SUBJECT ERASING AND PRONOMINALIZATION IN ITALIAN TEXT GENERATION

Fiammetta Namer
LADL
Université Paris VII
2, place Jussieu
75251 Paris Cedex 05
France

ABSTRACT

Certain Romance languages such as Italian, Spanish and Portuguese allow the subject to be erased in tensed clauses. This paper studies subject erasing in the framework of a text generation system for Italian. We will prove that it is first necessary to try to pronominalize the subject. Therefore, we are led to study the synthesis of subject and complement personal pronouns. In Romance languages, personal pronouns raise many syntactic problems, whose solution is complex in a generation system. We will see that pronominalization plays a fundamental role in the order in which the elements of a clause are synthesized, and consequently in the synthesis of this clause. Moreover, the synthesis of a clause must take into account the fact that subject erasing and the synthesis of complements are phenomena which depend on each other. The complex algorithm that must be used for the synthesis of a clause will be illustred in various examples.

1 Presentation of the generation system

In a generation system, two questions must be answered: "What to say?" (in order to decide on the content of the message to be produced) and "How to say it?" (producing the text which carries this content). We are interested only in the "How to say it?" question. We have adapted for Italian the generation system developped by L.Danlos (1987a,1987b) which produces texts in French and English. This generator includes two components: the strategic component and the syntactic component.

1) The strategic component takes both conceptual and linguistic decisions. It selects a discourse structure which determines the order of information, the number and form of the sentences of the text to be generated. It returns a text template which is a list of the form:

(Sent1 Punct1 ... Senti Puncti ... Sentn Punctn)

where Puncti is a punctuation mark and Senti a sentence template. For the sake of simplification, only sentence templates which are clause templates without adverbial phrases will be considered here. This means that adverbial phrases (e.g. subordinate clauses) and coordinations of sentence templates are put aside (L. Danlos 1987b). In a clause template (without adverbial phrases), which will be noted Cl, the elements are in the canonical order:

subject - verb - dir_object - prep_object(s)

In particular, the subject appears always before the verb although the subject can be placed after the verb in Italian:

Ha telefonato Gianni (Gianni has phoned)

Subject-verb inversion has been described by L. Rizzi (1982) as a phenomenon which is correlated with subject erasing. This approach may be suitable for an analysis system which has to identify the subject of a clause. However it is not for a generation system which has to synthesize an identified subject.

This is an example of text template:

(1) ((:Cl1 (:subject MAN1) (:verb *amare*)
         (:dir_object MISS1)) .
     (:Cl2 (:subject MAN2) (:verb *odiare*)
         (:dir_object MISS2)) .)

It is made up of two clause templates Cl1 and Cl2. Cl1 includes the tokens MAN1 and MISS1, Cl2 the tokens MAN2 and MISS2. These tokens may be defined as follows:

MAN1 =: PERSON		MISS1 =: PERSON	
NAME	: *Max*	NAME	: *Lia*
SEX	: masc	SEX	: fem

MAN2 =: PERSON		MISS2 =: PERSON	
NAME	: *Ugo*	NAME	: *Eva*
SEX	: masc	SEX	: fem

2) The syntactic component synthesizes a text template into a text. From the text template (1), it produces the following text if the verbs are conjugated in the present tense:

Max ama Lia. Ugo odia Eva. (Max loves Lia. Ugo hates Eva)

Given the following simplified text template, where the functional categories (eg. :Cl , :subject) are omitted for the sake of readibility:

(2) (MAN1 *amare* MISS2 . MAN2 *odiare* MISS2)
    (MAN1 love MISS2 . MAN2 hate MISS2)

the syntactic component synthesizes the first Cl as:

Max ama Eva. (Max loves Eva)

Then it synthesizes the second one according to the left-hand context, i.e. the first synthesized clause. Among other things, it computes that the second occurrence of MISS2 can be synthesized as a personal pronoun:

Max ama Eva. Ugo la odia.   (Max loves Eva.Ugo hates her)

The different steps required for the synthesis of a personal pronoun will be described in section 5.1. In the same way, the synthesis of the simplified text template:

(3) (MAN2 *essere cattivo.* MAN2 *odiare* MISS2)
      (MAN2 be nasty. MAN2 hate MISS2)

gives the following text in which the subject position is empty (see section 5.2):

Ugo è cattivo. Odia Eva.   (Ugo is nasty. He hates Eva)

and the synthesis of the text template:

(4) (MAN2 *picchiare* MISS2. MAN2 *odiare* MISS2)
      (MAN2 beat MISS2. MAN2 hate MISS2)

gives the following text, in which the subject position is empty and the direct object synthesized as a personal pronoun:

Ugo picchia Eva. La odia.   (Ugo beats Eva. He hates her)

2   Synthesis of a clause template

In a generation system producing texts in Romance languages, a syntactic component has to handle three different orders for the synthesis of a Cl:
- The order in which the elements appear in a Cl (this order is supposed here to be the canonical order).
- The order in which the elements of a Cl must be synthesized (see below).
- The order in which the synthesized elements must be placed in the final clause (eg. for Italian, subject-verb inversion). This order will not be discussed here.

The order in which the elements of a Cl must be synthesized is determined by "non-local dependencies" and "cross dependencies" (L.Danlos & F.Namer 1988, L.Danlos 1988). A non-local dependency is to be found when the synthesis of an element X depends on that of another element Y. A cross dependency is to be found when the synthesis of X depends on that of Y and when the synthesis of Y depends on that of X. For example, there is a cross dependency between the synthesis of a direct object and that of the verb[1]. First, let us show that the synthesis of the direct object depends upon that of the verb. Consider the following text template:

(5) (MAN1 e MISS1 *essersi sposati ieri.* MAN2 *adorare*
      MISS1.)
      (MAN1 and MISS1 get married yesterday. MAN2 adore MISS1.)

The pronominalisation of the second occurrence of MISS1 is attempted. The foreseen pronoun is *la*, which

is the feminine singular form of a direct object pronou This pronoun must be placed directly before the verb ar must be elided into *l'* since the verb *adorare* conjugat in the past begins with the vowel **a**. Howeve synthesizing the second occurrence of MISS1 as *l'* lea to an ambiguous text:

Max e Lia si sono sposati ieri. Ugo l'adorava.

since *l'* could also be the result of the elision of *l* which is the masculine singular form of a preverb direct object pronoun. The interpretation of this text either:

Max and Lia got married yesterday. Ugo adored hei
or:
Max and Lia got married yesterday. Ugo adored hii

The second occurrence of MISS1 must therefore synthesized not as a personal pronoun, but as a nomir phrase:

Max e Lia si sono sposati ieri. Ugo adorava Lia.
(Max and Lia got married yesterday. Ugo adored Lia.)

This example shows 1) that the synthesis of the dir object depends upon that of the verb, 2) that elisic which is a morphological operation, could not handled in the final step of the syntactic component the generator.

On the other hand, the synthesis of the verb depen on that of the direct object, since a verb conjugated the perfect tense agrees in number and gender with t direct object if the latter is synthesized as a preverl pronoun:

I ragazzi sono morti. Ugo li ha uccisi
(The boys are dead. Ugo killed them)
Le ragazze sono morte. Ugo le ha uccise
(The girls are dead. Ugo killed them)

The cross dependency between the verb and t direct object can be handled with the following sequer of partial syntheses:
1 - Partial synthesis (conjugation) of the ve without taking into account a possible agreeme between a past participle and a direct object pronoun.
2 - Synthesis of the direct object, eventua according to the first letter of the verb.
3 - Second partial synthesis of the verb: gen agreement with the direct object, if a) the verb conjugated in a compound tense, b) the direct object l been synthesized as a personal pronoun.

The phenomena of non-local and cross dependenc make that the synthesis of a Cl requires a comp algorithm which has nothing to do with a lin processus where the elements of a Cl are synthesiz from left to right. We are going to show that synthesis of the subject involves also a number of n local and cross dependencies where pronominalizat plays a fundamental role.

---

[1] Synthesizing a verb means conjugating it.

3 Introduction to subject erasing

First of all, it should be noted that subject erasing does not affect the other elements of the clause: the verb, for example, always agrees with its subject even if erased.

A subject can be erased only if it can be pronominalized since the synthesis of a subject token always comes under one of the three following cases:
1) The token is neither pronominalizable nor erasable.
2) It is both pronominalizable and erasable.
3) It is pronominalizable but not erasable.
In other words, there exists no Cl in which the subject token is erasable yet not pronominalizable.

1) In the text template:

(6) (MISS1 *e* MISS2 *tornare da Londra*. MISS2 *imparare l'inglese.*)
(MISS1 and MISS2 be back from London. MISS2 learn English)

the second occurrence of the token MISS2 can be neither pronominalized[2] (a):

(a) **Lia ed Eva sono tornate da Londra. Lei ha imparato l'inglese.*
(*Lia and Eva are back from London. She has learnt English)

nor erased (b):

(b) **Lia ed Eva sono tornate da Londra. Ha imparato l'inglese.*
(*Lia and Eva are back from London. She has learnt English)

2) In the text template:

(7) (MISS2 *tornare*. MISS2 *stare bene.*)
(MISS2 be back. MISS2 be well.)

the second occurrence of MISS2 can be either pronominalized (a) or erased (b):

(a)*Eva è tornata.Lei sta bene.* (Eva is back. She, she is well)
(b)*Eva è tornata. Sta bene.*　　　(Eva is back. She is well)

The presence of the pronoun *lei* in the second clause of (a) marks insistence on the entity the pronoun represents.

3) In the text template:

(8) (MISS2 *e* MAN2 *tornare da Londra*. MISS2 *imparare l'inglese.*)
(MISS2 and MAN2 be back from London. MISS2 learn English.)

the second occurrence of MISS2 can be pronominalized (a) but not erased (b):

(a) *Eva e Ugo sono tornati da Londra. Lei ha imparato l'inglese.*
(Eva and Ugo are back from London. She has learnt English)

(b) **Eva e Ugo sono tornati da Londra. Ha imparato l'inglese.*
(Eva and Ugo are back from London. (She+ he) has learnt English)

From the three previous examples, it must be clear that there is no Cl in which a subject token is erasable yet no pronominalizable.

Dialogue subject pronouns (i.e. first and second person) come under case 2 provided that the verb is not conjugated in the subjunctive[3]. A verb conjugated in a non-subjunctive form indicates always the number and person of its subject[4]. As a result, a dialogue subject pronoun is always erased in non-subjunctive clauses:

(9) *Verrai domani.*　　　　　(You will come tomorrow)

unless the speaker wishes to insist on the entity the pronoun represents:

(10)*Tu verrai domani.*　　(You, you will come tomorrow)

On the other hand, third person singular subject pronouns come under either case 1 or case 2 or case 3. For human entities, there are two pronominal forms, one masculine *lui*, and the other feminine *lei*[5]. For non human entities, there are also two singular pronominal forms: *esso* (masculine) and *essa* (feminine). Therefore erasing one of these four forms entails the loss of information about both the gender of the subject and its human nature (i.e. human or non-human). This loss of information can give rise to ambiguity.

Third person plural subject pronouns also come under either case 1 or case 2 or case 3. For human entities, there is one pronominal form *loro* used for both masculine and feminine. For non human entities, there are two forms: *essi* (masculine) and *esse* (feminine). Erasing a third person plural subject pronoun thus raises similar problems than erasing a third person singular subject pronoun. Therefore subject erasing will be illustrated only with third person singular token examples.

---

[2] An asterisk * placed in front of a text means that this text is unacceptable because ambiguous.

[3] Only clauses where the verb is conjugated in the indicative will be studied here.

[4] L.Rizzi (1982) associates morphological properties (i.e. number & person) to the verbal suffix. These properties are activated when the subject position is empty. The suffix then acts as subject pronoun.

[5] Two other forms can be used: *egli* (masculine singular) and *ella* (feminine singular). These forms have the same behaviour as *lui* and *lei*, they are simply used at a more literary stilistic level. Therefore only the forms *lui* and *lei* will be used in this paper.
A sentential subject can be pronominalized as the pronoun *ciò*. The synthesis of sentential subjects will not be discussed here.

4 Erasing a third person singular subject which can be pronominalized

The subject pronoun is always erasable in examples such as (7) where the left-hand context of the subject whose erasing is foreseen contains only one singular token. Apart from this trivial case, let us examine when erasing a subject pronoun is possible, i.e. when information about the gender of the subject and its human nature are both recoverable.

4.1 Recoverability of the human nature of the erased pronoun

The human nature of an erased subject pronoun is recoverable when the verbal predicate takes only a human subject or only a non-human subject. In

> *Ugo ha piantato un ciliegio. Esso fruttifica.*
> (Ugo planted a cherry-tree. It fructifies.)

the non-human subject pronoun *esso* can be erased:

> *Ugo ha piantato un ciliegio. Fruttifica.*
> (Ugo planted a cherry-tree. It fructifies.)

since the verb *fruttificare* can take only a non-human subject:

> *(Ugo + lui) fruttifica.*   (*(Ugo + he) fructifies)

On the other hand, in

> *Ugo ha piantato un ciliegio. Esso è ammirevole.*
> (Ugo planted a cherry-tree. It is admirable.)

the pronoun *esso* cannot be erased:

> *Ugo ha piantato un ciliegio. E' ammirevole.*
> (Ugo planted a cherry-tree. (It+he) is admirable.)

since *essere ammirevole* takes both human and non-human subject:

> *(Ugo + lui + questo ciliegio + esso) è ammirevole.*
> ((Ugo + he + this cherry-tree + it) is admirable)

4.2 Recoverability of the gender of the erased pronoun

To study when the gender of the subject is recoverable, we will suppose that the human nature of the subject is recoverable. In the examples below, the verb predicate can take only human subjects.

4.2.1 The gender of the erased pronoun is marked by another element of the clause

If the gender of the subject pronoun whose erasing is foreseen is marked by another element of the clause, then erasing this pronoun does not give rise to ambiguity. Consider the discourses (11) and (12) in which erasing the feminine singular pronoun *lei* (subject of the second clause) is attempted:

(11)*Ugo non vedrà più Eva. Lei è stata condannata all'ergastolo.*
(Ugo will not see Eva anymore. She's been condemned for li

(12)*Ugo non vedrà più Eva. Lei è in prigione per omicidio.*
(Ugo will not see Eva anymore. She's in jail for murder)

Erasing the subject pronoun in (11) does not give rise ambiguity, since the verb marks the gender of t subject[6]. *Ugo*, which is masculine, is thus a prohibit antecedent. The only possible antecedent of the eras subject is *Eva* and the following discourse where *lei* erased is unambiguous:

> *Ugo non vedrà più Eva. E' stata condannata all'ergastol*
> (Ugo will not see Eva anymore. She's been condemned for lif

On the other hand, if the pronoun *lei* is erased in (1. the information about subject gender is lost sin neither the verb nor any other element of the clau indicates it. The antecedents of the erased subject a *Ugo* and *Eva*. The following discourse where *lei* erased is ambiguous:

> **Ugo non vedrà più Eva. E' in prigione per omicidio.*
> (Ugo will not see Eva anymore. (He + she) is in jail for murd

Subject pronoun erasing is therefore prohibited.
The elements of a clause that mark the subje gender are the following:
- either a nominal or adjectival attribute which inflected for gender[7]:

> *Ugo non vedrà più Eva. E' troppo cattivo*
> (Ugo will not see Eva anymore. He is too nasty)

- or the verb, if it satisfies one of the followi conditions:
  a) it is conjugated in the passive (see example (11
  b) it is conjugated in a compound tense with t verb *essere* (be):

> *Ugo non vedrà più Eva. E' andata in Giappone.*
> (Ugo will not see Eva anymore. She's gone to Japan)

c) it is conjugated in a compound tense at t pronominal voice, for example because there is reflexive pronoun:

> *Ugo non ballerà con Eva stasera. Si è ferito.*
> (Ugo will not dance with Eva tonight. He's wounded hims

---

[6] The suffix a of its past participle marks the femini singular. Recall that a past participle agrees in gen(and number with the subject when the verb is conjugat with the auxiliary *essere* (be).

[7] Two classes of adjectives must be distinguished: th(which are inflected for gender, eg.
cattivo: masc.sing / *cattiva*: fem.sing. (nasty)
and those which are not, eg.
gentile: masc. sing. & fem. sing. (nice)
Several classes of nouns must be also distinguished.

4.2.2 The gender of the erased pronoun is computable from the synthesis of other elements of the clause

We are going to show that erasing a subject pronoun depends on the synthesis of complements of the clause (i.e. direct object and prep-objects) because of the constraint of no-coreferentiality between a subject and a complement personal pronoun. This constraint is based on the fact that a complement which is coreferential to the subject is synthesized as a reflexive pronoun. Therefore, in a clause such as *Eva le ha sparato* (Eva shot her), the indirect complement feminine singular personal pronoun *le* cannot be coreferential to the feminine singular subject *Eva* because if it were it would be a reflexive pronoun: *Eva si è sparata* (Eva shot herself). Let us illustrate the use of this constraint for erasing a subject pronoun with the following text:

(13) *Eva è stata uccisa da Ugo. Lui le ha sparato durante la notte.*
(Eva was killed by Ugo. He shot her during the night)

In (13), there is no subject attribute and the verb is conjugated with the auxilary *avere* (have). Therefore the subject gender is only marked in the subject pronoun *lui*. However if this pronoun is erased, the resulting text is not ambiguous:

Eva è stata uccisa da Ugo. Le ha sparato durante la notte.
(Eva was killed by Ugo. He shot her during the night)

The only interpretation (the only possible antecedent) of the erased subject is *Ugo*. The indirect complement pronoun *le* can only have a feminine singular antecedent, here *Eva*. The subject and this pronoun cannot be coreferent. Therefore the antecedent of the erased subject is the only other human which appears in the context: *Ugo*.

Similarly, consider text (14):

(14) *Ugo non ama più Eva. Lui l'ha abbandonata.*
(Ugo does not love Eva anymore. He abandoned her)

The direct object pronoun *l'* (elided form of the masculine singular *lo* or of the feminine singular *la*) does not indicate the gender of its antecedent. However, this gender is marked in the feminine past participle *abbandonata* [8]. The pronoun *l'* thus refers to *Eva*. Since the antecedent to this pronoun is necessarily different from that of the subject, *Eva* cannot be an antecedent of the subject. Erasing the subject pronoun does not give rise to ambiguity:

Ugo non ama più Eva. L'ha abbandonata.
(Ugo does not love Eva anymore. He abandoned her)

---

[8] Recall that the past participle of a verb conjugated with the auxiliary *avere* agrees in gender and number with its direct object if this object is in preverbal position.

5 Synthesis of a third person singular subject token

Since a third person singular subject token can be synthesized as an empty element only if its pronominalization is possible, the synthesis of such a token will take place as follows:
If the token has never been mentioned (see 5.1.b): synthesis of a nominal phrase (not described here)
Else check if pronominalizing it is allowed
    in this case, check if erasing it is allowed
        if it is, synthesis of an empty subject,
        else synthesis of a subject pronoun
    else redescription or repetition of the token (not described here).
We present below:
1) the different steps to be gone through for the synthesis of a subject pronoun, and more generally, for the synthesis of a personal pronoun.
2) the peculiar operations which are necessary in Italian for synthesizing a subject pronoun and erasing it.

5.1 Synthesis of a personal pronoun

The list of operations required for the synthesis of a personal pronoun is as follows:

a) 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.

b) Otherwise, we consider synthesizing a token as a third person personal pronoun only if it has already been synthesized (because it occurs in a previous clause template). In other words, we do not consider the left pronominalization phenomena[9]. Determining whether a token which has already been synthesized has to be synthesized as a pronoun requires the following steps to be gone through:

1° Compute the form of the foreseen pronoun. The form of a third person pronoun may depend on its syntactical position (subject, direct object, ...), on the number and human nature of the token (this semantic information is given in the token definition) and on the gender of the nominal phrase of the synthesis of the previous occurrence of the token. Gender in Italian is either masculine or feminine, and it is lexical and not

---

[9] In fact, the left pronominalization phenomena do rarely take place in a system of text generation, except in the synthesis of the first sentence, as in:
Each time he feels bad, Ugo is preoccupied.
where the pronoun *he* refers to *Ugo*, right-hand antecedent (see among others T.Reinhard (1883)). In the nth sentence, the left pronominalization is generally forbidden, as shown in the following example:
Max is feverish. Each time he feels bad, Ugo is preoccupied.
The pronoun *he* of the second sentence can only refer to *Max* (left-hand antecedent) and not to *Ugo* (right-hand antecedent). As our study is concerned with the synthesis of the nth (n > 1) Cl of a text template, we put aside left pronominalization phenomena.

semantic information. Let us consider the following definition of the token TABL1:

TABL1 =: TABLE
NUMBER: 1
DEFINITE: yes

It can be synthesized as a feminine nominal phrase: *la tavola* (the table) or as a masculine nominal phrase: *il tavolo* (the table). The gender of a pronoun is usually the same as the gender of the previous occurrence of the token:

(15)   *La tavola è rotta. Ugo la ripara.*
       Il tavolo è rotto. Ugo lo ripara.
       (The table is broken. Ugo repairs it)

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°. If L1 has only one element, go to step 5° with L3 = L1; otherwise:

3° Compute the sublist L2 of L1 that contains the elements that are semantically compatible with the foreseen pronoun. For the pronouns whose form indicates the human nature of the antecedent (eg. the subject pronoun *lui* indicates a human antecedent), the semantically compatible tokens are those with the right human nature. Moreover, the semantic features of each non human token of L1 may be checked on with regard to the relevant constraints of the verb. For example, in:

The book is on the table. It was published yesterday.

the subject of the verb *publish* used in the passive must be something *publishable*. This semantic information is not compatible with the token which represents *the table*, but only with the one which represents *the book*; the latter is thus the only element which is semantically compatible with the pronoun *it*. If L2 has only one element go to step 5° with L2 = L3; otherwise:

4° Compute the sublist L3 of L2 that contains the elements which are syntactically compatible with the foreseen pronoun. An example of coreferential syntactic incompatibility is the constraint of no-coreferentiality between a subject and a complement personal pronoun (see section 4.2.2). Another one is the following constraint:

If a personal 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 (L.Danlos 1988):*Maryᵢ wants that Maryⱼ leaves --> Maryᵢwants to leave.* An illustration of this constraint is that in *Mary wants that she leaves*, the pronoun *she* cannot refer to *Mary*.

5° As a first approach, if L3 contains one element synthesizing a pronoun is possible since this synthesis involves no ambiguity. Otherwise the foreseen pronoun is not synthesized. Counting the number of elements in L3 is not enough in determining the possibility of synthesizing or not a pronoun: pragmatical considerations, focus (C.Sidner 1981, B.Grosz, 1982 and parallelism (L.Danlos, 1987a) are phenomena that must be taken into account. They are not studied here.

As an illustration of these five steps, consider the following text template:

(16) (MISS1 *e* MISS2 *tornare*. MAN2 *dare un bacio a* MISS2.)
     (MISS1 and MISS2 be back. MAN2 give a kiss to MISS2)

The synthesis of the second occurrence of MISS2 as pronoun is attempted.
1° The form of the preverbal dative pronoun is *le*, third person feminine singular.
2° L1 contains the tokens which appear in the left-hand context that have been synthesized as feminine singular nominal phrases, i.e. L1 = (MISS1, MISS2).
3° All the elements of L1, which are humans, are semantically compatible with the foreseen pronoun, so L2=L1.
4° All the elements of L2 are syntactically compatible with the foreseen pronoun, so L3=L2.
5° L3 contains more than one element, so the pronoun is not synthesized. The resulting discourse will be:

Lia e Eva sono tornate. Ugo ha dato un bacio a Eva
(Lia and Eva are back. Ugo gave a kiss to Eva)

Another illustration is given by the following text template where TABL1 is supposed to be synthesized as the masculine nominal phrase *il tavolo*:

(17) (MAN1 *riparare* TABL1. MAN2 *dare un bacio a* MAN1.)
     (MAN1 repair TABL1. MAN2 give a kiss to MAN1.)

The synthesis of the second occurrence of MAN1 as pronoun is attempted:
1° The form of the preverbal dative pronoun is *gli*, third person masculine singular.
2° L1 contains the tokens of the left-hand context that have been synthesized as masculine singular nominal phrases, i.e. L1=(MAN1, TABL1, MAN2).
3° TABL1, which is not human, is semantically incompatible with the pronoun *gli* since the dative complement of *dare un bacio* must be human, hence L2=(MAN1,MAN2).
4° MAN2, which is the subject of the second Cl, syntactically incompatible with *gli* because of the constraint of no-coreferentiality between the subject and a complement personal pronoun. Hence L3=(MAN1). Since L3 contains only one element, the pronoun can be formed:

Max ha riparato il tavolo. Ugo gli ha dato un bacio.
(Max repaired the table. Ugo gave him a kiss)

- 230 -

5.2 Synthesizing and erasing an Italian subject pronoun

The synthesis of an Italian subject pronoun follows the operations described above, except that erasing the pronoun is attempted at the same time, as shown below:

1) A list L'1 is computed parallely to the computation of the list L1 (see step 2° of the section 5.1). L'1 contains the morphological antecedents of the foreseen erased pronoun. Two cases must be distinguished (see section 4.2.1):

a There is an element X in the clause which marks the subject gender (see example (11)). In this case, L'1 contains third person singular tokens of the same gender as this pronoun. In other words, L'1=L1.

b There is no element X which marks the gender of the pronoun whose erasing is foreseen (see example (12)). The morphological antecedents are then the tokens of the third person feminine and masculine singular (L'1 ⊃ L1).

If L'1 has only one element (this means that L1 has also only one element) go to step 4) with L'3=L'1; otherwise:

2) A sub-list L'2 of L'1 is computed parallely to the computation of the list L2 (step 3° of the section 5.1). L'2 contains the tokens which are semantically compatible to the foreseen erased subject. If L'2 (and hence L2) contains only one element, go to step 4) with L'2=L'3; otherwise:

3) If the list L3 (step 4° of the section 5.1) contains only one element[10], the sub-list L'3 of L'2 is computed. L'3 contains the tokens which are syntactically compatible to the foreseen erased subject. As shown in section 4.2.2, computing the list L'3 of the syntactic antecedents of the pronoun whose erasing is foreseen in a Cl depends on the synthesis of other tokens in Cl.

4) Pronoun erasing is usually allowed if list L'3 contains only one element.

6 Example of the synthesis of a clause template

Consider the following text template, where LOC1 is to be synthesized as the nominal phrase *il bosco* (the wood):

(18) (MAN1 *vedere* MISS1 *in* LOC1. MAN1 *abbracciare* MISS1)
(MAN1 see MISS1 in LOC1 . MAN1 kiss MISS1)

To begin with, suppose the verbs are conjugated in a compound tense (i.e. perfect). The synthesis of the first Cl is then:

Max ha visto Lia nel bosco.   (Max saw Lia in the wood)

---

10 If L3 contains more than one element, the subject token is not pronominalizable and thus not erasable.

Let us examine the synthesis of the second Cl. First, a partial synthesis of the subject MAN1 is carried out. Since MAN1 has already been mentioned, both synthesizing this token as a pronoun and erasing this pronoun are attempted.

6.1 First partial synthesis of the subject MAN1:

1) The form of the foreseen pronoun is *lui*, human masculine singular.

2) The list L1 contains the tokens which appear in the left-hand context that have been synthesized as masculine singular nominal phrases, i.e. L1=(MAN1,LOC1).

The list L'1 contains the tokens that have been synthesized as both masculine and feminine nominal phrases, since there is no element in the Cl which marks the subject gender; so L'1=(MAN1,LOC1,MISS1).

3) LOC1 is semantically incompatible with the pronoun *lui* which can have only human antecedents. So L2=(MAN1).
LOC1 is also semantically incompatible with an erased pronoun since the subject of the verb *abbracciare* must be human, so L'2=(MAN1,MISS1).

4) As L2=L3=(MAN1) contains only one element, the synthesis of the pronoun *lui* is possible. The computation of L'3 depends upon the synthesis of other elements of the Cl. Therefore the final synthesis of the subject (i.e. the decision to erase the pronoun *lui* according to the number of elements of L'3) is postponed.

6.2 First partial synthesis of the verb *abbracciare*

This verb is conjugated at the third person singular of the perfect tense with the auxiliary *avere*. In this first partial synthesis of the verb the possible agreement with a direct object is postponed. Thus the result of this partial synthesis is the form *ha abbracciato* where the past participle is in the masculine singular form which is the default value.

6.3 Synthesis of the direct object MISS1

The token MISS1 has been mentioned in the previous Cl, so synthesizing it as a personal pronoun is attempted:

1) Because of the conjugation of the verb, the feminine singular direct object pronoun *la* must be elided into *l'*.

2) The form *l'* does not mark the gender. However, the gender of the pronoun *la* will be marked in the past participle of the verb which is conjugated with the auxiliary *avere*. Therefore, L1 contains the tokens that have been synthesized as feminine singular nominal phrases, i.e. L1=(MISS1). Since L1 contains only one element, the pronoun *l'* is synthesized. Let us underline that the synthesis of the pronoun *l'* is based only upon morphological criteria and thus does not involve the constraint of no-coreferntiality between the subject and a complement. Therefore this constraint can be used for the second (and last) partial synthesis of the subject as shown in 6.5.

6.4 Second (and last) partial synthesis of the verb

Since the direct object pronoun *l'* has been synthesized, the past participle agrees in gender and number with this pronoun. The final result of the synthesis of the verb is: *ha abbracciata* where the past participle is in the feminine singular form.

6.5 Second (and last) partial synthesis of the subject

At this stage, the second Cl of (18) is foreseen to be synthesized as either *Lui l'ha abbracciata* or *L'ha abbracciata*. The last step to be carried out is the computation of the sub-list L'3 of L'2=(MAN1,MISS1) to determine if the subject pronoun can be erased. Since the direct object MISS1 has been synthesized as the pronoun *l'* only thanks to morphological criteria, the constraint of no-coreferentiality between a subject and a direct object can be used to state that MISS1 is a syntactically incompatible antecedent for the foreseen erased subject pronoun. So L'3 contains only one element: MAN1 and the subject pronoun can be erased. The synthesis of the second Cl of (18) is:

L'ha abbracciata.   (He kissed her)

Now, suppose that the verbs of (18) are conjugated in a simple tense (eg. present) and examine again the synthesis of the second Cl. The reader will check that the direct object MISS1 can be synthesized as the pronoun *l'* not thanks to morphological criteria (there is no past participle) but thanks to the constraint of no-coreferentiality. Therefore this constraint cannot be used again in computing L'3. Consequently L'3=L'2=(MAN1,MISS1) and the subject pronoun cannot be erased; the synthesis of this Cl is:

Lui l'abbraccia.   (He kisses her)

7 Future research

The sequential order of the operations for the synthesis of a Cl we have just described makes that the constraint of no-coreferentiality is called on as a priority for the synthesis of a complement, and if not used for any complement, it is called on for subject erasing. Our future research (L. Danlos, F. Namer, forthcoming) leads us to design a more global approach in which the constraint of no-coreferentiality is not called on as a priority for a complement. This approach will allow the second Cl of (18) (with the verb conjugated in the present) to be synthesized not only as *Lui l'abbraccia* but also as

Abbraccia la ragazza.   (He kisses the girl)

where the subject is erased and the direct object not pronominalized because the constraint of no-coreferentiality is used for the subject and not for the complement.

REFERENCES

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. Dortrecht/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.

Danlos, L., 1988, Some Pronominalization Issues in Generation of Texts in Romance Languages, *Electronic Dictionaries and Automata in Computational Linguistics*, D. Perrin Ed., Springler-Verlag publications, Berlin.

Danlos, L., Namer, F., forthcoming, A Global Approach for the Synthesis of a Personal Pronoun *Computers and Translation.*

Grosz, B., 1982, Focusing and Description in Natural Language Dialogues, *Elements of Discourse Understanding*, Cambridge University Press, Cambridge

Reinhart, T., 1983, *Anaphora and Semantic Interpretation* , Croom Helm, London.

Sidner, C., 1981, Focusing for Interpretation of Pronoun, *American Journal of Computational Linguistics*, vol. 7, no 4.

Rizzi, L., 1982, *Issues in Italian Syntax* , Foris publications, Dortrecht/Cinnaminson.