

Morphology and cross dependencies in the synthesis of personal pronouns in Romance languages

Laurence DANLOS and Fiammetta NAMER
LADL, Université de Paris 7
2, Place Jussieu
75251 Paris Cedex 05, France

Abstract

This paper describes some of the problems that arise from the synthesis of personal pronouns in a system that generates texts in Romance languages. It puts the emphasis first on the fact that the morphological level has to be taken into account early in the generation process, second on the numerous "cross dependency" phenomena which are to be found when the synthesis of an element X depends upon that of another element Y and when the synthesis of Y depends upon that of X. The linguistic examples are taken from French and Italian languages, for which a robust generation system has been implemented.

1 Introduction

It is generally believed that a generation system can be modularized into a sequence of components, the first one making the "high level" decisions (i.e. the conceptual decisions), the following ones making the linguistic decisions (e.g. lexical and syntactic construction choices), the penultimate one performing the "low level" operations (i.e. the syntactic operations), and the last one handling the morphological operations. We have shown in (L. Danlos 1985, 1987a) that the conceptual and linguistic decisions are operations that depend on each other. Therefore, we designed a generation system modularized in the following way: a "strategic component" makes the conceptual and linguistic decisions simultaneously and gives back "clause templates" which are synthesized into clauses by a "syntactic component". A simplified version of the clause template syntax is the following one (a more complete version is presented in (L. Danlos 1987b)) :

[C1] = (:C1 [subject] [verb] cplⁿ (0 ≤ n ≤ 2))
 [subject] = (:subject token)
 [verb] = (:verb *verb*)
 cplⁿ = [dir-object] / [à-object] / [de-object] / [prép-object]
 [dir-object] = (:dir-object token)
 [à-object] = (:à-object token)
 [de-object] = (:de-object token)
 [prép-object] = (:prép-object [prép] (:object token))
 [prép] = (:prép *preposition*)

The prepositional complements [à-object] and [de-object] are complements respectively introduced by *à*, and *de* in French, *a* and *di* 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 (cf. 3). An example of a clause template is :

(:C1 (:subject HUM1) (:verb *amare*) (:dir-object HUM2))

with

HUM1 =: PERSON	HUM2 =: PERSON
NAME : Ugo	NAME : Maria
SEX : masc	SEX : fem

According to the context (i.e. the clause templates that have been previously synthesized), the syntactic component, which handles pronominalization, produces one of the following Italian clauses (given that the verb

is in the present tense) :

<i>Ugo ama Maria</i>	(Ugo loves Mary)
<i>Ugo l'ama</i>	(Ugo loves her)
<i>Quest'uomo l'ama</i>	(This man loves her)
<i>Ama questa donna</i>	(He loves this woman)

It will be shown in 3 that pronominalization involves the morphological level. 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.

The second aim of this paper is to put forward "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. 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.

2 Non local and cross dependencies

The synthesis of the verb and direct object in French will be taken as an illustration of non local and cross dependencies. On the one hand, 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) :

<i>Ugo a detesté Marie</i>	(Ugo hated Mary)
<i>Ugo s'est detesté</i>	(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) :

<i>Ugo, je l'ai detesté</i>	(Ugo, I hated him)
<i>Maria, je l'ai detestée</i>	(Mary, I hated her)

On the other hand, the synthesis of the [dir-object] depends upon that of the verb in the following way which will be explained in detail in 3 : determining whether the [dir-object] has to be synthesized as a personal pronoun may depend upon the first letter and the form 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.

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 are checked over twice. Note that this is only for the synthesis of these two elements. The cross dependencies that arise from other elements imply that the direct object and the verb are checked over more than twice. Generally speaking, a clause template (i.e. a tree) is gone through several times in the syntactic component.

3 Synthesis of personal pronouns

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 the computation of this "dialogue" pronoun. Otherwise, we consider synthesizing a token as a third person 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 personal pronoun requires the following steps to be gone through:

- 1) Compute the form of the foreseen pronoun (cf. 3.1);
- 2) Compute the list L1 of tokens that have been synthesized in nominal phrases whose "morphological" features (i.e. gender and number) are compatible with the form of the foreseen pronoun (cf. 3.2).
- 3) Compute the sublist L2 of L1 that contains only the elements of L2 that are syntactically compatible with the foreseen pronoun. For example, in *Mary hated her*, *Mary* and the personal pronoun *her* cannot be coreferential. The token representing *Mary* is said to be syntactically incompatible with the pronoun *her*.
- 4) Compute the sublist L3 of L2 that contains only the elements of L2 that are semantically compatible with the foreseen pronoun. For example, in *The book is on the table, it was published recently*, the pronoun *it* and *the table* cannot be coreferential because the direct object of the verb *publish* in the active (its subject in the passive) cannot be a piece of furniture. The token representing *the table* is said to be semantically incompatible with the pronoun *it*.
- 5) According to the number of elements in L3, and maybe according to other considerations, decide actually if the foreseen pronoun has to be synthesized. 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, while it 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 non ambiguous even if L3 has more than one element (G. Hirst 1981, C. Sidner 1981, K. McKeown 1985, L. Danlos 1987a). Step 5 takes those

considerations into account to determine whether the foreseen pronoun has to be actually synthesized.

3.1 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 [prép-object] positions: the [subject] and [prép-object] positions generally give rise to pronouns that are similar to the English ones¹; the other positions may give rise to pronouns that must appear before the verb, such pronouns being noted *Ppv*.

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, 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]. The form of a pronoun corresponding to the [à-object] of one of these three verbs is given in the table below²:

TABLE 1	HUMAN			NON HUMAN		
	m-s	f-s	plur	m-s	f-s	plur
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).

¹ 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.

² The abbreviation "m-s" stands for masculine-singular, "f-s" for feminine-singular, "plur" for plural. The pronouns preceded by the preposition *a* are not placed before the verb.

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 :

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

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

Ugo, je l'ai vu et je l'entends (Ugo, I saw him and I hear him)

Marie, je l'ai vue et je l'entends (Mary, I saw her and I hear her)

This elision changes the computation of the morphological antecedents of the pronoun as shown in 3.2. Therefore, it has to be taken into account when determining if the [dir-object] has to be pronominalized.

6) The synthesis of other complements. This factor involves several non local dependencies. For example, in French, an [à-object] cannot be pronominalized as the *Ppv* =: *lui* if there is a [dir-object] synthesized as *te* (M. Gross 1968):

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)

Another example in Italian: 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] synthesized as the *Ppv* =: *lo*, 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)

3.2 Computation of the morphological 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 whose morphological features (i.e. gender and number) are compatible with the form of the foreseen pronoun. For example, if the foreseen pronoun is :

- the French [dir-objet] pronoun *la*, its morphological antecedents are the feminine singular noun phrases;
- 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 can also depend upon the synthesis of other elements, thereby involving non local dependencies. For example, when the foreseen pronoun is *l'*, 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: 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. It is an illustration of the claim that pronominalization involves the morphological level.

Conclusion

The cross dependencies and morphological interactions which were presented here concern only the synthesis of personal pronouns, putting aside the synthesis of sentential, subordinate and coordinated clauses. The reader can guess the complexity of a syntactic component for Romance languages. A robust French and Italian syntactic component has been implemented in a procedural Common-Lisp program. An English syntactic component has been implemented in a declarative formalism using functional descriptions (J.M. Lancel et alii 1988).

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.
- Elia, 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*, Bonn.
- Hirst, G., 1981, *Discourse oriented Anaphora resolution, American Journal of Computational Linguistics*, vol. 7, no 2.
- Lancel, J.M., Otani, M., Simonin, N., Danlos, 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.