

SOME ISSUES IN TEXT GENERATION

Laurence Danlos
LADL, CNRS
Université Paris VI

Resumen

A text generation system produces texts which translate the information given in a semantic representation. The ordering of the decision that have to be made is one of the major issues because most of the decisions are dependent one upon each other. Concrete examples of linguistic interactions will be presented: on one hand, interactions between ordering of the information, lexical choices and syntactic constructions, on the other hand, interactions between the synthesis of the constituents of a clause in regard to pronominalization.

**Some Pronominalization Issues
in generation of texts in Romance languages**

Laurence DANLOS
LADL, Université de Paris 7
2 Place Jussieu, 75251 Paris Cedex 07

1 Introduction

In a man-machine dialogue in natural language, a generation system produces the answers of the machine. A generation system is the counterpart of an analysis system which interprets the messages of the user.

The machine answers must 1) give the user the information he is expecting for, 2) formulate this information in a consistent style. These two tasks give rise to the following components of a generation system : a reasoning module (expert system) deals with the **What to say?** question and returns a semantic representation of the information that is to be conveyed to the user; this semantic representation is translated into a text by a linguistic generation component which deals with the **How to say it?** question. This paper is concerned only with the linguistic generation component.

It is generally believed that a linguistic generation component can be modularized into a sequence of modules, the first one making the conceptual decisions (e.g. ordering of the information), the following ones making the linguistic decisions (e.g. lexical and syntactic construction choices), the penultimate one performing the syntactic operations (e.g. subject-verb agreement), and the last one handling the morphological operations (e.g. conjugation of verbs). Given that "high level" decisions must be made before "low level" decisions, this way of modularizing a linguistic generation component relies on the following hypotheses:

conceptual decisions are "high level" decisions, linguistic decisions are "low level", syntactic operations are very low level and morphological operations are even lower level than the syntactic ones.

Our previous work has brought these hypotheses into question to the extent that we have shown in (L. Danlos 1985, 1987a) that the conceptual and linguistic decisions are operations that depend on each

none being "higher" than the other ones. Therefore, we designed a generation model modularized into a "strategic component" and a "syntactic component". The strategic component makes the conceptual and linguistic decisions simultaneously, as briefly presented in 2. It gives back a "text template" which is a list of the form:

(S1 Punct1 S2 Punct2 ... Si Puncti ... Sn Punctn)

where Puncti is a punctuation sign and Si a sentence template. The sentence template underlying the sentence

(1) *Ugo shot Mary and she was killed.*

(2)

S1 (:COORD-S (:coord-conj *and*)
 (:Cl (:subject HUM1) (:verb *shoot* (:tense past)) (:dir-object HUM2))
 (:Cl (:subject HUM2) (:verb *kill* (:tense past)(:voice passive))))

where HUM1 and HUM2 are tokens with the following definitions:

HUM1 =: PERSON

HUM2 =: PERSON

NAME : Ugo

NAME : Maria

SEX : masc

SEX : fem

Text templates are synthesized into texts by the syntactic component. On the one hand, this component synthesizes the tokens, e.g. it determines for each occurrence of HUM2 in S1 if it should be synthesized as a reflexive pronoun, a personal pronoun or a noun group and it produces the corresponding form. On the other hand, the syntactic component performs syntactic operations such as the subject-verb agreement, or the reduction of a sentential clause into an infinitive clause (L. Danlos 1987b).

We are going here to concentrate on some pronominalization issues. First, we are going to show that pronominalization involves the morphological level for the generation systems that produce texts in Romance languages. The decisions concerning pronominalization, which is a stumbling block for natural language processing, must certainly not be made last. Thus, the morphological level (level supposedly very "low") must not be taken into account only at the very last stage of the generation process.

Second, we will go one step further in our research that aims at showing the interaction of decisions in text generation. We will show that our modularization into two components - a strategic component and a syntactic component, the latter handling pronominalization - is even too modular: some pronominalization questions should be taken into account while the conceptual or linguistic decisions are made.

2 Brief presentation of the strategic component

The strategic component relies upon a linguistic data base that we have called "discourse grammar". A discourse grammar establishes a mapping between a type of semantic relationship (i.e. a causal relation) and the list of discourse structures enabling it to be expressed. Our notion of discourse grammar¹ integrates decisions on 1) the ordering of information, 2) the linearization into sentences (namely, the choice between juxtaposition, subordination, relativization or coordination as a sentence linearization procedure), 3) the form of the sentences (namely, the choice between the active, passive or ergative construction²). An example of discourse structure for causal relation is:

(A) (active <CAUSE>) and (passive-without-agent <RESULT>).

where <CAUSE> and <RESULT> stand for the sentences that will express respectively the cause and the result. The discourse *Ugo shot Mary and she was killed* is of structure (A). In this discourse, the cause precedes the result, the linearization mode is coordination with the conjunction *and*, the cause is expressed in the active and the result in the passive without agent. Another example of discourse structure for causal relation is:

(B) (active <CAUSE>) . (ergative <RESULT>) .

=: *Ugo knocked over the glass. It broke.*

The strategic component selects a discourse structure in the appropriate discourse grammar. This operation amounts in making simultaneously conceptual and some linguistic decisions. The other linguistic decision that is made by the strategic component is the lexical choices for verbs, more precisely, for predicative elements. The lexical choices and the selection of a discourse structure are also operations that depend on each other. For example, the selection of the discourse structure (B) is incompatible with choosing *kill* to express the result since this verb cannot undergo the ergative transformation. As a consequence, the following discourse of structure (B) is unacceptable: **Ugo shot Mary. She killed.*

3 Presentation of the syntactic component

A sentence is either a coordination of sentences (COORD-S) or a clause (CI). A simplified version of the clause template syntax is the following one³:

¹ Other researchers in text generation make use of a "discourse grammar", for example (K. McKeown 1985). However, her notion of discourse grammar gives only guidelines on the order of the information, thus is used only for conceptual decisions.

² The passive construction is examined both with and without agent. The ergative transformation when applied to the clause *Ugo broke the glass* gives back *The glass broke.*

³ A more complete version of the text and clause template syntax is given in (L. Danlos 1987b). It includes representations for subordination, attributes, adverbial phrases, and so on. A clause template may represent a

[Cl] = (:Cl [subject] [verb] cpltn ($0 \leq n \leq 2$))

[subject] = (:subject token)

[verb] = (:verb *verb*)

oplt = {dir-object} / [à-object] / [de-object] / [loc-object] / [prép-object]

[dir-object] = (:dir-object token)

[à-object] = (:à-object token)

[de-object] = (:de-object token)

[loc-object] = (:loc-object [loc] (:object token))

[prép-object] = (:prép-object [prép] (:object token))

[loc] = (:loc *locative-preposition*)

[prép] = (:prép *preposition*)

The prepositional complements [à-object], [de-object] and [loc-object] are complements respectively introduced by:

• *a*, *de* and a locative preposition in French

• *a*, *di* and a locative preposition in Italian.

They are separated from the prepositional complements [prép-object] introduced by other prepositions because they have a specific syntactic behaviour, especially in regard to pronominalization (see section 4).

Consider the following clause template:

(:Cl (:subject HUM1) (:verb *amare* (:tense present)) (:dir-object HUM2))

with

HUM1 =: PERSON

NAME : Ugo

SEX : masc

HUM2 =: PERSON

NAME : Maria

SEX : fem

According to the context (i.e. the clause templates that have been previously synthesized), the Italian syntactic component synthesizes it as one of the following clauses :

Ugo ama Maria

(Ugo loves Mary)

Ugo l'ama

(Ugo loves her)

Ama Maria

(He loves Mary)

L'ama

(He loves her)

Quest'uomo ama questa donna

(This man loves this woman)

Quest'uomo l'ama

(This man loves her)

Ama questa donna

(He loves this woman)

clause in the passive : the [subject] is then the surface subject, the verb being marked as having to be conjugated at the passive voice.

In a generation system producing texts in Romance languages, a syntactic component has to handle three different orders for the synthesis of a clause:

- the order of the elements in the clause template that is to be synthesized,
- the order in which the elements of the clause template must be synthesized,
- the order in which the synthesized elements must be placed in the final clause.

3.1 The order of the elements in the clause template that is to be synthesized

We will suppose that this order corresponds to the canonical order of a clause, i. e. :

subject - verb - direct object - indirect object(s)

3.2 The order in which the elements of the clause template must be synthesized

This order is determined by "non local dependencies" which are to be found when the synthesis of an element X depends upon that of another element Y. Such a dependency requires the synthesis of X to be carried out after that of Y, whatever the order of X and Y in the clause template and whatever their order in the final synthesized clause. Moreover, cases of "cross dependencies" are to be found when the synthesis of X depends upon that of Y and when the synthesis of Y depends upon that of X. A cross dependency leads to conflicting orderings, namely synthesis of X after that of Y and synthesis of Y after that of X. The solution to such conflicting orderings is to perform a sequence of incomplete syntheses of X and Y. To illustrate non local and cross dependencies, let us consider the synthesis of the verb and direct object in French.

a) The synthesis of the verb depends upon that of the [dir-object] for two reasons. First, there is a switch from the auxiliary *avoir* to the auxiliary *être* (when the verb is conjugated in a compound tense) if the [dir-object] is synthesized as a reflexive pronoun (which must appear before the verb):

Ugo a détesté Marie

(Ugo hated Mary)

Ugo s'est détesté

(Ugo hated himself)

Second, there is agreement in gender and number between the past participle of a verb conjugated in a compound tense and a [dir-object] synthesized as a personal pronoun (which must appear before the verb):

Ugo, je l'ai détesté

(Ugo, I hated him)

Marie, je l'ai détestée

(Mary, I hated her)

The synthesis of the [dir-object] depends upon that of the verb in the following way, which will be explained in detail in 4.1 and 4.2: determining whether the [dir-object] has to be synthesized as a personal pronoun may depend upon the first letter of the conjugated verb.

All In all, the synthesis of the verb depends upon that of the [dir-object] and the synthesis of the [dir-object] depends upon that of the verb. This cross dependency can be handled with the following sequence of incomplete syntheses:

- 1) Determine if the [dir-object] must be synthesized as a reflexive pronoun (by checking if its value is equal to the value of the subject). If it is, mark the verb as having to be conjugated with the auxiliary *être*.
- 2) Synthesize the verb (i.e. conjugate it) without taking into account a possible agreement between a past participle and a pronominalized [dir-object]. In Step 2, the verb is conjugated in a compound tense with the right auxiliary thanks to Step 1. Let us mention that the conjugation of a verb is a morphological operation.
- 3) Synthesize the [dir-object] if it has not been synthesized as a reflexive pronoun in Step 1. In Step 3, the form of the conjugated verb provided by Step 2 is used to determine if the [dir-object] has to be synthesized as a personal pronoun (see Sections 4.1 and 4.2).
- 4) Complete the synthesis of the verb if necessary, i.e. carry out the agreement in gender and number between a past participle if any (information given by Step 2) and a pronominalized [dir-object] if any (information given by Step 3).

These four steps imply that both the direct object and the verb have to be checked over twice. Note that these checkings are for the synthesis of these two elements only. The cross dependencies that arise from other elements (see Section 4.1) imply that the direct object and the verb have to be checked over more than twice. Generally speaking, a clause template (i.e. a tree) is gone through several times in the syntactic component processes.

3.3 The order in which the synthesized elements must be placed in the final clause

This order is determined by syntactic transformations that move synthesized elements. For example, the "length permutation transformation" moves a direct object after an indirect object when the former is "longer" than the latter:

I bought a book about Napoleon which was published in 1924 from this bookseller
--> *I bought from this bookseller a book about Napoleon which was published in 1924*

Another example of such a syntactic transformation is the English "dative transformation" illustrated in the following pair:

John gave a book to Mary

--> *John gave Mary a book*

The dative transformation cannot take place if the direct object is pronominalized:

John gave it to Mary

--> **John gave Mary it*

The decision to apply the dative transformation to a clause should therefore be made after the synthesis of the direct object.

4 Synthesis of personal pronouns⁴

If a token refers to the speaker(s) or the hearer(s), it must be synthesized as a first or second person pronoun. The only operation to be performed is then the computation of this "dialogue" pronoun. Otherwise, we consider synthesizing a token as a third person personal pronoun only if it has already been synthesized (because occurring in a previous clause template, for example). In other words, we do not consider the left pronominalization phenomena (T. Reinhart 1983). Determining whether a token which does not refer to the speaker(s) or hearer(s) and which has already been synthesized has to be synthesized as a pronoun requires the following steps to be gone through⁵:

- 1) Compute the form of the foreseen pronoun.
- 2) Compute the list L1 of tokens that have been synthesized in nominal phrases the "morphological" features (i.e. gender and number) of which are compatible with the form of the foreseen pronoun provided by Step 1.
- 3) Compute the sublist L2 of L1 that contains the elements that are syntactically compatible with the foreseen pronoun.
- 4) Compute the sublist L3 of L2 that contains the elements that are semantically compatible with the foreseen pronoun.
- 5) According to the number of elements in L3, and maybe according to other considerations, decide if actually the foreseen pronoun has to be synthesized.

⁴ This section has been written with F. Namer. Some parts of it are published in (L. Danlos and F. Namer 1988).

⁵ Steps 2,3,4 and 5 are the ones an analysis system should go through when searching the referent of a third person pronoun.

4.1 Computation of the form of the foreseen pronoun

This computation involves the following factors :

- 1) The syntactic position in which the token that could be synthesized as a pronoun appears. In English, it is enough to distinguish between the subject and complement positions. In French and Italian, it is necessary to distinguish between the [subject], [dir-object], [à-object], [de-object], [loc-object] and [prep-object] positions. On the one hand, the [subject] and [prep-object] positions give generally rise to pronouns that are similar to the English ones⁶. On the other hand, the other positions may give rise to pronouns that must stand before the verb, such pronouns being noted *Ppv* ("pronoms pré-verbaux").
- 2) The person and number of the token. Person and number are semantic information which are given in the definition of the token.
- 3) The gender of the nominal phrase that synthesizes the previous occurrence of the token. In French and Italian languages, which have only the masculine and feminine gender, gender is not semantic but lexical information. Consider the token TOK1 with the following definition :

TOK1 =: BICYCLE

NUMBER : 1

DEFINITE : yes

In French, it can be synthesized as a feminine noun group *la bicyclette* (the bicycle) or as a masculine noun group *le vélo* (the bike) . The gender of a pronoun which synthesizes a token is generally equal to the gender of the previous occurrence of the token :

*La bicyclette est cassée. (Elle + * Il) est au garage.* (The bicycle is broken. It is at the garage.)

*Le vélo est cassé. (Il + * Elle) est au garage.* (The bike is broken. It is at the garage.)

- 4) The human nature of the token along with the verb (in the infinitive form) of the clause template. As an example, let us consider the synthesis of an [à-object] in Italian. The verbs *dare*, *pensare* and *credere* can all take a human or non human [à-object], as shown in the following sentences:

Ugo diede un libro a Maria (Ugo gave a book to Mary)

Ugo diede una spolverata alla credenza (Ugo gave the sideboard a wipe)

Ugo crede (a Maria + alla teoria) (Ugo believes in (Mary + the theory))

Ugo ha pensato (a Maria + alla teoria) (Ugo thought of (Mary + the theory))

The form of a pronoun corresponding to the [à-object] of one of these verbs is given in the table below⁷ :

⁶ In fact, an Italian [subject] pronoun is erased when this erasing does not create any ambiguity. There is no room in this paper to discuss this complex phenomenon, which is also to be found in Spanish and Portuguese.

⁷ The pronouns preceded by the preposition *a* are not placed before the verb.

TABLE 1	HUMAN			NON HUMAN		
	mas-sing	fem-sing	plural	mas-sing	fem-sing	plural
dare	gli	le	loro	gli	le	loro
credere	gli	le	loro	ci	ci	ci
pensare	a lui	a lei	a loro	ci	ci	ci

In Italian as well as in French, the form of an [à-object] pronoun can only be obtained by consulting a "lexicon-grammar" (M. Gross 1975, 1986 ; A. Elia et alii 1981). For each verb, a lexicon-grammar records all its syntactic properties, among them those concerning pronominalization.

5) The synthesis of the verb. In French, a [dir-object] of the third person singular is pronominalized as *le* if the previous occurrence of the token is masculine, as *la* if feminine :

<i>Ugo, je le vois souvent</i>	(Ugo, I often see him)
<i>Marie, je la vois souvent</i>	(Mary, I often see her)

However, if the first letter of the conjugated verb is a vowel, there is elision of *le* or *la* into *l'* :

<i>Ugo, je l'ai vu et je l'entends</i>	(Ugo, I saw him and I hear him)
<i>Marie, je l'ai vue et je l'entends</i>	(Mary, I saw her and I hear her)

This elision changes the computation of the morphological antecedents of the pronoun as shown in 4.2. Therefore, it has to be taken into account when determining if the [dir-object] has to be pronominalized. Other elisions of *Ppv* are to be found when the first letter of the conjugated verb is a vowel, for example the elision of *me* into *m'* :

<i>Ugo me voit et m'entend</i>	(Ugo sees me and hears me)
--------------------------------	----------------------------

Contrary to the elision of *le* or *la* into *l'*, the elision of *me* into *m'* does not change the computation of the antecedents of the pronoun : *me* or *m'* can refer only to the speaker. Therefore, this elision can be one of the last operations performed by the generation system.

In Italian, the synthesis of the [dir-object] raises the same problems as it does in French, since there is elision of the pronouns *lo* and *la* into *l'* :

<i>Ugo, l'ho visto e lo sento</i>	(Ugo, I saw him and I hear him)
-----------------------------------	---------------------------------

This elision is ruled by more complicated conditions than it is in French. It does not only depend upon the first letter of the conjugated verb being a vowel, as shown in the following examples :

Questo profumo, Ugo lo offre spesso a Maria

(This perfume, Ugo often offers it to Mary)

Questo profumo, Ugo l'offrirà domani a Maria (This perfume, Ugo will offer it tomorrow to Mary)

b) The synthesis of other complements. This factor involves several non local dependencies, among them the following ones for French⁸ :

a) A [à-object] cannot be pronominalized as a *Ppv* if there is a [dir-object] synthesized as one of the following *Ppv* =: *me, te, nous, vous, se* :

Marie, je la présenterai à Ugo --> Je la lui présenterai

(Mary, I will introduce her to Ugo --> I will introduce her to him)

*Toi, je te présenterai à Ugo --> * Je te lui présenterai*

(You, I will introduce you to Ugo --> I will introduce you to him)

b) A [loc-object] designing a source locative cannot be pronominalized as the *PPv* =: *en* if there is a [dir-object] synthesized as the *Ppv* =: *en* :

Ces bonbons, je les ai sortis de la boîte --> Je les en ai sortis

(These candies, I pulled them out of the box --> I pulled them out of it)

*Des bonbons, j'en ai sortis de la boîte --> * J'en en ai sortis*

(Candies, I pulled some of them out of the box --> I pulled some of them out of it)

c) A [loc-object] designing a scenic locative cannot be pronominalized as the *PPv* =: *y* if there is a [à-object] synthesized as the *Ppv* =: *y* :

Marie, je l'ai rencontrée en Italie --> Je l'y ai rencontrée

(Mary, I met her in Italy --> I met her there)

*Marie, j'y ai pensé en Italie --> * J'y y ai pensé*

(Mary, I thought of her in Italy --> I thought of her there)

The non local dependencies described in a), b) and c) for French are observed in similar conditions for Italian. The following dependencies are observed only for Italian:

d) A [loc-object] designing a scenic locative cannot be pronominalized as the *Ppv* =: *ci* if there is a first person plural [dir-object] or [à-object] synthesized as the *Ppv* =: *ci* :

Ugo ti ha incontrato in Italia --> Ugo ti ci ha incontrato

(Ugo met you in Italy --> Ugo met you there)

*Ugo ci ha incontrati in Italia --> * Ugo ci ci ha incontrati*

(Ugo met us in Italy --> Ugo met us there)

⁸ (M. Gross 1968) presents a complete list of dependencies between the French *Ppv*.

e) An [à-object] of the third person singular can be pronominalized as *gli* if the previous occurrence of the token is masculine, as *le* if feminine (see Table 1). However, if there is a [dir-object] or [dc-object] synthesized as one of the following *Ppv* =: *lo*, *la*, *li*, *le*, *l'*, *ne*, the pronouns *gli* or *le* amalgamate with this *Ppv* and both become *glie* :

Diedi il libro a Maria --> *Le diedi il libro*

(I gave the book to Mary --> I gave the book to her)

Diedi il libro a Ugo --> *Gli diedi il libro*

(I gave the book to Ugo --> I gave the book to him)

Il libro, lo diedi a (Maria + Ugo) --> *Glielo diedi*

(The book, I gave it to (Mary + Ugo) --> I gave it to her/him)

4.2 Computation of the morphologically compatible antecedents of the foreseen pronoun

A token, which does not refer to the speaker(s) or hearer(s), corresponds to a morphological antecedent of the foreseen pronoun if it has been previously synthesized as a nominal phrase the morphological features (i.e. gender and number) of which are compatible with the form of the foreseen pronoun. For example, if the foreseen pronoun is :

- the French [subject] pronoun *il*, its morphological antecedents are the masculine singular noun
- the Italian [à-object] pronoun *gli*, its morphological antecedents are the masculine singular noun phrases;
- the Italian [à-object] pronoun *le*, its morphological antecedents are the feminine singular noun phrases;
- the Italian [à-object] pronoun *glie* (result of an amalgamation of *gli* or *le* with another *Ppv*), its morphological antecedents are the singular noun phrases.

In the cases mentioned above, the computation of the morphological antecedents of the foreseen pronoun (i.e. the computation of the list L1) only depends upon the form of the pronoun. The computation of L1 may also depend upon the synthesis of other elements, thereby involving non local dependencies. For example, when the foreseen pronoun is *l'* (result of an elision of *le* or *la*), its morphological antecedents are all the singular noun phrases if the conjugated verb does not include a past participle as in *Je l'entends* (I hear him/her/it) ; otherwise, its morphological antecedents are the singular noun phrases with the gender indicated by the past participle. Recall that there is agreement in gender (and number) between a past participle and a pronominalized [dir-object]. In *Je l'ai vu* (I saw him/it), the morphological antecedents of *l'* are the masculine singular noun phrases, while in *Je l'ai vue* (I saw her/it), the morphological antecedents of *l'* are the feminine singular noun phrases. This is why the synthesis of the [dir-object] depends upon that of the verb (cf. 3.2). In fact, the synthesis of the [dir-object] also depends upon that of a possible attribute of the direct object, as shown below:

- in *Je l'estime beau* (I consider him/it beautiful) with the masculine adjective *beau*, the morphological antecedents of *l'* are the masculine singular noun phrases;

• In *Je l'estime belle* (I consider her/it beautiful) with the feminine adjective *belle*, the morphological antecedents of *l'* are the feminine singular noun phrases;

• In *Je l'estime fragile* (I consider him/her/it fragile) with the adjective *fragile* which is both masculine and feminine, the morphological antecedents of *l'* are all the singular noun phrases.

This implies that the inflected forms of adjectives, which is morphological information, has to be taken into account when determining if a token can be synthesized as a personal pronoun. Similar morphological considerations have to be taken into account for the erasing of an Italian [subject] pronoun, a matter that is not discussed here (see note 6).

4.3 Computation of the morphologically and syntactically compatible antecedents of the foreseen pronoun

The computation of the morphologically compatible antecedents of the foreseen pronoun returns a list *L1* of tokens. Among those tokens, some may be syntactically incompatible with the envisioned pronoun. For example, in *Mary hated her*, the pronoun *her* cannot refer to *Mary*. The token representing *Mary* is said to be syntactically incompatible with the pronoun *her*. This coreferential syntactic incompatibility can be stated in the following rule:

if a personal pronoun synthesizes a complement in a clause template, then it does not refer to the subject of this clause template, because if it did, it would be synthesized as a reflexive pronoun (**Mary_i hates Mary_i* --> *Mary_i hates herself*).

Another example of coreferential syntactic incompatibility can be stated in the following rule:

if a pronoun synthesizes the subject of a sentential clause which must be reduced to an infinitive form when its subject is equal to the subject of the main clause, then this pronoun does not refer to the subject of the main clause, because if it did, the sentential clause would be reduced to an infinitive form (**Mary_i wants that Mary_i leaves* --> *Mary_i wants to leave*).

An illustration of this rule is that in *Mary wants that she leaves*, the pronoun *she* cannot refer to *Mary*.

4.4 Computation of the morphologically, syntactically and semantically compatible antecedents of the foreseen pronoun

The computation of the morphologically and syntactically compatible antecedents of the foreseen pronoun returns a list *L2* of tokens. Among these tokens, some may be semantically incompatible with the envisioned pronoun. The first reason why a token can be semantically incompatible with the foreseen pronoun is associated with the human nature of the token. Go back to TABLE 1 that records the

form of an Italian [\bar{a} -object] pronoun according to the verb and to the gender, number and human nature of the [\bar{a} -object]. TABLE1 shows that:

- the [\bar{a} -object] pronoun *ci* , which appears with verbs such as *credere* or *dare* , can refer only to non human tokens. The human tokens are said to be semantically incompatible with the [\bar{a} -object] pronoun *ci*.
- the [\bar{a} -object] pronoun *gli* can refer
 - either to human or non human tokens when the verb behaves as *dare* ,
 - or only to human tokens when the verb behaves as *credere* .

The second reason why a token can be semantically incompatible with the foreseen pronoun is associated with distributional constraints. An illustration of distributional constraints is the following one: the direct object of the verb *iron* in the active (its subject in the passive) must denote an item of clothing and not, for example, a piece of furniture. Therefore, in the discourse

The shirt is on the table. It was ironed recently .

the pronoun *it* refers compulsorily to ^{shirt}the *shirt* and not to *the table* . The token representing *the table* is said to be semantically incompatible with the pronoun *it* .

4.5 Last decisions

The computation of the morphologically, syntactically and semantically compatible antecedents of the foreseen pronoun returns a list L3 of tokens.

At a rough estimate, if the number of elements of L3 is one, then the foreseen pronoun can be synthesized since it does not lead to ambiguity, whereas the foreseen pronoun should not be synthesized if the number of elements in L3 is greater than one since it would lead to ambiguity. Yet, it is well known that pragmatic and structure parallelism considerations may allow a pronoun to be not ambiguous even if L3 has more than one element (G. Hirst 1981, C. Sidner 1981, K. McKeown 1985, L. Danlos 1987a)⁹. Those considerations, which will not be discussed here, have to be taken into account to determine whether the foreseen pronoun has to be actually synthesized.

5 Other interactions of decisions with the pronominalization issue

We have shown in 4.1 and 4.2 that pronominalization involves the morphological level. Hence, the following result: the morphological level should be taken into account before the very last stage of the generation process.

⁹ It may also happen that the foreseen pronoun should not be synthesized when the number of elements of L3 is one, because the distance between the antecedent and the pronoun is too great, for example.

In our generation model which is modularized into a strategic component and a syntactic component, the decisions made in the strategic component are independent from the pronominalization decisions which are all made in the syntactic component. We are going to show that this may be wrong.

First, let us point out that our generation model does not handle all the syntactic transformations in the same way: some transformations are decided in the strategic component, while others are decided in the syntactic component. For example, the passive transformation is decided in the strategic component since a discourse structure indicates if a sentence must be in the active or in the passive. This is justified by the fact that the choice between active and passive (with or without agent) is relevant to the discourse semantics (L. Danlos 1987a), as shown in the following pair of sentences:

John shot Mary who was killed.

Jon shot Mary whom he killed.

The first sentence, whose relative clause is in the passive without agent, expresses a causal relation between John's shooting and Mary's death. This is not the case for the second sentence whose relative clause is in the active. Similarly, the "ergative" transformation must be decided in the strategic component since it affects the discourse semantics, as shown in the following pair of sentences:

John knocked over the glass which broke.

John knocked over the glass which was broken.

The first sentence, whose relative clause is in the ergative, expresses a causal relation between the main clause and the relative clause. This is not the case for the second sentence, whose relative clause is in the passive without agent. Generally speaking, all the transformations that affect the discourse semantics should be decided "early", which means, in our generation model, in the strategic component.

On the other hand, the length permutation transformation and the dative transformation are decided in the syntactic component when placing the synthesized elements in the final clause. This is justified by the fact that these transformations can be decided only after the complements have been synthesized (cf. 1.3). Generally speaking, all the transformations that affect the sentence form and/or that depend upon the pronominalization decisions should be decided "late", which means, in our generation model, in the syntactic component.

Now, consider the clause templates that have a predicate as a subject, for example:

Cl.2 (:Cl (:subject PRED1) (:verb *plaire*) (à-object HUM2))

where the predicate PRED1 has the following definition:

PRED1 =: DRINK

AGENT = HUM1

OBJECT : wine

The token PRED1 can be synthesized as a pronoun if it has already been synthesized, like in the following discourse where the second sentence corresponds to the synthesis of CI2:

Ugo boit du vin. Cela plaît à Marie

(Ugo drinks wine. It pleases Mary.)

If PRED1 is not synthesized as a personal pronoun, it is synthesized as a sentential clause:

(1) *Qu'Ugo boive du vin plaît à Marie*

(That Ugo drinks wine pleases Mary)

The form (1) is stylistically unpleasant, as it is the case for most forms with a sentential subject. One wants thus to avoid generating such forms. One way to improve (1) is by applying the subject extraposition transformation:

Il plaît à Marie qu'Ugo boive du vin

(It pleases Mary that Ugo drinks wine)

The decision to apply subject extraposition must be made after the pronominalization decisions, because subject extraposition cannot take place when the subject is pronominalized:

Cela plaît à Ugo

(It pleases Ugo)

--> **Il plaît à Ugo cela*

(It pleases Ugo it)

Subject extraposition does not apply to any French verb (M. Gross 1975): roughly, it does not apply to transitive verbs:

(2) *Qu'Ugo boive du vin ravit Marie*

(That Ugo drinks wine delights Mary)

--> **Il ravit Marie qu'Ugo boive du vin*

(It delights Mary that Ugo drinks wine)

The form (2) is as stylistically unpleasant as (1), and thus should not be generated as such. One way to improve it is by applying the passive transformation:

Marie est ravie de ce qu'Ugo boive du vin (Mary is delighted by John drinking wine)

Hence, one would like to lay down the following rule:

if a [subject] predicate is not synthesized as a personal pronoun, synthesize it as a sentential clause and apply subject extraposition if possible, if not apply the passive transformation.

Each a rule can take place only after the pronominalization decisions, so in the syntactic component in our model. However, we have seen that the passive transformation, which is relevant to the discourse semantics, should be decided in the strategic component. Hence a contradiction which shows that our modularization into only two components - a strategic component and a syntactic component, the latter handling pronominalization - is even too modular.

Conclusion

A French and Italian syntactic component covering the phenomena described in Sections 3 and 4 has been implemented in a procedural Common-Lisp program. An English syntactic component covering some of these phenomena has been implemented in a declarative formalism using functional descriptions (J.M. Lancel et alii 1988). Note that the cross dependencies and morphological interactions presented in these sections concern only the synthesis of personal pronouns, putting aside the synthesis of sentential clauses, subordinate clauses and coordinated clauses (L. Danlos 1987b). The reader can guess the complexity of a robust syntactic component for Romance languages.

We have not yet designed a generation algorithm that would cover the phenomena described in Section 5, that is a generation algorithm which would handle a total interaction between the conceptual, linguistic, syntactic and morphological levels. Nevertheless and at least, we hope to have underlined the linguistic complexity of text generation, a complexity which is still underestimated.

Bibliography

- Danlos, L., 1985, *Génération automatique de textes en langues naturelles*, Masson, Paris.
- Danlos, L., 1987a, *The linguistic basis of text generation*, Cambridge University Press, Cambridge.
- Danlos, L., 1987b, A French and English Syntactic Component for Generation, *Natural Language Generation: New results in Artificial Intelligence, Psychology and Linguistics*, Kempen G. ed, Dordrecht/Boston, Martinus Nijhoff Publishers.
- Danlos, L., Namer, F., 1988, Morphology and cross dependencies in the synthesis of personal pronouns in Romance languages, *Proceedings of COLING-88*, Budapest.
- Ellin, A., Martinelli, M., D'Agostino, E., 1981, *Lessico e strutture sintattiche. Introduzione alla sintassi del verbo italiano*, Napoli Liguori, Napoli.
- Gross, M., 1968, *Grammaire transformationnelle du français: syntaxe du verbe*, Larousse, Paris.
- Gross, M., 1975, *Méthodes en syntaxe*, Hermann, Paris.
- Gross, M., 1986, Lexicon-Grammar, The Representation of Compound Words, in *Proceedings of Coling'86, 11th International Conference on Computational Linguistics*, Bonn.

- Hirst, G., 1981, Discourse oriented Anaphora resolution, *American Journal of Computational Linguistics*, vol. 7, no 2.
- Lancel, J.M., Otani, M., Simonin, N., Daňlos, L., 1988, Sentence Parsing and Generation with a Semantic Dictionary and a Lexicon-Grammar, in *Proceedings of Coling'88*, Budapest.
- McKeown, K., 1985, *Text generation*, Cambridge University Press, Cambridge.
- Reinhart, T., 1983, *Anaphora and semantic interpretation*, Croom Helm, London.
- Sidner, C., 1981, Focusing for Interpretation of Pronouns, *American Journal of Computational Linguistics*, vol. 7, no 4.