling that part of the sentence the whole sentence. The CD-FORM field gets the CD representation associated with the verb. The MODE and TENSE fields get any MODE or TENSE field that may be associated with the verb. The SUBJECT field gets the CD-FORM associated with the subject, the SUBSTANTIVE. The SAVE-PREDICATES command tells PHRAN to output any predicate information associated with the subject. Finally, the P-O-S field assigns SENTENCE as the part of speech of this pattern. As an example of a SIMPLE sentence, consider

```
Input:     TA1 XUE2 XI2
~     (He  studies )
Output:
((REFERENT (OBJECT HUMAN7)) (PERSON (OBJECT HUMAN7)))

($STUDY (ACTOR HUMAN7))
```

A SINGLE-type sentence takes a single object. It may be a direct object, as in

```
Input:     WO3 YOU3 SAN1   BEN3 BEN3 ZI3
~          (I  have three     notebooks)
Output:
((REFERENT (OBJECT EG04))
(PERSON (OBJECT EG04))
(NOTEBOOK (OBJECT NOTEBOOK2))
(PHYSOB (OBJECT NOTEBOOK2)))

(GROUP (OBJECT NOTEBOOK2) (DEGREE 3))

(POSS (ACTOR EG04) (OBJECT NOTEBOOK2))
```

or it may be an indirect object, as in

```
Input:     NI3 PENG2 YOU3 JIA01 WO3
~          (You friend    teach I
~     ie. Your friend teaches me)
Output:
((PERSON (OBJECT PERSON2))
(FRIEND (OBJECT PERSON2))
(REFERENT (OBJECT EG05))
(PERSON (OBJECT EG05))
(FAMILIAR (OBJECT YOU5))
(REFERENT (OBJECT YOU5))
(PERSON (OBJECT YOU5)))
```

(RELATION (ACTOR YOU5) (FRIEND PERSON2))

($TEACH (ACTOR PERSON2) (STUDENT EGO5))

To distinguish between the two above cases, PHRAN has two different patterns associated with the verb "JIAO1" under the SINGLE category. The first pattern, [(FORM SUBSTANTIVE) (NAME JIAO1) (PERSON)], requires that the object be a person, corresponding to the indirect object case. The second pattern, [(FORM SUBSTANTIVE) (NAME JIAO1) (AND (NOT (PERSON)) (FORM SUBSTANTIVE))], requires that the object be any normal substantive (noun) except for a person. It corresponds to the direct object case.

A DOUBLE sentence takes two objects and will always be of the form SUBJECT VERB INDIRECT-OBJECT DIRECT-OBJECT:

Input: NEI4 GE4 XIAN1 SHENG1 JIAO1 NEI4 GE4 XUE2 SHENG1 ZHONG1 WEN2
~ (That teacher teach that student Chinese)
Output:
((TEACHER (OBJECT TEACHER2))
(PERSON (OBJECT TEACHER2))
(STUDENT (OBJECT STUDENT2))
(PERSON (OBJECT STUDENT2))
(LANGUAGE (OBJECT CHINESE)))

($TEACH (ACTOR TEACHER2) (STUDENT STUDENT2) (OBJECT CHINESE))

The pattern PHRAN associates with "JIAO1" under the category DOUBLE to achieve the above example looks like [(FORM SUBSTANTIVE) (NAME JIAO1) (PERSON) (FORM SUBSTANTIVE)]. It restricts only the form of the indirect object, requiring that it be a person.

The final basic sentence type, SENTOBJ, deals with embedded phrases. The verb "xiang3" (to think) has the following pattern associated with it under the category SENTOBJ: [(FORM SUBSTANTIVE) (NAME XIANG3) (P-O-S SENTENCE)]. This pattern requires that a clause that could stand alone as a sentence follow the verb, as in

Input: TA1 XIANG3 WO3 YOU3 TA1 DE1 SHU1

```
~        (He think I  have his   book)
Output:
((REFERENT (OBJECT HUMAN8))
(PERSON (OBJECT HUMAN8))
(REFERENT (OBJECT EGO6))
(PERSON (OBJECT EGO6))
(BOOK (OBJECT BOOK2))
(LITERATURE (OBJECT BOOK2))
(PHYSOB (OBJECT BOOK2))
(REFERENT (OBJECT HUMAN9))
(PERSON (OBJECT HUMAN9)))

(OWN (TENSE PRESENT) (ACTOR HUMAN9) (OBJECT BOOK2))

(MBUILD (ACTOR HUMAN8) (MOBJECT (POSS (ACTOR EGO6) (OBJECT BOOK2))))
```

Here WO3 YOU3 TA1 DE1 SHU1 is a full sentence of type SINGLE that forms the object of XIANG3 (think).

There is one additional type of sentence in Chinese not covered by the basic types because it does not involve verbs. It has the form SUBJECT DEGREE ADJECTIVE. The degree word is a term describing the degree of an adjective, such as "HEN3" (very) or "FEI1 CHANG2" (extremely). So

```
Input:     NEI4 GE4 REN2     HEN3 HAO3
~          (That      person very good)
Output:
((MAN (OBJECT PERSON5)) (PERSON (OBJECT PERSON5)))

(IS (OBJECT PERSON5)
  (STATE-NAME QUALITY)
  (VALUE GOOD)
  (DEGREE VERY))
```

and

```
Input:     TA1 FEI1 CHANG2 LEI4
~          (He extremely     tired)
Output:
((REFERENT (OBJECT HUMAN11)) (PERSON (OBJECT HUMAN11)))

(IS (OBJECT HUMAN11)
  (STATE-NAME ALERTNESS)
  (VALUE TIRED)
  (DEGREE EXCEPTIONALLY))
```

are both normal Chinese sentences, matching the pattern

[(FORM SUBSTANTIVE) (P-O-S DEGREE) (P-O-S ADJECTIVE)].

5.5. Rewriting rules

All Chinese sentences have one of the above forms or some simple rewriting of one of them. There are in Chinese, for example, several ways to form a question. The simplest way is to append the particle MA1 to the sentence, as in

```
Input:      NI3 HUI4 XI1 BAN1 YA2 WEN2 MA1
~           (You know Spanish    ? )
((FAMILIAR (OBJECT YOU6))
(REFERENT (OBJECT YOU6))
(PERSON (OBJECT YOU6))
(LANGUAGE (OBJECT SPANISH)))

(MLOC (MOOD INTERROGATIVE) (OBJECT SPANISH) (LTM (ACTOR YOU6)))
```

To recognize this case, PHRAN uses the pattern [(P-O-S SENTENCE) MA1], that is, a sentence followed by the special question particle MA1. When PHRAN sees this pattern, it takes the CD information now associated with the SENTENCE and adds the field "(MOOD INTERROGATIVE)." Another slightly more complicated method is the "choice question." It gives both the main verb and the negation of the verb. The negation may immediately follow the positive form, as in

```
Input:      NI3 GEI3 BU4 GEI3 TA1 NEI4 ZHI1 QIAN1 BI3
~           (You give not give he  that    pencil
~      ie. Did you give him that pencil?)
Output:
((FAMILIAR (OBJECT YOU7))
(REFERENT (OBJECT YOU7))
(PERSON (OBJECT YOU7))
(REFERENT (OBJECT HUMAN10))
(PERSON (OBJECT HUMAN10))
(PENCIL (OBJECT PENCIL3))
(PHYSOB (OBJECT PENCIL3)))

(ATRANS (MOOD INTERROGATIVE)
     (ACTOR YOU7)
     (OBJECT PENCIL3)
     (FROM YOU7)
     (TO HUMAN10))
```

or it can come at the end of the sentence, as in

```
Input:      NI3 REN4 SHI4 ZHEI4 GE4 REN2  BU2 REN4 SHI4
~           (You recognize that      person not recognize
~      ie. Do you recognize that person?)
Output:
(FAMILIAR (OBJECT YOU8))
(REFERENT (OBJECT YOU8))
(PERSON (OBJECT YOU8))
(MAN (OBJECT PERSON3))
(PERSON (OBJECT PERSON3)))

(MLOC (MOOD INTERROGATIVE)
   (OBJECT (DESCRIPTION (ITEM PERSON3)))
   (LTM (ACTOR YOU8)))
```

PHRAN can obtain patterns for these questions by using rewriting rules. Both the above "choice" examples are of type SINGLE, which has the form SUBJECT VERB OBJECT. The entry for the first choice pattern is indexed under the target pattern (SUBSTANTIVE VERB NEGATIVE VERB SUBSTANTIVE). This entry then uses the rewrite rule [1 2 (FORM NEGATIVE) 2 3] to rewrite the pattern associated with the particular verb, where 1, 2, and 3 match the conditions on the SUBJECT, VERB, and OBJECT for the particular verb. This rule changes the normal pattern for the verb in the example from [(FORM SUBSTANTIVE) (NAME GEI3) (FORM SUBSTANTIVE)] to [(FORM SUBSTANTIVE) (NAME GEI3) (FORM NEGATIVE) (NAME GEI3) (FORM SUBSTANTIVE)]. The second choice form, indexed under the pattern (SUBSTANTIVE VERB SUBSTANTIVE NEGATIVE VERB), uses the rewriting rule [1 2 3 (FORM NEGATIVE) 2].

The descriptive, verbless sentence can also use the choice pattern to form a question. In this case, the adjective is repeated; so PHRAN contains the pattern [(FORM SUBSTANTIVE) (P-O-S ADJECTIVE) (FORM NEGATIVE) (P-O-S ADJECTIVE)]. This choice question does not permit the use of a degree word. To ask a question with a degree word, such as "Is he extremely ill," one must append the question word MA1 to the end of the sentence.

Simply negating a sentence also uses a rewriting rule. Again, for the case of a SINGLE type sentence, one writes [1 (FORM NEGATIVE) 2 3]. One might want

to have a single pattern something like [(FORM NEGATIVE) (P-O-S VERB)] to create a new verb term with a field marking it as negative, thus avoiding the necessity of using rewrites for all the different possible negative sentence forms. Having multiple patterns, however, does not increase the time it takes PHRAN to parse a sentence, whereas if a field were used in the verb to note the negative, that field would have to be checked every time a verb is processed to see whether to make the representation negative.

The descriptive type sentence does not require a separate pattern for the negative form because the negative particles, BU2 and BU4, are DEGREE words themselves. So the sentence

```
Input:     NEI4 ZHANG1 ZHUO1 ZI3 BU4 GAN1 JING4
~          (That        table    not clean    )
Output:
((TABLE (OBJECT TABLE1)) (PHYSOB (OBJECT TABLE1)))

(IS (OBJECT TABLE1)
  (STATE-NAME CLEANLINESS)
  (VALUE CLEAN)
  (DEGREE NOT))
```

describes the table's cleanliness with degree "not," using the pattern [(FORM SUBSTANTIVE) (P-O-S DEGREE) (P-O-S ADJECTIVE)].

5.6. Consecutive verbs

Frequently multiple verbs appear together. Chinese PHRAN combines these verbs to a single verb, adjusting the conceptual representation to account for the composing. For instance, in English one can say "I like to eat." The Chinese equivalent of "like to eat" is "xi3 huan1 chi1," where "xi3 huan1" means "like" and "chi1" means "eat." When PHRAN sees "xi3 huan1 chi1," it sees a "VERB VERB" combinatin, activating the following entry:

```
(INDEX-UNDER-PATTERN (VERB VERB)
[GET (FROM-END 2 NAME)]
 [1 (* AND 2)]
 [P-O-S 'VERB
```

```
NAME (VALUE 2 NAME)] VERBOBJ)
```

This pattern creates a new term, setting the part of speech to "verb" and the name to the name of the second verb, "chi1." It gets the rest of the information by calling up the entry listed under the category VERBOBJ of the verb second from the end, "xi3 huan1":

```
((VERBOBJ)
   [(NAME XI3HUAN1) (P-O-S VERB)]
   [CONCEPT '(ATTITUDE (TENSE ?TENSE) (MOOD ?MOOD) (MODE ?MODE)
              (ACTOR ?SUBJECT) (SCALE FONDNESS)
              (CONCEPT ?CONCEPT) (LEVEL HIGH))
    TENSE 'FUTURE
    CD (OR (VARSUB (VALUE 1 CD) (TERM'S CONCEPT) '?CONCEPT)
       (TERM'S CONCEPT))]))
```

This entry says that whenever the phrase "xi3 huan1 verb" appears, the conceptual respresentation is that the actor has a high attitude of fondness toward the concept associated with the verb following "xi3 huan1." The "CD" entry just checks whether there is some concept already associated with just the "xi3 huan1" term. If so, then there was a previous phrase of the form "VERB xi3 huan1," and this ATTITUDE concept gets combined with that concept, just as the concept associated with "chi1" will be embedded within the ATTITUDE concept.

Consider the Chinese sentence:

```
Input:     WO3 XI3 HUAN1 CHI1 FAN4
~          (I  like   eat  food)
Output:
((REFERENT (OBJECT EGO1))
(PERSON (OBJECT EGO1))
(PHYSOB (OBJECT FOOD1))
(FOOD (OBJECT FOOD1)))

(ATTITUDE (ACTOR EGO1)
      (SCALE FONDNESS)
      (CONCEPT (INGEST (ACTOR EGO1) (OBJECT FOOD1)))
      (LEVEL HIGH))
```

After composing the two verbs "xi3 huan1 chi1," the result matches the SINGLE pattern [(FORM SUBSTANTIVE) (NAME CHI1) (FORM SUBSTANTIVE)]. Unlike in

English, the Chinese verb for "to eat," chi1, requires an object. The resulting INGEST conceptualization is embedded in the ATTITUDE conceptualization associated with "xi3 huan1."

English PHRAN does not use this type of verb composition. Instead, it has an "infinitive-phrase" clause. In the sentence "John wants to eat some food," for example, English PHRAN classifies "to eat some food" as an infinitive-phrase and matches the sentence against the pattern [(PERSON) (ROOT WANT) (P-O-S INF-PHRASE)]. Chinese PHRAN avoids this method for two reasons. First, unlike in English where the special infinitive verb form sets apart this type of sentence, Chinese uses only one verb form, making it difficult to recognize an "infinitive-phrase" clause. Secondly, Chinese PHRAN uses the method to be consistent with the way Spanish PHRAN handles consecutive verbs. As seen previously, Spanish pronouns come before the verb. So the Spanish translation of "John wants to eat it" places the pronoun "it" before the verb "wants," even though the pronoun applies to the verb "eat." So Spanish PHRAN has to compose the verbs "want to eat," giving the new term the name "eat." Then the pronoun comes in the right position.

5.7. Chinese modifiers

Rewriting rules can be used on the basic sentence types to create some types of modifiers: those that are equivalent to the English relative clause. In English these clauses come in two types: 1) the relative clause shares its subject, as with "teacher" in "The teacher who knows Chinese teaches me" and in "I know the teacher who teaches you Russian." and 2) the relative clause shares an object, as with "book" in "The book that you do not want to read is very interesting" and "I do not like the book that you gave me."

In Chinese relative clauses come in the same two types, but they act like adjectives, always preceding the noun, as if in English the first example above

were "The knows Chinese teacher teaches my friend." Apart from several idiomatic exceptions, all adjectival phrases in Chinese are followed by the particle "de1," so for the first relative clause type, which shares it subject, the patterns for SIMPLE, SINGLE, DOUBLE, and SENTOBJ sentence types are used, minus the subject and followed by "de1." Consider, for example, the following pattern used with the SINGLE type:

```
(INDEX-UNDER-PATTERN (VERB SUBSTANTIVE DE1)
(GET (FROM-END 3 NAME))
[(AND 2 (NOT (PREVIOUS FORM SUBSTANTIVE))) 3 (* AND DE1)]
(OBJECT (VALUE 2)
 MODE (VALUE 1 MODE)
 CD-FORM (NEWSYM SPEC)
 SUBJECT (TERM'S CD-FORM)
 FORM 'SUBSTANTIVE
 TENSE (VALUE 1 TENSE)
 DO (SAVE-PREDICATES (TERM'S OBJECT) (VALUE 2 PREDS))
 DO (ADD-TO-*SC* '?CD)
 P-O-S 'SELECTOR) SINGLE)
```

The trigger (VERB SUBSTANTIVE DE1) activates this entry when it sees a verb followed by a substantive and the particle "de1," possibly the verb and object of a modifier sharing its subject. The entry looks at the SINGLE category listed under the particular verb, creating for the verb HUI4 ("to know" as "to know a language"), for example, the rewritten pattern [(AND (NAME HUI4) (NOT (PREVIOUS FORM SUBSTANTIVE)))
(FORM SUBSTANTIVE) (* AND DE1)]. The concept associated with HUI4 under the SINGLE category looks like

```
(MLOC (TENSE ?TENSE) (MOOD ?MOOD) (MODE ?MODE)
   (OBJECT ?OBJECT) (LTM (ACTOR ?SUBJECT)))
```

The modifier entry tries to fill in the MODE and TENSE slots from those respective fields of the verb. It takes the OBJECT as the value of the substantive in the expression. The subject, however, required to fill the ACTOR slot, has not yet appeared. Information associated with the subject must come from the noun

that follows, but the relative clause is one of possibly multiple modifiers of that noun.

Recall that a modifier list is part of a SUBSTANTIVE. As mentioned above, PHRAN generates names of the form SPECn when dealing with SUBSTANTIVES, so when outputting CD information about the relative clause, PHRAN uses the previously generated name for the subject or, as is the case for the above example, generates one if the clause is the first pattern of the SUBSTANTIVE. After filling the slots, PHRAN outputs the conceptualization as a supplementary concept. If the SUBSTANTIVE has an actual noun associated with it, PHRAN later replaces the SPECn name by the appropriate noun-associated name in the CD information, as with "teacher4" in the MLOC conceptualization and "teacher1" in the $TEACH conceptualization in the following examples.

In Chinese the two shared-subject relative clause examples given in English above are

```
Input:      HUI4  ZHONG1 WEN2 DE1 XIAN1 SHENG1 JIAO1 WO3
~           (Knows Chinese          teacher      teach I
~     ie. The teacher who knows Chinese teaches me)
~ Uses the SINGLE rewriting shown above to change the modifier
~   "Hui4 zhong1 wen2 de1" into a selector, where HUI4 is the
~   verb and ZHONG1 WEN2 is the substantive, followed by DE1,
~   creating a new selector term and outputing the supplemental
~   LOC conceptualization.
~ Next the pattern [(P-O-S SELECTOR) (P-O-S NOUN)] matches the
~   resulting selector and the noun XIAN1 SHENG1.
~ Then the SINGLE sentence pattern
~   [(FORM SUBSTANTIVE) (NAME JIAO1) (PERSON)] matches the resulting
~   noun-phrase, the verb JIAO1, and the person WO3.
Output:
((TEACHER (OBJECT TEACHER4))
(PERSON (OBJECT TEACHER4))
(REFERENT (OBJECT EGO8))
(PERSON (OBJECT EGO8))
(LANGUAGE (OBJECT CHINESE)))

(LOC (OBJECT CHINESE) (MLOC (LTM (ACTOR TEACHER4))))

($TEACH (ACTOR TEACHER4) (INDDOBJ EGO8))
```

and

```
Input:      WO3 REN4 SHI4 JIAO1 NI3 E4 WEN2 DE1 XIAN1 SHENG1
~           (I  recognize teach you Russian    teacher
~      ie. I know the teacher who teaches you Russian)
~ A modifier entry similar to the one above but operating on the
~   DOUBLE category creates the pattern
~   [(NAME JIAO1) (PERSON) (FORM SUBSTANTIVE) DE1] to match the modifier
~   consisting of the verb JIAO1, the person NI3, the substantive
~   E4 WEN2, and the particle DE1, producing a new selector term and
~   outputing the supplemental $TEACH conceptualization.
~ Next, the selector term and the noun XIAN1 SHENG1 match
~   the pattern [(P-O-S SELECTOR) (P-O-S NOUN)] to create a new
~   noun-phrase substantive.
~ Finally the SINGLE pattern
~   [(FORM SUBSTANTIVE) (NAME REN4SHI4) (FORM SUBSTANTIVE)] matches
~   the substantive WO3, the verb JIAO1, and the newly created
~   substantive to finish parsing the sentence.
Output:
((REFERENT (OBJECT EGO1))
(PERSON (OBJECT EGO1))
(TEACHER (OBJECT TEACHER1))
(PERSON (OBJECT TEACHER1))
(FAMILIAR (OBJECT YOU1))
(REFERENT (OBJECT YOU1))
(PERSON (OBJECT YOU1))
(LANGUAGE (OBJECT RUSSIAN)))

($TEACH (ACTOR TEACHER1) (STUDENT YOU1) (OBJECT RUSSIAN))

(LOC (OBJECT (DESCRIPTION (ITEM TEACHER1)))
   (MLOC (LTM (ACTOR EGO1))))
```

The fifth, descriptive Chinese sentence type, corresponding to the English "SUBJECT IS ADJECTIVE," also maps to the shared-subject relative clause. This mapping is the same as normal adjectives in English. So "the short pencil" and "the pencil that is short," two different but equivalent English phrases, both map to the same MODIFIER NOUN form in Chinese:

```
Input:      DUAN3 DE1 QIAN1 BI3
~           (Small      pencil  )
~ Uses the pattern [(P-O-S ADJECTIVE) DE1] to match DUAN3 DE1
~   and create a selector.
~ Then the pattern [(P-O-S SELECTOR) (P-O-S NOUN)] matches the
~   newly created selector and the noun QIAN1 BI3 to create
~   the final noun-phrase term.
Output:
((PENCIL (OBJECT PENCIL1)) (PHYSOB (OBJECT PENCIL1)))

(SHORT (OBJECT PENCIL1))
```

PENCIL1

The pattern [(P-O-S DEGREE) (P-O-S ADJECTIVE) (P-O-S NOUN)] can also be used to permit specifying the degree of an adjective.

The second relative clause type, that which shares an object, uses a similar rewriting scheme to share the direct object. The only sentence types with applicable direct objects are SINGLE and DOUBLE, so they are the only types used to rewrite the patterns. This scheme applies also to an indirect object when it is the only object, that is, it is the SINGLE sentence type. The slot filled by the object in the CD representation uses the applicable SPECn name, possibly replaced later by the name associated with the noun. The Chinese translations for the above English examples are

```
Input:      NI3 BU2 YAO4 KAN4 DE1 SHU1 HEN3 YOU3 YI4 SI1
~           (You not want see      book very interesting
~      ie. The book that you do not want to see is
~           very interesting)
~ First the two verbs YAO4 KAN4 are combined under the name KAN4.
~ Then a rewrite rule on a SINGLE type, grabbing the subject and verb,
~   produces the pattern
~   [(FORM SUBSTANTIVE) (FORM NEGATIVE) (NAME KAN4) DE1] to match the
~   substantive NI3, the negative BU2, the verb YAO4 KAN4, and the
~   particle DE1, creating a new selector term and outputing the
~   supplementary INTENT/ATTEND conceptualization.
~ Next the pattern [(P-O-S SELECTOR) (P-O-S NOUN)] matches this new
~   selector and the noun SHU1 to produce a new noun-phrase substantive.
~ Finally the pattern
~   [(FORM SUBSTANTIVE) (P-O-S DEGREE) (P-O-S ADJECTIVE)] matches the
~   new substantive, the degree HEN3, and the adjective YOU3 YI4 SI1.
Output:
((BOOK (OBJECT BOOK1))
(LITERATURE (OBJECT BOOK1))
(PHYSOB (OBJECT BOOK1))
(FAMILIAR (OBJECT YOU2))
(REFERENT (OBJECT YOU2))
(PERSON (OBJECT YOU2)))

(NOT (INTENT (TENSE FUTURE)
        (ACTOR YOU2)
        (INTENTION (ATTEND (TENSE FUTURE)
                (ACTOR YOU2)
                (OBJECT EYES)
                (TO BOOK1)))))
```

```
(IS (OBJECT BOOK1)
  (STATE-NAME QUALITY)
  (VALUE INTERESTING)
  (DEGREE VERY))
```

and

```
Input:      WO3 BU4 XI3 HUAN1 NI3 GEI3 WO3 DE1 SHU1
~           (I  not like      you give I      book
~     ie. I do not like the book that you gave me)
~ The modifier pattern [(FORM SUBSTANTIVE) (NAME GEI3) (PERSON) DE1],
~   created by rewriting the DOUBLE entry for GEI3, matches the subject
~   substantive NI3, the verb named GEI3, the person WO3, and the
~   particle DE1, creating a new selector term and outputing an
~   ATRANS conceptualization
~ The pattern [(P-O-S SELECTOR) (P-O-S NOUN)] then matches this selector
~   and the noun SHU1 to create a new noun-phrase substantive.
~ Finally the SINGLE pattern
~   [(FORM SUBSTANTIVE) (FORM NEGATIVE) (NAME XI3HUAN1) (FORM SUBSTANTIVE)]
~   matches the substantive WO3, the negative BU4, the verb XI3 HUAN1, and
~   the newly created substantive to parse the sentence.
Output:
((REFERENT (OBJECT EGO2))
(PERSON (OBJECT EGO2))
(BOOK (OBJECT BOOK2))
(LITERATURE (OBJECT BOOK2))
(PHYSOB (OBJECT BOOK2))
(FAMILIAR (OBJECT YOU3))
(REFERENT (OBJECT YOU3))
(PERSON (OBJECT YOU3))
(REFERENT (OBJECT EGO3))
(PERSON (OBJECT EGO3)))

(ATRANS (ACTOR YOU3) (OBJECT BOOK2) (FROM YOU3) (TO EGO3))

(NOT (CONCEPT (ATTITUDE (ACTOR EGO2)
                (SCALE FONDNESS)
                (OBJECT BOOK2)
                (LEVEL HIGH))))
```

Rewritings must also include the negative form, as demonstrated by the first example.

Relative clauses with a direct object which share an indirect object, such as "The man to whom you lent the book has finished reading it," have a different type of mapping to Chinese, and PHRAN does not presently accept them.

5.8. Supplementary phrases

As mentioned previously, Chinese PHRAN handles phrases that supply supplementary information, akin to prepositional phrases in English, with its "adverb" mechanism. For example, "cong2 SOMEPLACE" and "dao4 SOMEPLACE" correspond to "from" and "to" someplace, respectively. In

```
Input:    TA1 CONG2 BEI3 JING1 DAO4 DONG1 JING1 QU4
~         (He from Peking    to  Tokyo     go )
Output:
((REFERENT (OBJECT HUMAN1))
(PERSON (OBJECT HUMAN1))
(LOCATION (OBJECT TOKYO))
(CITY (OBJECT TOKYO))
(LOCATION (OBJECT PEKING))
(CITY (OBJECT PEKING)))

(PTRANS (ACTOR HUMAN1) (OBJECT HUMAN1) (FROM PEKING) (TO TOKYO))
```

"cong2 Bei3 Jing1" and "dao4 Dong1 Jing1" are ignored after matching the patterns [CONG2 (LOCATION)] and [DAO4 (LOCATION)]. PHRAN creates the CD information for the remaining basic sentence

```
Input:    TA1 QU4
~         (He  go )
~ Matches the SIMPLE pattern [(FORM SUBSTANTIVE) (NAME QU4)]
Output:
((REFERENT (OBJECT HUMAN2)) (PERSON (OBJECT HUMAN2)))

(PTRANS (ACTOR HUMAN2) (OBJECT HUMAN2) (FROM *HERE*))
```

Then the "cong2 Bei3 Jing1" term adjusts the FROM field and the "dao4 Dong1 Jing1" term independently adjusts the TO field. The actual pattern entry matching "cong2 Bei3 Jing1" looks like

```
(INDEX-UNDER-PATTERN (CONG2 LOCATION)
[(NIL
 [CONG2 (* LOCATION)]
 [P-O-S 'ADVERB
  DO (SAVE-PREDICATES (VALUE 2) (VALUE 2 PREDS))
  MODIFIES-IF (EQ (CAR (TERM'S CONCEPT)) 'PTRANS)
  MODIFIED-CONCEPT '(TOPSUBST 'FROM '@(VALUE 2) (OLD CONCEPT))])])
```

It tries to modify a pattern if it is a PTRANS primitive. In this case it modifies

the "FROM" field using the value of the location. The "save-predicates" line just outputs any predicate information associated with the location.

In some cases, the representation of the adverb might depend on the particular phrase that it supplements. We have not yet run into that problem in Chinese or Spanish, but consider the English phrase "to Mary." For the sentence "John gave the book to Mary," the representation for "to Mary" just fills the TO slot of the ATRANS with "Mary." In "John ran to Mary," however, the representation should fill the TO slot of the PTRANS with "(LOCATION MARY)." To correctly handle these cases, the adverb mechanism would have to not only check that the candidate for modification meet the restriction that it be an ATRANS or a PTRANS, but also make the representation dependent on the particular primitive.

Another important phrase specifies location: "ZAI4 location." Unlike English, in which the preposition specifies the location, such as "on," "in," or "under," "ZAI4" indicates only that an location follows. The location itself must specify all the necessary information. For a normal location, such as "Peking," nothing additional is necessary. Nouns that are not normally locations, however, require a "place word" following them to turn them into a location, as in

```
Input:      SHU1 LI3
~           (Book inside)
~ The pattern [(OR (P-O-S NOUN) (P-O-S NOUN-PHRASE)) (P-O-S LOCATOR)]
~    matches the noun SHU1 and the locator LI3.
Output:
((PHYSOB (OBJECT (IN BOOK1)))
(LITERATURE (OBJECT (IN BOOK1)))
(BOOK (OBJECT (IN BOOK1))))

(IN BOOK1)
```

"ZAI4 location" terms are handled just like the "from" and "to" terms above:

```
Input:      TA1 ZAI4 XIANG1 GANG3 XUE2 XI2
~           (He  at  Hong Kong   study  )
~ The adverb pattern [ZAI4 (LOCATION)] matches the word ZAI4 and
~    the location XIANG1 GANG3.
```

```
~ The remaining sentence matches the SIMPLE pattern
~   [(FORM SUBSTANTIVE) (NAME XUE2XI2)], where TA1 is the
~   substantive and XUE2 XI2 is the verb name XUE2XI2.
~ Then the adverb entry adds the PLOC field to the resulting
~   $STUDY conceptualization.
Output:
((REFERENT (OBJECT HUMAN1))
(PERSON (OBJECT HUMAN1))
(LOCATION (OBJECT HONG-KONG))
(CITY (OBJECT HONG-KONG)))

($STUDY (ACTOR HUMAN1) (PLOC HONG-KONG))
```

PHRAN must distinguish, however, between the phrase "ZAI4 location" and the use of "ZAI4" as an actual verb. As a verb, "ZAI4" also indicates location:

```
Input:    NI3 DE1 QIAN1 BI3 ZAI4 ZHUO1 ZI3 SHANG4
~         (Your    pencil  at   table     on
~    ie. Your pencil is on the table)
~ The noun ZHUO1 ZI3 and the locator SHANG4 match the pattern
~   [(OR (P-O-S NOUN-PHRASE) (P-O-S NOUN)) (P-O-S LOCATOR)]
~   to create a new location term.
~ The pattern [(FORM SUBSTANTIVE) (NAME ZAI4) (LOCATION)]
~   match the substantive NI3 DE1 QIAN1 BI3, the verb ZAI4, and
~   the location ZHUO1 ZI3 SHANG4.
Output:
((PENCIL (OBJECT PENCIL1))
(PHYSOB (OBJECT PENCIL1))
(PHYSOB (OBJECT (ON TABLE1)))
(TABLE (OBJECT (ON TABLE1)))
(FAMILIAR (OBJECT YOU1))
(REFERENT (OBJECT YOU1))
(PERSON (OBJECT YOU1)))

(OWN (TENSE PRESENT) (ACTOR YOU1) (OBJECT PENCIL1))

(LOC (OBJECT PENCIL1) (PLOC (ON TABLE1)))
```

PHRAN can distinguish that ZAI4 is a verb here because the sentence matches the sentence pattern [(FORM SUBSTANTIVE) (NAME ZAI4) (LOCATION)] in addition to the adverb pattern [ZAI4 (LOCATION)], and when two patterns match at the same point, PHRAN takes the longer of the patterns.

In Chinese, supplementary phrases must always appear between the SUBJECT and the VERB, though the adverb mechanism can handle such information appearing throughout the sentence.

6. Spanish PHRAN

PHRAN accepts Spanish words as they are normally written, with the exception of accents. Since Spanish frequently uses an accent to differentiate between two words, PHRAN understands an up-arrow [] as an accent over the preceding vowel. Thus, for example, "tu" is "you," and "tu" is "your." Also, PHRAN excludes the tilde [~], which sometimes appears over the letter "n" for pronunciation purposes.

6.1. Noun-phrases in Spanish

Spanish PHRAN treats noun-phrases much the same way as English PHRAN. Since a plural noun can appear alone, as in "Libros son interesantes" ("Books are interesting"), plurals are classified as having a NOUN-PHRASE part of speech. A singular noun is classified as a plain NOUN and requires an article to turn it into a NOUN-PHRASE. Thus PHRAN has the entry

```
(INDEX-UNDER-PATTERN (ARTICLE NOUN)
[(NIL

 [(P-O-S ARTICLE) (* P-O-S NOUN)]
 [P-O-S 'NOUN-PHRASE
  CD-FORM (NEW-NAME (CAR (VALUE 2 DESCRIPTION)))
  REF (VALUE 1 REF)
  DESCRIPTION (VALUE 2 DESCRIPTION)
  PREDS (VALUE 2 PREDS)
  WORD (VALUE 2 WORD)
  DO (COPY-TERM 2)])])
```

An article followed by a noun activates this entry. It creates a new term that is exactly the same as the old noun term except that it becomes a NOUN-PHRASE. Although plural nouns are already noun-phrases, they can also take articles, so PHRAN has a similar entry with the pattern [(P-O-S ARTICLE) (P-O-S NOUN-PHRASE)]. Adjectives often come after the noun in Spanish, so PHRAN has the pattern [(P-O-S NOUN-PHRASE) (P-O-S ADJECTIVE)]. This entry creates a noun-phrase term that is the same as the original noun-phrase, and it also outputs

adjectival information about the noun-phrase, frequently of the form (adjective (OBJECT noun-phrase)), as in

```
Input:    EL  RESTAURANTE GRANDE VENDE HAMBURGUESAS
~         (The restaurant  large  sells hamburgers  )
~ Uses the pattern [(P-O-S NOUN-PHRASE) (P-O-S ADJECTIVE)] to match
~   "el restaurante grande" and collapse it to a single noun-phrase.
~   This entry outputs the "large" information.
~ The "group" information is output on seeing the plural word
~   "hamburguesas."
~ Then it uses the OBJECTIVE pattern  (see later)
~     [(OR (INSTITUTION) (PERSON)) (ROOT VENDER) (PHYSOB)]
Output:
((INSTITUTION (OBJECT RESTAURANT1))
(LOCATION (OBJECT RESTAURANT1))
(RESTAURANT (OBJECT RESTAURANT1))
(PHYSOB (OBJECT HAMBURGUESAS1))
(FOOD (OBJECT HAMBURGUESAS1))
(HAMBURGER (OBJECT HAMBURGUESAS1)))

(LARGE (OBJECT RESTAURANT1))
(GROUP (OBJECT HAMBURGUESAS1) (MEMBER HAMBURGUESA))
(ATRANS (OBJECT MONEY) (TO RESTAURANT1))

(ATRANS (TENSE PRESENT)
     (ACTOR RESTAURANT1)
     (OBJECT HAMBURGUESAS1)
     (FROM RESTAURANT1))
```

Sometimes the adjective may have a special meaning, as is the case in "el restaurante Frances" ("the French restaurant"). So in

```
Input:    EL  RESTAURANTE FRANCES NO  TIENE HAMBURGUESAS
~         (The restaurant  French  not has  hamburgers  )
~ Uses the pattern [(RESTAURANT) (STATE NATIONALITY)] to match
~   "el restaurante Frances" and produce the type information about
~   the restaurant, turning the phrase into a noun-phrase.
~ The pattern [(P-O-S NOUN-PHRASE) (P-O-S ADJECTIVE)] also
~   matches, producing the (French (Object Restaurant)) information.
~ The result uses the pattern
~     [(OR (INSTITUTION) (PERSON)) NO (ROOT TENER) (PHYSOB)]
Output:
((INSTITUTION (OBJECT RESTAURANT2))
(LOCATION (OBJECT RESTAURANT2))
(RESTAURANT (OBJECT RESTAURANT2))
(PHYSOB (OBJECT HAMBURGUESAS2))
(FOOD (OBJECT HAMBURGUESAS2))
(HAMBURGER (OBJECT HAMBURGUESAS2)))

(FRENCH (OBJECT RESTAURANT2))
(RESTAURANT (OBJECT RESTAURANT2) (TYPE FRENCH))
```

(GROUP (OBJECT HAMBURGUESAS2) (MEMBER HAMBURGUESA))

(NOT (POSS (TENSE PRESENT) (ACTOR RESTAURANT2) (OBJECT HAMBURGUESAS2)))

PHRAN outputs the information that it is a restaurant of type French, as opposed to, perhaps, a restaurant located in France. Unfortuneately, PHRAN also outputs the adjective-like form, simply stating "(French (Object Restaurantn))", because given no way to decide between two competing patterns (the specific restaurant pattern versus the general "noun adjective" pattern) of the same length, PHRAN presently just uses them both. Theoretically, however, a specific case could be given preference over the general case, thus producing the correct result.

6.2. The Spanish basic sentence categories

Spanish PHRAN uses five basic sentence types: SIMPLE, OBJECTIVE, INDOBJECTIVE, COMPLEX, and SENTOBJ. SIMPLE and SENTOBJ are the same as in Chinese PHRAN. OBJECTIVE and COMPLEX are approximately SINGLE and DOUBLE, respectively, but OBJECTIVE requires a direct object always. INDOBJECTIVE is a new type which requires an indirect object with no direct object. The Spanish patterns use the root of a verb to look up entries since in Spanish, as in English, verbs can have many conjugations per root.

The following sentences illustrate the basic Spanish sentence types:

```
SIMPLE --
Input:     MARI^A COME
~          (Mary    eats)
Output:
((FEMALE (OBJECT MARIA1)) (PERSON (OBJECT MARIA1)))

(INGEST (TENSE PRESENT) (ACTOR MARIA1) (OBJECT FOOD))

~ Since the sentence specifies no object, Phran defaults the
~ object to "food."


OBJECTIVE --
Input:     ELLA QUIERE LA  COCHE
```

~ (She wants the car)
~ Uses pattern [(PERSON) (ROOT QUERER) (PHYSOB)]
Output:
((REFERENT (OBJECT HUMAN2))
(FEMALE (OBJECT HUMAN2))
(PERSON (OBJECT HUMAN2))
(PHYSOB (OBJECT AUTO1))
(AUTO (OBJECT AUTO1)))

(GOAL (TENSE PRESENT)
(PLANNER HUMAN2)
(OBJECTIVE (POSS (TENSE FUTURE) (ACTOR HUMAN2) (OBJECT AUTO1))))

INDOBJECTIVE --
Input: HAMBURGUESAS ME GUSTAN
~ (Hamburguesas me please
~ ie. I like hamburgers)
~ Uses pattern [(P-O-S NOUN-PHRASE) (INDIR) (ROOT GUSTAR)]
~ An INDIR is an indirect pronoun.
Output:
((PHYSOB (OBJECT HAMBURGUESAS1))
(FOOD (OBJECT HAMBURGUESAS1))
(HAMBURGER (OBJECT HAMBURGUESAS1))
(REFERENT (OBJECT EGO1))
(PERSON (OBJECT EGO1)))

(GROUP (OBJECT HAMBURGUESAS1) (MEMBER HAMBURGUESA))

(ATTITUDE (TENSE PRESENT)
(ACTOR EGO1)
(SCALE FONDNESS)
(OBJECT HAMBURGUESAS1)
(LEVEL HIGH))

COMPLEX --
Input: ELLOS NOS DAN SUS LIBROS
~ (They us give their books)
~ Uses the pattern [(P-O-S POSSESSIVE-PRONOUN) (NOUN-PHRASE)]
~ to match "sus libros," outputing the ownership information.
~ Uses pattern [(PERSON) (INDIR) (ROOT DAR) (PHYSOB)]
Output:
((REFERENT (OBJECT HUMAN6))
(MALE (OBJECT HUMAN6))
(PERSON (OBJECT HUMAN6))
(GROUP (OBJECT HUMAN6))
(PERSON (OBJECT EGO2))
(REFERENT (OBJECT EGO2))
(GROUP (OBJECT EGO2))
(PHYSOB (OBJECT LIBROS1))
(BOOK (OBJECT LIBROS1))
(REFERENT (OBJECT HUMAN8))
(PERSON (OBJECT HUMAN8)))

```
(GROUP (OBJECT LIBROS1) (MEMBER LIBRO))
(OWN (ACTOR HUMAN8) (OBJECT LIBROS1))

(ATRANS (TENSE PRESENT)
     (ACTOR HUMAN6)
     (OBJECT LIBROS1)
     (FROM HUMAN6)
     (TO EGO2))


SENTOBJ --
Input:     GUILLERMO DICE QUE  E^L NO  TIENE TU  BICICLETA
~          (William  says that he  not have  your bicycle  )
~ Uses the pattern [(P-O-S POSSESSIVE-PRONOUN) (NOUN-PHRASE)]
~  to match "tu bicicleta," outputing the ownership
~  information.
~ Uses pattern [(PERSON) (ROOT DECIR) QUE (P-O-S SENTENCE)]
Output:
((MALE (OBJECT GUILLERMO1))
(PERSON (OBJECT GUILLERMO1))
(REFERENT (OBJECT HUMAN10))
(MALE (OBJECT HUMAN10))
(PERSON (OBJECT HUMAN10))
(PHYSOB (OBJECT BICYCLE1))
(BICYCLE (OBJECT BICYCLE1))
(FAMILIAR (OBJECT YOU1))
(REFERENT (OBJECT YOU1))
(PERSON (OBJECT YOU1)))

(OWN (ACTOR YOU1) (OBJECT BICYCLE1))

(MTRANS (TENSE PRESENT)
     (ACTOR GUILLERMO1)
     (MOBJECT (NOT (POSS (TENSE PRESENT)
                    (ACTOR HUMAN10)
                    (OBJECT BICYCLE1))))
     (FROM GUILLERMO1))
```

6.3. Rewriting rules

Very free grammatical positioning rules in Spanish make for many rewritten patterns for these basic sentence types. The subject, for example, can come almost anywhere in the sentence or even be omitted entirely. So, for the OBJECTIVE sentence type, which has the basic form [SUBJECT] [VERB] [DIRECT-OBJECT], rewriting must also give the additional possibilities of [VERB] [SUBJECT] [DIRECT-OBJECT], [VERB] [DIRECT-OBJECT] [SUBJECT], and

[VERB] [DIRECT-OBJECT]. Where the subject is omitted, PHRAN must deduce a pronoun-like subject from the verb conjugation.

An alternate method to rewriting could specify the subject as an optional part of the sentence. This method would still require at least two different cases, though, because PHRAN does not permit an optional part at the beginning of a pattern. The subject was originally implemented as an optional part, but we ran into problems with PHRAN correctly handling the patterns when more than one pattern was suggested containing the same optional part. That approach was abandoned in favor of the rewriting approach which, while it takes more space, runs faster since it need do no optional part checking.

As an example of using the rewriting method, consider the following OBJECTIVE sentences:

```
Input:      JUAN COME LA  MANZANA
~          (John eats the apple  )
~ Uses pattern [(PERSON) (ROOT COMER) (PHYSOB)]
Output:
((MALE (OBJECT JUAN1))
(PERSON (OBJECT JUAN1))
(PHYSOB (OBJECT APPLE1))
(FOOD (OBJECT APPLE1))
(APPLE (OBJECT APPLE1)))

(INGEST (TENSE PRESENT) (ACTOR JUAN1) (OBJECT APPLE1))


Input:      COME JUAN LA  MANZANA
~          (Eats John the apple  )
~ Uses rewrite rule [2 1 3] to get the pattern
~     [(ROOT COMER) (PERSON) (PHYSOB)]
Output:
((MALE (OBJECT JUAN2))
(PERSON (OBJECT JUAN2))
(PHYSOB (OBJECT APPLE2))
(FOOD (OBJECT APPLE2))
(APPLE (OBJECT APPLE2)))

(INGEST (TENSE PRESENT) (ACTOR JUAN2) (OBJECT APPLE2))


Input:      COME LA  MANZANA JUAN
~          (Eats the apple  John)
~ Uses rewrite rule [2 3 1] to get the pattern
```

```
~     [(ROOT COMER) (PHYSOB) (PERSON)]
Output:
((MALE (OBJECT JUAN3))
(PERSON (OBJECT JUAN3))
(PHYSOB (OBJECT APPLE3))
(FOOD (OBJECT APPLE3))
(APPLE (OBJECT APPLE3)))

(INGEST (TENSE PRESENT) (ACTOR JUAN3) (OBJECT APPLE3))


Input:      COME LA  MANZANA
~           (Eats the apple
~      ie. He eats the apple)
~ Uses rewrite rule [2 3] to get the pattern
~      [(ROOT COMER) (PHYSOB)]
Output:
((REFERENT (OBJECT HUMAN18))
(PERSON (OBJECT HUMAN18))
(PHYSOB (OBJECT APPLE4))
(FOOD (OBJECT APPLE4))
(APPLE (OBJECT APPLE4)))

(INGEST (TENSE PRESENT) (ACTOR HUMAN18) (OBJECT APPLE4))
```

They are all different possibilities obtained simply by moving or omitting the subject. In the last example, PHRAN looks at the PERSON field of the verb to decide which pronoun to use since the subject does not appear in the sentence.

The OBJECTIVE sentence type must also be rewritten to allow for pronouns used as direct objects. Normally the direct object comes after the verb, but when a pronoun is used the pronoun must come directly in front of the verb. So substituting "it" for "the car" in the OBJECTIVE example above produces

```
Input:      ELLA LA QUIERE
~           (She  it wants -- She wants it)
~ Uses the rewrite rule [1 (AND 3 (DIR)) 2] to produce the
~  pattern [(PERSON) (AND (PHYSOB) (DIR)) (ROOT QUERER)]
~   A DIR is a direct pronoun
Output:
((REFERENT (OBJECT HUMAN20))
(FEMALE (OBJECT HUMAN20))
(PERSON (OBJECT HUMAN20))
(REFERENT (OBJECT IT6)))

(GOAL (TENSE PRESENT)
   (PLANNER HUMAN20)
   (OBJECTIVE (POSS (TENSE FUTURE) (ACTOR HUMAN20) (OBJECT IT6))))
```

Since the COMPLEX type also has a direct object, it can also be rewritten with a pronoun. The direct pronoun in this case goes between the indirect pronoun and the verb. Replacing "the books" by "them" in the above COMPLEX example gives

```
Input:     ELLOS NOS LOS    DAN
~          (They  us  them give
~     ie. They give them to us)
~ Uses the rewriting rule [1 2 (AND 4 (DIR)) 3] to produce the
~  pattern [(PERSON) (INDIR) (AND (PHYSOB) (DIR)) (ROOT DAR)]
Output:
((REFERENT (OBJECT HUMAN23))
(MALE (OBJECT HUMAN23))
(PERSON (OBJECT HUMAN23))
(GROUP (OBJECT HUMAN23))
(PERSON (OBJECT EG03))
(REFERENT (OBJECT EG03))
(GROUP (OBJECT EG03))
(GROUP (OBJECT IT7))
(REFERENT (OBJECT IT7)))

(ATRANS (TENSE PRESENT)
     (ACTOR HUMAN23)
     (OBJECT IT7)
     (FROM HUMAN23)
     (TO EG03))
```

English and Chinese do not have this rewriting requirement because direct object pronouns come in the same place as normal direct objects, after the verb.

Many verbs that take embedded phrases, as with "decir" (to say), can take an indirect object. Spanish requires that in this case the sentence must use an indirect pronoun, even though the actual indirect object, such as "to Maria," may be specified elsewhere in the sentence. So the SENTOBJ type [SUBJECT] [VERB] [QUE] [SENTENCE] ("que" means "that" and is required after the verb) must be rewritten as [SUBJECT] [INDIR-PRONOUN] [VERB] [QUE] [SENTENCE] to allow, for example,

```
Input:
 GUILLERMO ME DICE QUE  NO  TIENE TU   BICICLETA
~          (William  me says that not has  your bicycle
~     ie. William tells me that he does not have your
~        bicycle)
```

```
~ Uses the pattern [(P-O-S POSSESSIVE-PRONOUN) (NOUN-PHRASE)]
~   to match "tu bicicleta," outputing the ownership
~   information.
~ Uses rewrite rule [1 (INDIR) 2 3 4] to produce the pattern
~       [(PERSON) (INDIR) (ROOT DECIR) QUE (P-O-S SENTENCE)]
Output:
((MALE (OBJECT GUILLERMO2))
(PERSON (OBJECT GUILLERMO2))
(REFERENT (OBJECT EGO4))
(PERSON (OBJECT EGO4))
(REFERENT (OBJECT HUMAN27))
(PERSON (OBJECT HUMAN27))
(PHYSOB (OBJECT BICYCLE2))
(BICYCLE (OBJECT BICYCLE2))
(FAMILIAR (OBJECT YOU2))
(REFERENT (OBJECT YOU2))
(PERSON (OBJECT YOU2)))

(OWN (ACTOR YOU2) (OBJECT BICYCLE2))

(MTRANS (TENSE PRESENT)
     (ACTOR GUILLERMO2)
     (MOBJECT (NOT (POSS (TENSE PRESENT)
                    (ACTOR HUMAN27)
                    (OBJECT BICYCLE2))))
     (FROM GUILLERMO2)
     (TO EGO4))
```

These pronoun rewritings must also allow for the various locations of the SUBJECT, and all rewritings are duplicated with a negative version, as in

```
Input:     DANIEL NO LOS TIENE
~          (Daniel not them has
~       ie. Daniel does not have them)
~ Uses rewritten pattern [(PERSON) NO (DIR) (ROOT TENER)]
Output:
((MALE (OBJECT DANIEL1))
(PERSON (OBJECT DANIEL1))
(GROUP (OBJECT IT8))
(REFERENT (OBJECT IT8)))

(NOT (POSS (TENSE PRESENT) (ACTOR DANIEL1) (OBJECT IT8)))
```

As mentioned in the Chinese section, a different method might combine the negative and the verb, as with the pattern [NO (ROOT TENER)], creating a new verb term with a field denoting the negative state. Because indirect and indirect pronouns must come between the negative word "no" and the verb, however,

PHRAN would also require a pattern such as [NO (DIR) (ROOT TENER)]. The new term would have to include not only the verb information plus the fact that it is negative, but it would also need to include pronoun information. The rewriting method does not increase the parsing time and provides a closer match of the pattern to the sentence structure.

6.4. Consecutive verbs

Multiple verbs can appear together in Spanish, just as in Chinese, such as "querer ir" (to want to go). Again, PHRAN composes these verbs to one verb term. With "querer ir," PHRAN creates a new verb term with the associated concept that the actor has the goal of accomplishing the conceptualization associated with "ir," a PTRANS. The following example demonstrates this composition:

```
Input:      JUAN QUIERE IR
~           (John wants  to go)
~ Uses the pattern [(ROOT QUERER) (FORM INFINITIVE)]
~  to compose the verbs "quiere ir," creating a
~  new verb term with root "ir."
~ Then the pattern [(PERSON) (ROOT IR)] matches JUAN
~  as the PERSON and the new term as ROOT IR.
((MALE (OBJECT JUAN1)) (PERSON (OBJECT JUAN1)))

(GOAL (TENSE PRESENT)
   (PLANNER JUAN1)
   (OBJECTIVE (PTRANS (TENSE PRESENT)
              (ACTOR JUAN1)
              (OBJECT JUAN1)
              (FROM *HERE*))))
```

The use of infinitives in combined verb forms creates another place for pronoun positioning and thus, more rewriting. As shown above, pronouns come before the verb. When using an infinitive, however, pronouns may be placed at the end of the infinitive, as part of the same word. So PHRAN accepts both

```
Input:      JUAN LA TIENE QUE COMER
~           (John it has      to eat)
~ The pattern [(ROOT TENER) QUE (FORM INFINITIVE)] composes
~  "tiene que comer" to a single term of root "comer."
~ Then the normal direct object pronoun rewriting rule
~  [1 (AND 3 (DIR)) 2] applied to the OBJECTIVE category
~  of "tener" produces the pattern
```

~ [(PERSON) (AND (PHYSOB) (DIR)) (ROOT COMER)], where
~ Juan matches the PERSON, "la" is the physical object
~ direct pronoun, and the new term matches the ROOT COMER
((MALE (OBJECT JUAN2)) (PERSON (OBJECT JUAN2)) (REFERENT (OBJECT IT1)))

(OBLIGATION (TENSE PRESENT)
(ACTOR JUAN2)
(REQUIREMENT (INGEST (TENSE FUTURE)
(ACTOR JUAN2)
(OBJECT IT1))))

and

Input: JUAN TIENE QUE COMERLA
~ (John has to eat it)
~ Here, the pattern [(ROOT TENER) QUE (FORM INFINITIVE)]
~ matches "tiene que comerla" and creates a new verb
~ term of root "comer." Since "comerla" contains both verb
~ and pronoun information, the new term also contains verb
~ and pronoun information.
~ Then the rewrite rule [1 (AND 2 3 (DIR))] applied to the
~ OBJECTIVE category of "comer" produces the pattern
~ [(PERSON) (AND (ROOT COMER) (PHYSOB) (DIR))],
~ where "Juan" is the PERSON, and the new term matches the
~ second requirement.
((MALE (OBJECT JUAN3)) (PERSON (OBJECT JUAN3)) (REFERENT (OBJECT IT2)))

(OBLIGATION (TENSE PRESENT)
(ACTOR JUAN3)
(REQUIREMENT (INGEST (TENSE FUTURE)
(ACTOR JUAN3)
(OBJECT IT2))))

6.5. Relative clauses

As with Chinese, rewriting rules can rewrite basic sentence types to create patterns to match relative clauses. Unlike in Chinese, however, Spanish relative clauses do not appear as modifiers but are set off with relative pronouns such as "quien" (who) and "que" (that), as in English. The following sentences illustrate Spanish relative clauses that share subject, direct object, and indirect object:

SHARED SUBJECT --
Input: LA PERSONA QUIEN TIENE LA NARANJA VA A COMERLA
~ (The person who has the orange goes to eat it
~ ie. The person who has the orange is going to eat it)
~ Uses rewriting rule [1 (P-O-S RELATOR) 2 3] on the OBJECTIVE
~ type to get the pattern:

~ [(PERSON) (P-O-S RELATOR) (ROOT TENER) (PHYSOB)]
~ which turns the relative clause into a noun-phrase
~ and outputs the possession information.
~ Then it uses the pattern [(ROOT IR) A (FORM INFINITIVE)]
~ to compose the verbs "va a comerla" to a term indicating
~ the the INGEST will occur in the future.
~ Finally it uses [(PERSON) (AND (ROOT COMER) (DIR))] to
~ complete the sentence, matching the noun-phrase as the
~ PERSON and the composed term as the term with
~ ROOT COMER and containing a direct object pronoun.
Output:
((PERSON (OBJECT PERSON1))
(REFERENT (OBJECT IT13))
(PHYSOB (OBJECT ORANGE1))
(FOOD (OBJECT ORANGE1))
(ORANGE (OBJECT ORANGE1)))

(POSS (TENSE PRESENT) (ACTOR PERSON1) (OBJECT ORANGE1))

(INGEST (TENSE FUTURE) (ACTOR PERSON1) (OBJECT IT13))

SHARED DIRECT OBJECT --
Input: EL REGALO QUE TU~ ME DASTE ME GUSTA
~ (The present that you me gave me pleases
~ ie. I like the present that you gave me)
~ Uses rewriting rule [4 (P-O-S RELATOR) 1 2 3] on the COMPLEX type
~ to produce the pattern
~ [(PHYSOB) (P-O-S RELATOR) (PERSON) (INDIR) (ROOT DAR)]
~ which composes "el regalo" (PHYSOB), "que" (RELATOR),
~ "tu~" (PERSON), "me" (INDIR -- indirect pronoun), and
~ "daste" (ROOT DAR) into a single noun-phrase, outputing
~ the ATRANS representation.
~ Then the pattern [(P-O-S NOUN-PHRASE) (INDIR) (ROOT GUSTAR)] matches
~ the result.
Output:
((PHYSOB (OBJECT PRESENT1))
(PRESENT (OBJECT PRESENT1))
(REFERENT (OBJECT EGO6))
(PERSON (OBJECT EGO6))
(REFERENT (OBJECT EGO5))
(PERSON (OBJECT EGO5))
(REFERENT (OBJECT YOU3))
(PERSON (OBJECT YOU3))
(FAMILIAR (OBJECT YOU3)))

(ATRANS (TENSE PAST)
(ACTOR YOU3)
(OBJECT PRESENT1)
(FROM YOU3)
(TO EGO5))

(ATTITUDE (TENSE PRESENT)
(ACTOR EGO6)

```
      (SCALE FONDNESS)
      (OBJECT PRESENT1)
      (LEVEL HIGH))


SHARED INDIRECT OBJECT --
Input:     LA  PERSONA A     QUIEN NOSOTROS NO  LE  QUEREMOS
~          (The person  to whom  we       not him want
     DAR EL  ANILLO HA      LLEGADO
~    to give the ring   has arrived
~    ie. The person to whom we do not want to give
~       the ring has arrived)
~ Uses the pattern [QUERER (FORM INFINITIVE)] to compose the
~  verbs "queremos dar" to one of root "dar," retaining the
~  GOAL representation.
~ Uses rewriting rule [(PERSON) A (P-O-S RELATOR) 1 NO 2 3 4]
~  to produce from the COMPLEX type the pattern
~ [(PERSON) A (P-O-S RELATOR) (PERSON) NO (INDIR) (ROOT DAR) (PHYSOB)]
~  which turns "la persona" (PERSON), "a" (A), "quien" (RELATOR),
~  "nosotros" (PERSON), "no" (NO), "le" (INDIR), "queremos dar"
~  (ROOT DAR), and "el anillo" (PHYSOB) into a single noun-phrase,
~  associates the ATRANS representation as a component of the GOAL,
~  and outputs the information.
~ It uses the pattern [HABER (FORM PERFECTIVE)] to compose the
~  verbs "ha llegado" to one of root "llegar" and tense "occurence."
~ Finally, it uses the pattern [(PERSON) (ROOT LLEGAR)].
Output:
((PERSON (OBJECT PERSON1))
(REFERENT (OBJECT EGO1))
(PERSON (OBJECT EGO1))
(GROUP (OBJECT EGO1))
(PHYSOB (OBJECT RING1))
(RING (OBJECT RING1)))

(NOT (GOAL (TENSE PRESENT)
       (PLANNER EGO1)
       (OBJECTIVE (ATRANS (TENSE PRESENT)
                  (ACTOR EGO1)
                  (OBJECT RING1)
                  (FROM EGO1)
                  (TO PERSON1)))))

(PTRANS (TENSE OCCURRENCE-PRESENT)
     (ACTOR PERSON1)
     (OBJECT PERSON1)
     (TO *HERE*))
```

6.6. Passive sentences

Patterns to match passive Spanish sentences can be obtained by applying rewriting rules to the OBJECTIVE and COMPLEX sentence types. The OBJECTIVE

and COMPLEX forms are just rewritten with the direct object at the beginning, with the helping verb "ser" (is) and an optional "por SUBJECT" (by SUBJECT) at the end, as in [DIR-OBJECT] [ROOT SER] [VERB] [POR] [SUBJECT] for the OBJECTIVE form. Examples with and without the subject are

```
Input:     LA MANZANA FUE COMIDA
~          (The apple  was eaten )
~ Uses the rewriting rule [3 (ROOT SER) (AND 2 (FORM PERFECTIVE))]
~  on the OBJECTIVE type to produce the pattern
~    [(PHYSOB) (ROOT SER) (AND (ROOT COMER) (FORM PERFECTIVE))]
Output:
((PHYSOB (OBJECT APPLE1))
(FOOD (OBJECT APPLE1))
(APPLE (OBJECT APPLE1))
(REFERENT (OBJECT HUMAN6)))

(INGEST (TENSE PAST) (ACTOR HUMAN6) (OBJECT APPLE1))
```

and

```
Input:     EL ANILLO ME FUE DADO POR MARI~A    [COMPLEX]
~          (The ring  me was given by Mary )
~ Uses the rewriting rule
~  [4 2 (ROOT SER) (AND 3 (FORM PERFECTIVE)) POR 1]
~  on the COMPLEX sentence type to produce the pattern
~  [(PHYSOB) (INDIR) (ROOT SER) (AND (ROOT DAR) (FORM PERFECTIVE))
~    POR (PERSON)]
Output:
((FEMALE (OBJECT MARIA1))
(PERSON (OBJECT MARIA1))
(REFERENT (OBJECT EGO3))
(PERSON (OBJECT EGO3))
(PHYSOB (OBJECT RING2))
(RING (OBJECT RING2)))

(ATRANS (TENSE PAST)
     (ACTOR MARIA1)
     (OBJECT RING2)
     (FROM MARIA1)
     (TO EGO3))
```

PHRAN's optional-part mechanism could also be used to handle the optional "por SUBJECT" phrase that can come at the end of the sentence.

6.7. Supplementary phrases

As in Chinese, Spanish PHRAN uses the adverb mechanism to handle phrases which provide supplementary information. Unlike in Chinese, however, these phrases can usually appear in many placese throughout the Spanish sentence. When using an indirect object, for example, the indirect pronoun must ALWAYS be used. The particular person, however, can also be specified somewhere in the sentence, as in

```
Input:     SE  LO DA    A  WILAMINA JUAN
~          (Her it gives to Willa       John
~     ie. John gives it to Willa)
~ Uses pattern [A (PERSON)] to match and then ignore
~  "a Wilamina."
~ Next it uses the rewriting rule [2 (AND 4 (DIR)) 3 1] on the
~  COMPLEX category of the verb "dar" to produce the pattern
~     [(INDIR) (AND (PHYSOB) (DIR)) (ROOT DAR) (PERSON)]
~ Then the [A (PERSON)] term updates the "TO" field of the ATRANS.
Output:
((MALE (OBJECT JUAN1))
(PERSON (OBJECT JUAN1))
(PERSON (OBJECT HUMAN1))          ~ Human1 is left over after being
(REFERENT (OBJECT HUMAN1))        ~ replaced by Wilamina1
(REFERENT (OBJECT IT1))
(FEMALE (OBJECT WILAMINA1))
(PERSON (OBJECT WILAMINA1)))

(ATRANS (TENSE PRESENT)
     (ACTOR JUAN1)
     (OBJECT IT1)
     (FROM JUAN1)
     (TO WILAMINA1))
```

as opposed to just

```
Input:     SE  LO DA    JUAN
~          (Her it gives John
~     ie. John gives it to her)
~ Uses same pattern as above minus the [A (PERSON)] pattern.
Output:
((MALE (OBJECT JUAN2))
(PERSON (OBJECT JUAN2))
(PERSON (OBJECT HUMAN4))
(REFERENT (OBJECT HUMAN4))
(REFERENT (OBJECT IT2)))

(ATRANS (TENSE PRESENT)
     (ACTOR JUAN2)
```

(OBJECT IT2)
(FROM JUAN2)
(TO HUMAN4))

This mechanism also handles direction information:

```
Input:    VA  DE  CHICAGO A MADRID DANIEL
~         (Goes from Chicago to Madrid Daniel)
~ The adverb mechanism uses the patterns [DE (LOCATION)] and
~  [A (LOCATION)] to match and then ignore "de Chicago" and
~  "a Madrid."
~ Then it uses the pattern [(ROOT IR) (PERSON)] to match
~   "Va Daniel."
~ Finally the adverb mechanism updates the FROM field from the
~  [DE (LOCATION)] entry and the TO field from the
~  [A (LOCATION)] entry.
Output:
((MALE (OBJECT DANIEL1))
(PERSON (OBJECT DANIEL1))
(LOCATION (OBJECT MADRID))
(CITY (OBJECT MADRID))
(LOCATION (OBJECT CHICAGO))
(CITY (OBJECT CHICAGO)))

(PTRANS (TENSE PRESENT)
     (ACTOR DANIEL1)
     (OBJECT DANIEL1)
     (FROM CHICAGO)
     (TO MADRID))
```

versus simply

```
Input:    VA  DANIEL
~         (Goes Daniel)
~ Uses [(ROOT IR) (PERSON)]
Output:
((MALE (OBJECT DANIEL2)) (PERSON (OBJECT DANIEL2)))

(PTRANS (TENSE PRESENT) (ACTOR DANIEL2) (OBJECT DANIEL2) (FROM *HERE*))
```

Normal grammatical adverbs also use this mechanism. "Con despacio," for instance, means "slowly," giving

```
Input:    CON DESPACIO ELLOS COMIERON
~         (Slowly      they ate    )
~ The adverb mechanism uses the pattern [CON DESPACIO] to match
~  and remove the adverb "con despacio."
~ The remainder matches the SIMPLE pattern [(PERSON) (ROOT COMER)]
~ The adverb mechanism associated with [CON DESPACIO] then updates
~  the MODE field of the resulting INGEST to "slow."
```

Output:
((REFERENT (OBJECT HUMAN9))
(MALE (OBJECT HUMAN9))
(PERSON (OBJECT HUMAN9))
(GROUP (OBJECT HUMAN9)))

(INGEST (TENSE PAST) (ACTOR HUMAN9) (OBJECT FOOD) (MODE SLOW))

7. Statistics

Chinese PHRAN knows the following number of one-word "words":

Numbers: 12
Pronouns: 4
Nouns: 6
Verbs: 13
Degree words: 4
Adjectives: 11
Locators (correspond to "on," "in," etc.): 5
Negative particles: 2
Interrogative particle: 1
Demonstratives: 4
Measure words: 4

Total one-word entries: 66

It also knows the following number of multi-word "words":

Locations: 7
Nouns: 15
Verbs: 4
Degree words: 1
Adjectives: 2

Total multi-word entries: 29

These patterns use the basic Chinese sentence types:

Normal basic sentences: 4
Basic sentences rewritten to give negative sentences: 5
Basic sentences rewritten to give questions: 10
Descriptive, verb-less sentence, including the question form: 3
Relative clause modifiers, including the negative form: 32

Total sentence-type patterns: 55

Chinese also has 1 pattern to handle verb composition, 13 patterns to construct substantives, and 3 patterns that use the adverb mechanism.

Total Chinese entries: 150

Although PHRAN does not have information associated with all of them, it contains entries for all 1596 one-word Chinese words.

Spanish PHRAN knows about the following number of items:

Articles: 8
Pronouns: 34
Adjectives: 34
Numbers: 10
Prepositions: 6
Nouns: 43
Names: 5
Verbs: 24
Relators: 2
Locations: 7

Total number of Spanish words: 173

Spanish PHRAN also has the following distribution of patterns to construct terms:

Patterns to construct noun-phrases: 9
Patterns that use the basic sentence types, plus those rewrites that move the subject and make negative sentences: 38
Patterns that rewrite the basic sentences to move the direct object pronoun before the verb, including the negative versions and the different subject locations: 18
Patterns that compose verbs: 13
Patterns that use the adverb mechanism: 7
Patterns to handle the passive voice: 8
Patterns to handle relative clauses, including negative clauses: 32

Total number of Spanish phrase patterns: 125

Total number of Spanish entries: 298

8. Conclusion

Chinese and Spanish PHRAN demonstrate that a pattern-based mechanism can be extended to non-English languages. While sharing an Indo-European base with English, Spanish offers a freer word ordering and strict pronoun requirements, all of which PHRAN handles. With Chinese, PHRAN handles a language which shares with English only the most basic concept of subject, verb, and object.

This basic subject-verb-object structure constitutes a key portion of the philosophy behind Spanish and Chinese PHRAN. The pattern mechanism uses the basic sentence types to break a sentence apart, drawing the conceptual dependency representation from information assiciated with the particular verb. PHRAN can use rewriting rules on the basic sentence types to form patterns to match phrases that require a different word order, such as some types of questions, negations, and relative clauses.

Supplementary prepositional phrase type structures are associated with conceptual primitives rather than verbs. The pattern representing "to LOCATION," for instance, in both Chinese and Spanish, tries to update the "TO" field of an associate representation. If it were associated instead with the verb, then all verbs giving rise to this representational element, such as "run," "walk," "hop," "skip," "jump," and others would have to account for the expression as an optional part of all patterns associated with the verbs, appearing in all the positions permitted by the particular language's word ordering. The possibility of yet other supplementary phrases, such as "from LOCATION," would greatly increase the complexity of patterns associating these phrases directly with the verb, whereas with the adverb method, just one pattern exists per supplementary phrase.

The ability to construct words from a given pattern of words and to

recognize special phrases complements the above basic and supplementary mechanisms, providing conceptual information at the subject and object rather than sentence level. Together, these mechanisms demonstrate that PHRAN is a sufficiently powerful tool to handle a good portion of Spanish and Chinese language constructions.

References

1) Arens, Y. (1981). Using Language and Context in the Analysis of Text. In *Proceedings of the Seventh International Joint Conference on Artificial Intelligence*. Vancouver, B. C., Canada.

2) Schank, R. C. (1975). *Conceptual Information Processing*. American Elsevier Publishing Company, Inc., New York.

3) Schank, R. C., Lebowitz, M. and Birnbaum, L. (1980). An Integrated Understander. In *American Journal of Computational Linguistics*, vol. 6 no. 1, January-March 1980.

4) Wilensky, R. (1981). A Knowledge-Based Approach to Language Processing: A Progress Report. In *Proceedings of the Seventh International Joint Conference on Artificial Intelligence*. Vancouver, B. C., Canada.

5) Wilensky, R. and Arens, Y. (1980). PHRAN: A Knowledge-Based Approach to Natural Language Analysis. Berkeley Electronic Research Laboratory Memorandum No. UCB/ERL/M80/34.