# I, me, mine (1)
## Psycholinguistic Constraints of French Clitics in Sentence Generation

Michael ZOCK
Gérard SABAH

LIMSI, Langues Naturelles
B.P. 30 - 91406 ORSAY Cédex/France

Abstract :

This paper describes an implemented tutoring system, designed to show various ways of converting a given meaning structure into its corresponding surface expression. The system is meant to be a teaching tool for students who learn French as a foreign language.

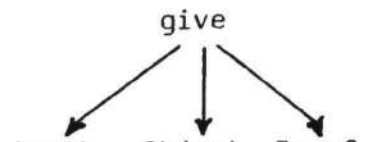
While showing various ways of converting a given meaning structure into its corresponding surface expression, the system helps not only to discover WHAT data to process but also HOW this information processing should take place. In other words, we are concerned with efficiency in verbal planning (flexibility and economy of performance).

Recognizing that the same result can be obtained by various methods, the student should find out which one is best suited to the circumstances (what is known, task demands etc). Informational states, hence the processor's needs, may vary to a great extent, as may his STRATEGIES or cognitive styles. In consequence, in order to become an efficient processor, the student has to acquire not only STRUCTURAL or RULE-KNOWLEDGE but also PROCEDURAL-KNOWLEDGE (skill).

With this in mind we have designed three modules in order to foster a reflective, experimental attitude in the learner, helping him to discover insightfully the most efficient strategy.

## 1 THE PROBLEM :

It is well known that children or students learning French as a foreign language need a great deal of practice before they master the French pronoun system well enough to fluently produce correct sentences with 1 or 2 pronoun complements, such as:




2 3 1

a) Tu me le donnes? S-IO-DO-V
Do you give it to me ?

b)Donne-le moi ! V-DO-IO
Give it to me !

c) Ne me le donnes pas! neg-IO-DO-V-neg
Dont' give it to me

The student's problem can be stated in the following terms: he has to learn how to determine both form and position for the French pronouns. Basically he is faced with the following problem:

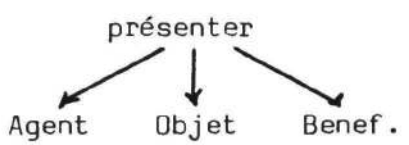
- he has to LEARN two lists, one for morphology the other for syntax (frames)
- he has to DISCOVER under what conditions each of these elements applies.

| List of MORPHOLOGY | | list of SYNT. FRAMES |
|---|---|---|
| SPEAKER: | je, me, moi, nous | S-DO-IO-V |
| LISTENER: | tu, te, toi, vous | S-IO-DO-V |
| 3d PERSON: | le, la, les, lui, leur, eux | S-DO-V-prep-IO |
| | il, elle, ils, elles, on, se soi | etc. |

As we will see, the student's task is not all that easy. These are some of the reasons why:

A) SYNTAX depends upon MORPHOLOGY:

By modifying the form of a given element one may also alter its position. In other words, certain features influence simultaneously morphology and syntax (see d-e).




3 1 3 d) Il me présente à elle S-DO-V-prep-IO
He presents me to her

3 3 3 e) Il le lui présente S-DO-IO-V
He presents him to her

B) The determination of FORM and POSITION requires complex FEATURE-CLUSTERS:

Any form or sentence-frame depends upon a number of conditions or features (2). Even a simple concept such as SPEAKER requires quite a lot of proces-

sing. It may be expressed by a variety of forms (je,me,moi); each may affect the choice of a sentence frame.

Furthermore, features such as PERSON, NUMBER, CASE and GENDER (3) which come readily into mind, are by no means sufficient. As the examples "a-c" clearly show, many other variables like SENTENCE-MODE (a,b), NEGATION (b,c), etc. come into play.

C) Some MORPHEMES are MULTIPLE DEPENDANTS (4):

The form of some morphemes depends not only upon features inherent in the coreferent (vertical dependancy), but also upon features coming from another referent (horizontal dependancy). This is the case of the indirect object, whose form depends upon the value of the direct object (see d,e). This fact is interesting for its procedural implications, namely it excludes any word-to-word processing.

In the light of these facts, one must admit, that what looked easy at first sight turned out to be a complex enterprise.

## 2 OBJECTIVE :

The system described here (5) is an attempt to help the student to acquire the necessary structural and procedural knowledge in order to economically generate pronoun-constructions in French (6).

While converting a given meaning structure into its corresponding surface expression, the student should not only learn WHAT data to process, but also HOW this information processing should take place. Recognizing that the same result can be obtained by various methods (strategies), the student should find out which one is best suited to the circumstances.

Particular emphasis is placed upon the discovery of operating principles (7) and the building of larger blocks (schematas). This chunking method should not only help to avoid unnecessary disruptions and memory load but also allow evolution from serial to simultaneous processing.

## 3 DESCRIPTION OF THE SYSTEM :

The heart of the system is a knowledge base which contains, in the form of production rules, the structural information necessary to incrementally determine form as well as position. Furthermore the system contains an inference-mechanism, i.e. a set of rules, whose function is to deduce new facts from information given to the system.

The base can be accessed in various ways, thus allowing for varying usage of the knowledge according to the objectives.

3.1 THE SOCRATIC METHOD :

The system guides the student in the form of a dialogue, by showing him what and how to process in order to get from an input to the output. The user starts by providing the input (verb-pattern composed of a verb, its complements and eventual prepositions):

donner (qn,qc,à qn)  give (so, sth, to so)

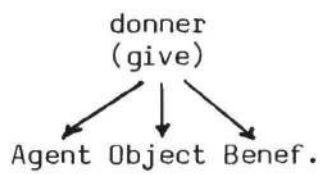
Then the system takes over, asking for more information about these basic elements. By asking specific questions (person, gender, number), the system shows which information is relevant when determining form as well as position. While answering these questions the student incrementally determines the final form of the sentence.

3.2 GUIDED DISCOVERY :

The system still controls the nature of the operations but no longer controls their order. The latter is controlled, via strategies, by the user. He decides in what order to process the data. Having determined the subject, whose position is invariable, one can choose from three strategies:

- a syntactical one (syntactic-driven processing)
- and two morphological ones (lexical-driven processing).

If priority is given to syntax (strategy 1), no further reordering of syntactic constituents is meant to take place, i.e. all information necessary to determine word order will have been processed. The result is an ordered categorial structure or syntactical frame (T1) which will be filled in by the morphological values determined later (T2), for example:




(T1) sentence frame :Subject - Dir.Obj.- Ind.Obj.- Verb
(T2) morphology : il - le - lui - donne

3 3 3 (he gives it to her)

If priority is given to morphology (lexically driven generation), the form is determined before the relative order of the constituent elements. In this case two strategies are possible: either one processes the direct (strategy 2), or the indirect object (strategy 3).

This experimental method should make the student aware of the fact that there are several ways of arriving at a particular solution (sentence).It is precisely his task to find out which strategy is the best suited. By applying performance criteria such as:

- number of steps necessary to generate the sentence

- what is known when? (form/position)
- congruence of input/output order (are permutations necessary? LIFO/FIFO)
- are there any conceptual disruptions?

the student hopefully discovers the procedural knowledge necessary for economic production of pronoun constructions.

### 3.3 USER DRIVEN EXPERIMENTATION :

This method, like the previous one, is empirical. By playing with the system the student may gain certain insights about processing order.

A matrix appears on the screen, whose blank spaces have to be filled in by the student. The horizontal line shows the syntactic information given with the input (verb, subject, object, preposition), -more information is needed about those elements- the vertical line shows the nature of the information necessary to arrive at the output.

Thus the processing once again consists of the specification of the values of a list of attributes. However there is a fundamental difference between this approach and the former, namely that the system has an inference mechanism. Each item of information given to the system is considered for its meaning potential, i.e. the system tries to find out whether some new facts can be deduced from the old fact.

It should be noted that the inference power varies with the nature of the data as well as with their order. There are cases where a single fact enables 3 other facts (reflexives) to be deduced. A given inference may allow further deductions (inference-chain/knowledge propagation). This has of course an effect on the process, namely the greater the inference power, the greater the economy of processing. This suggests the following operating principle:

> the greater the inference power of a given piece of information, the earlier it should be processed.

This method is interesting in that, by testing different items and different orders, it makes it possible to see on the screen which items allow what inferences. Since those inferences depend upon the nature of the input as well as on the moment at which that information is given, we believe that this system is particularly useful in helping to discover the optimal order of processing.

Furthermore we think that this method has another virtue, namely that it can simulate literally any knowledge state, thus making it possible, by experimental means to discover the shortest path between a given informational state (input) and the solution (output).

## 4 CONCLUSIONS :

We have stressed the need for teaching procedural knowledge (strategies) as well as structural knowledge (linguistic rules). In order to achieve this goal we have designed three modules, one demonstrative, the other two experimental.

Each module is intended for a different learning stage or learner type, as the cognitive styles may vary both among individuals and within the same individual. By progressively moving the control from outside (system) to inside (student), i.e. in integrating the student into the learning-process (8) we hope to make him:

- actively curious (testing of hypotheses - learning by discovery);
- conscious of the need for planning (how far should one plan ahead?),
- selective about the means he should use (Which strategy is best under what circumstances ?).

The whole idea of having different strategies compete has been largely ignored by current work on language generation. While this aspect may be only of secondary interest for automatic generation in general, it certainly is not an unimportant issue in cognitive modelling, whether it be language learning or usage.

NOTES :

1° In memory of one of the Beatle's songs.

2° The average number for a given pronoun is about three.

3° We call these inherent features, as opposed to features like sentence-mode, negation etc. which also determine form but whose information is not explicited in the morpheme.

Note that only PERSON and NUMBER are inherent and well marked in all pronouns. A feature like GENDER is only present in some forms (mostly 3d person), whereas the feature CASE is ambiguous for many direct and indirect object pronouns, in particular for 1st and 2nd person.

4° The way we use the term dependancy here is not to be confounded with Tesnière's dependancy theory (1959). Tesnière uses the word dependacy to signify "valency" whereas we use the word in its literal meaning.

5° The modules described are written in Simula and Prolog. They were implemented by G.Sabah and C.Alviset.

6° By economy we mean: number of operations, necessary permutations and the number of items to be stored.

7° Among those operating principles are the following:

- avoid disruptions by grouping together what belongs conceptually together;
- start with the most informative items
  (feature hierarchy: PERSON > CASE > NUMBER > GENDER);
- avoid unnecessary storage - start with the leftmost item.

8° The system could be greatly improved if it contained a module, capable to analyse the students performances (see Woolf & McDonald).

REFERENCES :

L.Tesnière
1959 Eléments de Syntaxe Structurale, Klinksieck, Paris

B.Woolf & D. McDonald
1984 Context-dependent Transitions in Tutoring Discourse
Proceedings of AAAI