Université Paris VII

1990
NAM

PRONOMINALISATION ET EFFACEMENT DU SUJET EN GENERATION AUTOMATIQUE DE TEXTES EN LANGUES ROMANES

Fiammetta Namer

Thèse de doctorat en informatique-linguistique
Directeur de thèse : Laurence DANLOS
Mai 1990

UNIVERSITE de MARNE-la-VALLEE ★ BIBLIOTHEQUE ★

Jury :

A. ELIA
L. DANLOS
M. GROSS
M. NIVAT
G. SABAH

UNIVERSITÉ de MARNE - la - VALLÉE * BIBLIOTHEQUE *

Remerciements

Je remercie Annibale Elia, Maurice Gross, Maurice Nivat, Gérard Sabah qui m'ont fait l'honneur de composer mon jury de thèse et de s'intéresser à ce travail.

Je remercie Anne Abeillé, Fulvia Airoldi Namer, Christian Leclère, Maurice Nivat et Patrice Ringot pour leur relecture attentive, leurs conseils et leurs encouragements.

Je remercie Maurice Gross qui a bien voulu m'accueillir au CERIL.

Je dois une reconnaissance particulière aux membres d'Eurotra avec qui j'ai eu de nombreux et fructueux échanges scientifiques, en particulier : Sergio Balari et Nuria Bel (Barcelonne), Mounira Loughraïeb (Nancy), Erich Steiner (Sarrebrück), Béatrice Daille, Anna Firenze, Anne Guyon, Tita Kyriacopoulou, Eric Laporte, Amid Nehmé et Pollet Samvilian (Paris).

Je tiens enfin à adresser mes plus vifs remerciements à Laurence Danlos qui m'a guidée dans ce travail. Je lui suis infiniment reconnaissante pour sa grande disponibilité à mon égard, ses conseils avisés et ses encouragements répétés. Qu'elle trouve ici l'expression de ma profonde gratitude.

TABLE DES MATIERES

TABLE DES MATIERES 1

1 PRESENTATION GENERALE 1

2 SCHEMA DE TEXTE 11

2.1 INTRODUCTION 11
2.2 REPRESENTATION DES CONNAISSANCES 13
2.3 COMPOSANT STRATEGIQUE 15
2.3.1 DE LA REPRESENTATION SEMANTIQUE AU SCHEMA DE TEXTE 15
2.3.1.1 CHOIX LEXICAUX 16
2.3.1.2 LINEARISATION EN PHRASES 17
2.3.1.3 GRAMMAIRE DE DISCOURS ET LEXIQUE-GRAMMAIRE 18
2.3.2 ORDRE DES DECISIONS 20
2.3.3 COMPLETIVES 21
2.4 SCHEMA DE TEXTE 24
2.5 CONCLUSION 27

3 PRONOMINALISATION 29

3.1 INTRODUCTION 29
3.1.1 FORMES REFLEXIVES 31
3.1.2 DESCRIPTION LINGUISTIQUE DES PRONOMS PERSONNELS 35
3.1.3 DISTINCTION SUJET ET OBJET 40
3.2 SYNTHESE DES PRONOMS PERSONNELS OBJETS 40
3.2.1 DEPENDANCES NON LOCALES ET DEPENDANCES CROISEES 41
3.2.2 PREMIERES OPERATIONS A EFFECTUER POUR LA SYNTHESE D'UN PRONOM 44
3.2.2.1 CALCUL DE LA FORME DU PRONOM 47
3.2.2.2 ANTECEDENTS DE LA FORME CHOISIE 56
3.2.2.2.1 ANTECEDENTS MORPHOLOGIQUES DU PRONOM 57
3.2.2.2.2 ANTECEDENTS SEMANTIQUES DU PRONOM 57
3.2.2.2.3 ANTECEDENTS SYNTAXIQUES DU PRONOM 58
3.2.2.2.4 AUTRES OPERATIONS 58
3.3 EFFACEMENT DU SUJET 60
3.3.1 LIEN ENTRE PRONOMINALISATION ET EFFACEMENT DU SUJET 60
3.3.2 EFFACEMENT D'UN PRONOM SUJET 62
3.3.2.1 PRONOM SUJET 62
3.3.2.2 RECUPERATION DE LA NATURE SEMANTIQUE DU PRONOM EFFACE 64
3.3.2.3 RECUPERATION DU GENRE DU PRONOM EFFACE 65
3.4 CONCLUSION 67

4 CONTRAINTES DE NON COREFERENCE ET ALGORITHME DE SYNTHESE D'UNE PHRASE 71

4.1 CONTRAINTES DE NON COREFERENCE 71
4.1.1 CONTRAINTE DE COREFERENCE ENTRE SUJET ET PRONOM OBJET 71
4.1.2 CONTRAINTE DE NON COREFERENCE ENTRE DEUX OBJETS 72
4.1.3 CONTRAINTE DE NON COREFERENCE ENTRE OBJET ET SUJET EFFACE 73
4.1.4 CONSEQUENCE : ELABORATION D'UN ALGORITHME 74
4.2 L'APPROCHE GLOBALE 80
4.2.1 TOUTES LES FORMES ENVISAGEABLES POUR UNE ETIQUETTE 81
4.2.2 TOUTES LES VERSIONS POSSIBLES POUR UN SCHEMA DE PHRASE 83
4.3 CONCLUSION 89

5 CONSIDERATIONS INTERPHRASTIQUES DE PRONOMINALISATION : CHOIX D'UN ALGORITHME 93
5.1 PARALLELISME DE STRUCTURE 93
5.1.1 PRESENTATION DU PROBLEME 93
5.1.2 PARALLELISME 95
5.1.2.1 CRITERES 95
5.1.2.2 CONSEQUENCES DU PARALLELISME DE STRUCTURE 101
5.1.3 COREFERENCE DES SUJETS 105
5.1.3.1 ETUDE DES DIFFERENTS CAS 106
5.1.3.2 CONSEQUENCES 108
5.1.4 ORIENTATION DES RECHERCHES 109
5.2 DISTANCE 112
5.2.1 DISTANCE LINEAIRE 112
5.2.2 DISTANCE ET PARALLELISME 114
5.3 INTEGRATION DANS L'ALGORITHME DES CONDITIONS INTERPHRASTIQUES 115
5.3.1 QUAND APPLIQUER LES REGLES DE PREFERENCE INTERPHRASTIQUES ? 116
5.3.2 INTEGRATION DES REGLES DE PREFERENCE INTERPHRASTIQUES DANS LES DEUX APPROCHES 118
5.4 CONCLUSION : AVANTAGES DE L'APPROCHE GLOBALE 120

6 ARGUMENTS COMPLEXES DE LA PHRASE : SYNTHESE DES COMPLETIVES 133
6.1 INTRODUCTION 133
6.2 DESCRIPTION D'UNE COMPLETIVE 135
6.2.1 NOMBRE DE COMPLETIVES 135
6.2.2 MODE DE LA COMPLETIVE ET SUBSTANTIFS QUI LA REGISSENT 135
6.2.3 REDUCTION DES COMPLETIVES A L'INFINITIVE 137
6.3 CONTRAINTES DE NON COREFERENCE DANS LES PHRASES A COMPLETIVES 140
6.3.1 COMPLETIVES NON REDUITES A L'INFINITIVE DONT LE SUJET EST UNE FORME PRONOMINALE 140
6.3.2 COMPLETIVES REDUITES A L'INFINITIVE 141
6.4 SYNTHESE D'UNE COMPLETIVE 143
6.4.1 APPLICATION RECURSIVE DE L'APPROCHE GLOBALE 143
6.4.2 GESTION DE L'HISTORIQUE DES ETIQUETTES 148
6.4.3 AUTRES CONSEQUENCES DE LA SYNTHESE DES COMPLETIVES 150
6.5 CONCLUSION : ADEQUATION DE L'APPROCHE GLOBALE RECURSIVE 151

7 PROGRAMME 155
7.1 INTRODUCTION 155
7.2 REPRESENTATION SEMANTIQUE NECESSAIRE AU DOMAINE 156
7.3 OBTENTION D'UN SCHEMA DE TEXTE 160
7.3.1 TEXTE 161
7.3.2 EVENEMENT INDEPENDANT 163
7.3.2.1 MOYEN 163
7.3.2.2 CONSEQUENCE 164
7.3.2.3 CAUSE 165
7.3.2.4 BUT 166
7.3.3 EVENEMENT DE BASE 167
7.3.4 CONCLUSION : EXEMPLE DE SCHEMA DE TEXTE 172
7.4 EXEMPLES DE TEXTES GENERES 174
7.4.1 DON CARLOS 174
7.4.2 UN BALLO IN MASCHERA 180
7.4.3 I VESPRI SICILIANI 181
7.5 CONCLUSION 182

CONCLUSION GENERALE 185

BIBLIOGRAPHIE 191

1 PRESENTATION GENERALE

Dans un système de dialogue entre l'homme et la machine en langage naturel, on peut distinguer trois modules principaux :

- un **analyseur automatique,** qui transforme la question de l'utilisateur en une formule compréhensible par la machine ;
- un **module de raisonnement formel,** qui effectue des calculs à partir de cette formule et produit une représentation sémantique de la réponse à fournir ;
- **un générateur,** qui traduit en langage naturel la représentation formelle ainsi obtenue.

De façon similaire, on distingue trois modules dans un système ayant pour but la production de résumés de textes :

- l'analyseur rend le texte de départ compréhensible pour la machine ;
- le module de raisonnement extrait les informations essentielles de la représentation de départ ;
- le générateur synthétise un résumé du texte original à partir de cette nouvelle représentation.

La production de textes en langage naturel est un domaine qui a suscité un intérêt assez tardif chez les chercheurs, contrairement à l'analyse automatique. Les recherches en génération ont été développées depuis une quinzaine d'années. Le module de raisonnement formel, qui répond à la question "Que dire ?", est parfois considéré comme faisant partie de la génération, au même titre que le générateur proprement dit, qui répond lui à la question "Comment le dire ?". Mais c'est seulement cette dernière question qui nous intéresse : le travail que nous décrirons ici concerne uniquement la traduction en langage naturel d'une représentation sémantique.

Les systèmes de génération automatique qui ont été développés à ce jour sont souvent découpés en plusieurs modules :

- le **premier** module prend des décisions de "haut niveau" (i.e. les décisions conceptuelles) ;
- le **second**, les décisions linguistiques (i.e. choix lexicaux et choix des constructions syntaxiques). Il produit un "schéma de texte" qui constitue un squelette du texte à synthétiser. La forme et le nom donné à ce schéma de texte diffèrent d'un système à l'autre. Nous en présentons certains exemples au chapitre suivant ;
- le **troisième** effectue les opérations syntaxiques ;
- le **quatrième** s'occupe des opérations de "bas niveau", c'est à dire les calculs morphologiques (accord des verbes, élision ...).

L. Danlos (1985, 1987) a montré que les deux premiers modules étaient dépendants l'un de l'autre, et a donc élaboré un modèle de génération de textes constitué de deux composants :

- le composant stratégique qui regroupe les deux premiers modules, c'est à dire qui prend les décisions conceptuelles et linguistiques,

- le composant syntaxique qui regroupe les deux derniers modules, c'est à dire qui effectue les opérations syntaxiques et morphologiques.

Ces deux composants s'articulent autour d'un schéma de texte, produit par le composant stratégique et transformé en texte par le composant syntaxique. C'est de ce modèle (décrit en détail au chapitre 2) que nous nous sommes inspirée. Nous nous sommes intéressée surtout à la traduction d'un schéma de texte en un texte en langage naturel, c'est à dire à l'étude du composant syntaxique. Ce travail a été effectué pour le français et l'italien. Il est illustré par un programme qui génère des textes en italien.

Un schéma de texte contient une suite de représentations schématiques de phrases. Par exemple, (ST1) est un schéma de texte contenant deux schémas de phrases juxtaposées et dont une version simplifiée est :

(ST1) (MAX *amare* PIA. UGO *amare* EVA.)
(MAX aimer PIA. UGO aimer EVA.)

A partir de (ST1), un seul texte peut être généré, en supposant que les verbes sont conjugués au présent :

(1) *Max ama Pia. Ugo ama Eva.*
(Max aime Pia. Ugo aime Eva.)

Mais la synthèse du $i^{ème}$ schéma de phrase ($i > 1$) peut, en fonction du contexte, (i.e. des i-1 phrases précédentes) déboucher sur un certain nombre de phrases. Ainsi, examinons le schéma de texte (ST2) dont le deuxième schéma de phrase est égal à celui de (ST1) :

(ST2) *(MAX odiare EVA. UGO amare EVA.)*
(MAX détester EVA. UGO aimer EVA.)

La synthèse du deuxième schéma de phrase donne lieu ici à plusieurs possibilités :

(2a) *Max odia Eva. Ugo ama questa donna.*
(Max déteste Eva. Ugo aime cette femme.)
(2b) *Max odia Eva. Ugo l'ama.*
(Max déteste Eva. Ugo l'aime.)

Il en va de même pour la synthèse du deuxième schéma de phrase de (ST3) (qui diffère de ST2 et ST1 par son premier schéma de phrase uniquement) :

(ST3) *(UGO odiare PIA. UGO amare EVA.)*
(UGO détester PIA. UGO aimer EVA.)

(3a) *Ugo odia Pia. Ama Eva.*
(Ugo déteste Pia. [Ileffacé] aime Eva.)
(3b) *Ugo odia Pia. Egli ama Eva.*
(Ugo déteste Pia. Il aime Eva.)
(3c) *Ugo odia Pia. Quest'uomo ama Eva.*
(Ugo déteste Pia. Cet homme aime Eva.)

Citons encore la synthèse du deuxième schéma de phrase de (ST4) (on suppose que le verbe de la première phrase est conjugué au passé composé) :

(ST4) *(UGO sposare EVA. UGO amare EVA.)*
(UGO épouser EVA. UGO aimer EVA.)

(4a) *Ugo ha sposato Eva. L'ama.*
(Ugo a épousé Eva. [Ileffacé] l'aime.)

(4b) *Ugo ha sposato Eva. Egli l'ama.*
(Ugo a épousé Eva. Il l'aime.)

(4c) *Ugo ha sposato Eva. Quest'uomo l'ama.*
(Ugo a épousé Eva. Cet homme l'aime.)

(4d) *Ugo ha sposato Eva. Ama questa donna.*
(Ugo a épousé Eva. [Ileffacé] aime cette femme.)

(4e) *Ugo ha sposato Eva. Egli ama questa donna.*
(Ugo a épousé Eva. Il aime cette femme.)

(4f) *Ugo ha sposato Eva. Quest'uomo ama questa donna.*
(Ugo a épousé Eva. Cet homme aime cette femme.)

Indépendamment du contexte, il est possible à priori d'obtenir en tout douze versions du deuxième schéma de phrase de STi ($1 \leq i \leq 4$) : *UGO amare EVA*. Ces douze versions diffèrent par la façon dont ont été synthétisés les personnages symbolisés par UGO et EVA dans le schéma de phrase : ceux-ci sont synthétisés soit en groupes nominaux (e.g. *Eva*, ou *questa donna* pour EVA), soit en pronoms personnels (e.g. *egli* pour UGO), soit en élément vide (e.g. le sujet dans la deuxième phrase de l'exemple (4d)). Mais si l'on tient compte du contexte (e.g. *Max ama Pia, Max odia Eva, Ugo odia Pia, Ugo ha sposato Eva*), seul un sous-ensemble de ces douze versions sont acceptables. Ces exemples montrent que la production de textes cohérents, bien formés, précis et non ambigus[1] nécessite une étude poussée des phénomènes de pronominalisation et d'anaphore nominale. De plus, elle nécessite, pour l'italien, celle de l'effacement du sujet. Nous avons consacré la plus grande partie de notre travail à ces questions. La synthèse des pronoms, ainsi que l'effacement du sujet, sont des problèmes de base pour la génération de texte, comme la résolution d'anaphore l'est pour l'analyse. La synthèse des pronoms en génération et la résolution d'anaphores en analyse ne se posent pas exactement dans les mêmes termes, mais nous verrons que la synthèse des pronoms nécessite une certaine forme de résolution anaphorique. Les résultats auxquels nous avons abouti en génération pourraient donc être utilisés en analyse.

Les phénomènes d'anaphore sont importants non seulement du fait de leur omniprésence dans un texte, mais également à cause de leur rôle prépondérant dans l'algorithme de synthèse. Nous nous sommes intéressée plus particulièrement à l'étude de la synthèse des pronoms personnels, ce que nous désignerons à partir de maintenant par le terme de **pronominalisation**. Cette étude peut être divisée en trois points :

- Rôle de la pronominalisation dans l'algorithme de génération de texte,
- Conditions de pronominalisation,

1 La production de textes non ambigus est l'un de nos principaux soucis, contrairement à certaines démarches en génération de textes. Dans la mesure où il est possible d'éviter une ambiguïté, nous considérons qu'en introduire ou en laisser une équivaut à une perte inutile d'information.

- Intégration des conditions de pronominalisation dans l'algorithme de génération de textes.

Ces trois parties, que nous introduisons ici, seront développées tout au long de notre travail :

A) Rôle de la pronominalisation dans l'algorithme de génération de texte :

Nous mettrons en évidence que la pronominalisation intervient au moins à quatre niveaux dans la conception d'un système de génération de texte :

1) la pronominalisation joue un rôle prépondérant dans l'ordre de synthèse des éléments d'une phrase,

2) l'effacement du sujet en italien joue aussi un rôle capital dans l'ordre de synthèse des éléments d'une phrase,

3) la pronominalisation établit une dépendance entre les opérations morphologiques et syntaxiques,

4) la pronominalisation établit aussi une dépendance entre le composant stratégique et le composant syntaxique.

Le contenu de chacun de ces points est résumé ci-dessous :

1) Il est admis, dans certains systèmes, que la synthèse d'une phrase puisse être effectuée de gauche à droite. Or, si cette méthode est éventuellement envisageable pour la synthèse de l'anglais, elle aboutit régulièrement à des résultats inacceptables dans le cas des langues romanes. Considérons, par exemple, la séquence [sujet] [verbe] [objet-direct], où le [verbe] est conjugué au passé composé et où l'[objet direct] est pronominalisé : la synthèse dans l'ordre [sujet] < [verbe] < [objet-direct] a toutes les chances de conduire à la phrase (5) :

(5) **Ugo li ha visto.* [2]
(*Ugo les$_{masc}$ a vu.)

qui est inacceptable, puisque le participe passé devrait s'accorder avec le pronom masculin-pluriel objet direct *li*. Plus généralement, la pronominalisation met en jeu de nombreuses dépendances entre les éléments d'une phrase, avec pour conséquence l'élaboration d'un algorithme de synthèse de phrase qui n'a rien à voir avec un algorithme linéaire (synthèse de gauche à droite) ; en effet, nous verrons :

- que la synthèse d'un élément d'une phrase peut dépendre de celle d'un autre élément de la même phrase (dépendance non locale),

- que les synthèses de deux éléments d'une même phrase peuvent dépendre l'une de l'autre (dépendances croisées).

Ce sont ces dépendances non locales et croisées qui déterminent l'ordre de synthèse des éléments d'une phrase.

2 L'étoile (*) placée devant un texte indique l'inacceptabilité de ce texte, soit au niveau formel (i.e. grammatical), soit parce qu'il est ambigu. De la même façon, les symboles (*?) placés devant un texte indiquent que ce texte est d'une acceptabilité douteuse. Enfin, le signe (?) indique que l'acceptabilité du texte est litigieuse.

2) Cette étude concerne les langues romanes, plus spécifiquement le français et l'italien ; une attention particulière a été portée à un phénomène que l'on rencontre en italien, en espagnol, en portugais et en roumain, mais qui n'existe pas en français : l'effacement du sujet dans les phrases à verbe conjugué, dont l'exemple (6) est une illustration[3]

(6) *Vado al cinema.*
([Je$_{effacé}$] vais au cinéma.)

Nous montrerons que l'effacement du sujet est un phénomène lié à la pronominalisation, et qu'il met en jeu lui aussi de nombreuses dépendances non locales et croisées. Les résultats ont été établis sur l'italien, mais on peut s'attendre à des résultats similaires pour l'espagnol, le portugais et le roumain.

3) La morphologie joue un rôle dans la synthèse des pronoms personnels, ce qui établit une dépendance entre les opérations syntaxiques et morphologiques. Cette dépendance implique que opérations morphologiques doivent être prises en compte rapidement dans un système de génération, et non pas en dernière étape (comme le suggère le terme de "bas niveau"), comme c'est le cas dans les systèmes composés de quatre modules. Observons par exemple comment intervient l'élision du pronom féminin-singulier *la* en *l'* dans la deuxième phrase des textes (7) et (8), qui sont conjugués respectivement au passé composé et au passé simple :

(7) *Ugo a amené Eva à la maison à 10 heures. Je l'ai accueillie avec enthousiasme.*
(8) **Ugo amena Eva à la maison à 10 heures. Je l'accueillis avec enthousiasme.*

Dans (7), l'élision n'a pas d'incidence sur la compréhension du texte écrit[4] : le participe passé (*accueillie*), conjugué avec *avoir*, s'accorde en genre et en nombre avec le pronom objet direct *l'* placé en position préverbale : il permet donc de calculer la forme de base de ce pronom élidé. Puisque le participe marque le féminin-singulier, on sait que *l'* provient du féminin-singulier *la*. Le pronom *l'* n'est donc pas ambigu. Le texte (8) par contre est ambigu, et la synthèse du pronom est interdite : l'élision rend la pronominalisation impossible. En d'autres termes, la pronominalisation, qui met en jeu des opérations morphologiques et des opérations syntaxiques, fait que ces deux types d'opérations sont dépendantes.

4) La pronominalisation remet en question le cloisonnement existant entre les opérations conceptuelles et linguistiques d'une part, et les opérations morpho-syntaxiques de l'autre. Récapitulons en effet les problèmes de modularité : les systèmes de génération sont souvent découpés en quatre modules :

- le premier prend les décisions conceptuelles,
- le second, les décisions linguistiques,
- le troisième effectue les opérations syntaxiques,

3 La notion d'effacement du sujet ne concerne pas le sujet effacé dans les coordinations, phénomène qui existe lui, bien sûr, aussi en français.

4 Nous nous intéressons ici aux textes écrits uniquement. Il est évident que (7) et (8) présentent le même niveau d'ambiguïté à l'oral : la production de messages oraux à partir de représentations sémantiques (L. Danlos, F. Emerard & C. Sorin (1985) et L. Danlos, F. Emerard & E. Laporte (1986)) nécessite la prise en compte de la phonétique pour la résolution des ambiguïtés.

- le quatrième effectue les opérations morphologiques.

L. Danlos a montré que les deux premiers modules dépendent l'un de l'autre : elle en fait un seul composant, le composant stratégique. Par ailleurs, la pronominalisation est responsable de la dépendance entre le troisième et le quatrième module : les opérations syntaxiques et morphologiques sont interdépendantes, et donc regroupées dans le composant syntaxique. Donc, à ce stade, il ne reste que deux modules : le composant stratégique et le composant syntaxique. Et c'est l'indépendance de ces deux composants qui est mise en cause. En effet, la structure du schéma de texte fourni par le composant stratégique peut être guidée par la pronominalisation. Considérons les exemples suivants :

(9a) **Ugo e Max sono gemelli. Eva lo confonde con lui.*
(*Ugo et Max sont jumeaux. Eva le confond avec lui.)
(9b) --> *Ugo e Max sono gemelli. Eva li confonde.*
(Ugo et Max sont jumeaux. Eva les confond.)

(10a) *?Ugo è riuscito a (Ø + far sì) che Max stringa amicizia con lui.*
(?Ugo a réussi à ce que Max lui donne son amitié.)
(10b) --> *Ugo è riuscito ad ottenere l'amicizia di Max.*
(Ugo a réussi à obtenir l'amitié de Max.)

Dans (9a), la double pronominalisation de la deuxième phrase rend le texte inacceptable, et implique de modifier la structure des arguments du verbe *confondere*, de façon à obtenir (9b). Dans (10a), il est souhaitable de modifier les choix lexicaux du complément phrastique du verbe *riuscire* pour alléger le texte qui est stylistiquement désagréable, afin d'obtenir (10b). Ces deux exemples reflètent un fait courant pour toute personne écrivant un texte : lorsqu'une phrase manuscrite ne nous plaît pas, et que tous les changements d'ordre local ont été effectués en vain, on décide alors de "tourner" la phrase autrement, ce qui a parfois pour conséquence le choix de nouveaux items lexicaux ou de nouvelles constructions syntaxiques. Cette attitude humaine équivaut, dans un système automatique, à un "retour en arrière" ("backtracking"). Or, chez l'humain, le retour en arrière est assez rare. Dans un système de génération à deux composants distincts, il risque d'être plus fréquent. L'idéal serait donc un système de génération conçu d'un seul bloc, avec seulement quelques rares opérations de "backtracking". Mais nous n'avons pas la prétention de construire un générateur qui prenne toutes les décisions en même temps. Il n'est même pas certain qu'un tel système soit concevable actuellement. Nous gardons donc le principe de deux modules distincts, en apportant quelques conditions, en rapport avec la pronominalisation, pour la construction des schémas de textes.

B) Conditions de pronominalisation :

La synthèse des pronoms personnels repose sur un certain nombre de contraintes, qui sont, comme nous le verrons, d'ordre morphologique, sémantique, syntaxique, stylistique et conceptuel. Ces différentes contraintes ne doivent pas être prises en compte au même niveau, et on peut les classer selon trois paramètres. En effet, elles n'ont pas le même rôle, ni le même mode d'application, ni le même degré de certitude. Ainsi, la morphologie a pour rôle le calcul d'une forme pronominale, alors que les règles stylistiques donnent un jugement sur l'opportunité de synthétiser cette forme ; de même, la syntaxe a un mode d'application intraphrastique, à l'inverse du mode d'application (interphrastique) des règles basées sur le contexte. Enfin, le degré de fiabilité des règles sémantiques ou morphologiques est supérieur à celui des règles stylistiques ou contextuelles, qui reposent sur des idées subjectives. Ces différentes contraintes doivent être appliquées dans un certain ordre qu'il faudra déterminer en fonction de ces trois paramètres.

C) Intégration des conditions de pronominalisation dans l'algorithme de génération de textes :

Les points A et B conduisent respectivement aux conclusions suivantes :

- la pronominalisation joue un rôle dans l'**ordre** de synthèse des éléments d'une phrase,
- la synthèse d'un pronom est le résultat d'un certain nombre d'opérations qu'il faut effectuer dans un **ordre** donné.

L'intégration de ces deux points doit donc aboutir à l'élaboration d'un algorithme de synthèse qui tienne compte de chacun des deux ordres rappelés ci-dessus.

La solution qui vient à l'idée en premier consiste en une "approche séquentielle" : celle-ci revient tout simplement à synthétiser les éléments d'une phrase dans l'ordre établi au point A, en respectant pour chaque élément dont la pronominalisation est envisagée l'ordre établi au point B. On s'apercevra que cette approche présente un certain nombre d'inconvénients. Par exemple, les conditions de synthèse du $i^{ème}$ élément d'une phrase ne prennent pas en compte les conditions de synthèse du $i+k^{ème}$ élément. D'autre part, cette approche ne permet pas d'utiliser les éléments lexicaux résultant de la synthèse d'un élément pour engendrer les autres éléments.

Nous avons donc imaginé une autre approche : "l'approche globale". Comme son nom l'indique, elle envisage la synthèse de tous les éléments de la phrase en même temps, et pour chacun d'eux elle propose toutes les formes (pro)nominales possible, ce qui a pour résultat la production de plusieurs versions d'un même schéma de phrase. En d'autres termes, plutôt que d'ordonner les différentes décisions, cette approche opère un choix final entre les différentes formes. Ce choix final, ainsi que le fait de manipuler en même temps tous les éléments de la phrase présente de nombreux avantages que nous décrirons dans la suite. L'un d'eux, que nous présentons ici brièvement, est la possibilité d'utiliser le contexte droit (i.e. les phrases non encore synthétisées) dans le choix d'une version d'une phrase. Examinons en effet le schéma de texte (ST11), où les verbes sont conjugués au passé composé :

(ST11) *(MAX e UGO far amicizia : EVA presentare MAX a UGO. LIA presentare poi MAX a FAMIGLIA. ...)*
(MAX et UGO faire connaissance : EVA présenter MAX à UGO. LIA présenter ensuite MAX à FAMIGLIA. ...)

On suppose que MAX est un vieillard appelé *Max*, et que UGO est un jeune homme appelé *Ugo*. Le premier schéma de phrase est synthétisé en :

Max e Ugo hanno fatto amicizia :
(Max et Ugo ont fait connaissance :)

L'approche globale permet de produire deux versions du deuxième schéma de phrase :

(a) *Eva lo ha presentato al ragazzo.*
(Eva l'a présenté au jeune homme.)

(b) *Eva gli ha presentato il vecchietto.*
(Eva lui a présenté le vieillard.)

à l'inverse de l'approche séquentielle, qui ne permet d'obtenir que (a) ou (b).

La décision de retenir la version (a) ou (b) est déterminée, dans l'approche globale, par la structure et le contenu du troisième schéma de phrase. Ceux-ci imposent le choix de (a), ce qui conduit à :

(11a) *Max e Ugo hanno fatto amicizia : Eva lo ha presentato al ragazzo. Lia lo ha poi presentato al resto della famiglia....*
(Max et Ugo ont fait connaissance : Eva l'a présenté au jeune homme. Lia l'a ensuite présenté au reste de la famille ...)

alors que la version (b) aurait débouché sur (11b), qui est stylistiquement désagréable :

(11b) *Max e Ugo hanno fatto amicizia : Eva gli ha presentato il vecchietto. Lia ha poi presentato (Max + il vecchietto) al resto della famiglia....*
(Max et Ugo ont fait connaissance : Eva lui a présenté le vieillard. Lia a ensuite présenté (Max + le vieillard) au reste de la famille ...)

étant donné que la pronominalisation de MAX dans la dernière phrase engendrerait une ambiguïté :

(11b') **Max e Ugo hanno fatto amicizia : Eva gli ha presentato il vecchietto. Lia lo ha poi presentato al resto della famiglia....*
(Max et Ugo ont fait connaissance : Eva lui a présenté le vieillard. Lia l'a ensuite présenté au reste de la famille ...)

Nous voyons ici l'un des avantages de l'approche globale sur l'approche séquentielle, qui ne permet pas de bénéficier du choix entre (a) et (b) : il est clair que dans l'approche séquentielle, on aboutira à l'une des versions (11a) ou (11b) indépendamment des contraintes linguistiques imposées par le troisième schéma de phrase.

Rôle de la pronominalisation, mécanisme de la pronominalisation et élaboration d'un algorithme guidé par les phénomènes de pronominalisation, sont donc les trois principaux points qui seront évoqués tout au long de notre travail. Plus précisément, voici l'ordre dans lequel ces différentes parties seront présentées ; comme nous le verrons, cet ordre répond à des raisons absolument indépendantes de l'ordre d'application des contraintes qui sont traitées :

- dans le chapitre 3, après une description linguistique des phénomènes de pronominalisation et d'effacement du sujet, nous décrivons les opérations morphologiques et sémantiques liées à la synthèse des pronoms, et les dépendances qui sont mises en jeu ;

- les opérations syntaxiques qui interviennent dans la synthèse des pronoms et dans l'effacement du sujet sont décrites dans le chapitre 4. Les algorithmes de synthèse correspondant à l'approche séquentielle et à l'approche globale seront alors présentés ;

- nous abordons ensuite, au chapitre 5, des règles de préférence liées au contexte qui interviennent dans la décision de former un pronom, donc dans l'algorithme : l'étude du parallélisme de structure, des conditions de coréférence obligatoire des sujets et le calcul de la distance entre deux entités. Nous conclurons ce chapitre en comparant les deux approches, et en exposant nos arguments en faveur de l'approche globale ;

- enfin, au chapitre 6, nous présentons les travaux auxquels nous nous consacrons actuellement, et qui concernent la synthèse des constituants complexes d'une phrase (ces

constituants sont les arguments phrastiques d'un verbe, les relatives et les compléments du nom). Nous nous limiterons à la synthèse des arguments phrastiques dont nous décrivons les propriétés et le comportement, pour en étudier ensuite l'intégration dans l'algorithme de synthèse. Nous apporterons ainsi un nouvel argument en faveur de l'approche globale.

Ces différents points ont été testés par un programme de génération, qui produit des résumés d'opéras en italien : dans le dernier chapitre, nous décrirons des textes issus du générateur, et nous en dégagerons les faits pertinents du point de vue de la syntaxe. Nous y présenterons également également des récits d'attentats en italien, résultant de l'adaptation du composant stratégique de L. Danlos à notre générateur.

Mais avant tout, il est nécessaire de présenter le modèle de génération de L. Danlos, dont nous nous sommes inspirée. Le prochain chapitre est par conséquent consacré aux étapes débouchant sur la formation d'un schéma de texte. La syntaxe d'un schéma de texte italien est le point de départ de notre travail. C'est pourquoi nous en proposerons la description détaillée.

2 SCHEMA DE TEXTE

2.1 INTRODUCTION

Les systèmes de génération, dont le but est de traduire un message en un texte bien formé ont tous en commun la production, tôt ou tard, d'un schéma de texte. Ce schéma de texte est appelé de différentes manières et peut prendre des formes diverses ; dans tous les cas cependant, il contient le message à traduire et la forme globale du futur texte. Ainsi, dans le système de K. McKeown (1985) qui produit des réponses à des questions posées pour interroger une base de données, le programme construit des propositions à partir de schémas propres au domaine concerné. Ces propositions sont ensuite transformées, grâce à un ensemble de règles, en schémas de textes, appelées **structures profondes**, qui déterminent la voix du verbe, la nature des déterminants (i.e. sont-ils définis ou indéfinis), etc ... en attribuant à chaque élément de cette structure une fonction thématique donnée. Ainsi, la phrase :

La maison a été abandonnée par Max.

peut provenir de la structure profonde suivante :

```
[(verbe          ((v = abandonner)))
 (AGENT          ((n = Max))
 (OBJET          ((article = défini)
                  (n = maison)))
 (FOCUS = OBJET)]
```

D. McDonald (1983) a une approche similaire à celle de K. McKeown, mais les schémas de textes, qui seront transformés en textes par un composant syntaxique, apparaissent sous forme **d'arbres syntagmatiques**. Plus récemment, le programme GESTER de L. Pemberton (1989) génère des récits épiques : il produit un **niveau de discours**, qui est une forme élaborée de schéma de texte (e.g. les verbes sont conjugués) mais qui ne peut pas être considéré comme un texte (e.g. aucune opération de pronominalisation n'est effectuée) : le passage du niveau du discours au niveau textuel est assuré par un ensemble de règles de lexicalisation. Voici un extrait du niveau de discours produit par GESTER :

a) *Charles and Aymeri broke into Narbonne.*

b) *Charles succeded in getting Narbonne.*

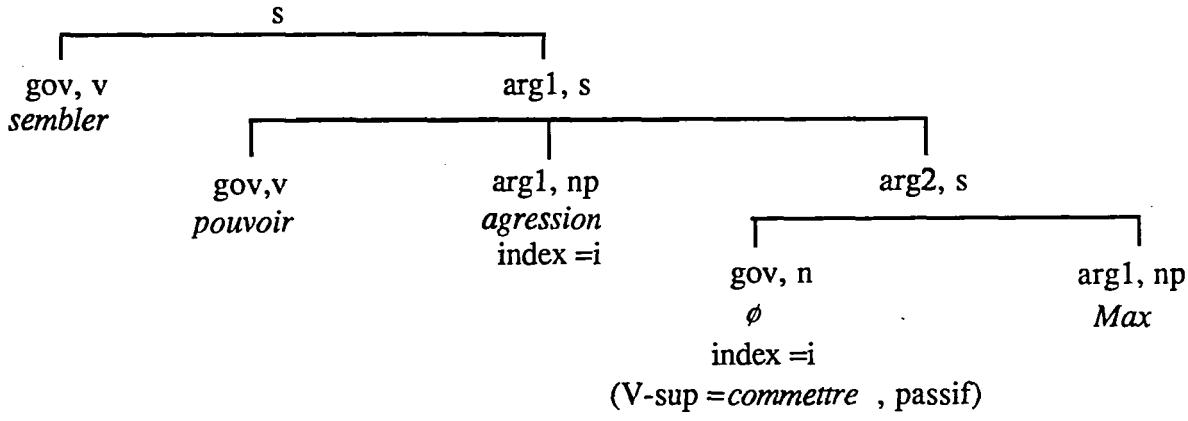
c) *Charles praised God.*

d) *Charles forgot to reward Aymeri.*

Un autre exemple de schéma de texte est celui qui constitue le point de départ du module de génération du système de traduction de textes EUROTRA. Il est appelé **structure d'interface** ; c'est une représentation formelle, sous forme d'arbre, du texte dans la langue cible, contenant tous les traits nécessaires à la synthèse du texte correspondant. Ainsi, la phrase ci-dessous :

Des agressions semblent pouvoir être commises par Max.

est dérivée de la représentation suivante :



Enfin, un dernier exemple de **schéma de texte** est celui du système de génération de L. Danlos (1985, 1987a) dont nous nous sommes inspirée. Ainsi,

Max odia Pia. Ugo ama Eva.
(Max déteste Pia. Ugo aime Eva.)

est un texte tiré du schéma suivant, dont la syntaxe sera largement décrite dans ce chapitre :

```
(       (:cl    (:table 42)
                (:sujet UGO)
                (:verbe odiare)
                (:dir-objet PIA))
        "."
        (:cl    (:table 42)
                (:sujet UGO)
                (:verbe amare)
                (:dir-objet EVA))       )
```

Le schéma de texte, tel qu'il est défini par L. Danlos, constitue pour l'essentiel le point de départ de notre travail. Ce chapitre est consacré à l'exposé des différentes opérations nécessaires à la formation d'un schéma de texte dans le modèle de génération de textes de L. Danlos.

L'une des originalités du système élaboré par L. Danlos est qu'il n'est formé que de deux composants distincts, dont nous rappelons ci-dessous les contenus respectifs :

- le composant stratégique construit un schéma de texte à partir d'une représentation sémantique, en effectuant un certain nombre de choix qui sont développés dans la section 2.3 ;
- le composant syntaxique prend en entrée le schéma de texte et le synthétise en un texte, en effectuant un certain nombre d'opérations (e.g. la pronominalisation).

Le composant stratégique effectue des opérations dont L. Danlos a prouvé qu'elles dépendaient les unes des autres. Cette approche la distingue de chercheurs comme K. McKeown (1985), D.

McDonald (1983) ou encore L. Pemberton (1989) qui ont séparé ces opérations en différents modules (cf. section 2.3).

Ces opérations seront décrites dans ce chapitre, ainsi que la syntaxe d'un schéma de texte et celle des schémas de phrases, qui en sont les constituants. Nous résumerons donc ici une partie du travail de L. Danlos, et nous donnerons un aperçu des modifications que nous avons apportées à ce modèle ; ces modifications seront détaillées au chapitre 7. Mais avant tout, nous allons décrire la façon dont est représenté, dans le modèle dont nous nous inspirons, le message à traduire en texte.

2.2 REPRESENTATION DES CONNAISSANCES

La représentation des connaissances adoptée dans le générateur de L. Danlos s'inscrit dans un formalisme objet. Ainsi, elle satisfait les critères des langages objets :

- les unités de base de la connaissance (personne, objet, action, état ...), sont caractérisées par des champs. Par exemple, une personne est caractérisée par les champs **NOM** et **SEXE** ;

- un champ est la généralisation de la notion de propriété, i.e. de couple attribut-valeur ;

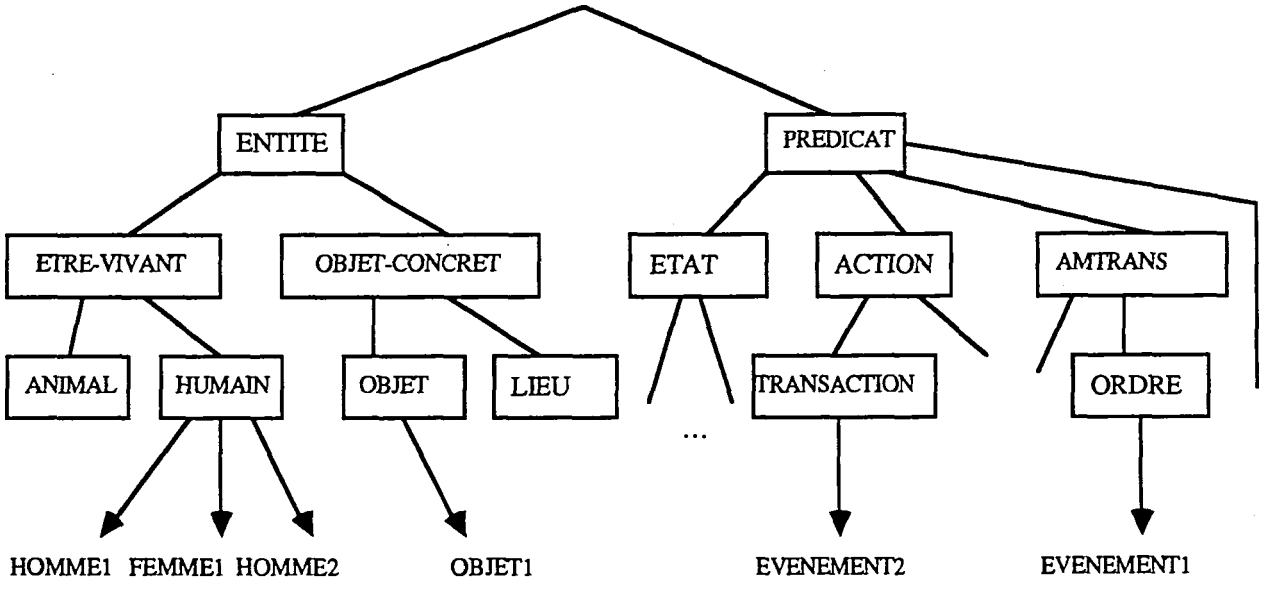
- les unités de base se divisent en classes et en étiquettes : les classes sont des types génériques dont les étiquettes sont des instances. Pour indiquer que HOMME1 est une instance de la classe **HUMAIN** de sexe masculin dont le nom est *Max,* on note :

HOMME1 =: **HUMAIN**
NOM : *Max*
SEXE : masculin

La représentation sémantique donnée en entrée au générateur est une étiquette.

- les classes sont organisées hiérarchiquement. Ainsi, la classe **HUMAIN** peut être considérée comme une sous-classe de la classe **ETRE-VIVANT**, dont elle hérite des propriétés. Une sous-classe possède évidemment des propriétés plus spécifiques que les classes situées plus haut dans la hiérarchie.

Un exemple de hiérarchie des classes est représenté dans l'arbre suivant :



Le nombre et la nature des classes génériques sont dépendants du domaine. Dans la figure ci-dessus, on distingue essentiellement deux classes : les **ENTITES** et les **PREDICATS** : toute classe est soit une sous-classe de **ENTITE** soit une sous-classe de **PREDICAT**. La classe **ENTITE** contient la classe des êtres vivants (e.g. les êtres humains (HOMME1), les animaux), et la classe des objets concrets (e.g. les objets et les lieux). La classe **PREDICAT** regroupe les éléments prédicatifs (au sens logique du terme). La représentation sémantique de la phrase (1) :

(1) *Max ha ordinato a Eva di vendere la casa a Ugo.*
(Max a ordonné à Eva de vendre la maison à Ugo.)

ainsi que celle de la phrase (1') :

(1') *Max ha dato ordine a Eva di vendere la casa a Ugo.*
(Max a donné l'ordre à Eva de vendre la maison à Ugo.)

est la suivante :

(RS)

EVENEMENT1 =: **ORDRE**
AGENT : HOMME1
INTERLOCUTEUR : FEMME1
OBJET : EVENEMENT2

EVENEMENT2 =: **TRANSACTION**
ACHETEUR : HOMME2
VENDEUR : FEMME1
MARCHANDISE : OBJET1

OBJET1 =: **OBJET**
TYPE : *casa* (maison)
DEFINITUDE : defini
NOMBRE : 1

HOMME1 =: **HUMAIN**
NOM : *Max*
SEXE : masculin

FEMME1 =: **HUMAIN**
NOM : *Eva*
SEXE : féminin

HOMME2 =: **HUMAIN**
NOM : *Ugo*
SEXE : masculin

La représentation sémantique EVENEMENT1 est une instance de la classe prédicative **ORDRE** (**ORDRE** est une sous-classe des prédicats **AMTRANS**[1], qui contiennent les concepts où un agent intervient sur un interlocuteur pour effectuer un procès); elle a pour champs l'**AGENT**, dont la valeur est l'étiquette HOMME1, l'**INTERLOCUTEUR**, qui est la femme FEMME1, et l'**OBJET**, qui est le prédicat EVENEMENT2 de la classe **TRANSACTION** (appartenant à la classe **ACTION**). Ce prédicat est caractérisé par trois champs : l'**ACHETEUR** (dont la valeur est la personne HOMME2, de nom *Ugo*), le **VENDEUR** (FEMME1) et la **MARCHANDISE**, qui est un objet concret du **TYPE** *casa* (maison), défini et de nombre singulier. Notons que la représentation sémantique (RS) peut déboucher non seulement sur les phrases (1) ou (1'), mais encore sur les phrases (2) ou (2') :

(2) *?Max ha ordinato a Eva che Ugo le comperi la casa.*
(?Max a ordonné à Eva qu'Ugo lui achète la maison.)

(2') *?Max ha dato ordine a Eva che Ugo le comperi la casa.*
(?Max a donné l'ordre à Eva qu'Ugo lui achète la maison.)

qui sont stylistiquement désagréables, et qui posent, en plus, un problème au niveau de leur contenu logique. Comme nous allons le voir, éviter l'obtention de constructions incorrectes comme (2) ou (2') est le rôle du composant stratégique. C'est également le composant stratégique qui détermine à laquelle des phrases (1) ou (1') le générateur va aboutir en fonction de critères que nous précisons ci-dessous.

Avant d'examiner le composant stratégique, soulignons le point suivant : la représentation des connaissances est un domaine à part entière. Il est connu qu'à l'heure actuelle, chaque système utilise sa propre représentation des connaissances. Certaines recherches sont consacrées exclusivement à l'élaboration d'une représentation sémantique qui se voudrait universelle. Nous n'avons pas cette préoccupation. La représentation que nous proposons n'a pour but que de constituer le point de départ d'une étude de génération de texte du point de vue linguistique, ce qui est notre seul objectif. Le lecteur intéressé trouvera dans G. Sabah (1988) un panorama des principales représentations sémantiques utilisées à ce jour.

2.3 COMPOSANT STRATEGIQUE

2.3.1 DE LA REPRESENTATION SEMANTIQUE AU SCHEMA DE TEXTE

Le passage de la représentation sémantique au schéma de texte a pour but d'exprimer les concepts présents dans la représentation sémantique par une ébauche de texte contenant toutes les informations sémantiques, lexicales et syntaxiques requises pour aboutir, après l'application du composant syntaxique, à un texte bien formé et reproduisant fidèlement les informations contenues dans cette représentation sémantique. Un tel passage requiert deux types d'opérations :

- choisir les items lexicaux pour sélectionner une phrase simple exprimant un concept donné. Une phrase simple est une forme du type :

1 Nous avons emprunté le terme AMTRANS à R. Schank (1975).

SUJET *verbe* $COMPL_i$, $i \geq 0$

où SUJET et COMPLi représentent la complémentation du verbe principal *verbe*. Par exemple, une phrase simple associée au prédicat EVENEMENT2 est la suivante :

(PS2a) *(HOMME2 comperare OBJET1 a FEMME1.)*
(HOMME2 acheter OBJET1 à FEMME1.)

Une autre phrase simple possible est (PS2b) :

(PS2b) *(FEMME1 vendere OBJET1 a HOMME2.)*
(FEMME1 vendre OBJET1 à HOMME2.)

- effectuer la linéarisation en phrases des concepts exprimés sous forme de phrases simples.

Ces deux séries d'opérations ont été mises en évidence par L. Danlos (1985, 1987a), et illustrées dans ses travaux sur les relations causales. Elles sont résumées ci-dessous.

2.3.1.1 CHOIX LEXICAUX

Ces décisions sont effectuées en fonction d'un certain nombre de facteurs :

- les informations sémantiques concernant l'agent et les objets du concept concerné ; supposons par exemple que l'on veuille exprimer un concept **AMOUR** dont l'agent est un homme : on peut choisir l'expression verbale *être amoureux* , si l'objet est une femme, mais ce choix est interdit si l'objet est un aliment (en dehors de toute connotation métaphorique) :

(3) *Paul est amoureux de Eva.*
(4) *Paul (aime + *est amoureux de) la glace au chocolat.*

- le domaine choisi : l'ensemble des choix lexicaux correspondant à un concept donné doit être restreint au domaine étudié ; ainsi, dans un texte traitant d'attentats, on peut supprimer, parmi les choix correspondant au prédicat MORT, tous les verbes et expressions verbales qui ne sont pas considérés comme le résultat direct d'un attentat, e.g. les prédicats exprimant un accident ou une mort naturelle.

- les contraintes syntaxiques :

choix d'un item lexical permettant la transformation passive, si l'agent est nul : considérons par exemple le concept **AMOUR**, dont l'objet est FEMME1. Si l'agent est inconnu, l'une des formulations les plus naturelles de ce concept est alors d'utiliser une phrase au passif sans agent. Cette transformation est incompatible avec le choix de l'item *être amoureux* qui n'est pas passivable. On choisira alors par exemple le verbe *aimer* :

Eva est aimée.

choix d'un item lexical en fonction des propriétés distributionnelles de ses arguments : pour exprimer le concept **TUER**, on ne pourra utiliser le verbe *lyncher* que si l'agent de cette action représente un ensemble d'êtres humains :

*(La foule + *Le commerçant) a lynché le voleur.*

Similairement, le verbe *massacrer* ne s'emploie, au sens propre, qu'avec un objet pluriel[2] :

*Les soldats ont massacré (les habitants de cette ville + *cette dame).*

- les relations sémantiques concernées : L. Danlos (1985, 1987a) a prouvé que dans une relation causale, par exemple, le choix de la phrase simple exprimant l'acte dépend de celui de la phrase simple exprimant le résultat, et vice-versa.

- la stylistique : Dans la représentation sémantique d'une histoire longue, le choix d'un verbe exprimant un concept peut dépendre du nombre d'occurrences de ce concept, ce, dans le but d'éviter les répétitions. Ainsi, si le concept AMOUR entre un homme et une femme est présent à plusieurs reprises dans le récit, il peut être exprimé la première fois par *être amoureux*, qui contient les informations sémantiques les plus précises, et la seconde fois par *aimer*.

2.3.1.2 LINEARISATION EN PHRASES

La linéarisation en phrases est le résultat de la combinaison des trois options suivantes :

- choix de la construction syntaxique pour une phrase simple, e.g. transformation passive (avec ou sans agent) ou nominale. Ainsi, les modifications de la phrase simple (PS2a) exprimant le concept EVENEMENT2 :

(PS2a) *(HOMME2 comperare OBJET1 a FEMME1.)*
(HOMME2 acheter OBJET1 à FEMME1.)

sont entre autres les suivantes :

(PS2'a)
[passif avec agent] *(OBJET1 essere comperato da HOMME2 a FEMME1.)*
(OBJET être acheté par HOMME2 à FEMME1.)

(PS2"a)
[passif sans agent] *(OBJET1 essere comperato a FEMME1.)*
(OBJET1 être acheté à FEMME1.)

(PS2"'a)
[nominalisation] *(l'acquisto di OBJET1 fatto da HOMME2 a FEMME1.)*
(l'achat de OBJET1 par HOMME2 à FEMME1.)

De même, à partir de l'autre phrase simple (PS2b) exprimant EVENEMENT2 :

2 Au sens figuré, l'objet peut être singulier :
Le boxeur a massacré son adversaire.

(PS2b) *(FEMME1 vendere OBJET1 a HOMME2.)*
(FEMME1 vendre OBJET1 à HOMME2.)

on obtient :

(PS2'b)
[passif avec agent] *(OBJET1 essere venduto da FEMME1 a HOMME2.)*
(OBJET1 être vendu par FEMME1 à HOMME2.)

(PS2''b)
[passif sans agent] *(OBJET1 essere venduto a HOMME2.)*
(OBJET1 être vendu à HOMME2.)

(PS2'''b)
[nominalisation] *(la vendita di OBJET1 fatta da FEMME1 a HOMME2.)*
(la vente de OBJET1 par FEMME1 à HOMME2.)

- choix de l'ordre des informations : dans quel ordre doit on exprimer les concepts formant une représentation sémantique ? Cette décision dépend de la relation sémantique liant les concepts ;

- choix du nombre de phrases, i.e. de l'articulation de deux phrases simples exprimant chacune un concept :

  - par juxtaposition,
  - par coordination,
  - par relativation ou subordination de l'une de ces phrases simples par rapport à l'autre.

  Le choix du nombre de phrases, ainsi que le mode d'articulation entre ces phrases, dépend de la relation existant entre les phrases simples concernées.

2.3.1.3 GRAMMAIRE DE DISCOURS ET LEXIQUE-GRAMMAIRE

Pour une représentation sémantique donnée, on obtient, on l'a vu, un certain nombre de combinaisons possibles *a priori*. Par exemple, dans le cas des relations causales, L. Danlos obtient 36 combinaisons CAUSE / RESULTAT. Elle montre que parmi celles-ci, seules 15 combinaisons sont formellement réalisables et sémantiquement acceptables : c'est le cas de la structure

(actif <ACTE>) . (actif <RESULTAT>)
(Max a tiré sur Eve. Il l'a tuée.)

où :

- les deux phrases simples sont construites à l'actif,
- la phrase simple exprimant l'acte apparaît avant celle exprimant le résultat,
- les deux phrases sont juxtaposées.

Par contre, ce n'est pas le cas de la structure :

(actif <RESULTAT>) [relative (passif-sans-agent <ACTE>)
*(*Max a tué Eve qui a été matraquée.)*

où :

- la phrase simple exprimant l'acte est construite au passif sans agent,
- le résultat apparaît avant l'acte,
- la phrase exprimant l'acte est une relative qui modifie l'objet de la phrase exprimant le résultat.

Plus généralement, nous arrivons à la conclusion suivante : les décisions concernant l'ordre des informations, le nombre de phrases et les constructions syntaxiques sont dépendantes les unes des autres. Les combinaisons autorisées pour un type donné de relation sémantique constituent une **grammaire de discours** pour cette relation sémantique. Les décisions sur les critères de linéarisation de phrase reviennent alors à sélectionner une structure de discours.

La deuxième conclusion de L. Danlos est qu'une structure de discours n'engendre un texte correct que si elle est réalisée avec des phrases simples présentant les propriétés syntaxiques adéquates : en d'autres termes, le choix lexical et le choix d'une structure de discours (i.e. choix d'un mode de linéarisation en phrase) sont dépendants l'un de l'autre. Ainsi :

- si on effectue d'abord le choix lexical (i.e. sélection de la phrase simple), les structures de discours autorisées dépendent des propriétés syntaxiques des phrases simples retenues,
- si on sélectionne d'abord la structure de discours, il faut choisir la phrase simple de façon à ce qu'elle présente les propriétés syntaxiques requises par la structure de discours retenue.

Dans un cas comme dans l'autre, on remarque l'importance des propriétés syntaxiques d'une phrase simple. Stocker ces propriétés est le rôle du **lexique-grammaire.** Celui-ci contient les phrases simples décrivant les principaux concepts du domaine avec leurs conditions d'emploi et leurs propriétés syntaxiques : la nature sémantique des arguments, les transformations syntaxiques autorisées (passivation, nominalisation), ainsi qu'éventuellement les conditions nécessaires à ces transformations. Le lexique-grammaire fournit également des renseignements sur les catégories syntaxiques des arguments : nombre et structure des compléments, nature et valeur de la préposition qui les introduit (cf. section 2.4), forme des pronoms auxquels ils donnent lieu (cf. chapitre 3). Un lexique-grammaire a été développé pour le français par l'équipe du L.A.D.L. et les résultats sont regroupés dans M. Gross (1975, 1982) et dans J.P. Boons, A. Guillet et Ch. Leclère (1976a, 1976b). Pour l'italien, on peut citer les travaux de A. Elia et alii (1981), ceux de A. Elia (1982) et de E. D'Agostino (1983). Les verbes et expressions verbales qui ont été étudiés sont représentés, tant en italien qu'en français, sous forme de tables : les verbes appartenant à la même table partagent un certain nombre de propriétés ; d'autres propriétés sont propres à chaque entrée. La figure ci-dessous reproduit un fragment de la table 4 pour l'italien, qui regroupe les verbes dont la construction est : **sujet verbe *a* objet**. En abscisse se trouvent les entrées verbales, en ordonnée les propriétés possibles pour ces verbes :

N0			N1			
N0 = hum N0 = non-hum	Table 4	N0 V	N1 = hum N0 gli V		N1 = non-hum N0 ci V	
- +	appartenere	-	+	+	+	-
- +	mancare	+	+	+	+	+
- +	sorridere	+	+	+	+	-

2.3.2 ORDRE DES DECISIONS

Les deux conclusions dégagées par L. Danlos et présentées ci-dessus peuvent être résumées de la manière suivante. Les quatre opérations :

- sélection d'une phrase simple,
- éventuel apport d'une modification à cette phrase simple (e.g. le passif),
- détermination de l'ordre des phrases simples,
- articulation des phrases simples,

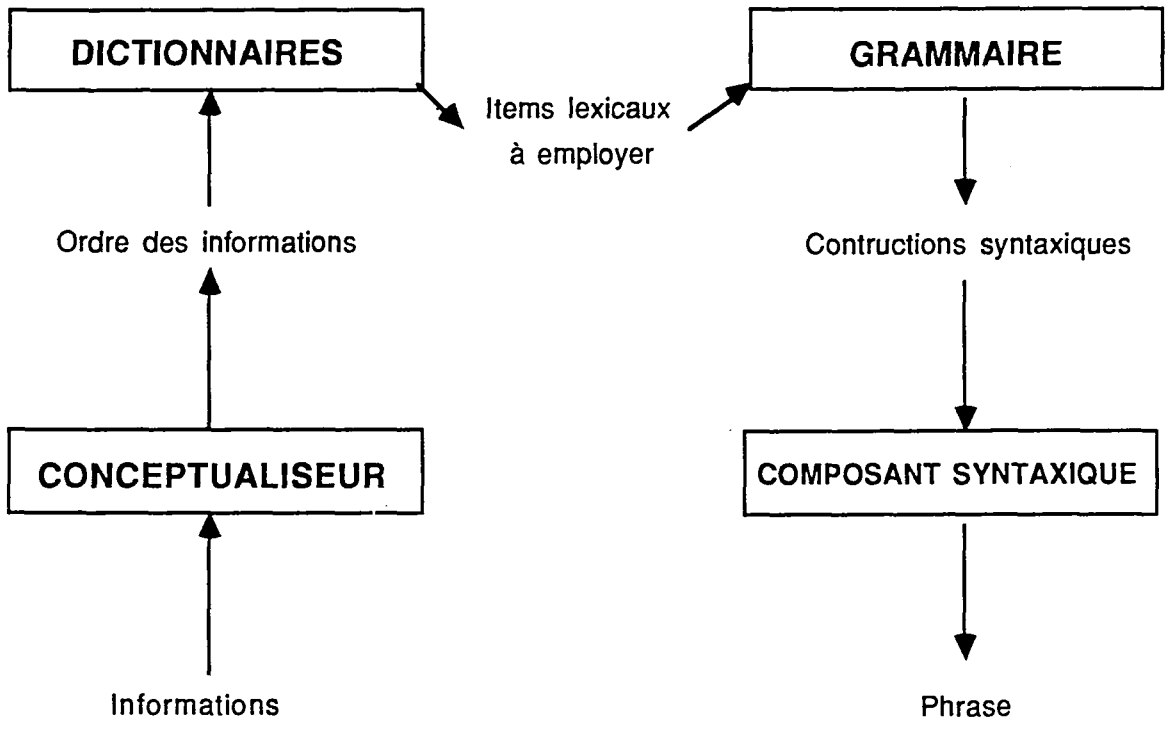
dépendent les unes des autres. Il est donc difficile d'établir une hiérarchie de ces opérations, i.e. de donner *a priori* un ordre de priorité dans les décisions.

Par cette dernière conclusion , L. Danlos (1985) s'oppose aux approches de L. Pemberton (1989), K. McKeown (1985) et D. McDonald (1983). En effet :

Le générateur de L. Pemberton est découpé en plusieurs niveaux d'analyse : au plus haut niveau se trouve la grammaire de narration, qui produit un modèle abstrait composé d'unités appelées fonctions et rôles. Ce modèle de narration peut donner lieu à n'importe quel type de récit. Les fonctions et les rôles sont donc des unités génériques : une fonction sert à créer des liens (cause, tentative, ...) entre deux fragments de l'histoire ; quant à la notion de rôle, elle est vue en termes de cas (sujet, bénéficiaire ...). Une deuxième grammaire, dépendante du genre de l'histoire, construit ensuite à partir de ce modèle une "trame"[3] qui est la version chronologique du récit. Cette trame est transformée par une troisième grammaire en niveau de discours, où les événements subissent une restructuration . Cette transformation est d'ordre conceptuel (modification éventuelle de l'ordre des informations) et stylistique (décision du style de texte à adopter en fonction du genre littéraire représenté par l'histoire). De plus, certaines informations sont supprimées, quand leur contenu est supposé connu (les connaissances pragmatiques impliquées sont calculées à partir d'une base de données liée au domaine). Enfin, un dernier module (règles de lexicalisation) prend les décisions lexicales et syntaxiques, et transforme le modèle de discours en texte.

3 Le terme anglais est : "*story line* ".

Pour K. McKeown et D. McDonald aussi les décisions conceptuelles précèdent les choix lexicaux, comme l'illustre le schéma ci-dessous :



L. Danlos a montré que cet ordre arbitraire des décisions peut conduire à des textes inacceptables *a priori*, parce que sémantiquement imprécis : si, par exemple, pour une représentation sémantique décrivant un suicide, l'ordre choisi est RESULTAT-CAUSE, alors le verbe *se suicider* ne peut par être employé :

**Max s'est jeté de la Tour Eiffel se suicidant.*

et l'emploi d'un autre verbe a pour conséquence de supprimer une information sémantique capitale.

L'ordre de priorité varie en fait d'un domaine à l'autre, d'un type de texte à l'autre, d'un cas à l'autre. L'algorithme de génération doit permettre l'interaction entre les différentes décisions ; il doit en particulier déterminer celles qui sont sacrifiées en cas de conflit. Au chapitre 7, nous décrirons les choix effectués et les méthodes employées pour déterminer des structures de discours dans le domaine d'un récit où les événements sont relatés dans l'ordre chronologique.

2.3.3 COMPLETIVES

Les phrases simples étudiées par L. Danlos ont des verbes ou expressions verbales construits avec des arguments nominaux. Il existe cependant des concepts complexes dont l'agent ou/et l'objet est un autre prédicat : les phrases simples qui expriment ces concepts complexes ont alors un verbe qui se construit avec un sujet ou/et un (des) objet(s) phrastique(s). Un tel argument est généralement appelé **complétive**. Reprenons l'exemple (RS) de la section

2.2, dans lequel le concept EVENEMENT1 a un champ objet prédicatif : le concept EVENEMENT2. Rappelons que EVENEMENT1 peut être exprimé par la phrase suivante :

(1) *Max ordina a Eva di vendere la casa a Ugo.*
(Max ordonne à Eva de vendre la maison à Ugo.)

La phrase simple (PS) qui permet d'obtenir (1) est réalisée en réunissant une phrase simple exprimant EVENEMENT1 :

(HOMME1 ordinare EVENEMENT2 a FEMME1.)
(HOMME1 ordonner EVENEMENT2 à FEMME1.)

et la phrase simple PS2b exprimant EVENEMENT2 :

(PS2b) *(FEMME1 vendere OBJET1 a HOMME2.)*
(FEMME1 vendre OBJET1 à HOMME2.)

ce qui permet d'obtenir :

(PS) *(HOMME1 ordinare [FEMME1 vendere OBJET1 a HOMME2] a FEMME1.)*
(HOMME1 ordonner [FEMME1 vendre OBJET1 à HOMME2] à FEMME1.)

Un prédicat en position sujet ou objet d'un autre prédicat possède certaines propriétés :

a) il peut être synthétisé en une proposition infinitive, si le sujet de cette proposition est égal à l'un des arguments du verbe principal, appelé contrôleur. Cette propriété, qui dépend du verbe de la principale, a été étudiée en particulier dans M. Gross (1975) et A. Elia et alii (1981). Certains résultats pertinents à notre étude sont exposés en détail au chapitre 6. Dans certains cas, la réduction de la complétive à l'infinitif est obligatoire. C'est le cas de la phrase (5) :

**Max vuole che Max venda la casa a Ugo.*
(*Max que Max vende la maison à Ugo.)

**Max*$_i$ *vuole che egli*$_i$ *venda la casa a Ugo.*
(*Max$_i$ veut qu'il$_i$ vende la maison à Ugo.)[4]

=> (5) *Max vuole vendere la casa a Ugo.*
(Max veut vendre la maison à Ugo.)

Par ailleurs, on a généralement recours à ce genre de transformation, quand elle est possible, dans le but d'obtenir des textes plus légers et élégants. Ainsi, au lieu de :

(2) *?*Max ordina a Eva che Ugo le comperi la casa.*
(?*Max ordonne à Eva que Ugo lui achète la maison.)

où la complétive est obtenue à partir de la phrase simple (PS2a) :

4 L'inacceptabilité (i.e. l'agrammaticalité), indiquée par *, n'est plus vraie dans une interprétation où *Max* et le pronom sujet *egli* (il) ne représentent pas la même personne : c'est pourquoi l'indice i rappelle que les deux termes sont coréférents.

(PS2a) *(HOMME2 comperare OBJET1 a FEMME1.)*
(HOMME2 acheter OBJET1 à FEMME1.)

on préfère la phrase (1)

(1) *Max ordina a Eva di vendere la casa a Ugo.*
(Max ordonne à Eva de vendre la maison à Ugo.)

dans laquelle la complétive provient de la phrase simple (PS2b), et dont la position des arguments rend possible la réduction à l'infinitive, contrôlée par l'égalité entre le sujet de la complétive et l'objet prépositionnel *Eva* de *ordinare*, introduit par *a*. Cet exemple montre que le choix lexical pour la complétive dépend de l'item lexical qui a été choisi dans la phrase simple de la principale. De la même façon, à la place de :

(6a) *?Eva vieta a Ugo che Max lo accompagni a scuola.*
(?Eva interdit à Ugo que Max l'accompagne à l'école.)

où la complétive objet direct du verbe *consigliare* est à l'actif, on préfère écrire :

(6b) *Eva vieta a Ugo di essere accompagnato a scuola da Max.*
(Eva interdit à Ugo d'être accompagné à l'école par Max.)

où la complétive objet direct a subi une transformation passive, ce qui permet la réduction à l'infinitive, contrôlée par l'égalité entre le sujet de la complétive et l'objet prépositionnel *Ugo* du verbe *consigliare* introduit par la préposition *a*. C'est ici la construction syntaxique de la phrase simple complétive qui dépend de l'item lexical choisi dans la principale.

b) il peut être synthétisé en un groupe nominal prédicatif. Un nom prédicatif a des arguments, au même titre qu'un verbe ou une expression verbale. La notion d'élément prédicatif, et plus particulièrement de nom prédicatif, a été introduite par Z.S. Harris (1968), et est étudiée en particulier par L. Danlos (1988b), J. Giry-Schneider (1978, 1987), A. Meunier (1980) et G. Gross (1989). Comme on l'a vu pour la réduction à l'infinitif, cette transformation est parfois préférable d'un point de vue stylistique. Ainsi, à la place de :

(7a) *Ugo convince i giurati che è innocente.*
(Ugo convainc les jurés qu'il est innocent.)

où l'objet prédicatif du verbe *convincere*, exprimant le concept d'innocence, est synthétisé en une complétive autour de l'expression verbale *essere innocente* , on préfère :

(7b) *Ugo convince i giurati della propria innocenza.*
(Ugo convainc les jurés de son innocence.)

où ce concept est synthétisé en un groupe nominal prédicatif autour du nom *innocenza*. Dans cet exemple, on voit que le choix d'une formulation exprimant le concept en position objet dépend de l'item lexical contrôlant ce concept.

D'autres critères stylistiques entrent en jeu lors du choix d'une structure de discours exprimant un concept complexe composé d'un concept principal et d'un ou plusieurs arguments prédicatifs : ainsi, on évite la synthèse des complétives sujet. A la place de :

(8a) *?Il fatto che Paul l'abbia tradita delude Eva.*
(?Le fait que Paul l'ait trompée déçoit Eva.)

on tourne la principale au passif (si l'item lexical choisi permet une telle construction) :

(8b) *Eva è delusa dal fatto che Paul l'abbia tradita.*
(Eva est déçue du fait que Paul l'ait trompée.)

De plus, quand c'est possible, on réduit la complétive à l'infinitif, après l'avoir mise à la voix passive :

(8c) *Eva è delusa di essere stata tradita da Paul.*
(Eva est déçue d'avoir été trompée par Paul.)

Un autre exemple est celui de la phrase (9) :

(9a) *Il fatto che Paul sia tornato dalla madre porta a tutti un barlume di speranza.*
(Le fait que Paul soit revenu chez sa mère apporte à tous une lueur d'espoir.)

qui ne permet pas la passivation de *portare*, mais que la transformation nominale de la complétive sujet rend plus lisible :

(9b) *Il ritorno di Paul dalla madre porta a tutti un barlume di speranza.*
(Le retour de Paul chez sa mère apporte à tous une lueur d'espoir.)

En conclusion, le choix d'une phrase simple représentant un argument prédicatif et le choix d'une phrase simple exprimant le prédicat principal dépendent l'un de l'autre : les exemples (9a) et (9b) illustrent la dépendance entre la décision lexicale pour l'argument prédicatif et le choix de la construction syntaxique de la principale. Les exemples (8a-c) montrent la dépendance existant entre les transformations syntaxiques subies par les deux propositions. En d'autres termes, la synthèse des phrases à complétive demande de prendre des décisions qui viennent s'ajouter à celles mises en évidence par L. Danlos et exposées dans la section précédente. Ces décisions seront décrites en détail au chapitre 7.

2.4 SCHEMA DE TEXTE

Le schéma de texte est réalisé à partir de la représentation sémantique en fonction des choix décrits ci-dessus. Il contient toutes les informations de la représentation de départ et il est formellement réalisable : sa réalisation (pronominalisation, effacement d'éléments dans des coordinations, conjugaison et accord des verbes, etc ...) est effectuée par le composant syntaxique.

Un schéma de texte T est constitué d'un ensemble de schémas de phrase Ph_i, séparés par des signes de ponctuation $Punct_i$:

$$T = (Ph_1\ Punct_1\ Ph_2\ Punct_2\ ...\ Ph_i\ Punct_i\ ...\ Ph_n\ Punct_n)$$

Un schéma de phrase Ph a la syntaxe suivante, présentée en Normal Backus Form (formalisme souvent utilisé pour décrire les langages de programmation) :

[ph] = [phrase] / [phrase-coordonnée] / [phrase-avec-subordonnée]
[phrase-coordonnée] = ([conj-coord] [ph] [ph]$^{+}$)
[phrase-avec-subordonnée] = ([sub] [ph]) / ([ph] [sub])
[sub] = ([conj-sub] [ph])
[conj-coord] = (:conj-coord *conjonction de coordination*)
[conj-sub] = (:conj-sub *conjonction de subordination*)

[phrase] = (:Cl [classe] [sujet] [verbe] complement^{n} $(0 \leq n \leq 2)$ [adverbe])
[classe] = (:table *numéro de table*)
[sujet] = (:sujet étiquette)
[verbe] = (:verbe *verbe*)
complement = [dir-objet] / p-objet
[dir-objet] = (:dir-objet étiquette)
p-objet = [a-objet] / [di-objet] / [da-objet] / [loc-objet] / [prep-objet] / [attribut]
[a-objet] = (:a-objet étiquette)
[di-objet] = (:di-objet étiquette)
[da-objet] = (:da-objet étiquette)
[loc-objet] = (:loc-objet [prep] étiquette)
[prep-objet] = (:prep-objet [prep] étiquette)
[attribut] = (:attribut *adjectif*) (:attribut étiquette)
[prep] = (:prep *préposition*)
[adverbe] = (:adverbe *adverbe*)

La barre oblique / symbolise l'alternative.

Le symbole $^{+}$ signifie que l'élément apparait au moins une fois.

L'étiquette :cl est mise pour le terme anglais *clause*. L'anglais, contrairement au français, distingue *clause*, i.e. phrase sans subordonnée ni coordination, de *sentence*. C'est pourquoi nous employons ce symbole pour désigner le phrase "minimale".

La [classe] est le numéro de table auquel appartient le verbe. Ce renseignement permet d'accéder aux informations syntaxiques associées à ce verbe dans le lexique-grammaire.

Les termes en italiques représentent des items lexicaux. En particulier, *verbe* est le verbe à l'infinitif.

Une étiquette est soit une instance de la classe des ENTITES, soit (exception faite de l'[attribut]) une instance de la classe des PREDICATS. Dans ce dernier cas, l'expression de cette étiquette peut être un nouveau schéma de phrase.

Une illustration de la notion de schéma de textes est donnée ci-dessous par l'exemple (T10), qui est obtenu à partir de la représentation sémantique (RS) :

(T10) (:cl (:table 47)
(:sujet HOMME1)
(:verbe *ordinare*)
(:dir-objet
(:cl (:table 23)
(:sujet FEMME1)
(:verbe *vendere*)
(:dir-objet OBJET1)
(:a-objet HOMME2)))
(:a-objet FEMME1))

Les différentes catégories d'objets prépositionnels [p-objet] sont propres à la langue : ces distinctions dépendent des propriétés syntaxiques (e.g. la pronominalisation) de ces objets en fonction du verbe avec lequel ils se construisent. En français comme en italien, le lexique-grammaire (cf. section 2.3.1.3) a servi de base à ces distinctions. En italien, les différentes catégories d'objets prépositionnels sont les suivantes :

- un [a-objet] est un objet introduit par la préposition *a*. Dans la littérature, il est appelé parfois objet datif ou encore objet indirect ;

- un [di-objet] est un objet introduit par la préposition *di*. Le pronom correspondant à cet objet est la forme préverbale *ne* (cf. chapitre suivant) ;

- un [da-objet] est un objet introduit par la préposition *da*. Un [da-objet] est soit un locatif de provenance (et répond dans ce cas à la question *Da dove ?* (D'où ?)), soit un locatif de provenance métaphorique ou abstrait (et répond alors à la question *Da chi ?* ou *Da che cosa ?*). La pronominalisation en *ne* n'est pas systématique comme c'est le cas pour le [di-objet]. Les conditions de pronominalisation seront exposées au prochain chapitre ;

- un [loc-objet] est un objet qui indique soit la destination, soit l'endroit où l'on se trouve; il répond à la question *Dove ?* (Où ?), et le pronom correspondant est la forme préverbale *ci*. Un [loc-objet] peut être introduit par n'importe quelle préposition locative :

[prep = *a*] *Ugo va a Londra.*
(Ugo va à Londres.)
[prep = *da*] *Ugo va da Eva.*
(Ugo va chez Eva.)

- la notion de [prep-objet] regroupe tous les objets prépositionnels qui ne peuvent pas être rangés dans les précédentes catégories, parce que la pronominalisation ne donne pas de résultats homogènes, ou que ces objets ne répondent pas aux questions correspondant aux catégories ci-dessus. Ainsi, un [prep-objet] peut être introduit par la préposition *da*, et la pronominalisation en *ne* de cet objet peut être interdite :

[prep = *da*] *Ugo è stato investito da une macchina.*
(Ugo a été écrasé par une voiture.)
**Ugo ne è stato investito.*
(*Ugo en a été écrasé.)

Dans un schéma de phrase, français ou italien, les différents objets sont placés dans un ordre fixe (l'ordre canonique) ; en italien, c'est le suivant :

sujet - verbe - dir-objet - a-objet - di-objet - da-objet - loc-objet - prep-objet - adverbiaux

Dans un système de génération de textes, on doit manipuler trois ordres différents :

- l'ordre dans lequel sont placés les éléments d'un schéma de phrase, ou ordre canonique, que nous venons de voir ;
- l'ordre dans lequel ces éléments sont synthétisés ; cet ordre dépend en grande partie des phénomènes de pronominalisation, comme nous l'avons annoncé au chapitre 1 et comme nous le verrons dans les deux prochains chapitres ;
- l'ordre d'apparition de ces éléments dans la phrase synthétisée. Cet ordre dépend d'un certain nombre de facteurs stylistiques (e.g. la longueur des compléments). Mais il peut dépendre également de l'ordre de synthèse des éléments de la phrase, comme nous le verrons au chapitre 6.

2.5 CONCLUSION

Le schéma de texte constitue l'entrée du composant syntaxique. Il est formé d'une suite ordonnée de schémas de phrases. Le composant syntaxique synthétise les schémas de phrases les uns après les autres. Nous avons vu que, pour plusieurs raisons, le composant stratégique était dépendant du domaine. A l'inverse, un composant syntaxique doit être indépendant du domaine, pour couvrir tous les phénomènes syntaxiques d'une langue donnée. Notre travail est consacré essentiellement à l'étude de ce composant, et plus particulièrement à un phénomène capital au niveau du discours : la pronominalisation.

Nous avons évoqué dans le chapitre précédent le fait que le choix d'une structure de discours pouvait dépendre de la pronominalisation ; en d'autre termes que les décisions qui relèvent du composant stratégique pouvaient dépendre des décisions qui relèvent du composant syntaxique. Nous donnerons des illustrations de cette affirmation tout au long de notre exposé, et nous proposons au chapitre 7 une méthode pour résoudre certains des problèmes ainsi soulevés. Nous ne prétendons cependant pas apporter ici une solution globale au problème de la modularité. Ceci fera l'objet de recherches futures.

Les exemples que nous présentons ont pour point de départ un schéma de texte ou de phrase. Par souci de lisibilité, les schémas de textes qui serviront d'exemples seront simplifiés : les étiquettes importantes pour la démonstration seront en majuscules, et les autres éléments en italiques. Par exemple, le schéma de texte (T10) :

```
(T10)   (:cl (:table 47)
             (:sujet HOMME1)
             (:verbe ordinare )
             (:dir-objet
                  (:cl (:table 23)
                       (:sujet FEMME1)
                       (:verbe vendere)
                       (:dir-objet OBJET1)
                       (:a-objet HOMME2)))
             (:a-objet FEMME1))
```

sera présenté de la façon simplifiée suivante:

(T10') *(HOMME1 ordinare (FEMME1 vendere OBJET1 a HOMME2) a FEMME1.)*

Notre étude étant consacrée à l'italien, nous serons parfois amenée à donner une traduction littérale de nos exemples (notée [litt]), si celle-ci est nécessaire à la compréhension. Par ailleurs, si un élément est effacé dans l'exemple italien, la traduction de cet élément sera indiquée entre crochets. Ainsi, dans l'exemple ci-dessous, le sujet est effacé dans la deuxième phrase, et la traduction de cet élément vide est le pronom *il* :

Ugo vede Maria. La saluta.
(Ugo voit Maria. [Il] la salue.)

De même, si un mot contient une information morphologique ou sémantique en italien qui n'a pas d'équivalent en français, cette information est indiquée en indice, quand cela est nécessaire (i.e. pour éviter des ambiguïtés). Ainsi, on indique que le pronom *egli* (il) ne peut représenter qu'un humain en marquant la traduction par l'indice : hum

Ugo pianta un ciliegio. Egli è ammirevole.
(Ugo plante un cerisier. Il$_{hum}$ est admirable.)

3 PRONOMINALISATION

3.1 INTRODUCTION

Le terme pronominalisation indique dans notre étude l'ensemble des opérations qui aboutissent à la synthèse d'un pronom personnel. Ce phénomène joue un rôle capital en génération : c'est l'un des facteurs qui déterminent la qualité d'un texte. Ainsi, regardons l'extrait (A) produit par le programme GESTER de L. Pemberton (1989) de génération d'histoires épiques, qui a été brièvement présenté au chapitre 2 :

(A) a) *Charles and Aymeri broke into Narbonne.*
b) *Charles succeded in getting Narbonne.*
c) *Charles praised God.*
d) *Charles forgot to reward Aymeri.*

Notons (A') la traduction de (A) :

(A') a) *Charles et Aymeri pénétrèrent dans Narbonne.*
b) *Charles réussit à conquérir Narbonne.*
c) *Charles loua Dieu.*
d) *Charles oublia de récompenser Aymeri.*

Ce résultat, comme nous l'avons déjà dit, peut être considéré comme un schéma de texte élaboré où certaines opérations syntaxiques ont déjà été effectuées. Mais si (A') est présenté comme un texte, alors la succession des répétitions en fait un texte inacceptable.

Maintenant, supposons que des transformations sont effectuées sur (A'), afin de supprimer les répétitions :

(A'') a) *Charles et Aymeri pénétrèrent dans Narbonne.*
b) *Il réussit à la conquérir.*
c) *Il loua Dieu.*
d) *Il oublia de le récompenser.*

Le résultat (A'') est inacceptable, car les pronoms rendent ce texte ambigu. En d'autres termes, un texte sans pronom est de mauvaise qualité, et un texte contenant trop de pronoms est également incorrect. Puisque notre but est la génération de textes bien formés et cohérents les conditions de synthèse d'un pronom constituent un point essentiel de notre travail.

La synthèse d'un pronom personnel nécessite trois étapes. La première étape consiste à déterminer quels éléments doivent être pronominalisés. En un deuxième temps, on décide de la forme pronominale à engendrer. Ensuite, la démarche est la même qu'en analyse pour la résolution d'anaphores : on calcule les antécédents de ce pronom. Le résultat de ces calculs est déterminant dans la décision de former le pronom envisagé : ainsi, (A''') est l'un des textes que l'on souhaiterait obtenir par transformation de (A') :

(A''') a) *Charles et Aymeri pénétrèrent dans Narbonne.*
b) *Charles réussit à conquérir cette ville.*
c) *Il loua Dieu.*

d) *Il oublia de récompenser son ami.*

Cet exemple montre que la synthèse d'une entité peut parfois aboutir à un pronom (*Il*), mais également à une répétition (*Charles*, dans la phrase b de (A'")) ou à une redescription (*cette ville*, *son ami*), que l'on appelle généralement anaphore nominale. Les transformations nécessaires à l'obtention des anaphores nominales sont souvent évoquées (entre autres par G. Fauconnier (1974), F. Corblin (1980) et J.C. Milner (1982)) à cause de certaines similitudes existant entre ces éléments anaphoriques et les pronoms, notamment en ce qui concerne les contraintes liées à leur emploi. Nous n'insisterons pas sur les conditions de synthèse des anaphores nominales, que nous réservons à des recherches futures : nous nous limitons ici à l'étude de la pronominalisation. Notre travail étant consacré plus particulièrement à l'étude de l'italien, nous verrons que la synthèse d'une entité, quand elle est en position sujet, peut aussi aboutir à l'effacement de ce sujet : ce phénomène est caractéristique de toutes les langues romanes, excepté le français[1].

Le rôle de la pronominalisation en génération de textes a été étudié, entre autres, par G. Sabah et M. Zock, qui ont élaboré un système didactique d'aide à la génération des constructions à pronoms préverbaux en français (G. Sabah., M. Zoch, C. Alviset (1986)). Plus généralement les buts de ce système sont :

- d'assister les étudiants à l'apprentissage de génération de phrase ;
- d'étudier les mécanismes de production d'une phrase en temps réel (i.e. quels sont les choix effectués par le locuteur, et dans quel ordre (M. Zock, G. Francopoulo, A. Laroui (1988), M. Zock (1987), M. Zock (1988))).

Cette approche psychologique de la génération de textes met en évidence de nouveaux aspects de l'importance de la pronominalisation. Cependant la nature même de ce système (i.e. la psychologie), et le but didacticiel qu'il s'est donné, font que ces résultats ne peuvent pas être intégrés dans notre générateur.

La pronominalisation est non seulement un facteur capital dans la production de textes bien formés, c'est également un élément majeur dans la détermination de l'algorithme de synthèse d'une phrase. Ainsi, les données linguistiques qui seront exposées dans ce chapitre et dans les deux chapitres suivants aboutissent à une discussion sur l'algorithme à adopter. Dans ce chapitre, nous présenterons une description linguistique des pronoms en italien, effectuée à partir de différentes grammaires (M. Sensini (1988) et L. Serianni (1988)). De cette description, on dégagera les faits pertinents en génération de textes. Ensuite, on mettra en évidence les raisons pour lesquelles il faut distinguer la synthèse des objets de celle du sujet (une des raisons principales étant l'effacement de ce dernier). Nous présenterons alors la synthèse des pronoms objets, pour lesquels nous donnerons des exemples de dépendances non locales et de dépendances croisées. Enfin, nous étudierons la synthèse, et plus particulièrement l'effacement, du sujet.

1 Nous ne nous occupons pas des phénomènes d'ellipse dans les cas de coordination ou d'énumération de phrases :

Paul voit Eva et la salue.

ou

Paul voit Eva, la salue.

Mais avant tout, nous allons étudier la synthèse des formes réflexives, qui contrairement à celle des pronoms personnels, ne nécessite pas d'autres opérations que le calcul de la forme à synthétiser.

3.1.1 FORMES REFLEXIVES

"Forme réflexive" est un terme qui regroupe les formes énumérées dans le tableau ci-dessous :

	FORMES ATONES	FORMES TONIQUES
1ps	*mi* (me)	*me* (moi)
2ps	*ti* (te)	*te* (toi)
3ps	*si* (se)	*sè* (soi / lui / elle)
1pp	*ci* (nous)	*noi* (nous)
2pp	*vi* (vous)	*voi* (vous)
3pp	*si* (se)	*sè* (soi / eux / elles)

Une forme réflexive réfère exclusivement au sujet. Parmi les termes qui apparaissent dans le tableau ci-dessus, seules les formes de la troisième personne du singulier et du pluriel sont exclusivement réflexives. Toutes les autres (cf. tableau page 34) peuvent référer à une entité autre que le sujet.

Les formes atones, que l'on nomme en général pronoms réflexifs, sont placées dans ce que nous appellerons le **noyau verbal**, i.e. l'entourage immédiat du verbe : l'espace compris entre l'emplacement du sujet et celui du premier objet, ou de la première préposition, si l'objet en question est prépositionnel. Nous dirons, par abus de langage, que les formes atones sont préverbales, bien qu'elles soient parfois placées après le verbe :

Eva vuole ammirarsi.
(Eva veut s'admirer.)

Les formes toniques sont, elles, non préverbales : elles occupent toujours la place du groupe nominal qu'elles remplacent. D'après certaines grammaires (L. Serianni (1988)), elles doivent leur nom à ce que leur emploi peut correspondre à un désir de mise en valeur de l'entité représentée, contrairement aux formes atones :

Eva si ammira. — *Eva ammira solo sè.*
(Eva s'admire.) — ([litt] Eva n'admire que soi.
=> Eva n'admire qu'elle-même.)

Une forme tonique peut toujours être suivie de l'adjectif *stesso* (même). Ceci ne sera pas discuté ici, puisque n'ayant aucune conséquence sur la synthèse d'un tel pronom. Par contre, la synthèse de l'adjectif *stesso* peut jouer un rôle dans l'effacement du sujet, puisqu'il en indique le genre, étant donnée la coréférence obligatoire entre le sujet et le réflexif[2] :

2 Je tiens à remercier A. Abeillé qui a attiré mon attention sur ce point.

Max odia Eva. Ama solo se stessa.
(Max déteste Eva. [Elle + *il] n'aime qu'elle-même.)

Max odia Eva. Ama solo se stesso.
(Max déteste Eva. [Il + *elle] n'aime que lui-même.)

Cette propriété sera vue plus en détail dans la section 3.3.2.3, à propos des conditions d'effacement du sujet.

Les termes "tonique" et "atone" se retrouvent également dans la description des pronoms personnels, pour lesquels une étude plus détaillée de la position des formes "atones" dans le noyau verbal sera présentée dans la section 3.1.2.

Une forme réflexive ne pose pas de gros problème, que ce soit en analyse ou en génération :

- en analyse, la forme du pronom, atone ou tonique, de la troisième personne (*se* ou *soi* en français, *si* ou *sè* en italien) indique s'il s'agit ou non d'un réflexif, et, par conséquent, d'un pronom référant exclusivement au sujet. Aux premières et deuxièmes personnes, il suffit de comparer les propriétés morphologiques du pronom à celles du verbe ou du sujet : le pronom est réflexif en cas d'égalité en nombre et personne ;

- de façon analogue, la seule condition nécessaire à la synthèse d'un pronom réflexif, en génération de textes, est de pouvoir déterminer que le sujet d'une phrase et l'un des objets de cette phrase représentent le même individu ou le même objet. Le calcul de la forme réflexive est alors la seule opération à effectuer, contrairement à ce qui se passe pour la synthèse d'un pronom personnel, où le calcul de la forme est suivie, nous le verrons, d'une série d'opérations pour déterminer si la synthèse de la forme envisagée est possible.

Notons cependant que le calcul de la forme réflexive n'est pas une opération simple. Nous n'insisterons pas ici sur les différents aspects syntaxiques et sémantiques des verbes pronominaux, ni des conséquences pour la synthèse des réflexifs. Nous nous contenterons d'illustrer par les schémas de phrase simplifiés (S1), (S2), (S3) et (S4) les différents facteurs qui influencent le calcul des formes réflexives, en dehors des valeurs, en nombre et en personne, du sujet de la phrase :

(S1) *(EVA ammirare EVA.)*
(EVA admirer EVA.)
(S2) *(EVA pensare a EVA.)*
(EVA penser à EVA.)
(S3) *(EVA agire per EVA.)*
(EVA agir pour EVA.)
(S4) *(EVA dare una pettinata a EVA.)*
(EVA donner un coup de peigne à EVA.)

Les quatre schémas contiennent un sujet et un objet qui représentent la même personne i.e. *Eva*, 3ème personne féminin-singulier ; pour chaque Si, on envisage alors la synthèse de l'objet en une forme réflexive de la 3ème personne :

- le pronom synthétisant l'objet direct EVA de (S1) est la forme atone *si*, qui est placée devant le verbe :

(1) *Eva si ammira.*
(Eva s'admire.)

- par contre, le pronom synthétisant l'objet prépositionnel de (S2) ou de (S3) est la forme tonique *sè* qui occupe la même position qu'un groupe nominal prépositionnel :

(2) *Eva pensa a sè.*
[litt] (Eva pense à soi.)
(Eva pense à elle-même.)
(3) *Eva agisce per sè.*
[litt] (Eva agit pour soi.)
(Eva agit pour elle-même.)

- la conclusion qui consiste à dire que la synthèse d'un réflexif dépend de la catégorie syntaxique de l'objet (dans (S1) l'objet EVA est direct, dans (S2) et (S3), il est prépositionnel) n'est pas suffisante, comme l'indique la phrase (4), issue de la synthèse de (S4) :

(4) *Eva si dà una pettinata.*
(Eva se donne un coup de peigne.)

L'expression verbale *dare una pettinata* se construit avec un objet prépositionnel en *a*, comme c'est le cas pour le verbe *pensare* en (S2) ; pourtant la forme du réflexif est la même qu'en (1), où l'objet est direct.

En fait, le calcul d'une forme réflexive dépend :

- du nombre et de la personne du sujet,
- de la position syntaxique de l'objet à synthétiser,
- du verbe utilisé[3].

A partir des exemples (S1), (S2), (S3) et (S4), on peut alors dégager les conclusions suivantes :

si l'objet est direct, alors la forme réflexive est une forme atone, donc placée en position préverbale ;

---

3 Le calcul peut dépendre aussi du substantif N utilisé, s'il y a égalité entre le sujet et le complément de ce substantif N (la position syntaxique de l'objet à synthétiser est alors complément du nom). Ainsi :
(EVA mette CAPPOTTO (di EVA).)
(EVA met MANTEAU (de EVA).)
débouche sur :
Eva si mette il cappotto.
([litt] Eva se met le manteau.)
plutôt que sur :
?Eva mette il suo cappotto.
(Eva met son manteau.)

	SUJET			
	HUMAIN		NON-HUMAIN	
	masc	fem	masc	fem
1ère ps	io		-	
2ème ps	tu		-	
3ème ps	egli,lui	lei,ella	esso	essa
1ère pp	noi		-	
2ème pp	voi		-	
3ème pp	loro		essi	esse

	DIR-OBJET					
	FORME ATONE		FORME TONIQUE			
			HUMAIN		NON-HUMAIN	
	masc	fem	masc	fem	masc	fem
1ère ps	mi		me		-	
2ème ps	ti		te		-	
3ème ps	lo	la	lui	lei	esso	essa
1ère pp	ci		noi		-	
2ème pp	vi		voi		-	
3ème pp	li	le	loro		essi	esse

	DI-OBJET et DA-OBJET				
	FORME ATONE	FORME TONIQUE			
		HUMAIN		NON-HUMAIN	
		masc	fem	masc	fem
1ère ps	ne	me		-	
2ème ps	ne	te		-	
3ème ps	ne	lui	lei	esso	essa
1ère pp	ne	noi		-	
2ème pp	ne	voi		-	
3ème pp	ne	loro		essi	esse

	A-OBJET						
	FORME ATONE			FORME TONIQUE			
	HUMAIN		NON-HUM	HUMAIN		NON-HUMAIN	
	masc	fem		masc	fem	masc	fem
1ère ps	mi			me		-	
2ème ps	ti			te		-	
3ème ps	gli	le	ci	lui	lei	esso	essa
1ère pp	ci			noi		-	
2ème pp	vi			voi		-	
3ème pp	loro			loro		essi	esse

	LOC-OBJET				
	FORME ATONE	FORME TONIQUE			
		HUMAIN		NON-HUMAIN	
		masc	fem	masc	fem
1ère ps	ci	me		-	
2ème ps	ci	te		-	
3ème ps	ci	lui	lei	esso	essa
1ère pp	ci	noi		-	
2ème pp	ci	voi		-	
3ème pp	ci	loro		essi	esse

	PREP-OBJET			
	FORME TONIQUE			
	HUMAIN		NON-HUMAIN	
	masc	fem	masc	fem
1ère ps	me		-	
2ème ps	te		-	
3ème ps	lui	lei	esso	essa
1ère pp	noi		-	
2ème pp	voi		-	
3ème pp	loro		essi	esse

TABLEAU DES FORMES PRONOMINALES

si l'objet est prépositionnel introduit par une autre préposition que *a*, alors le pronom est une forme tonique ; il occupe la place du groupe nominal qu'il remplace ;

si, enfin, l'objet est prépositionnel introduit par *a*, alors la forme du pronom dépend du verbe utilisé : on distinguera les verbes du type *pensare* (synthèse d'une forme tonique) des verbes du type *dare* (synthèse d'une forme atone).

3.1.2 DESCRIPTION LINGUISTIQUE DES PRONOMS PERSONNELS

Avant d'aborder les différentes opérations nécessaires à la synthèse d'un pronom personnel, il est utile de présenter quelques faits linguistiques liés à la pronominalisation en italien et dont l'importance peut varier, selon que l'on se situe du point de vue de l'analyse ou de la génération.

Tableaux des formes pronominales (cf. page 34)

- Chaque tableau représente une catégorie syntaxique, et regroupe les formes pronominales propres à cette catégorie (les catégories [di-objets] et [da-objet] partagent le même tableau) ;
- Dans chaque tableau, la personne est portée en abscisses ('ps' symbolise "personne du singulier" ; 'pp' symbolise "personne du pluriel") ;
- Les valeurs portées en ordonnées dépendent de la catégorie syntaxique considérée ; on peut trouver :
  - le genre (masc = masculin, fem = féminin). La distinction, quand elle existe, n'a lieu qu'aux troisièmes personnes,
  - la nature sémantique (humain / non humain). Une forme non humain n'existe qu'à la troisième personne.

Position préverbale et non préverbale

Comme il a été rappelé, deux formes peuvent être utilisées en italien : les formes atones et les formes toniques. Parfois, l'emploi de l'une de ces formes a un but emphatique, mais la plupart du temps son utilisation correspond à une contrainte syntaxique (comme pour les pronoms réflexifs). Une forme tonique occupe la position syntaxique du groupe nominal qu'il remplace. Par contre une forme atone, appelée particule préverbale et notée ppv[4], remplace un constituant entier (groupe nominal ou prépositionnel), et est placée dans le noyau verbal. La position d'un pronom dans la phrase pose un problème à la fois en analyse, où il est question de recherche de constituants dans un arbre, et en génération, comme on le verra dans les sections 3.2.1 et 3.2.2.1.

Position d'une ppv dans le noyau verbal

1) ppv = *loro*

4 Le terme "préverbal" est un abus de langage (comme pour les réflexifs) : la notation "ppv", qui est héritée du français (M. Gross (1969)), ne doit pas être prise au pied de la lettre, ainsi que nous allons le voir ci-dessous.

La ppv *loro* constitue toujours, quand elle est présente, le dernier élément du noyau verbal, quel que soit le mode de conjugaison du verbe :

[indicatif] *Ugo ha fatto piacere ai bambini. Ha regalato* ***loro*** *questo libro.*
(Ugo a fait plaisir aux enfants. [Il] leur a offert ce livre.)

[gérondif] *Ugo ha fatto piacere ai bambini regalando* ***loro*** *questo libro.*
(Ugo a fait plaisir aux enfants en leur offrant ce livre.)

2) ppv ≠ *loro*

Quand le verbe est conjugué à l'indicatif, au subjonctif ou au conditionnel, la ppv est placée entre l'élément de négation (s'il existe) et le verbe (ou auxiliaire) conjugué :

Ugo ha fatto piangere il bambino. [Non] ***gli*** *ha regalato questo libro.*
(Ugo a fait pleurer l'enfant. [Il] [ne] lui a pas offert ce livre.)

Par contre, si le verbe est à l'infinitif, au gérondif ou à l'impératif, alors la ppv se comporte comme un suffixe. Si le verbe est à l'infinitif, le *e* final de sa terminaison tombe :

[gérondif] *Ugo ha fatto piacere al bambino regalando****gli*** *questo libro.*
(Ugo a fait plaisir à l'enfant en **lui** offrant ce livre.)
[infinitif] *Il bambino piange. Ugo non vuole regalar****gli*** *questo libro.*
(L'enfant pleure. Ugo ne veut pas **lui** offrir ce livre.)
[impératif] *Il bambino piange. Regala****gli*** *questo libro !*
(L'enfant pleure. Offre-**lui** ce livre !)

De plus, si le radical du verbe est monosyllabique, alors il y a doublement de la consonne initiale de la ppv :

di' + lo --> dillo (dis-le)

Les phénomènes de "suffixation" de la ppv peuvent poser des problèmes en analyse. En effet, il faut tenir compte de toutes les formes résultant de l'agglutination verbe + ppv ; en d'autres termes, le système doit :

- reconnaître une forme verbe + ppv,
- retrouver à quelle catégorie appartient la ppv.

Pour effectuer ces opérations, deux approches sont possibles :

a) On peut envisager la constitution d'un dictionnaire contenant toutes les formes agglutinées compatibles avec un verbe donné, mais cette solution est très coûteuse : si on considère que pour une entrée verbale, on peut avoir jusqu'à 11 possibilités verbe + ppv[5] par conjugaison concernée (i.e. infinitif, gérondif et chacune des trois personnes de l'impératif), cela donne au maximum 5 x 11 = 55 entrées pour un verbe. Par exemple, pour le verbe *guardare* (regarder),

5 Il existe en tout 12 formes préverbales distinctes en italien, mais la ppv *l'* n'apparaît jamais en suffixe verbal : **regalandol'* (offrant + l')

qui se construit avec un objet direct (et qui peut avoir un modifieur de lieu), les formes suivantes sont possibles :

INFINITIF :

guardarmi (me regarder)	*guardarti* (te regarder)	*guardarlo* (le regarder)
guardarla (la regarder)	*guardarli* (les$_{masc}$ regarder)	*guardarle* (les$_{fem}$ regarder)
guardarci (nous/y regarder)	*guardarvi* (vous / y regarder)	*guardarne* (en regarder)

GERONDIF :

guardandomi (me regardant)	*guardandoti* (te regardant)	*guardandolo* (le regardant)
guardandola (la regardant)	*guardandoli* (les$_{masc}$ regardant)	*guardandole*(les$_{fem}$ regardant)
guardandoci (nous/y regardant)	*guardandovi* (vous / y regardant)	*guardandone* (en regardant)

IMPERATIF :

guardami (regarde-moi)	*guardati* (regarde-toi)	*guardalo* (regarde-le)
guardala (regarde-la)	*guardali* (regarde-les$_{masc}$)	*guardale* (regarde-les$_{fem}$)
guardaci (regarde-nous/y)	*guardane* (regarde-en)	
guardiamoti ([litt] : regardons-toi)	*guardiamolo* (regardons-le)	*guardiamola* (regardons-la)
guardiamoli (regardons-les$_{masc}$)	*guardiamole* (regardons-les$_{fem}$)	
guardiamoci (regardons-nous/y)	*guardiamovi* (regardons-vous / y)	
guardiamone (regardons-en)		
guardatemi (regardez-moi)	*guardatelo* (regardez-le)	*guardatela* (regardez-la)
guardateli (regardez-les$_{masc}$)	*guardatele* (regardez-les$_{fem}$)	
guardateci (regardez-nous/y)	*guardatevi* (regardez-vous / y)	
guardatene (regardez-en)		

De plus, il faut tenir compte des combinaisons issues de la présence simultanée de plusieurs ppv. Ce phénomène est décrit au paragraphe suivant.

Exemple : ***regalarglielo*** (**le lui** offrir)

b) La deuxième solution consiste à rétablir dans un mot les unités lexicales verbe et ppv(s), par des opérations qui reconnaissent :

1) le radical du verbe,

2) le ou les objets représentés dans la terminaison de ce mot..

Par contre, en génération, ce problème ne se rencontre pas : le "suffixage" n'est qu'une concaténation verbe + ppv; le système effectue cette opération, qui n'a aucune influence sur la pronominalisation, en dernière étape : cette opération de "bas niveau" est l'un des derniers calculs à effectuer pour obtenir la phrase finale.

Présence simultanée de plusieurs ppv

La cooccurrence de plusieurs ppv est un phénomène important en génération (cf. 3.2.2.1) car la synthèse d'une ppv peut être conditionnée par la présence d'une autre ppv. Ainsi, la présence simultanée de *ci*, ppv correspondant à un [dir-objet] de la première personne du pluriel, et de *ci*, ppv représentant un locatif de destination, est interdite :

**Se andremo a Parigi, Ugo ci ci vedrà.*
(Si nous allons à Paris, Ugo nous y verra.)

Quand deux ppv sont présentes simultanément, la voyelle finale de la première des deux est parfois altérée. Ainsi, *ci* devient *ce* devant *lo, la, le, li, l'* ou *ne :*

	Ugo mette il libro in biblioteca.
	(Ugo met le livre dans la bibliothèque.)
[dir-objet -> lo]	*Ugo lo mette in biblioteca.*
	(Ugo le met dans la bibliothèque.)
[loc-objet -> ci]	*Ugo ci mette il libro.*
	(Ugo y met le livre.)
[double pro.]	*Ugo ce lo mette.*
	(Ugo l'y met.)

Devant une ppv représentant un [dir-objet] de la 3ème personne, ou un [di-objet], il y a également altération des ppv *mi, ti, vi*, et du réflexif *si*, respectivement en : *me, te, ve* et *se*. Ces altérations n'ont pas de conséquence en génération ; elles ne posent pas d'avantage de problèmes en analyse.

Par contre, la présence simultanée de l'une des ppv [a-objet] *le* ou *gli* et de l'une des ppv *lo, la, le, li* ou *ne*, conduit à une altération qui pose des problèmes en génération et en analyse. En effet, en présence de *lo, la, le, li* ou *ne*, la ppv *gli* ou *le* devient *glie*. *Glie* s'"agglutine" alors à la ppv [dir-objet] ou [di-objet] ; cette forme représente aussi bien une entité masculin-singulier qu'une entité féminin-singulier :

	A-OBJET = EVA	**A-OBJET = UGO**
	Max parla di Capri a Eva. (Max parle de Capri à Eva)	*Max parla di Capri a Ugo.* (Max parle de Capri à Ugo)
[di-objet -> ne]	*Max ne parla a Eva.* (Max en parle à Eva)	*Max ne parla a Ugo.* (Max en parle à Ugo)
[a-objet -> ppv]	*Max le parla di Capri.* (Max lui(fem) parle de Capri)	*Max gli parla di Capri.* (Max lui(mas) parle de Capri)
[double pro.]	*Max gliene parla.* (Max lui(mas + fem) en parle)	

L'ordre dans lequel les différentes ppv sont placées dans le noyau verbal est fixe en italien comme en français. En italien, cet ordre est le suivant :

mi, ti, ci1, vi1 < gli, le < ci2, vi2 < si < ne < lo, la, le, li < VERBE *< loro*

- *mi, ti, ci1* et *vi1* sont les ppv de la première et deuxième personne [dir-objet] ou [a-objet],
- *gli* et *le* sont les ppv [a-objet] de la troisième personne du singulier,
- *ci2* et *vi2* sont les deux formes possibles pour une ppv [loc-objet],

- *si* est le pronom réflexif de la troisième personne (singulier ou pluriel),
- *ne* est la ppv [di-objet] ou [da-objet],
- *lo*, *la*, *le* et *li* sont les ppv [dir-objet] de la troisième personne.

Cet ordre a été étudié en particulier par V. Lo Cascio (1969) pour l'italien, et par M. Gross (1969) pour le français. Il est important en analyse. En particulier, il est utile pour la résolution d'ambiguïtés comme la distinction entre le *si* réflexif et le *si* pronom indéfini (qui n'est pas étudié ici) : sachant que

position (si_{refl}) < position [dir-objet]$_{3ep}$ < position (si_{indef})

alors il est possible d'analyser les phrases suivantes, qui diffèrent uniquement par la position de *si* :

Se lo mangia.
([il + elle] se le mange.)
Lo si mangia.
(On le mange.)

Par contre, l'ordre de placement des différentes ppv ne joue aucun rôle en génération : il s'agit encore une fois d'une opération morphologique de bas niveau.

Enfin, le problème du suffixage de plusieurs ppv à une forme verbale à l'infinitif, au gérondif ou à l'impératif a été évoqué au paragraphe précédent, pour un système d'analyse. Quand plusieurs ppv jouent le rôle d'une terminaison verbale, elles occupent, l'une par rapport à l'autre, la même place qu'en position préverbale :

*Ugo **me lo** regala.* (Ugo **me l'masc** offre.)
*Ugo vuole regalar**melo**.* (Ugo veut **me l'masc** offrir.)

Par conséquent, cette opération est analogue à celle qui a été présentée au paragraphe précédent, sur la concaténation d'une ppv à un verbe : il s'agit ici aussi d'une opération de bas niveau.

Elision d'une ppv

Elle concerne les ppv *mi, ti, lo, la, vi, ci* et le réflexif *si*. Dans le cas des ppv *mi, ti, vi, ci* et du réflexif *si*, l'élision ne se produit que si la voyelle initiale du mot suivant la ppv est un *i* (cf. J. Brunet (1985)). Cette élision n'est pas appliquée systématiquement à l'écrit, à l'inverse du français. Elle ne pose aucun problème, que ce soit en analyse ou en génération, puisque l'élision de l'une de ces ppv n'engendre aucune perte d'information sur le genre, le nombre ou la personne, et n'entraîne donc pas d'ambiguïté : par exemple, la forme élidée *m'* ne peut correspondre qu'à la ppv de la 1ère personne du singulier *mi*.

Par contre, l'élision des pronoms *lo* et *la* en *l'* ne relève pas d'une simple substitution : elle soulève des problèmes d'ambiguïté puisque *l'* = *lo* ou *la*. En fait, l'élision de *la* en *l'* est assez systématique si la voyelle initiale du mot suivant est un *a*. Des études (voir entre autres J. Brunet (1985)) ont montré que l'élision de *la* en *l'* est beaucoup plus fréquente que celle de *lo*. La première est souvent obligatoire devant la voyelle *a* :

**Max la aspetta* --> *Max l'aspetta* .

(Max l'attend.)

alors que la seconde, dans les mêmes conditions (i.e. devant la voyelle *o*) est beaucoup moins représentée à l'écrit :

Max (lo + l') offrirà a Ugo.
(Max l'offrira à Ugo.)

Ces problèmes en génération (on retrouve les mêmes en analyse) sont exposés à la section 3.2.1.

3.1.3 DISTINCTION SUJET ET OBJET

La description ci-dessus a permis de cerner les différents phénomènes liés à la pronominalisation et qui sont pertinents en génération de texte. Le dernier point que nous soulèverons ici, avant de commencer l'étude proprement dite de la synthèse des pronoms, concerne la distinction qu'il faut faire entre la synthèse du sujet et celle des objets. Des écoles linguistiques, comme celle des générativistes (cf. N. Chomsky (1981,1982)) mettent l'accent sur la différence entre le sujet et les objets, en mettant en évidence des propriétés particulières des sujets. Elles s'opposent, entre autres, à la théorie HPSG (Head-Driven Phrase Structure Grammar, cf. C. Pollard (1984, 1985)), qui insistent plutôt sur les ressemblances. Nous sommes en accord avec la GBT (Government-Binding Theory), car certaines des propriétés spécifiques du sujet affectent directement la génération. Il s'agit notamment de l'effacement du sujet : le sujet, contrairement aux objets, peut être effacé (cf. section 3.3) dans les phrases non infinitives en italien comme dans les autres langues romanes, sauf le français (cf. note 1), des phénomènes de focus et ceux de parallélisme de structure, qui seront étudiés au chapitre 5.

Parmi les autres propriétés qui différencient sujet et objets, et qui sont d'un ordre plus général, on peut citer : le niveau des relations existant avec le verbe (le sujet est toujours en accord avec celui-ci, contrairement à l'objet), et le rôle spécifique du sujet dans la structure de la phrase (le sujet est en effet le seul complément de verbe qui soit obligatoire, si on exclut, bien sûr, les prédicats "météorologiques" comme *Piove* (Il pleut)).

Dans la suite de ce chapitre, nous présenterons successivement certains aspects de la synthèse des pronoms personnels objets, puis nous nous intéresserons à la synthèse du sujet ; en particulier nous mettrons en évidence les liens existant, en génération au moins, entre pronominalisation et effacement du sujet (pour L. Rizzi (1982), l'effacement du sujet n'est pas directement relié à la pronominalisation. Sa théorie est très brièvement rappelée dans la note 18 de ce chapitre).

3.2 SYNTHESE DES PRONOMS PERSONNELS OBJETS

Nous avons évoqué (cf. chapitre 1) que la pronominalisation met en jeu de nombreuses dépendances non locales et dépendances croisées. Nous présentons donc tout d'abord quelques exemples de ces dépendances, puis nous indiquerons les différentes opérations à effectuer pour la synthèse d'un pronom personnel.

3.2.1 DEPENDANCES NON LOCALES ET DEPENDANCES CROISEES

Rappelons qu'une dépendance non locale s'observe quand la synthèse d'un élément X dépend de celle d'un autre élément Y, et que l'on parle de dépendance croisée quand la synthèse de X et de Y dépendent l'une de l'autre. Une présentation de ces phénomènes peut être trouvée dans L. Danlos & F. Namer (1988). Nous présentons ci-dessous un exemple de dépendance non locale et un exemple de dépendance croisée en italien. Nous traiterons au chapitre 5 des dépendances croisées spécifiques du français.

Exemple 1 :

La synthèse du [a-objet] et du [di-objet] est une illustration de la notion de dépendance non locale. Comme nous l'avons vu, la forme d'un pronom préverbal [a-objet] de la 3ème personne du singulier est *gli* si ce pronom représente une entité masculine, et *le* s'il s'agit d'une entité féminine :

[a-objet] masculin :	*Max parla di Capri a Ugo.* -> *Max **gli** parla di Capri.* (Max parle de Capri à Ugo -> Max **lui**masc parle de Capri.)
[a-objet] féminin :	*Max parla di Capri a Eva.* -> *Max **le** parla di Capri.* (Max parle de Capri à Eva. -> Max **lui**fem parle de Capri.)

Cependant, si le [di-objet] est synthétisé en un pronom préverbal (*ne*), alors la forme de la ppv [a-objet] de la 3ème personne est *glie*, quel que soit le genre de l'entité représentée par le pronom ; cette forme s'agglutine à la ppv [di-objet] :

	Max parla di Capri a (Ugo + Eva). (Max parle de Capri à (Ugo + Eva).)
[di-objet] & [a-objet] --> ppv	*Max **gliene** parla.* (Max **lui**masc + fem en parle.)

Notons que l'altération *gli / le* --> *glie* doit être prise en compte tout de suite : nous verrons (section 3.2.2.2) que le fait que *glie* est invariable en genre, contrairement à *gli* ou *le*, influe sur la décision de synthétiser ce pronom, puisqu'il y a augmentation du nombre de ses antécédents possibles. La synthèse du [a-objet] dépend donc bien de celle du [di-objet].

Exemple 2 :

Illustrons la notion de "dépendances croisées" en prenant l'exemple de la synthèse du verbe et de l'objet direct[6].

a) La synthèse du verbe dépend de celle de l'objet direct pour deux raisons. Premièrement, l'auxiliaire de conjugaison d'un verbe qui se construit avec un objet direct est *avere* (avoir) à un temps composé, sauf si l'objet direct est synthétisé en un un pronom réflexif atone ; dans ce cas, l'auxiliaire est *essere* (être) :

(5) *Ugo **ha** odiato Eva.*
(Ugo a détesté Eva.)

6 Cet exemple est donné ici en italien. Mais les résultats présentés sont vrais également pour le français.

(5a) *Ugo si è odiato.*
(Ugo s'est détesté.)

Deuxièmement, il y a accord en genre et nombre entre le participe passé d'un verbe conjugué à un temps composé et un objet direct synthétisé en un pronom personnel préverbal :

(6) *Ugo, Max l'ha odiato.*
(Ugo, Max l'a détesté.)
(6a) *Eva, Max l'ha odiata.*
(Eva, Max l'a détestée.)

Notons que les altérations dans la conjugaison du verbe (changement d'auxiliaire, accord en genre et nombre) ne sont observées, en italien, qu'en cas de synthèse d'un pronom (réflexif ou personnel) objet direct préverbal. En d'autres termes, elles n'ont pas lieu en cas de synthèse d'une forme tonique :

(5b) *Ugo **ha** odiato molto se stesso.*
[litt] (Ugo a beaucoup détesté soi-même.)
(?Ugo a beaucoup détesté lui-même.)
(6b) *Max ha odiato solo lei.*
(Max n'a détesté qu'elle.)

b) La synthèse de l'objet direct dépend de celle du verbe, car la conjugaison du verbe peut déterminer la forme du pronom à synthétiser, ainsi que les informations morphologiques contenues dans la forme pronominale. Prenons l'exemple de la synthèse d'un objet direct de la 3ème personne du singulier en un pronom personnel. Comme l'indique le tableau, page 34, la forme d'un pronom masculin-singulier est *lo*, celle d'un pronom féminin-singulier est *la*.

*Ugo, Max **lo** vede.*
(Ugo, Max le voit.)
*Eva, Max **la** vede.*
(Eva, Max la voit.)

Cependant, si le verbe est conjugué à un temps composé avec *avere* (avoir), ou si sa voyelle initiale est *a*, alors les pronoms *lo* et *la* sont élidés en *l'* :

Ugo, Max l'ha visto.
(Ugo, Max l'a vu.)
Eva, Max l'ha vista.
(Eva, Max l'a vue.)

A première vue, cette élision n'affecte en rien le processus de pronominalisation, puisque l'information concernant le genre, préalablement contenue dans les formes *lo* et *la*, est maintenant portée par le participe passé. Soulignons qu'un participe passé est toujours variable en genre et en nombre (le suffixe *a* indique le féminin-singulier, *o* : le masculin-singulier, *e* : le féminin-pluriel et *i* : le masculin-pluriel).

Cependant, considérons le schéma de texte (S7) :

(S7) *(UGO portare PIA a casa. MAX accogliere PIA con gentilezza.)*
(UGO amener PIA à la maison. MAX accueillir PIA avec gentillesse.)

Si les verbes sont conjugués à un temps composé, la synthèse de (S7) aboutit au texte (7a) :

(7a) *Ugo ha portato Pia a casa. Max l'ha accolta con gentilezza.*
(Ugo a amené Pia à la maison. Max l'a accueillie avec gentillesse.)

qui est non ambigu. Par contre, si les verbes sont conjugués à un temps simple, la synthèse de (S7), où la deuxième occurrence de PIA serait pronominalisée, conduirait à (7b) :

(7b) **Ugo porterà Pia a casa. Max l'accoglierà con gentilezza.*
(*Ugo amènera Pia à la maison. Max l'accueillera avec gentillesse.)

qui est ambiguë, puisqu'ici la forme élidée *l'*, dans la deuxième phrase, peut référer aussi bien à *Ugo* qu'à *Pia*. Il n'y a pas ici de participe passé qui permette de retrouver le genre du pronom élidé. Par conséquent, la synthèse du pronom personnel est interdite. La synthèse de (S7) peut être alors :

(7c) *Ugo porterà Pia a casa. Max accoglierà la ragazza con gentilezza.*
(Ugo amènera Pia à la maison. Max accueillera la jeune fille avec gentillesse.)

où la 2ème occurrence de PIA est synthétisée en un groupe nominal.

La différence de synthèse observée entre (7a) et (7c), due uniquement à la conjugaison du verbe, illustre non seulement le fait que la synthèse de l'objet direct dépend de celle du verbe, mais également que les opérations morphologiques (e.g. l'élision) doivent être prises en compte assez tôt dans un système de génération. Exécuter de telles opérations à la dernière étape d'un système de génération revient à produire des textes comme (7b), qui sont ambigus.

Les points a) et b) conduisent à la conclusion suivante : puisque la synthèse du verbe et celle de l'objet direct dépendent l'une de l'autre, la seule solution envisageable est de "morceler" la synthèse de ces constituants. Ainsi :

1) On détermine si l'objet direct doit être synthétisé en un pronom réflexif préverbal (cf. section 3.1.1). Si c'est le cas, on marque le verbe comme devant être conjugué avec l'auxiliaire *essere* (être).

2) On conjugue le verbe sans tenir compte de l'accord possible entre un participe passé et un pronom objet direct. Le résultat de l'étape 1 permet de conjuguer le verbe à un temps composé avec l'auxiliaire approprié.

3) On synthétise l'objet direct si cela n'a pas déjà été fait en 1 (i.e. il a été synthétisé en une forme réflexive). La forme du verbe calculée en 2 est utilisée pour déterminer si l'objet direct sera synthétisé en un pronom personnel.

4) On complète éventuellement la synthèse du verbe, i.e. on effectue l'accord en genre et nombre entre le participe passé, s'il existe (information fournie par 2), et l'objet direct synthétisé en une ppv, s'il existe (information provenant de 3).

Cet exemple, qui illustre la notion de dépendances croisées, montre combien la pronominalisation est un phénomène complexe, et combien le mécanisme de formation d'un pronom, qui est décrit ci-dessous, doit tenir compte de la synthèse des autres éléments de la phrase.

3.2.2 PREMIERES OPERATIONS A EFFECTUER POUR LA SYNTHESE D'UN PRONOM

La première question que l'on se pose en ce qui concerne la pronominalisation dans un système de génération est : quand envisager la synthèse d'un pronom personnel ? En théorie, la réponse est double :

1 - Rappelons que si T est un schéma de texte tel que :

$$T = Ph_1\ punct_1\ Ph_2\ punct_2 \ldots Ph_i\ punct_i \ldots punct_{n-1}\ Ph_n\ punct_n$$

alors chaque schéma de phrase est synthétisé à tour de rôle. En d'autres termes :

$$\text{synthèse}\ (Ph_1) < \text{synthèse}\ (Ph_2) < \ldots < \text{synthèse}\ (Ph_i) < \ldots < \text{synthèse}\ (Ph_n)$$

On envisage de synthétiser une étiquette TOKi, apparaissant dans Phi, en un pronom personnel quand TOKi a été mentionnée dans le contexte gauche de Phi (i.e. dans l'une des Phj, avec $j \leq i$).

2 - On envisage de synthétiser TOKi en un pronom personnel quand une autre occurrence de TOKi apparaît dans la phrase Phi en cours de synthèse, pourvu que les deux occurrences de TOKi respectent certaines contraintes. Ainsi, à partir de :

(LUC parler à MERE (de LUC).)

où apparaissent deux occurrences de LUC, nous engendrons :

Luc$_i$ a parlé à sa$_i$ mère.[7].

tout en évitant :

**Il$_i$ a parlé à la mère de Luc$_i$..*

dont l'inacceptabilité a été largement décrite dans la littérature (cf. R. Langacker (1969)). De même, à partir de :

(MOI parler de MERE (de MARIE) à MARIE.)

nous engendrons :

J'ai parlé à Marie$_i$ de sa$_i$ mère.

et non (cf. R. Kayne (1971)) :

**Je lui$_i$ ai parlé de la mère de Marie$_i$.*

7 L'indice i, commun à *Luc* et à *sa* est mis pour indiquer que les deux éléments sont coréférents.

Par contre, nous ne traitons pas les phénomènes dits de "pronominalisation gauche", étudiés entre autres par G. Fauconnier (1974) et T. Reinhard (1983) :

(a1) *Chaque fois qu'il$_i$ est malade, Jean$_i$ est inquiet.*
(b1) *Près de lui$_i$, Jean$_i$ voit un serpent.*

En effet :

- Dans un système de génération de textes, les phénomènes de pronominalisation gauche ne se rencontrent que rarement : exception faite de la première phrase d'un texte, la pronominalisation gauche est interdite dès qu'elle débouche sur un pronom qui a un antécédent possible dans le contexte gauche. Ceci est illustré par les exemples ci-dessous, dans lesquels on interdit respectivement la pronominalisation gauche du sujet (a2) et celle de l'objet (b2) :

(a2) *Max est absent. Chaque fois qu'il est malade, Jean est inquiet.*
(b2) *Max est allongé dans l'herbe. Près de lui, Jean voit un serpent.*

Le pronom sujet *il* de la 2ème phrase de (a2) peut référer *Max* (antécédent gauche) ou à *Jean* (antécédent droit). Quant au pronom objet *lui* de la deuxième phrase de (b2), il a pour antécédent le plus probable *Max*.[8]

- Contrairement à un module d'analyse, qui opère sur des phrases ou des textes où l'ordre des constituants est fixe, l'ordre d'apparition des éléments synthétisés est décidé par le système de génération. Dans les exemples (a1) et (b1), les adverbiaux (i.e. la subordonnée de (a1) et le groupe prépositionnel modifieur de (b1)) apparaissent en tête de phrase dans un système d'analyse. En génération, c'est le système qui décide où les placer. A priori, les adverbiaux sont en fin de phrase : les conditions stylistiques qui aboutiraient à la décision de mettre un groupe adverbial en tête de phrase dépassent le cadre de ce travail.

Pour les raisons qui viennent d'être évoquées, nous laissons donc de côté pour l'instant le phénomène de pronominalisation gauche: on essaie de synthétiser TOKi en un pronom seulement s'il a déjà été mentionné.

Avant d'aller plus loin, citons un problème qui ne sera pas, lui non plus, traité ici : il s'agit du calcul des antécédents conjoints d'un pronom pluriel. Le terme "conjoints" indique que l'antécédent du pronom est soit une étiquette pluriel, de même nombre que ce pronom, soit un ensemble d'étiquettes singulier ou pluriel. Ainsi, l'antécédent d'un pronom (sujet ou objet) de la première personne du pluriel est l'association du pronom de la première personne du singulier et d'une, ou plusieurs entités d'une autre personne (du singulier ou du pluriel) :

[1ère pers. sing + 2ème pers. sing] *Andrò al cinema con te . Max ci seguirà.*
([J'] irai au cinéma avec toi. Max nous suivra.)

8 Bien entendu, l'introduction d'une sémantique dans le contexte gauche influe sur l'interprétation préférentielle du pronom. Ainsi, dans (a3) :
(a3) *Max a la varicelle. Chaque fois qu'il est malade, Jean est inquiet.*
le seul antécédent possible de *il* est *Max*, alors que dans (a4) il s'agit de *Jean* :
(a4) *Le docteur est déjà venu six fois aujourd'hui. Chaque fois qu'il est malade, Jean est inquiet.*

[1ère pers. sing + 3ème pers. sing] *Ugo mi ha presentato a Eva. Siamo andati al cinema.*
(Ugo m'a présenté à Eva. Nous sommes allés au cinéma.)

Dans cet exemple, l'antécédent du pronom sujet effacé de la première personne du pluriel, dans la deuxième phrase, peut représenter l'un des ensembles suivants :

mi + Ugo ; *mi + Ugo + Eva* ; *mi + Eva*

De la même façon, l'antécédent d'un pronom de la troisième personne du pluriel est soit une entité de la troisième personne du pluriel, soit l'association de plusieurs entités de la troisième personne, au pluriel ou au singulier. Des études ont été faites à ce sujet, entre autres, pour la résolution d'anaphores dans un système d'analyse, par C. Eschenbach, C. Habel, M. Herweg et K. Rehkämper (1989).

Deux conclusions se dégagent finalement des remarques ci-dessus :

- si l'étiquette à synthétiser représente le(s) locuteur(s) ou l' (les) interlocuteur(s), alors la seule opération à effectuer est le calcul d'un pronom de la première ou de la deuxième personne (noté pronom du dialogue) ;

- si l'étiquette TOKi à synthétiser représente une entité de la troisième personne, on considère alors la synthèse de TOKi en un pronom personnel de la 3ème personne seulement si TOKi apparaît dans un schéma de phrase préalablement synthétisé, et que les deux occurrences de TOKi représentent exactement le même objet ou le même individu : on ne s'occupe que de la synthèse des pronoms définis. La synthèse d'un pronom représentant une partie d'un objet, ou un sous-ensemble d'un groupe :

Ugo ha tre gattini. Ne dà uno a Eva.
(Ugo a trois chatons. Il en donne un à Eva.)
Ugo compra un libro. Ne legge la metà.
(Ugo achète un livre. Il en lit la moitié.)

ou encore faisant intervenir des quantifieurs :

Ugo ha molti amici. Ne ha presentati alcuni a Max.
(Ugo a beaucoup d'amis. Il en a présenté certains à Max.)

n'est pas étudiée ici.

Les opérations à effectuer pour déterminer si un pronom peut être synthétisé sont :

1) calculer la forme de ce pronom,

2) déterminer si la synthèse de cette forme n'engendre pas d'ambiguïté ; en d'autres termes, compter le nombre d'antécédents (ou d'interprétations) de ce pronom.

3.2.2.1 CALCUL DE LA FORME DU PRONOM

Les facteurs intervenant dans la forme d'un pronom de la troisième personne sont :

1) sa position syntaxique ;

2) le nombre et de la nature sémantique de l'étiquette à synthétiser. Ces informations sont contenues dans la définition de l'étiquette. Le rôle de la nature sémantique de l'étiquette dans le calcul de la forme du pronom est parfois liée au verbe de la phrase, comme on le verra dans les exemples ci-dessous ;

3) le genre du groupe nominal qui a été synthétisé pour l'occurrence précédente de l'étiquette. En italien comme en français, le genre est soit masculin, soit féminin, et il s'agit d'une information lexicale et non sémantique. Considérons en effet la définition de l'étiquette TABLE1 :

TABLE1 := **TABLE**
NOMBRE = 1
DEFINI = oui

TABLE1 peut être synthétisé en un groupe nominal féminin : *la tavola* ou masculin : *il tavolo* (la table). Le genre d'un pronom est généralement le même que celui de la précédente occurrence de l'étiquette[9] :

*La tavola è rotta. Ugo **la** ripara.*
*Il tavolo è rotto. Ugo **lo** ripara.*
(La table est cassée. Ugo la répare.)

4) la synthèse du verbe (e.g. le choix entre les formes *lo* et *l'* dépend de la voyelle initiale du verbe, comme il a été montré dans l'exemple 2 de la section 3.2.1) ;

5) la synthèse des autres compléments (e.g. le choix entre les formes *glie* et *gli* dépend de la synthèse de l'objet prépositionnel introduit par *di*, comme on l'a vu dans l'exemple 1 de la section 3.2.1).

Supposons maintenant que TOK, dont la synthèse en un pronom personnel est envisagée, soit une étiquette représentant une entité de la 3ème personne en position objet. Voici, pour chaque fonction syntaxique, les conditions liées au calcul des formes pronominales :

Dir-objet (objet direct): Si l'objet direct est construit avec un verbe symétrique comme *confondere*, alors la forme du pronom dépend de la synthèse de l'objet prépositionnel (cette dépendance croisée entre le [dir-objet] et le [prep-objet] sera développée au chapitre 5). Sinon, deux cas de figure se présentent :

- le nombre de TOK est pluriel, alors la forme du pronom ne dépend que du genre de TOK : *le* au féminin, *li* au masculin :

9 Si l'étiquette représente un humain, alors le genre du pronom est calculé en fonction du sexe de cet humain : *Ugo è tornato dall'America. **Questa canaglia** è rimasta laggiù per un anno. Pare che si sia arricchito.*
(Ugo est revenu d'Amérique. **Cette canaille** est restée un an là-bas. Il paraît qu'[il] s'est enrichi.)

Ho visto i fratelli Rossi. Li ho salutati.
([J'] ai vu les frères Rossi. [Je] les ai salués.)

- le nombre de TOK est singulier, alors la forme du pronom dépend du genre de TOK et de la voyelle initiale du verbe (cf. exemple 2 de la section 3.2.1) :

si le verbe est *avere*, alors la forme est *l'*, quel que soit le genre de TOK :

Ho visto Ugo. L'ho salutato.
([J'] ai vu Ugo. [Je] l'ai salué.)

si la voyelle initiale du verbe est *a* et que le genre de TOK est féminin, alors la forme est également *l'* :

Eva è carina. Max l'adora.
(Eva est mignonne. Max l'adore.)

sinon, la forme du pronom est *la* au féminin et *lo* au masculin.

Ugo è carino. Max lo adora.
(Ugo est mignon. Max l'adore.)
Eva è odiosa. Max la evita quando può.
(Eva est odieuse. Max l'évite quand il peut.)

Di-objet (objet introduit par la préposition *di*) : Si TOK représente un individu (i.e. la nature sémantique de TOK a pour valeur "humain"), alors on a le choix entre la forme préverbale *ne*, qui est invariable en genre et en nombre et qui peut représenter un humain ou un non humain, et l'une des formes toniques : *lui, lei, loro* (précédée de *di*), qui ne représentent que des humains :

Ugo è innamorato di Eva.
(Ugo est amoureux de Eva.)
[di-objet --> *ne*] *Ugo ne è innamorato.*
(Ugo en est amoureux.)
[di-objet --> *lei*] *Ugo è innamorato di lei.*
(Ugo est amoureux d'elle.)

Le problème des conditions syntaxiques du choix entre *ne* et *di lui, di lei, di loro* a été discuté dans B. Lamiroy (1988) et N. Ruwet (1990). Mais ce problème concerne l'analyse, pas la génération : on peut choisir systématiquement de synthétiser un [di-objet] humain en *di lui / di lei / di loro*, qui est plus précis que *ne*, et souvent stylistiquement plus élégant. Par conséquent, dans le cas d'une étiquette humaine pluriel, la forme calculée est *loro ;* si le nombre est singulier, alors elle dépend du genre de TOK (*lei* au féminin , *lui* au masculin) :

Ugo ha parlato (dei Polacchi + delle Polacche) => Ugo ha parlato di loro.
(Ugo a parlé (des Polonais + des Polonaises) => Ugo a parlé d'(eux + elles).)

Ugo ha parlato di Max => Ugo ha parlato di lui.
(Ugo a parlé de Max => Ugo a parlé de lui.)

Ugo ha parlato di Eva => Ugo ha parlato di lei.
(Ugo a parlé de Eva => Ugo a parlé d'elle.)

Si, par contre, TOK est non-humain, alors on a le choix entre la forme préverbale *ne* et une des formes toniques *esso, essa, essi, esse* variables en genre et en nombre, et qui ne représentent que des non-humains :

Ugo parla di questo libro.
(Ugo parle de ce livre.)
[di-objet --> *ne*] *Ugo ne parla.*
(Ugo en parle.)
[di-objet -> *esso*] *Ugo parla di esso.*
([litt] Ugo parle de lui$_{non\text{-}hum.}$)
(=> Ugo parle de celui-ci.)

Contrairement à ce qui se passe pour les humains, on préfère utiliser ici la forme préverbale :

Ugo parla (della città + delle città + del paese + dei paesi) => Ugo ne parla.
(Ugo parle (de la ville + des villes + du pays + des pays) => Ugo en parle.)

Cependant, ce choix n'est pas toujours possible :

a) Si un objet direct est synthétisé en ppv de la troisième personne, alors la synthèse de *ne* ne peut être envisagée, puisque la présence simultanée de ces ppv est jugée incorrecte par plusieurs auteurs (L. Serianni (1988), A. Elia et alii (1981)[10]) :

Ugo accusa Eva dell'omicidio.
(Ugo accuse Eva de l'homicide.)
[di-objet -> ne] *Ugo ne accusa Eva.*
(Ugo en accuse Eva.)
[dir-objet -> lo] *Ugo l'accusa dell'omicidio.*
(Ugo l'accuse de l'homicide.)
[dir-objet & di-objet -> ppv] **Ugo ne l'accusa.*
(Ugo l'en accuse.)

b) Puisque la ppv *ne* est invariable en genre et en nombre, et qu'elle représente aussi bien des humains que des non humains, elle est souvent ambiguë :

**Questa pianta dà dei bei fiori. Max ne parla spesso.*
(*Cette plante donne des belles fleurs. Max en parle souvent.)

Quand on se trouve en présence de l'un des cas (a) ou (b), on envisage alors la synthèse de la forme tonique *esso, essa, essi, esse :*

Questa pianta dà dei bei fiori. Max parla spesso di essa.
(Cette plante donne des belles fleurs. Max parle souvent d'elle.)

10 Cette interdiction est liée à la fonction syntaxique occupée par *ne*. En effet, V. Lo Cascio (1969) ne refuse pas la présence simultanée de *lo* + *ne* quand *ne* représente un [da-objet] complément locatif de provenance :
Ugo trae il vino dalla botte => *Ugo ne lo trae.*
(Ugo tire le vin du tonneau => Ugo l'en tire.)

Mais cette solution n'est pas toujours possible. L'acceptabilité dépend de la nature sémantique de l'entité non humain considérée :

Ugo accusa Eva dell'omicidio --> ?Ugo l'accusa di esso.
(Ugo accuse Eva de l'homicide --> Ugo l'accuse de celui-ci.)

Si aucune forme pronominale n'est envisageable, alors on utilise un groupe nominal :

Ugo accusa Eva del delitto.
(Ugo accuse Eva du crime.)

A-objet (objet introduit par la préposition *a*): La forme du pronom personnel dépend du verbe utilisé et de la nature sémantique (humain / non-humain) de TOK. Les verbes peuvent être rangés dans l'une des trois catégories représentées par les modèles *credere, dare* et *pensare*[11] ; les informations concernant la forme du pronom qu'ils autorisent sont fournies par le lexique-grammaire (cf. chapitre 2) :

dare :

[a-objet = hum] *Eva dà un bacio a Ugo.*
=> *Eva **gli** dà un bacio.*
(Eva donne un baiser à Ugo.
=> Eva lui donne un baiser.)

[a-objet = n-hum] *Eva dà una spolverata alla casa.*
=> *Eva **le** dà una spolverata.*
(Eva donne un coup de torchon à la maison.
=> [litt] Eva lui donne un coup de torchon.
Eva y donne un coup de torchon.)

credere :

[a-objet = hum] *Eva crede alla veggente => Eva **le** crede.*
(Eva croit à la voyante => [litt] Eva lui croit.)

[a-objet = n-hum] *Eva crede a questa storia => Eva **ci** crede.*
(Eva croit à cette histoire => Eva y croit.)

pensare :

[a-objet = hum] *Eva pensa a Ugo => Eva pensa **a lui**.*
(Eva pense à Ugo => Eva pense à lui.)

[a-objet = n-hum] *Eva pensa a questa storia => Eva **ci** pensa.*
(Eva pense à cette histoire => Eva y pense.)

Les formes envisageables sont résumées dans le tableau ci-dessous, où le symbole 'ms' est mis pour masculin-singulier, 'fs' pour féminin-singulier, et 'pl' pour pluriel :

11 Cette notion a déjà été présentée dans la section 3.1.1 pour la synthèse des réflexifs, où on regroupe les verbes en deux catégories : ceux qui se comportent comme *dare*, et ceux qui se comportent comme *pensare*. La présence d'un troisième modèle est liée à la distinction entre pronoms humains et pronoms non humains.

	HUMAINS			NON-HUMAINS		
	ms	fs	pl	ms	fs	pl
dare	*gli*	*le*	*loro*	*gli*	*le*	*loro*
credere	*gli*	*le*	*loro*	*ci*	*ci*	*ci*
pensare	*a lui*	*a lei*	*a loro*	*ci*	*ci*	*ci*

La forme du pronom [a-objet] dépend également de la synthèse du [dir-objet] :

- si le [dir-objet] est synthétisé en une ppv de dialogue, alors la forme préverbale est interdite[12].

*Ugo mi presenta a Max => *Ugo **mi gli** presenta => Ugo **mi** presenta **a lui**.*
(Ugo me présente à Max => *Ugo me lui présente => Ugo me présente à lui.)

*Ugo mi presenta ai direttori => *Ugo **mi** presenta **loro** => Ugo **mi** presenta **a loro**.*
(Ugo me présente aux directeurs => *Ugo me leur présente => Ugo me présente à eux.)

- si le [dir-objet] est synthétisé en une ppv de la 3ème personne, alors le pronom [a-objet] envisagé à la 3ème personne du singulier est la forme préverbale invariable en genre *glie*, qui s'agglutine au pronom [dir-objet] :

	A-OBJET = EVA	A-OBJET = MAX
	Ugo dà la chiave a Eva. (Ugo donne la clé à Eva)	*Ugo dà la chiave a Max.* (Ugo donne la clé à Max)
[dir-objet -> ppv]	*Ugo la dà a Eva.* (Ugo la donne à Eva)	*Ugo la dà a Max.* (Ugo la donne à Max)
[a-objet -> ppv]	*Ugo le dà la chiave.* (Ugo lui(fem) donne la clé)	*Ugo gli dà la chiave.* (Ugo lui(mas) donne la clé)
[double pro.]	*Ugo gliela dà.* (Ugo la lui(mas + fem) donne)	

12 V. Lo Cascio (1969) n'interdit pas les combinaisons :
mi (gli + le) *ti (gli +le)* *ci (gli + le)* *vi (gli + le)*
Elles sont cependant jugées incorrectes dans la plupart des grammaires de l'italien moderne (comme par exemple M. Sensini (1988), L. Serianni (1988)) et on ne les trouve presque jamais dans la littérature (J. Brunet (1985)).

L'agglutination du pronom [a-objet] s'observe également quand le [di-objet] est synthétisé en ppv *ne*, comme on l'a vu dans le premier exemple de la section 3.2.1 :

Ugo parla di questa storia a (Eva + Max).
(Ugo parle de cette histoire à (Eva + Max).)
[double pro.] *Ugo gliene parla.*
(Ugo lui$_{mas + fem}$ en parle.)

Ceci amène, rappelons-le, à la conclusion que la synthèse du [a-objet] dépend également de celle du [di-objet].

Da-objet (objet introduit par la préposition *da*) : Si TOK est humain, le pronom envisagé est la forme tonique *lui, lei* ou *loro* en fonction du nombre et du genre de TOK (cf. di-objet) :

Il successo dipende da Max => Il successo dipende da lui.
(Le succès dépend de Max => Le succès dépend de lui.)

Sinon, la forme envisagée est la ppv *ne*, si le verbe de la phrase accepte cette pronominalisation (information fournie par le lexique-grammaire), ce qui est le cas, par exemple, des verbes régissant les locatifs de provenance, et des verbes comme *dipendere* :

I risultati dipendono dall'allenamento => I risultati ne dipendono.
(Les résultats dépendent de l'entraînement => les résultats en dépendent.)

mais pas des verbes comme *cominciare* :

*I guai sono cominciati dal ritorno di Ugo => *I guai ne sono cominciati.*
(Les ennuis ont commencé avec le retour de Ugo => *Les ennuis en ont commencé.)

Loc-objet (objet introduit par une préposition locative indiquant la destination ou le lieu où l'on est) : Parmi les prépositions introduisant un locatif, se trouve la préposition *da*, quand elle signifie *chez :*

Ugo va da Eva.
(Ugo va chez Eva.)

Dans ce cas, *da* introduit uniquement un objet humain.

La forme pronominale envisagée pour un [loc-objet] humain est *lui, lei* ou *loro :*

Ugo va da Eva => Ugo va da lei.
(Ugo va chez Eva => Ugo va chez elle.)

Cependant, il existe une catégorie de verbes et de prépositions introduisant des locatifs, avec lesquels l'une des formes préverbales *gli, le* ou *loro* (qui synthétisent habituellement un pronom [a-objet]) peut être choisie :

Ugo è corso davanti a Eva.
(Ugo a couru devant Eva.)
[Eva --> lei] *Ugo è corso davanti a lei .*

(Ugo a couru devant elle.)
[Eva --> ppv] *Ugo **le** è corso davanti.*
([litt] *Ugo lui a couru devant.)

Ugo mette il piatto vicino a Eva.
(Ugo met l'assiette à côté de Eva.)
[Eva --> lei] *Ugo mette il piatto vicino **a lei.***
(Ugo met l'assiette à côté d'elle.)
[Eva --> ppv] *Ugo **le** mette il piatto vicino.*
([litt] *Ugo lui met l'assiette à côté.)

Les locutions prépositionnelles qui permettent un tel comportement sont composées de deux prépositions :

prep1 [prep2] TOK

où prep2 = *a* (à) et plus rarement *di* (de). La ppv substitue alors l'ensemble [prep2 TOK], et prep1 conserve sa position. Notons cependant qu'en fonction de la valeur de prep1, prep2 n'est pas toujours présente quand TOK est synthétisé en un groupe nominal : ainsi, si prep1 = *davanti* (devant) , alors prep2 = *a* (à), et cette préposition est obligatoire (exemple (8a)); par contre prep2 est absente si prep1 = *sopra* (au dessus) (exemple (9a)) :

(8a) [prep1 = *davanti*, prep2 = *a*] *L'acrobata è balzato davanti **a** Ugo.*
(L'acrobate a bondi devant Ugo.)

(9a) [prep1 = *sopra*, prep2 = ϕ] *L'acrobata è balzato sopra Ugo.*
(L'acrobate a bondi au dessus [de] Ugo.)

Mais quand TOK est synthétisé en pronom non préverbal (*lui, lei* ou *loro*), la présence de prep2 est obligatoire, quelle que soit la valeur de prep1. Ainsi, prep2 vaut *a* dans l'exemple (8b), comme dans l'exemple (8a) ; elle vaut *di* en présence de prep1 = *sopra* (exemple (9b)) :

(8b) [prep1 = *davanti*, prep2 = *a*] *L'acrobata è balzato davanti **a** lui.*
(L'acrobate a bondi devant lui.)

(9b) [prep1 = *sopra*, prep2 = *di*] *L'acrobata è balzato sopra **di** lui.*
(L'acrobate a bondi au dessus de lui.)

La ppv *gli* substitue alors respectivement le groupe *a lui* (exemple (8c)) et *di lui* (exemple (9c)) :

(8c) [*a lui* --> *gli*] *L'acrobata gli è balzato davanti.*
(*L'acrobate lui a bondi devant.)

(9c) [*di lui* --> *gli*] *L'acrobata gli è balzato sopra.*
(L'acrobate lui a bondi dessus.)

La possibilité de synthétiser une ppv [loc-objet] pour un TOK humain dépend non seulement de la valeur de prep1 :

Ugo è corso verso Eva.
(Ugo a couru vers Eva.)

[Eva --> lei] *Ugo è corso verso di lei .*
(Ugo a couru vers elle.)
[Eva --> ppv] **Ugo le è corso verso.*
(*Ugo lui a couru vers.)

mais aussi de la nature du verbe :

*Ugo ha mangiato davanti a Eva => *Ugo le ha mangiato davanti.*
(Ugo a mangé devant Eva => *Ugo lui a mangé devant.)

Ces deux conditions impliquent la constitution de la liste exhaustive des prépositions qui se comportent comme *davanti* (dont V. Lo Cascio (1969) donne quelques éléments) et l'étude des verbes qui acceptent ce type de pronominalisation. Ces recherches concernent d'ailleurs également la pronominalisation des entités non humaines :

*Ugo mette la forchetta vicino al piatto => Ugo **ci** mette la forchetta vicino.*
(Ugo met la fourchette à côté de l'assiette => [litt] Ugo y met la fourchette à côté.)

ainsi que la synthèse des [prep-objet] :

Il cane abbaia contro Eva.
(Le chien aboie contre Eva.)
[Eva -> lei] *Il cane abbaia contro **di lei** .*
(Le chien aboie contre elle.)
[Eva -> ppv] *Il cane **le** abbaia contro.*
([litt] le chien lui aboie contre.)

Ces recherches sont encore en cours, et nous dirons donc en première approximation que les seuls pronoms humains envisageables sont les formes toniques *lui, lei* et *loro*, selon le nombre et le genre de TOK. De même nous laissons de côté la synthèse en ppv des locatifs non-humains et des [prep-objet] qui présentent les caractéristiques décrites ci-dessus.

Si TOK est non-humain, alors la ppv *ci*, ou son équivalent *vi* (d'un style un peu plus littéraire), est envisagée si :

1) la préposition qui introduit le [loc-objet] est *a, in* ou *su* : plus généralement, si elle indique l'endroit où l'on va (ou celui où l'on est) et non pas la direction dans laquelle on se dirige :

*Max corre **al** mercato.* (Max court au marché.) *Max corre **sul** tetto* (Max court sur le toit.) *Max corre **in** cantina.* (Max court à la cave.)	==>	*Max **ci** corre.* (Max y court.)
*Max corre **verso** la casa.* (Max court vers la maison.)	==>	**Max **ci** corre.* (*Max y court.)

Dans le dernier exemple, l'emploi de la ppv *ci* modifie le sens de la phrase : il est donc interdit.

2) le verbe ne se construit pas avec un [a-objet] dont la ppv est *ci* (information fournie par le lexique-grammaire) :

(a) *Max pensa nella camera.*
(Max pense dans la chambre.)
(b) **Max ci pensa.*
(*Max y pense.)
(c) **Max vi pensa.*
(*Max y pense.)

(b) est ambiguë, donc à éviter, car l'interprétation de *ci* qui s'impose concerne l'objet de *pensare* et non pas un lieu.

(c) est également incorrecte par analogie avec (b) même si elle ne contient pas l'ambiguïté de (b)[13].

La forme du pronom locatif n'est pas uniquement fonction de la préposition et du verbe employés. Elle dépend également de la synthèse du [dir-objet] et du [a-objet] : si l'un d'eux est un pronom de la première personne du pluriel alors la ppv *ci* ne peut pas être envisagée[14] :

(10a) *Ugo ci vedrà a Roma.*
(Ugo nous verra à Rome.)
(10b)[loc-objet -> *ci*] **Ugo ci ci vedrà.*
(Ugo nous y verra.)

La forme *vi* est également inacceptable, à cause de la confusion qui s'ensuit dans l'interprétation du couple *vi ci*. Ainsi la phrase (10c), résultant de la pronominalisation du [loc-objet] *a Roma* en ppv *vi* :

(10c)[loc-objet -> *vi*] *Ugo vi ci vedrà.*

provient indifféremment de la dérivation de (10a) :

(10a) *Ugo **ci** vedrà a Roma.*
(Ugo nous verra à Rome.)

ou de (10d) :

(10d) *Ugo vi vedrà a Roma.*
(Ugo vous verra a Rome.)

Si la forme préverbale n'est pas envisageable, on peut parfois choisir de synthétiser un adverbe de lieu, qui remplace le constituant entier (préposition + entité) :

13 De plus, dans le cas précis de *pensare*, elle évoque le sens : *Max pense à vous*, exprimé en termes poétiques.

14 Le raisonnement est le même si on considère le locatif *vi* et le [dir-objet] ou [a-objet] de la 2ème personne du pluriel.

(10e) *Ugo ci vedrà li'.*
(Ugo nous verra là.)

ou, dans certains cas, une des formes toniques *esso, essa esse, essi* :

Ugo gira intorno alla casa => Ugo gira intorno ad essa.
(Ugo tourne autour de la maison => Ugo tourne autour d'elle$_{non\text{-}hum.}$)

Pour les mêmes raisons que celles évoquées pour le di-objet, la précision de la nature sémantique de TOK intervient dans le choix du pronom préverbal non humain :

Ugo si è fermato nei dintorni di Roma.
(Ugo s'est arrêté aux alentours de Rome.)
--> **Ugo si è fermato nei dintorni di essa.*
(*Ugo s'est arrêté aux alentours de celle-ci.)

Prep-objet (objet introduit par une autre préposition que celles présentées jusqu'ici) : Nous ne reviendrons pas sur le phénomène de pronominalisation préverbale lié aux "doubles" prépositions, qui a été présenté dans le paragraphe précédent concernant le loc-objet.
La forme du pronom [prep-objet] peut dépendre de celle du [dir-objet] (les dépendances croisées entre un [dir-objet] et un [prep-objet] introduit par *con* seront illustrées dans le chapitre 5). Sinon, pour un humain, le pronom envisagé est l'une des formes toniques *lui, lei* ou *loro* :

Max vive con Eva => Max vive con lei.
(Max vit avec Eva => Max vit avec elle.)

Pour un non-humain, on envisage de synthétiser l'une des formes e*sso, essa, esse* ou *essi :*

Le idee di Max coincidono colla teoria di Ugo.
(Les idées de Max coincident avec la théorie de Ugo.)
[la teoria di Ugo -> essa] *Le idee di Max coincidono con essa.*
(Les idées de Max coincident avec elle$_{non\text{-}hum.}$)

3.2.2.2 ANTECEDENTS DE LA FORME CHOISIE

Pour déterminer si la forme choisie peut être synthétisée, on calcule les antécédents possibles de ce pronom[15]. On peut découper ce travail en plusieurs opérations indépendantes, qui peuvent donc être effectuées dans n'importe quel ordre :

- calcul de la liste des antécédents morphologiques de la forme pronominale. Cette liste est constituée des étiquettes dont la synthèse est compatible en genre et en nombre avec le pronom ;

- calcul de la liste des antécédents sémantiques. Cette liste est constituée des étiquettes représentant les entités qui vérifient les contraintes sémantiques (ou distributionnelles) du pronom ;

15 En première approximation, on tiendra compte de tous les antécédents possibles de ce pronom, quelle que soit leur distance. Nous verrons l'utilisation de ce paramètre au chapitre 5.

- calcul de la liste des antécédents syntaxiques. Cette liste contient les éléments syntaxiquement compatibles avec le pronom ;

- les calculs sur la distance, le focus, les contraintes pragmatiques et de parallélisme, etc ... fournissent d'autres listes.

L'ensemble des antécédents possibles du pronom est alors le résultat de l'intersection de ces listes. La décision de synthétiser le pronom dépend du cardinal de cet ensemble. Dans la pratique, on choisit un ordre dans lequel sont effectués ces calculs (en l'occurrence, l'ordre décrit ci-dessus) : chaque liste est alors une sous-liste de la liste précédente. Quant à la liste des antécédents morphologiques, elle est calculée à partir des éléments qui ont été préalablement synthétisés.

3.2.2.2.1 ANTECEDENTS MORPHOLOGIQUES DU PRONOM

- Si la forme qui vient d'être calculée appartient à la liste :

{*esso, essa, essi, esse, lo, la, li, le, lui, lei, gli*}

alors elle porte la marque du genre et du nombre : la liste, notée L1, des antécédents morphologiques de ce pronom contient les éléments qui ont été synthétisés en groupes nominaux de même genre et nombre que ce pronom ;

- Si la forme pronominale est *l'*, et que le verbe est conjugué à un temps composé[16] alors le participe passé, qui s'accorde avec le [dir-objet] marque le genre de ce pronom. L1 contient alors les étiquettes synthétisées en groupes nominaux singulier de même genre que le pronom ;

- Si le verbe n'indique pas le genre de *l'*, ou si la forme pronominale calculée est *glie*, alors L1 est constituée des étiquettes synthétisées en groupes nominaux féminin-singulier et masculin-singulier ;

- La liste L1 des antécédents morphologiques de *loro* (forme tonique ou atone) contient les étiquettes qui ont été synthétisées en groupes nominaux féminin-pluriel et masculin-pluriel ;

- Quant à la liste L1 reliée aux formes invariables *ci, vi* et *ne*, elle contient toutes les étiquettes préalablement synthétisées.

3.2.2.2.2 ANTECEDENTS SEMANTIQUES DU PRONOM

La sous-liste L2 des antécédents sémantiques d'un pronom est calculable :

- à partir du pronom lui-même : par exemple, si celui-ci est *lui, lei* ou *loro*, alors l'antécédent est un humain. De même, les étiquettes représentant des humains sont exclues de L2 si le pronom est *esso, essa, essi* ou *esse ;*

- grâce aux propriétés distributionnelles du verbe. Dans le texte :

Ugo ha messo il lenzuolo nell'armadio. Eva lo ha stirato ieri.
(Ugo a mis le drap dans l'armoire. Eva l'a repassé hier.)

16 Un verbe transitif (i.e qui se construit avec un objet direct) se conjugue toujours avec l'auxiliaire *avere* (avoir), sauf s'il y a synthèse d'un pronom réflexif (i.e. d'une forme réflexive atone).

le pronom masculin-singulier *lo* de la deuxième phrase a *a priori* trois antécédents : *Ugo*, *il lenzuolo* et *l'armadio*. Mais les renseignements attachés au verbe *stirare* indiquent que l'objet direct ne peut être humain, ce qui exclut *Ugo*. Par ailleurs, si les distinctions sont assez fines du point de vue sémantique (e.g. on sait identifier les objets "repassables" de ceux qui ne le sont pas), on pourra également en déduire que *l'armadio* est un antécédent distributionnellement incompatible avec le pronom.

On voit donc que plus les classes sémantiques sont fines, plus le nombre des antécédents possibles d'un pronom diminue. Cependant, il est très difficile d'établir une classification sémantique fine qui couvre l'ensemble de la langue : on doit alors se restreindre à travailler dans des domaines précis. Par conséquent, établir une classification sémantique fine revient à trouver un codage adéquat du domaine choisi, et donc à constituer un dictionnaire spécifique à ce domaine.

3.2.2.2.3 ANTECEDENTS SYNTAXIQUES DU PRONOM

A partie de L2, on calcule la sous-liste L3 des antécédents syntaxiquement compatibles avec le pronom. Un exemple d'incompatibilité syntaxique est la contrainte de non coréférence entre un sujet et un pronom personnel objet. Une conséquence de cette contrainte, qui est présentée au chapitre 4, est que dans :

Eva la pettina.
(Eva la peigne.)

Le sujet *Eva* et l'objet *la* ne sont pas coréférents, puisque s'ils l'étaient, l'objet direct serait synthétisé en un pronom réflexif :

Eva si pettina.
(Eva se peigne.)

Un autre exemple d'incompatibilité syntaxique est la contrainte suivante (qui sera développée au chapitre 6) : si un pronom personnel synthétise le sujet d'un complément phrastique qui doit être réduit à l'infinitif en cas d'égalité de son sujet avec celui de la principale, alors ce pronom ne réfère pas au sujet de la principale, car si c'était le cas, le complément phrastique serait réduit à l'infinitif (L. Danlos (1988a)) :

Eva$_i$ veut que Eva$_i$ parte. ==> Eva veut partir.

Une illustration de cette contrainte est que dans :

Eva veut qu'elle parte.

le pronom *elle* ne peut pas référer à *Eva*, qui est donc exclue de la liste L3.

3.2.2.2.4 AUTRES OPERATIONS

Pragmatique :

La pragmatique est le domaine de recherche qui vise à intégrer aux analyses linguistiques l'ensemble des connaissances que l'on a du monde. Les connaissances pragmatiques permettent à un humain de désambiguer le texte suivant, ce que n'arrive pas à faire l'ensemble des règles

morphologiques, sémantiques et syntaxiques présentées ci-dessus, pour lesquelles le pronom *il* est ambigu :

Ugo est à la recherche de Paul. Il a disparu depuis deux jours.

En effet *il* référe de préférence à *Paul*, puisque l'on sait que la disparition de quelqu'un provoque généralement sa recherche. L'établissement de relations pragmatiques rend donc possible la désambiguation d'une forme pronominale quand elle a plusieurs antécédents possibles. Mais la formalisation de ces relations pragmatiques nécessite d'être dans un domaine restreint, et de trouver un moyen raisonnable de codage de ces connaissances. L'exemple suivant, de Ch. Leclère, montre à quel point ce codage peut être difficile à établir : il nécessite en effet d'associer au verbe *dormir* un certain nombre de relations qui désambiguent le pronom *elle* de la principale :

Quand Sue est arrivée chez Eve, elle dormait.

dont l'antécédent est *Eve* ; et ces relations ne doivent pas se retrouver dans la locution verbale *dormir debout*, puisque dans ce cas le pronom *elle* de la principale peut référer aussi bien à *Eve* qu'à *Sue* :

Quand Sue est arrivée chez Eve, elle dormait debout.

Le manque de fiabilité des critères pragmatiques peut être illustré également par le couple d'exemples suivants. Dans le texte de Pitrat (1985) :

Le professeur a puni l'élève. Il lui avait lancé des boulettes.

les relations pragmatiques permettent d'établir la coréférence entre le pronom *il* de la deuxième phrase et le groupe nominal *l'élève*. Mais cette coréférence n'est plus vraie dans :

Le professeur a puni l'élève. Il lui a lancé des boulettes.

où les verbes des deux phrases sont au même temps. Pour déterminer si la synthèse d'un pronom est possible, nous préférons tout d'abord exploiter autant que possible les critères morphologiques, sémantiques et syntaxiques, que l'on peut définir avec certitude, qui sont formalisables et qui donne des résultats fiables. Les critères pragmatiques, par contre, qui sont souvent difficiles à cerner et qui aboutissent le plus souvent à des interprétations préférentielles (et non pas à des coréférences indiscutables) ne seront pris en compte qu'ensuite.

Parallélisme :

Le parallélisme de structure permet d'établir la coréférence entre les sujets ou les objets de deux phrases, quand celles-ci sont jugées parallèles. Il s'en suit une modification de la liste des antécédents possibles d'un pronom, comme on le verra au chapitre 5, consacré en partie à ce phénomène.

Focus :

Le focus d'une phrase est défini intuitivement comme l'individu ou l'objet dont il est question. La notion de focus a été utilisée dans le système de génération développé par K.

McKeown (1983). Un ensemble de règles a été basé sur la recherche du focus pour ordonner les informations à synthétiser, et pour effectuer les transformations à l'intérieur des schémas de phrases (ces règles ont été critiquées par L. Danlos (1987), cf. chapitre 2). C. Sidner (1981), ainsi que B. Grosz et alii (1982) ont utilisé la notion de focus pour déterminer le référent d'un pronom. Ces chercheurs ont donné leur préférence à une approche utilisant le focus plutôt qu'à une approche utilisant le parallélisme. A l'inverse de leur point de vue, et en accord avec L. Danlos, nous pensons que la notion de parallélisme, qui est basée sur des critères syntaxiques et formalisables de comparaison de structures (ou d'arbres[17]), est plus abordable que celle de focus, qui est avant tout intuitive, et composée d'un grand nombre de sous-types : focus potentiel, focus explicite, focus local ...

Distance :

La distance peut être vue soit en termes de nombre de mots, soit en termes de nombre de phrases, soit en termes de géométrie dans un arbre, comme nous le verrons en détail au chapitre 5. Elle intervient non seulement dans la décision d'envisager la synthèse d'un pronom (la précédente occurrence de l'étiquette TOK, dont la synthèse en un pronom personnel est envisagée, est-elle trop "éloignée" ?), mais également pour déterminer l'inacceptabilité éventuelle d'un antécédent morphologiquement, sémantiquement, syntaxiquement et pragmatiquement compatible avec le pronom. La notion de distance, en particulier celle liée au concept de profondeur dans un arbre, est théoriquement calculable par un certain nombre de contraintes (mises au point essentiellement par les chercheurs génératives). Par ailleurs, le rapport entre distance et parallélisme sera évoqué au chapitre 5.

Ces calculs, et d'autres encore peut-être, permettent d'affiner la liste des antécédents du pronom envisagé : on décide de synthétiser ce pronom si la liste finale est réduite à un élément.

3.3 EFFACEMENT DU SUJET

La synthèse d'un pronom sujet suit les mêmes étapes que celles qui ont été présentées pour le pronom objet. Cependant, en italien, ainsi que dans toutes les langues romanes (sauf le français), le sujet est effacé, dans les propositions non infinitives, quand cette opération n'entraîne pas d'ambiguïté. Ce sujet a été évoqué dans F. Namer (1989). Cet effacement n'affecte pas les autres éléments de la phrase : par exemple, le verbe reste en accord avec son sujet même quand celui-ci est effacé. Dans un premier temps, nous allons montrer que pronominalisation et effacement du sujet sont deux phénomènes liés, en d'autres termes que l'effacement du sujet revient à l'effacement d'un pronom. Ensuite, après une brève présentation des pronoms sujet, nous verrons quelles sont les opérations nécessaires à l'effacement de ce pronom.

3.3.1 LIEN ENTRE PRONOMINALISATION ET EFFACEMENT DU SUJET

Cette section a pour but de montrer que l'effacement du sujet est un phénomène dépendant directement de la pronominalisation, tout du moins dans un système de

17 La structure d'un schéma de phrase peut être assimilée à un arbre, dans lequel le symbole :ps représente la racine, et chaque constituant entre parenthèses l'un des fils de :ps.

génération[18] : un sujet ne peut être effacé (sans création d'ambiguïté) que s'il est pronominalisable, c'est à dire remplaçable par un pronom dont la synthèse n'engendre pas d'ambiguïté. Cette affirmation concerne évidemment les prédicats à sujet non impersonnels ; i.e. nous ne nous occupons pas de verbes comme *piovere* (pleuvoir) où le sujet est obligatoirement effacé en italien, et où il est obligatoirement pronominalisé en *il* (non référencé) en français. Dans les autres cas, la synthèse d'une étiquette sujet TOK rentre toujours dans l'une des trois catégories :

1) TOK n'est ni pronominalisable ni effaçable,

2) TOK est pronominalisable et effaçable,

3) TOK est pronominalisable, mais pas effaçable.

En d'autres termes, il n'existe pas de schéma de phrase dans lequel le sujet soit effaçable, et non pronominalisable. Les trois catégories décrites ci-dessus sont illustrées respectivement par les exemples (T11), (T12) et (T13).

1) Considérons le schéma de texte (T11) :

(T11) *(EVA e LIA tornare da Londra. EVA imparare l'inglese.)*
(EVA et LIA revenir de Londres. EVA apprendre l'anglais.)

La seconde occurrence de l'étiquette EVA ne peut être ni pronominalisée (11a) :

(11a) **Eva e Lia sono tornate da Londra. Lei ha imparato l'inglese.*
(*Eva et Lia sont revenues de Londres. Elle a appris l'anglais.)

ni effacée (11b) :

(11b) **Eva e Lia sono tornate da Londra. Ha imparato l'inglese.*
(*Eva et Lia sont revenues de Londres. [Elle] a appris l'anglais.)

2) Par contre, dans le schéma de texte (T12) :

(T12) *(EVA tornare da Londra. EVA imparare l'inglese.)*
(EVA revenir de Londres. EVA apprendre l'anglais.)

La seconde occurrence de l'étiquette EVA peut être soit pronominalisée (12a) :

(12a) *Eva è tornata da Londra. Lei ha imparato l'inglese.*
(Eva est revenue de Londres. Elle a appris l'anglais.)

---

18 L. Rizzi (1982) ne relie pas pronominalisation et sujet effacé. Il considère ce dernier comme une trace, jouant parfois le rôle de pronom :

Sei venuto.
([tu] es venu.)

et parfois celui de "relai" avec un sujet inversé:

E' arrivato Gianni.
([litt] Est arrivé Gianni.)
(Gianni est arrivé.)

L'inversion du sujet, fréquente en italien, est en effet un phénomène problématique en analyse, puisqu'il s'agit d'identifier le sujet d'une phrase. L'approche de L.Rizzi peut donc être utile en analyse. Mais en génération elle n'est pas pertinente, puisqu'un système de génération doit, lui, synthétiser un sujet identifié.

soit effacée (12b) :

(12b) *Eva è tornata da Londra. Ha imparato l'inglese.*
(Eva est revenue de Londres. [Elle] a appris l'anglais.)

La présence du pronom *lei* dans la deuxième phrase de (12a) peut être due à un désir d'insistance sur l'entité représentée par ce pronom :

Elle, elle a appris l'anglais.

3) Enfin, dans le schéma de texte (T13) :

(T13) *(EVA e MAX tornare da Londra. EVA imparare l'inglese.)*
(EVA et MAX revenir de Londres. EVA apprendre l'anglais.)

La seconde occurrence de l'étiquette EVA peut être pronominalisée (13a) :

(13a) *Eva e Max sono tornati da Londra. Lei ha imparato l'inglese.*
(Eva et Max sont revenus de Londres. Elle a appris l'anglais.)

mais pas effacée (13b) :

(13b) **Eva e Max sono tornati da Londra. Ha imparato l'inglese.*
(*Eva et Max sont revenus de Londres. [Elle + il] a appris l'anglais.)

Il n'existe pas de cas où la synthèse du pronom sujet est interdite car ambiguë, et où le sujet est quand même effaçable. De plus, les exemples montrent que la synthèse du sujet repose sur la résolution de deux problèmes :

- le sujet est-il pronominalisable ? (i.e. la synthèse du pronom sujet n'entraîne-t-elle aucune ambiguïté ?)
- si oui, le pronom est-il effaçable ? (i.e. l'effacement de ce pronom n'entraîne-t-il aucune ambiguïté ?)

3.3.2 EFFACEMENT D'UN PRONOM SUJET

3.3.2.1 PRONOM SUJET

Synthèse d'un pronom sujet de la troisième personne :

Rappelons tout d'abord que la synthèse de l'étiquette sujet en un pronom personnel est envisagée à la troisième personne si et seulement si cette étiquette a déjà été mentionnée, comme c'est le cas pour les autres compléments. Le calcul de la forme d'un pronom sujet (qui est une forme tonique, cf. tableau page 34) ne dépend que de la personne et du nombre si l'étiquette sujet est de la première ou deuxième personne. Si l'étiquette est de la troisième personne, la forme dépend du nombre, du genre et de la nature humaine de l'étiquette à synthétiser ('ps' signifie personne du singulier et 'pp', personne du pluriel) :

	HUMAIN		NON HUMAIN	
	masc	fem	masc	fem
3ps	*lui, egli*	*lei, ella*	*esso*	*essa*
3pp	*loro*	*loro*	*essi*	*esse*

Les formes *lui* et *egli* sont équivalentes. On emploiera plutôt la seconde à l'écrit, la première étant généralement réservée à l'oral. A l'inverse, entre les formes *lei* et *ella* (qui sont aussi équivalentes) on préfère *lei*, *ella* étant de moins en moins utilisée (cf. J. Brunet (1985)). On remarque que le pronom sujet humain de la 3ème personne du pluriel *loro* est invariable en genre, à l'inverse des formes *essi* et *esse*. Ces formes tendent de plus en plus à être utilisées dans l'italien courant pour représenter des humains. Cependant, pour notre propos, nous nous en tiendrons aux règles des grammaires (M. Sensini (1988) et L. Serianni (1988)) et des dictionnaires (N. Zingarelli (1986)), qui les réservent aux non-humains.

Le calcul des antécédents du pronom, qui ne concerne que les formes de la 3ème personne (pour les raisons évoquées au début de la section 3.2.2) est identique à celui qui a été exposé pour les pronoms objets dans la section 3.2.2.2.

Effacement d'un pronom sujet de dialogue (1ère ou 2ème personne) :

Si le verbe indique le nombre et la personne de son sujet, alors le pronom de dialogue peut toujours être effacé: la synthèse d'un sujet de la 1ère ou 2ème personne rentre alors dans la catégorie des étiquettes qui sont pronominalisables et effaçables. Le verbe indique toujours le nombre et le genre de son sujet, à condition qu'il ne soit pas conjugué au subjonctif. En effet, les trois personnes du singulier d'un verbe au subjonctif présent peuvent être identiques :

Ugo vuole che (io + tu + egli + lei) parta.
(Ugo veut que (je + tu + il + elle) parte(s).)

Pour un verbe régulier, conjugué au subjonctif passé, les deux personnes du singulier sont les mêmes :

Ugo voleva che (io + tu) partissi.
(Ugo voulait que (je + tu) parte(s).)

En conséquence, un pronom sujet de dialogue peut toujours être effacé dans les phrases qui ne sont pas au subjonctif :

Verrai domani.
([Tu] viendras demain.)

sauf si le locuteur veux insister sur l'entité que le pronom représente :

***Tu** verrai domani.*
(Toi, tu viendras demain.)

Effacement d'un pronom de la 3ème personne :

Un pronom de la 3ème personne du singulier contient des informations concernant la nature humaine et le genre de l'entité qu'il représente. Il en va de même pour un pronom de la 3ème personne du pluriel, excepté le cas de *loro*, qui est invariable en genre. Donc l'effacement de l'une de ces formes pronominales entraîne une perte d'information concernant le genre et la nature sémantique du sujet. Cette perte d'information peut donner lieu à des ambiguïtés. En supposant, dans ce qui suit, que le sujet est pronominalisable, voyons dans quelles conditions ce pronom est effaçable[19], i.e. quand les informations concernant le genre et la nature sémantique du sujets sont récupérables.

Etant donné que les problèmes soulevés par l'effacement d'un pronom de la 3ème personne du pluriel sont les mêmes que ceux soulevés par l'effacement d'un pronom de la 3ème personne du singulier, les conditions d'effacement d'un pronom, présentées ci-dessous seront illustrées uniquement par des exemples d'étiquettes de la 3ème personne du singulier.

3.3.2.2 RECUPERATION DE LA NATURE SEMANTIQUE DU PRONOM EFFACE

Il est possible de retrouver la nature sémantique d'un sujet effacé quand le prédicat verbal se construit exclusivement avec un sujet humain ou un sujet non-humain. Dans :

Ugo ha piantato un ciliegio. Esso fruttifica.
(Ugo a planté un cerisier. $Il_{non\text{-}hum}$ fructifie.)

le pronom sujet non humain *esso* peut être effacé :

Ugo ha piantato un ciliegio. Fruttifica.
(Ugo a planté un cerisier. [Il] fructifie.)

puisque le verbe *fruttificare* se construit uniquement avec un sujet non humain, et que par conséquent l'exemple ci-dessous est agrammatical :

**(Ugo + egli) fruttifica.*
(*(Ugo + il_{hum}) fructifie.)

Par contre, dans :

Ugo ha piantato un ciliegio. Esso è ammirevole.
(Ugo a planté un cerisier. $Il_{non\text{-}hum}$ est admirable.)

le pronom *esso* ne peut pas être effacé, et le texte résultant serait ambigu :

**Ugo ha piantato un ciliegio. E' ammirevole.*
(Ugo a planté un cerisier. [il_{hum} + non-hum] est admirable.)

---

19 Nous ne nous occupons pas des cas triviaux comme (T12) :
(T12) *(EVA tornare da Londra. EVA imparare l'inglese.)*
(EVA revenir de Londres. EVA apprendre l'anglais.)
où le contexte gauche du sujet dont l'effacement est envisagé ne contient qu'un élément : le pronom sujet est donc effaçable.

puisque *essere ammirevole* prend indifféremment un sujet humain ou non humain :

(Ugo + il ciliegio + esso + egli) è ammirevole.
((Ugo + le cerisier + il$_{non\text{-}hum}$ + il$_{hum}$) est admirable.)

3.3.2.3 RECUPERATION DU GENRE DU PRONOM EFFACE

Pour cette question, supposons que l'on peut retrouver la nature sémantique du sujet. Pour ce faire, les prédicats verbaux utilisés dans les exemples ci-dessous se construisent exclusivement avec des sujets humains.

Le genre du pronom effacé est marqué par un autre élément de la phrase :

Si le genre du pronom sujet dont l'effacement est envisagé est marqué par un autre élément de la phrase, alors cette opération n'engendre pas d'ambiguïté. Considérons les textes (14a) et (15a) dans lesquels on envisage d' effacer le pronom *lei*, sujet de la deuxième phrase :

(14a) *Ugo non vedrà più Eva. Lei è stata condannata all'ergastolo.*
(Ugo ne verra plus Eva. Elle a été condamnée à perpétuité.)

(15a) *Ugo non vedrà più Eva. Lei è in prigione per omicidio.*
(Ugo ne verra plus Eva. Elle est en prison pour homicide.)

Si le pronom sujet de (14a) est effacé, le résultat n'est pas ambigu, puisque le participe passé du verbe marque le genre du sujet[20]. *Ugo*, qui est masculin, est donc un antécédent interdit. Le seul antécédent possible pour le sujet effacé est *Eva*, et le texte (14b) ou *lei* est effacé est non ambigu :

(14b) *Ugo non vedrà più Eva. E' stata condannata all'ergastolo.*
(Ugo ne verra plus Eva. [Elle] a été condamnée à perpétuité.)

Par contre, si le pronom *lei* est effacé en (15a), l'information concernant le genre du sujet est perdue, puisque ni le verbe ni aucun autre élément de la phrase ne l'indique. L'antécédent du pronom effacé est alors soit *Ugo* soit *Eva*. Le texte (15b), ou *lei* est effacé, est alors ambigu :

(15b) **Ugo non vedrà più Eva. E' in prigione per omicidio.*
(Ugo ne verra plus Eva. [Elle + il] est en prison pour homicide.)

L'effacement du sujet est donc interdit.

Les éléments d'une phrase qui indiquent le genre du sujet sont les suivants :

A) l'attribut adjectival ou nominal, quand il est variable en genre. Cette propriété sépare les adjectifs en deux catégories : ceux qui se comportent selon le modèle de *cattivo* (méchant), variable en genre (masculin-singulier *cattivo* ; féminin-singulier *cattiva* ; masculin-pluriel *cattivi* ; féminin-pluriel *cattive*) :

20 Le participe passé d'un verbe conjugué avec l'auxiliaire *essere* s'accorde en genre et nombre avec le sujet, comme c'est le cas en français.

Ugo non vedrà più Eva. E' troppo cattiva.
(Ugo ne verra plus Eva. [Elle] est trop méchante.)

et ceux qui, comme *gentile* (gentil), ne varient qu'en nombre (masculin et féminin singulier *gentile* ; masculin et féminin pluriel *gentili*) et conservent donc l'ambiguïté dans l'exemple ci-dessous :

**Ugo vedrà Eva domani. E' molto gentile.*
(Ugo verra Eva demain. [Il + elle] est très gentil(le).)

Il faut, de la même façon, distinguer deux classes de noms :

Ugo ha sposato Eva. E' la figlia del direttore.
(Ugo a épousé Eva. [Elle + c'$_{fem}$] est la fille du directeur.)

Rappelons que le genre des groupes nominaux est lexical et non sémantique en français et en italien, et c'est pourquoi le genre du déterminant du groupe nominal attribut n'est pas toujours suffisant pour indiquer celui du sujet. Ainsi, l'exemple ci-dessous est ambigu :

**Ugo ha sposato Eva. E' una bella canaglia.*
(*Ugo a épousé Eva. [C'$_{masc + fem}$] est une belle canaille.)

B) le verbe, s'il satisfait l'une des conditions suivantes :

1) il est conjugué au passif (cf. textes 14a et 14b) ;

2) il est conjugué à un temps composé avec l'auxiliaire *essere* (être) :

Ugo non vedrà più Eva. E' andata in Giappone.
(Ugo ne verra plus Eva. [Elle] est allée au Japon.)

3) il est conjugué à un temps composé à la voix pronominale, par exemple à cause de la présence d'un pronom réflexif :

Ugo non balla con Eva questa sera. Si è ferito.
(Ugo ne danse pas avec Eva ce soir. [Il] s'est blessé.)

Dans ces trois cas, le participe passé s'accorde en genre et nombre avec le sujet.

C) N'importe quel objet, s'il est égal au sujet et qu'il est synthétisé en une forme réflexive tonique, accompagnée de l'adjectif *stesso*. Comme on l'a vu, cet adjectif s'accorde en genre et nombre avec le sujet même quand celui-ci est effacé :

Ugo odia Eva. Conta solo su se stesso.
(Ugo déteste Eva. [Il + *elle] compte seulement sur lui-même.)

Ugo odia Eva. Conta solo su se stessa.
(Ugo déteste Eva. [*Il + elle] compte seulement sur elle-même.)

En conclusion :

- la synthèse du sujet et celle du verbe dépendent l'une de l'autre : la synthèse du sujet dépend de celle du verbe, puisque l'effacement du pronom sujet est conditionné par la conjugaison du verbe. D'autre part, la synthèse du verbe dépend de l'étiquette sujet pour l'accord en nombre et en personne (information sémantique), et elle dépend de la synthèse du sujet pour l'accord éventuel en genre (information lexicale) ;

- par le même raisonnement, la synthèse du sujet et celle de l'attribut sont mutuellement dépendantes ;

- enfin, la synthèse du sujet dépend de la présence et de la nature d'une forme réflexive : la présence d'un forme réflexive atone (cas B3) est liée à la synthèse de l'objet direct et du [a-objet] ; celle d'une forme tonique modifiée par *stesso* (cas C), à la synthèse d'un autre objet prépositionnel (ou de tout objet, puisque la synthèse d'une forme réflexive tonique peut être motivée par des raisons stylistiques).

Le genre du pronom effacé est calculable à partir de la synthèse d'autres éléments de la phrase :

La contrainte de non coréférence entre le sujet et un pronom objet a été évoquée dans la section 3.2.2.2.3, à propos du calcul des antécédents syntaxiques d'un pronom objet. Cette contrainte intervient également entre un pronom sujet que l'on veut effacer et un objet, pour calculer les antécédents syntaxiques du pronom sujet, comme on le verra en détail au prochain chapitre.

3.4 CONCLUSION

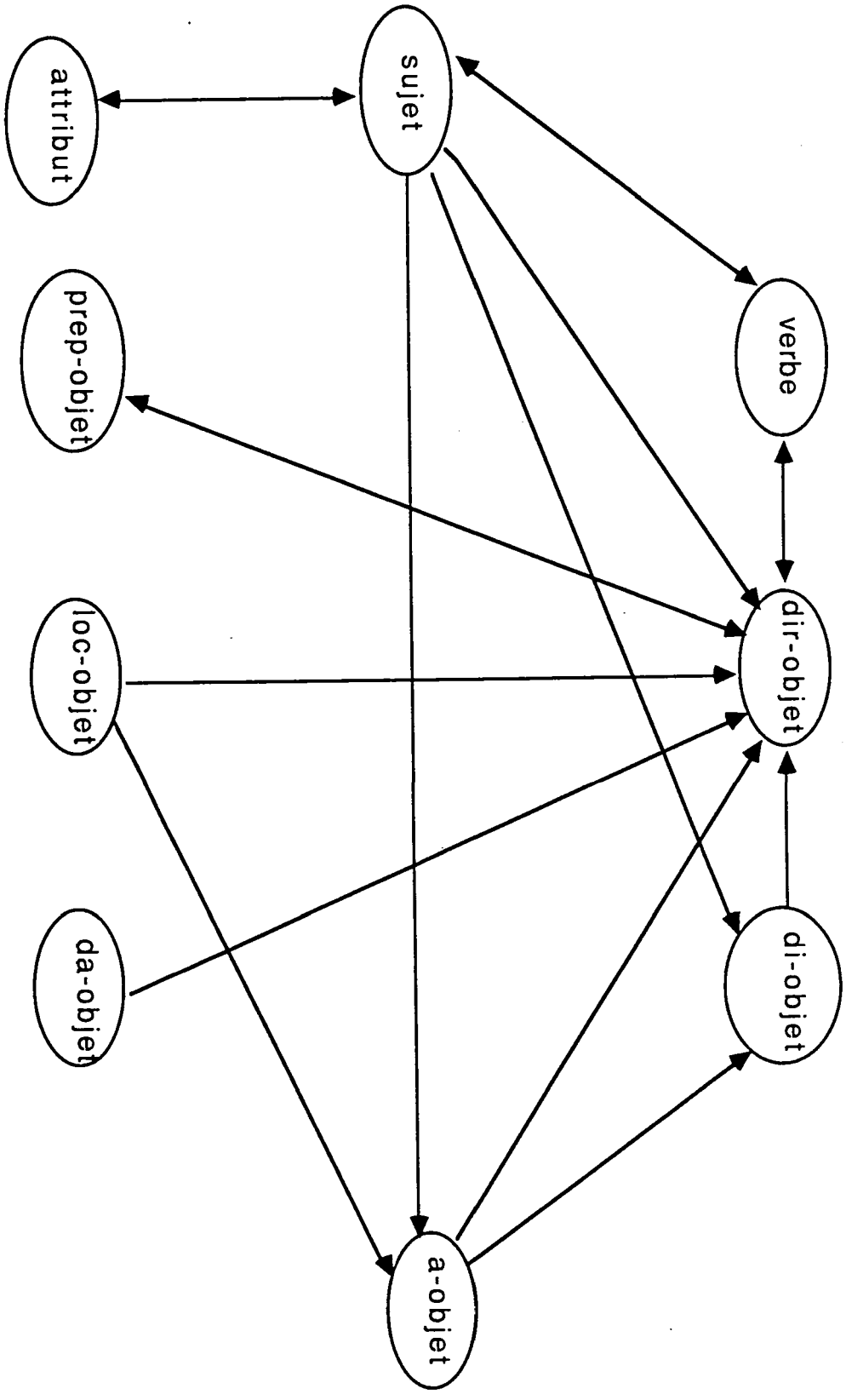
Pronominalisation et effacement du sujet imposent des conditions sur l'ordre de synthèse des éléments d'un schéma de phrase. Ces dépendances sont résumées dans le graphe ci-dessous, qui doit servir de base pour l'élaboration d'un algorithme de synthèse d'une phrase.

Nous venons de faire la description des opérations morphologiques et sémantiques nécessaires à la synthèse d'un pronom personnel (ou à l'effacement de ce pronom, dans le cas du sujet). Ces calculs sont basés sur des considérations solides : les notions d'accord en genre et nombre d'une part, et celles de contraintes distributionnelles de l'autre, ne peuvent pas être remises en question. Ces opérations peuvent être classées parmi les considérations intraphrastiques (à cause des dépendances croisées mises en jeu) et interphrastiques (à cause de la présence des antécédents possibles dont il faut tenir compte. Notons cependant que le contexte n'intervient que pour le calcul des antécédents morpho-sémantiques : ni leur position syntaxique, ni leur éloignement par rapport à l'étiquette que l'on est en train de synthétiser ne sont pris en compte.

Dans les deux chapitres suivants, nous allons présenter les opérations syntaxiques de pronominalisation, qui contiennent les divers aspects des contraintes de non coréférence entre deux arguments d'une même phrase, et qui reposent par conséquent sur des considération intraphrastiques (cf. chapitre 4) ; nous examinerons aussi les règles de préférence liées d'une part à la stylistique (cf. chapitre 4) et d'autre part au contexte (cf. chapitre 5). Ces règles de préférence permettent de choisir entre plusieurs formes pronominales calculées. Nous proposerons à la fin du chapitre 4, en guise de récapitulation des opérations morphologiques, sémantiques, syntaxiques et des règles stylistiques de préférence, deux algorithmes de synthèse d'une phrase qui tiennent compte des dépendances (non-locales et croisées) résumées dans le

graphe. Les règles interphrastiques de préférence serviront, elles, à déterminer notre choix entre ces deux approches.

Rappelons (cf. chapitre 1) que l'ordre de présentation de ces différentes contraintes de pronominalisation ne correspond pas forcément à l'ordre d'application de ces règles : celui-ci sera discuté à la fin du chapitre 5.



dépendance non locale ⟶

dépendance croisée ⟷

4 CONTRAINTES DE NON COREFERENCE ET ALGORITHME DE SYNTHESE D'UNE PHRASE

Ce chapitre est consacré à la présentation d'algorithmes pour la synthèse d'une phrase. L'élaboration de tels algorithmes est guidée par les dépendances mises en jeu par la pronominalisation exposées au chapitre précédent. Nous présenterons deux approches qui tiennent compte de ces dépendances : l'approche séquentielle et l'approche globale. Le choix entre les deux algorithmes, à savoir notre préférence pour l'approche globale, sera discuté au chapitre suivant ; la justification d'un tel choix résulte en effet de l'application de règles de préférence (i.e. les différents aspects du parallélisme de structure et le calcul de la distance) que nous exposerons au chapitre 5.

Mais tout d'abord nous allons examiner le calcul des antécédents syntaxiques d'un pronom dans une phrase simple. Ce calcul implique deux contraintes :

- la contrainte de non coréférence entre le sujet et les objets, qui est utilisée pour déterminer non seulement l'ensemble des antécédents syntaxiques d'un objet, mais également ceux du sujet,
- la contrainte de non coréférence entre les objets, qui sert au calcul des étiquettes syntaxiquement compatibles avec un pronom objet.

Nous verrons comment l'application de ces contraintes, dont le mécanisme est expliqué ci-dessous, joue un rôle dans la décision du choix de l'algorithme de synthèse.

4.1 CONTRAINTES DE NON COREFERENCE

4.1.1 CONTRAINTE DE COREFERENCE ENTRE SUJET ET PRONOM OBJET

L'utilisation de cette contrainte sert à déterminer quelles sont les étiquettes syntaxiquement compatibles avec le pronom envisagé. Cette contrainte est basée sur le fait qu'un objet coréférent au sujet est obligatoirement synthétisé en pronom réflexif. Comme nous l'avons annoncé à la section 3.2.2.2.3 du chapitre précédent, la première conséquence que l'on peut en tirer, est qu'un pronom personnel objet n'est pas coréférent avec le sujet : dans

(1a) *Eva la pettina.*
(Eva la peigne.)

la et *Eva* ne représentent pas le même individu.

Donc, si le sujet d'une phrase est exprimé de façon non ambiguë, (e.g. il est synthétisé en un nom propre, comme dans (1a)), l'étiquette qui représente le sujet est exclue de la liste des antécédents d'un pronom objet (de la même phrase) dont la synthèse est envisagée : l'étiquette sujet est syntaxiquement incompatible avec le pronom en question. Ainsi, à partir du schéma de texte (T1) :

(T1) *(LIA e EVA essere in cucina. EVA pettinare LIA.)*
(LIA et EVA être dans la cuisine. EVA peigner LIA.)

le texte synthétisé (1), qui contient (1a) :

(1) *Lia e Eva sono in cucina. Eva la pettina.*
(Lia et Eva sont dans la cuisine. Eva la peigne.)

est non ambigu : la liste des antécédents morpho-sémantiques du pronom *la* de (1a) contient les étiquettes qui représentent *Eva* et *Lia*. Cependant, par application de la contrainte de non coréférence, *Eva* devient antécédent syntaxiquement interdit du pronom qui par conséquent se retrouve avec un seul antécédent possible : *Lia*.

Si, par contre, le sujet de la phrase peut lui-même avoir plusieurs antécédents possibles, alors la contrainte de coréférence n'est pas applicable pour le calcul des antécédents syntaxiques du pronom objet. Ainsi, dans :

Lei la pettina.
(Elle la peigne.)

on ne peut rien affirmer de plus que le fait suivant : *lei* et *la* sont deux individus différents, ce qui ne permet pas de déterminer qui représente qui.

4.1.2 CONTRAINTE DE NON COREFERENCE ENTRE DEUX OBJETS

Il existe une autre contrainte, qui dans sa forme ressemble à la contrainte de non coréférence entre sujet et objet, et qui concerne, elle, l'incompatibilité syntaxique entre deux objets de la même phrase. Cette contrainte est basée sur le fait que généralement deux objets de la même phrase ne sont pas coréférents. Ainsi, dans :

Eva lo presenta a Max.
(Eva le presente à Max.)

Le [a-objet] qui représente *Max* et le pronom [dir-objet] *lo* sont deux individus différents. En d'autres termes, *Max* est syntaxiquement incompatible avec le pronom : l'étiquette qui représente *Max* est exclue de la liste des antécédents syntaxiques de ce pronom.

Il est clair cependant que dans certains domaines (comme, par exemple, un contexte psychanalytique), deux objets peuvent être coréférents. C'est le cas dans l'exemple ci-dessous, où *Eva* peut être une personne schizophrénique qui suit une psychanalyse :

Le psychiatre a présenté Eva$_i$ à elle-même$_i$.

Nous excluons à partir de maintenant de tels domaines.

Pour illustrer l'application de cette contrainte de non coréférence entre les objets considérons maintenant le schéma de texte (T2), dans lequel on suppose que les verbes sont conjugués au présent :

(T2) *(MAX non conoscere EVA. MOI presentare EVA a MAX.)*
(MAX ne pas connaître EVA. MOI présenter EVA à MAX.)

dans lequel l'étiquette MOI, qui représente le locuteur, est effacée en Italien et synthétisée en un pronom sujet de la première personne du singulier en Français (*je*). La synthèse de la deuxième occurrence de MAX en pronom est envisagée, ainsi que celle de la deuxième occurrence de EVA. On envisage la synthèse du [dir-objet] EVA en ppv féminin-singulier *la*. EVA est le seul élément présent dans le contexte gauche du deuxième schéma de phrase, qui soit morphologiquement compatible avec le pronom : la synthèse de *la* n'engendre donc pas d'ambiguïté. Une conséquence de cette synthèse est que le pronom envisagé pour le [a-objet] MAX est la forme invariable en genre *glie* (cf. chapitre précédent, section 3.2.2.1). La liste des antécédents morpho-sémantiques de cette forme pronominale est constituée des étiquettes MAX, masculin-singulier et EVA, féminin-singulier. Mais EVA, qui a été synthétisé de façon non ambiguë et qui apparaît en position objet dans le même schéma de phrase, est syntaxiquement incompatible avec le pronom *glie* : le seul antécédent syntaxique de la ppv est MAX ; la synthèse de ce pronom n'engendre pas d'ambiguïté, comme l'indique le texte (2) issu de la synthèse de (T2) :

(2) *Max non conosce Eva. Gliela presento domani.*
(Max ne connaît pas Eva. Je la lui présente demain.)

4.1.3 CONTRAINTE DE NON COREFERENCE ENTRE OBJET ET SUJET EFFACE

La contrainte de non coréférence entre un sujet et un objet peut être utilisée, comme nous l'avons vu en 4.1.1, pour calculer la liste des antécédents syntaxiques d'un pronom objet. Mais, de la même façon que l'on peut déduire de cette contrainte qu'un pronom personnel objet n'est pas coréférent avec le sujet, on obtient aussi la conclusion inverse, à savoir qu'une forme sujet, pronominale ou effacée, est syntaxiquement incompatible avec un objet qui n'est pas synthétisé en pronom réflexif. Comme nous l'avons annoncé à la section 3.2.2.2.3 du chapitre précédent, la contrainte de non coréférence entre sujet et objet peut s'appliquer pour rechercher les antécédents syntaxiques d'un pronom sujet effacé. Cette utilisation est illustrée par le texte suivant :

(3) *Eva è stata uccisa da Ugo. Egli le ha sparato durante la notte.*
(Eva a été tuée par Ugo. Il lui a tiré dessus pendant la nuit.)

Rappelons (sections 3.3.2.2 et 3.3.2.3 du précédent chapitre) que le pronom sujet *egli* est effaçable si sa nature humaine est récupérable (ce qui est le cas, car le verbe *sparare* ne se construit qu'avec un sujet humain), et s'il en va de même pour le genre de ce pronom. L'un des cas où le genre est récupérable est lorsqu'il est indiqué par le verbe ou l'attribut. Aucune de ces conditions n'est remplie dans la deuxième phrase de (3) : le genre, masculin, du sujet n'est marqué que par le pronom sujet *egli*. Pourtant si ce pronom est effacé, le texte (3') résultant n'est pas ambigu :

(3') *Eva è stata uccisa da Ugo. Le ha sparato durante la notte.*[1]
(Eva a été tuée par Ugo. [Il] lui a tiré dessus pendant la nuit.)

La seule interprétation (le seul antécédent possible) du sujet effacé est *Ugo*. En effet, le pronom [a-objet] *le* ne peut avoir qu'un antécédent féminin-singulier, ici *Eva*. Le sujet et ce pronom ne

1 Si l'on tient compte des critères pragmatiques, il est clair que (3') est non ambiguë. Mais rappelons que l'on ne considère ici que les résultats des calculs morphologiques, sémantiques et syntaxiques.

peuvent pas être coréférents. C'est pourquoi l'antécédent du sujet effacé est le seul autre humain présent dans le contexte gauche, c'est à dire *Ugo*.

De la même façon, considérons le texte (4) :

(4) *Ugo non ama più Eva. Egli l'ha abbandonata.*
(Ugo n'aime plus Eva. Il l'a abandonnée.)

La ppv objet direct *l'* (qui est, rappelons-le, la forme élidée de *lo*, masculin-singulier ou de *la*, féminin-singulier) n'indique pas le genre de son antécédent. Mais ce genre est marqué dans le participe passé féminin *abbandonata* (qui s'accorde en genre et nombre avec l'objet direct). Par conséquent, le pronom *l'* réfère exclusivement à *Eva*. Puisque l'antécédent de ce pronom est nécessairement différent de celui du sujet, *Eva* ne peut pas être un antécédent du sujet. C'est pourquoi l'effacement du sujet n'engendre pas d'ambiguïté :

Ugo non ama più Eva. L'ha abbandonata.
(Ugo n'aime plus Eva. [Il] l'a abandonnée.)

4.1.4 CONSEQUENCE : ELABORATION D'UN ALGORITHME

Les exemples (3) et (4) ci-dessus montrent que la synthèse du sujet dépend de celle de tous les autres éléments de la phrase. A l'inverse, l'exemple (1) de la section 4.1.1 illustre le fait que la synthèse des objets dépend de celle du sujet. Enfin, le texte (2) met en évidence la dépendance mutuelle de deux objets. L'application de la contrainte de non coréférence entre sujet et objet, ainsi que celle de la contrainte de non coréférence entre les objets, pose des problèmes pour l'élaboration d'un algorithme de synthèse d'une phrase qui doit tenir compte de toutes les dépendances mises en jeu.

La première méthode qui vienne à l'idée pour la synthèse d'une phrase, est d'imaginer un algorithme séquentiel, où les constituants sont synthétisés l'un après l'autre, dans un ordre qui respecte les dépendances exposées au chapitre 3. Mais nous allons voir, avec l'exemple de la synthèse du schéma de texte (T5), puis avec celui de la synthèse du schéma de phrase (S6), que cette approche est insuffisante pour tenir pleinement compte des contraintes qui viennent d'être présentées.

Considérons tout d'abord la synthèse de (T5) :

(T5) *(UGO portare ogni giorno REGALO a EVA. UGO adulare EVA.)*
(UGO apporter chaque jour REGALO à EVA. UGO aduler EVA.)

Nous supposerons que les verbes de (T5) sont conjugués au présent, et que l'étiquette REGALO est synthétisée en un groupe nominal masculin singulier : *un regalo* (un cadeau). Dans ces conditions, nous admettrons que le premier schéma de phrase de (T5) est synthétisé en :

Ugo porta ogni giorno un regalo a Eva.
(Ugo apporte chaque jour un cadeau à Eva.)

Avant d'aborder la synthèse de la deuxième phrase, nous rappelerons (cf. section 3.2.1 du chapitre 3) que la synthèse du verbe et celui de l'objet direct sont morcelés pour tenir compte des dépendances qui les lient. Quant à la synthèse du sujet, elle est guidée par les deux

questions propres au sujet : peut-on synthétiser un pronom ? peut-on effacer ce pronom ? La réponse à la première question suit les mêmes étapes que celles qui ont été exposées pour la synthèse d'un objet, dans le précédent chapitre (cf. section 3.2.2). Pour répondre à la seconde, on est amené à effectuer, parallèlement aux calculs des listes relatives à la synthèse d'un pronom, des opérations dans le but d'effacer le pronom sujet. Les détails de ces opérations sont présentés dans l'exemple ci-dessous. Examinons donc la synthèse du deuxième schéma de phrase de (T5).

a) En premier lieu, l'égalité entre l'étiquette sujet et l'une des étiquettes objet est vérifiée. Le résultat de cette recherche est négatif : pas de synthèse de pronom réflexif.

b) Synthèse du sujet UGO : synthèse d'un pronom

Il est nécessaire d'amorcer la synthèse du sujet avant celle du verbe, puisque la conjugaison de celui-ci dépend de la personne, du nombre et du genre de son son sujet (rappelons que le genre est lexical et non sémantique : il ne peut pas être prévu uniquement par la définition de l'étiquette). Le sujet UGO a déjà été mentionné, c'est pourquoi on envisage tout d'abord de le synthétiser un pronom. Si la synthèse du pronom est possible, alors on envisagera l'effacement de ce pronom :

- Le pronom éventuel est la forme humaine masculin-singulier *egli* ;
- La liste $L1_{[sujet]}$ des antécédents morphologiques du pronom contient les étiquettes préalablement synthétisées en groupes nominaux masculin-singulier : $L1_{[sujet]}$ = (UGO, REGALO)

Parallèlement à la liste $L1_{[sujet]}$, la liste $L'1_{[sujet]}$ est calculée : cette liste est constituée des antécédents morphologiques du pronom sujet effacé. Comme on l'a vu au chapitre 3 (cf. section 3.3.2.2) deux cas doivent être distingués :

- il existe un élément dans la phrase qui marque le genre du sujet effacé : dans ce cas, $L'1_{[sujet]} = L1_{[sujet]}$;
- sinon, $L'1_{[sujet]}$ contient les étiquettes qui ont été synthétisées en groupes nominaux féminin et masculin, du même nombre que le sujet. En d'autres termes $L'1_{[sujet]} \supset L1_{[sujet]}$.

Dans le schéma de phrase qui nous intéresse, il n'y a pas d'attribut et le verbe est conjugué à un temps simple, ce qui signifie qu'il ne s'accorde pas en genre avec son sujet. Nous nous retrouvons donc dans le deuxième cas pour le calcul de $L'1_{[sujet]}$, qui est composée des éléments féminin-singulier et masculin-singulier : $L'1_{[sujet]}$ = (UGO, REGALO, EVA) ;

- A partir de $L1_{[sujet]}$, on calcule la liste $L2_{[sujet]}$ des antécédents sémantiques du pronom *egli*. Cette forme est spécifique à un humain, donc REGALO est exclu de $L2_{[sujet]}$, qui est réduit à la liste (UGO).

  De la même façon, on calcule la sous-liste $L'2_{[sujet]}$ de $L'1_{[sujet]}$ des antécédents distributionnellement compatibles avec le sujet vide. Le verbe *adulare* n'accepte qu'un sujet humain, donc REGALO est exclu de $L'2_{[sujet]}$, qui est réduit à la liste (UGO, EVA) ;

- La sous-liste $L3_{[sujet]}$ de $L2_{[sujet]}$ des antécédents syntaxiques du pronom sujet n'est pas calculée puisque, dans notre exemple, $L2_{[sujet]}$ ne contient qu'un élément, et que par

conséquent la synthèse du pronom *egli* n'engendre pas d'ambiguïté. Comme la synthèse du pronom sujet est possible, on peut envisager l'effacement de ce pronom. La sous-liste L'3[sujet] de L'2[sujet] des antécédents syntaxiquement compatibles avec le sujet effacé qui met en jeu la contrainte de non coréférence entre sujet et objet, n'est calculable qu'à partir de la synthèse des autres compléments de la phrase. Le calcul de L'3[sujet], qui dépend de la synthèse de l' objet direct, est donc différée.

c) Début de la synthèse du verbe *adulare :*

La conjugaison du verbe, en tenant compte de l'accord avec le sujet mais pas de celui avec l'objet direct, aboutit à la forme 3ème personne du singulier du présent de l'indicatif : *adula.* Puisque le verbe est conjugué à un temps simple, la synthèse du verbe est achevée (pas d'accord avec l'objet à envisager).

d) Synthèse de l'objet direct EVA:

Le [dir-objet] EVA a déjà été mentionné dans la phrase précédente, et n'a pas été synthétisé en un pronom réflexif : c'est pourquoi on envisage de le synthétiser en un pronom personnel :

- Le pronom objet direct de la troisième personne féminin-singulier est *la.* Mais étant donné que le verbe conjugué commence par la voyelle *a*, *la* est obligatoirement élidé en *l'* ;

- Le verbe est conjugué à un temps simple, il n'indique donc pas le genre du pronom objet direct ; de plus, il n'y a pas d'attribut de l'objet. La liste L1[dir-objet] des antécédents morphologiques de *l'* contient donc les étiquettes préalablement synthétisées en groupe nominaux masculin-singulier et féminin-singulier : L1[dir-objet] = (UGO, REGALO, EVA) ;

- Les contraintes distributionnelles du verbe *adulare* excluent l'étiquette REGALO de la sous-liste L2[dir-objet] de L1[dir-objet]. Donc L2[dir-objet] = (UGO, EVA) ;

- La sous-liste L3[dir-objet] de L2[dir-objet] des antécédents syntaxiques du pronom objet direct est calculée en appliquant la contrainte de non coréférence entre le sujet et un pronom objet : il a été établi que la synthèse du pronom sujet *egli* était possible. Or, *egli*, qui représente un individu masculin-singulier, n'a qu'un antécédent morpho-sémantique : UGO. Cette étiquette est donc exclue de la liste L3[dir-objet]. Puisque L3[dir-objet] contient un seul élément, la synthèse du pronom *l'* est autorisée.

e) Synthèse du sujet UGO: effacement du pronom

Peut-on se servir de la contrainte de non coréférence pour calculer la liste L'3[sujet] des antécédents syntaxiques du sujet effacé, à partir de la liste L'2[sujet] = (UGO, EVA), obtenue en b) ? Cette contrainte indique que le sujet effacé ne représente pas la même entité que l'objet direct, synthétisé en pronom personnel, mais elle ne peut être utilisée pour exclure des antécédents que si le [dir-objet] est défini de façon non ambiguë. Or, dans notre cas, la synthèse du pronom objet n'a pas dépendu uniquement de critères morpho-sémantiques, mais a fait appel à la synthèse du pronom sujet : en d'autres termes, la contrainte de non coréférence a déjà été utilisée pour l'objet, elle ne peut donc pas l'être pour le sujet. C'est pourquoi l'étiquette objet direct EVA ne peut pas être exclue de L'3[sujet]. De ce fait, L'3[sujet], qui est égal à L'2[sujet], contient plus d'un élément. L'effacement du pronom sujet est à ce niveau interdit car il peut engendrer une ambiguïté. Le texte résultant de la synthèse de (T5) est donc :

(5a) *Ugo porta ogni giorno un regalo a Eva. Egli l'adula.*
(Ugo apporte chaque jour un cadeau à Eva. Il l'adule.)

Bien sûr, du point de vue de la pragmatique, du focus ou du parallélisme, l'effacement du pronom sujet de la deuxième phrase n'engendre pas d'ambiguïté. Rappelons (cf. note 1) que dans ce chapitre ces critères ne sont pas pris en compte (cf. chapitre 5) : le terme "ambigu" désigne ici une ambiguïté d'ordre morpho-sémantico-syntaxique.

L'obtention du texte (5a) illustre le fait que l'approche séquentielle a pour résultat de donner priorité à la synthèse des objets par rapport à celle du sujet : la contrainte de non coréférence est appelée d'abord pour le calcul des antécédents syntaxiques d'un pronom objet ; ensuite, si cette contrainte n'est pas appliquée pour la synthèse d'un pronom objet, elle est utilisée en vue de l'effacement du pronom sujet. L'approche séquentielle ne permet donc pas la synthèse de (5b) à partir de (T5) :

(5b) *Ugo porta ogni giorno un regalo a Eva. Adula questa ragazza.*
(Ugo apporte chaque jour un cadeau à Eva. [Il] adule cette fille.)

où la contrainte de non coréférence est appelée d'abord pour le sujet, qui est effacé, et ensuite pour l'objet, dont la pronominalisation est par conséquent impossible. Par contre, la synthèse de (5b) à partir de (T5) est possible avec l'approche globale, comme nous le verrons en 4.2.

Un deuxième exemple, qui souligne lui-aussi les limitations de l'approche séquentielle, illustre non seulement les restrictions sur l'application de la contrainte de non coréférence entre sujet et objet, mais également des restrictions sur l'application de la contrainte de non coréférence entre objets :

(S6) *(EVA presentare MAX a UGO.)*
(EVA présenter MAX à UGO.)

Nous poserons comme hypothèse que les étiquettes EVA, MAX et UGO apparaissent dans le contexte gauche de (S6), où EVA est synthétisée en *Eva*, MAX en *Max* et UGO en *Ugo*. De plus, nous supposerons que l'étiquette MAX peut être redécrite en *il vecchietto* (le vieillard), UGO en *il ragazzo* (le jeune homme) et EVA en *la ragazza* (la jeune fille). Examinons la synthèse de (S6) :

a) Il n'y a pas de synthèse de pronom réflexif.

b) Synthèse du sujet EVA : synthèse d'un pronom

Le sujet EVA ayant déjà été mentionné, on envisage tout d'abord de le synthétiser un pronom. Si la synthèse du pronom est possible, alors on envisagera l'effacement de ce pronom :

- Le pronom éventuel est la forme humain féminin-singulier *lei*.

- La liste $L1_{[sujet]}$ des antécédents morphologiques du pronom contient les étiquettes préalablement synthétisées en groupes nominaux féminin-singulier : $L1_{[sujet]}$ = (EVA). Parallèlement à la liste $L1_{[sujet]}$, la liste $L'1_{[sujet]}$ est calculée : cette liste est constituée des antécédents morphologiques du pronom sujet effacé. Comme dans l'exemple précédent,

il n'y a pas d'attribut et le verbe est conjugué à un temps simple, ce qui signifie qu'il ne s'accorde pas en genre avec son sujet. $L'1_{[sujet]}$ est donc composée des éléments féminin-singulier et masculin-singulier : $L'1_{[sujet]}$ = (UGO, MAX, EVA).

- La sous-liste $L2_{[sujet]}$ de $L1_{[sujet]}$ des antécédents sémantiques du pronom *lei* n'a pas à être calculée, puisque $L1_{[sujet]}$ est réduite à un élément : EVA. La synthèse du pronom sujet est donc possible. Par contre, on calcule la sous-liste $L'2_{[sujet]}$ de $L'1_{[sujet]}$ des antécédents distributionnellement compatibles avec le sujet vide. Les contraintes sémantiques du verbe *presentare* ne permettent d'exclure aucun élément de $L'1_{[sujet]}$ qui ne contient que des étiquettes représentant des humains : par conséquent, $L'2_{[sujet]}$ = $L'1_{[sujet]}$ = (UGO, EVA, MAX).

- La sous-liste $L'3_{[sujet]}$ de $L'2_{[sujet]}$ des antécédent syntaxiquement compatibles avec le sujet effacé est calculable, en utilisant la contrainte de non coréférence entre sujet et objet, à partir de la synthèse des autres compléments de la phrase. Le calcul de $L'3_{[sujet]}$, qui dépend de la synthèse de l' objet direct et de celle du [a-objet], est donc différée.

c) Synthèse du verbe *presentare :*

La conjugaison du verbe, en tenant compte de l'accord avec le sujet mais pas de celui avec l'objet direct, aboutit à la forme 3ème personne du singulier du présent de l'indicatif : *presenta*. Puisque le verbe est conjugué à un temps simple, la synthèse du verbe est achevée (pas d'accord avec l'objet à envisager).

Etant donné que la synthèse du [a-objet] dépend de celle de l'objet direct, comme on l'a vu au chapitre précédent, la synthèse du [dir-objet] précède celle de l'objet prépositionnel.

d) Synthèse de l'objet direct MAX :

Le [dir-objet] MAX a déjà été mentionné dans le contexte gauche, et n'a pas été synthétisé en un pronom réflexif : on envisage de le synthétiser en un pronom personnel.

- Le pronom objet direct de la troisième personne masculin-singulier est *lo*.

- La liste $L1_{[dir-objet]}$ des antécédents morphologiques de *lo* contient les étiquettes synthétisées en groupes nominaux masculin-singulier : $L1_{[dir-objet]}$ = (MAX, UGO).

- $L2_{[dir-objet]}$ = $L1_{[dir-objet]}$, car ni la forme du pronom, ni les propriétés distributionnelles du verbe *presentare* ne permettent d'exclure un des éléments de $L1_{[dir-objet]}$.

- La sous-liste $L3_{[dir-objet]}$ de $L2_{[dir-objet]}$ des antécédents syntaxiques du pronom objet direct est calculée en appliquant la contrainte de non coréférence entre les objets[2]: UGO, qui apparaît en position objet de (S6), est de ce fait syntaxiquement incompatible avec le pronom *lo* qui représente MAX, puisque deux objets de la même phrase représentent des entités différentes. Donc $L3_{[dir-objet]}$ = (MAX) et le pronom *lo* peut être synthétisé.

2 Puisque EVA n'appartient pas à $L2_{[dir-objet]}$, la contrainte de non coréférence entre sujet et objet n'intervient pas ici.

e) Synthèse du [a-objet] UGO :

On envisage de synthétiser UGO, qui a déjà été mentionné, en un pronom personnel.

- la forme d'un pronom [a-objet] masculin-singulier dans une construction avec le verbe *presentare* est *gli* (rappelons que la forme du pronom [a-objet] dépend du verbe) ; mais en présence de *lo*, le pronom est altéré en *glie* et cette forme s'agglutine avec le pronom [dir-objet] *lo*.

- La liste $L1_{[a\text{-objet}]}$ des antécédents morphologiques de *glie* contient les étiquettes synthétisées en groupes nominaux masculin-singulier et féminin-singulier : $L1_{[a\text{-objet}]}$ = (MAX, UGO, EVA)

- Pour les raisons qui ont été évoquées lors de la synthèse du [dir-objet] MAX, $L2_{[a\text{-objet}]} = L1_{[a\text{-objet}]}$ = (MAX, UGO, EVA).

- La liste $L3_{[a\text{-objet}]}$ des antécédents syntaxiques du pronom *glie* est calculée en utilisant la contrainte de non coréférence entre le sujet et l'objet (on obtient une liste provisoire $L3a_{[a\text{-objet}]}$) et la contrainte de non coréférence entre les objets :

a) contrainte de non coréférence entre le sujet et le [a-objet] : le sujet *lei* est défini de façon non ambigu, son seul antécédent est féminin-singulier : EVA. Donc EVA est syntaxiquement incompatible avec le pronom *glie*. $L3a_{[a\text{-objet}]}$ = (UGO, MAX).

b) contrainte de non coréférence entre l'objet direct et le [a-objet] : cette contrainte a déjà été utilisée pour calculer les antécédents syntaxiques du pronom objet direct. Elle ne peut pas être utilisée une deuxième fois pour le calcul de ceux du [a-objet]. Par conséquent, MAX ne peut pas être exclu de $L3a_{[a\text{-objet}]}$: $L3_{[a\text{-objet}]} = L3a_{[a\text{-objet}]}$ = (UGO, MAX). Puisque $L3_{[a\text{-objet}]}$ contient plus d'un élément, la synthèse du pronom est interdite.

On envisage alors la synthèse de UGO en un groupe nominal qui sert de redescription : *il ragazzo* et qui n'a qu'une interprétation possible.

f) Synthèse du sujet EVA : effacement du pronom

A partir de $L'2_{[sujet]}$ = (UGO, EVA, MAX), on calcule la sous-liste $L'3_{[sujet]}$ des antécédents syntaxiques du sujet effacé, en utilisant, si cela est possible, la contrainte de non coréférence entre le sujet et l'objet direct (on obtient une liste provisoire : $L'3a_{[sujet]}$), puis la contrainte de non coréférence entre le sujet et le [a-objet].

a) contrainte de non coréférence entre le sujet et l'objet direct : cette contrainte est appliquée, puisqu'elle n'a pas été utilisée pour la synthèse de l'objet direct. Celui-ci a été synthétisé, de façon non ambiguë, en *lo* ; le seul antécédent de ce pronom (i.e. MAX) est donc syntaxiquement incompatible avec le sujet effacé. Par conséquent, $L'3a_{[sujet]}$ = (UGO, EVA).

b) contrainte de non coréférence entre le sujet et le [a-objet] : cette contrainte a déjà été utilisée (à l'étape e) pour la synthèse du pronom [a-objet] *glie*. Cependant cette synthèse n'a pas abouti : le [a-objet] a été synthétisé un en groupe nominal, qui n'a qu'un antécédent possible : UGO. Cette étiquette est donc exclue de $L'3_{[sujet]}$. Donc $L'3_{[sujet]}$ = (EVA), et le pronom sujet est par conséquent effacé. La phrase obtenue à partir de (S6) est alors :

(6a) *Lo presenta al vecchietto.*
([Elle] le présente au vieillard.)

Comme c'est le cas pour la synthèse de (T5), l'approche séquentielle donne une priorité à l'application de la contrainte de non coréférence pour la synthèse de certains constituants : dans le cas de (T5), nous avons vu qu'elle donne priorité à la synthèse des objets par rapport à celle du sujet. L'exemple ci-dessus montre qu'elle donne priorité à la synthèse de l'objet direct par rapport à celle du [a-objet] : cette approche ne permet pas d'obtenir l'autre version à laquelle on peut s'attendre lors de la synthèse de (S6) :

(6b) *Gli presenta il ragazzo.*
([Elle] lui présente le jeune homme.)

où la contrainte de non coréférence entre objets est appelée d'abord pour le [a-objet], qui est pronominalisé, et ensuite pour le [dir-objet], dont la pronominalisation est alors impossible, comme nous allons le voir. En effet, pour synthétiser (6b) ou (5b) il est nécessaire d'aborder la synthèse de la phrase par une approche globale.

4.2 L'APPROCHE GLOBALE

La rigidité de l'approche séquentielle, qui a pour première conséquence d'imposer un ordre de priorité dans l'application des contraintes de non coréférence, nous a conduit à rechercher comment aborder la synthèse d'une phrase sans favoriser un constituant par rapport à un autre, pour aboutir ainsi à un choix entre plusieurs textes générés.

L'idée générale de l'approche globale est d'envisager en parallèle, pour chaque entité d'un schéma de phrase, l'ensemble de toutes ses formes nominales et pronominales possibles. En dressant toutes les combinaisons possibles, dont chaque élément est un constituant de la phrase, on obtient ainsi l'ensemble des phrases syntaxiquement acceptables qui sont issues à partir d'un même schéma. Par exemple, à partir du deuxième schéma de phrase de (T5), on génère en théorie cinq phrases :

(a) [UGO = pronom, EVA = pronom]	*Egli l'adula.* (Il l'adule.)
(b) [UGO = ø, EVA = groupe nominal]	*Adula questa ragazza.* ([Il] adule cette fille.)
(c) [UGO = g.n., EVA = pronom]	*Quest'uomo l'adula.* (Cet homme l'adule.)
(d) [UGO et EVA = g.n.]	*Quest'uomo adula questa ragazza.* (Cet homme adule cette fille.)
(e) [UGO = pronom, EVA = g.n.]	*Egli adula questa ragazza.* (Il adule cette fille.)
(f) [UGO = ø, EVA = pronom]	*L'adula.* ([Il] l'adule.)

parmi lesquelles, à la suite d'un certain nombre de critères de préférence, décrits plus loin, les phrases (c), (d) et (e) sont exclues. D'un point de vue strictement morpho-sémantico-syntaxique (qui est le seul traité dans ce chapitre), seules les phrases (a) et (b) seront retenues ; d'autres considérations, comme le parallélisme (cf. chapitre 5) permettent éventuellement de retenir également la phrase (f). Mais rappelons encore une fois que nous ne nous occupons dans ce chapitre que des résultats des critères morphologiques, sémantiques et syntaxiques, (i.e. les critères les plus fiables), contrairement à d'autres critères, moins fiables, tels que la pragmatique, ou encore le parallélisme et le calcul de la distance, qui seront pris en compte au prochain chapitre.

On peut décomposer l'approche globale en deux séries d'opérations :

1) Les opérations, effectuées pour chaque étiquette, qui aboutissent à ce que toutes les formes pronominales et nominales possibles sont envisagées.

2) Les opérations qui impliquent tous les constituants (étiquettes) du schéma de phrase, avec toutes leurs formes possibles.

4.2.1 TOUTES LES FORMES ENVISAGEABLES POUR UNE ETIQUETTE

Supposons que TOKi représente l'un des constituants du schéma de phrase à synthétiser. La liste Ui des formes possibles pour TOKi est calculée de la manière suivante :

- si TOKi est un objet, et qu'il est égal au sujet, alors Ui = (pronom-réflexif)
- si TOKi n'a jamais été mentionné, alors Ui = (GN)
- si TOKi a déjà été mentionné, et qu'il en existe d'autres descriptions, alors Ui contient un groupe nominal qui redécrit TOKi : Ui = (redescription).
- sinon, on engendre la liste Ui des possibilités de synthèse nominales et pronominales.

Le contenu de Ui est limité à un groupe nominal de redescription quand, par exemple, TOKi se définit aussi bien par son nom que par ses titres : ainsi, si TOKi représente un humain célèbre, on énumère ses titres avant d'employer une forme pronominale :

> *François Mitterrand est arrivé. **Le Président** a serré la main de ses ministres. Il a ensuite quitté la pièce.*

est un texte jugé meilleur que :

> *François Mitterrand est arrivé. **Il** a serré la main de ses ministres. Il a ensuite quitté la pièce.*

Dans les autres cas, la redescription est un groupe nominal utilisé pour synthétisé TOKi, quand toute forme pronominale est ambiguë :

> *Eva et Ugo sont arrivés. *Je **lui** ai dit de couper du bois.*
> --> *Eva et Ugo sont arrivés. J'ai dit **au garçon** de couper du bois.*

La constitution de la liste Ui est le résultat de deux opérations :

- le calcul des formes pronominales et nominales associées à TOKi,
- le calcul des antécédents morphologiques et sémantiques de TOKi.

Par exemple, reprenons le schéma de phrase (S6) de la section précédente avec les mêmes hypothèses concernant son contexte gauche, et les possibilités de redescription des étiquettes :

(S6) *(EVA presentare MAX a UGO.)*

(EVA présenter MAX à UGO.)

Les listes U[sujet], U[dir-objet] et U[a-objet] sont compilées comme suit :

U[sujet] = ([pro] : *lei* [EVA] ,
[eff] : ϕ [EVA, UGO, MAX],
[redes] : *la ragazza* [EVA])
U[dir-objet] = ([ppv] : *lo* [UGO, MAX] ,
[redes] : *il vecchietto* [MAX])
U[a-objet] = ([pro] : *lui* [UGO, MAX],
[ppv] : *glie* <dir-objet = *lo, la, le, li* ou di-objet = *ne*> [UGO, EVA, MAX],
[ppv] : *gli* <dir-objet ≠ *lo, la, le, li, mi, ti, ci, vi* et di-objet ≠ *ne*> [UGO, MAX],
[redes] : *il ragazzo* [UGO])

Le calcul des formes pronominales (i.e. des pronoms personnels) est effectué selon les critères décrits au chapitre précédent. Si l'acceptabilité d'une forme pronominale dépend de la présence d'un autre pronom, alors on associe cette dépendance à la forme envisagée. Ainsi, dans l'exemple ci-dessus, il est indiqué par :

glie <dir-objet = *lo, la, le, li* ou di-objet = *ne*>

que la synthèse de la ppv *glie* dépend de la synthèse du [dir-objet] (qui doit être l'une des ppv *lo, la, li, le*) ou de celle du [di-objet] qui doit valoir *ne*.

Le calcul des antécédents morpho-sémantiques qui a également été présenté au chapitre précédent, suppose dans notre exemple que le contexte gauche de (S6) est constitué des étiquettes EVA, MAX et UGO.

D'une manière plus formelle, une liste $U_{[constituant]}$, où

[constituant] = [sujet] / [dir-objet] / [a-objet] / [di-objet] / [da-objet] / [loc-objet] / [prep-objet]

peut être décrite récursivement comme suit :

$U_{[constituant]}$ = ($forme_1$, $forme_2$, ... , $forme_i$, ... , $forme_n$), avec $n \geq 1$
$forme_i$ = [type] : lex_i < conditions> [TOK1, ... , TOKj] /
[type] : lex_i [TOK1, ... , TOKj], avec $j \geq 1$
type = pro / eff / redes / ppv / gn
pro désigne un pronom non préverbal,
eff, un élément effacé,
redes, une redescription, i.e. un groupe nominal
ppv, une particule préverbale,
gn, un groupe nominal.
conditions = condition ou conditions / condition
condition = cond et condition / cond

lex_i est soit un élément lexical, soit un groupe nominal, soit le symbole Ø (i.e. l'élément vide)

cond est une condition concernant la valeur d'un autre constituant de la phrase, dont dépend l'acceptabilité de l'élément lex_i.

[TOK1 ... TOKj] est l'ensemble des antécédents morpho-sémantiques de la forme lex_i.

Par souci de clarté, nous conserverons à partir de maintenant la notation :

$U_{[constituant]}$ = ($forme_1$, $forme_2$, ... , $forme_i$, ... , $forme_n$), avec $n \geq 1$
$forme_i$ = lex_i < conditions> [TOK1, ... , TOKj] /
lex_i [TOK1, ... , TOKj], avec $j \geq 1$

dans laquelle [type] est supprimé. Voyons alors un autre exemple, où l'on suppose que les étiquettes EVA et UGO ont déjà été mentionnées :

(S7) *(EVA pensare a UGO.)*
(EVA penser à UGO.)

Lorsque la forme envisagée dépend de la nature du verbe, il est possible de vérifier tout de suite si cette forme doit être exclue de la liste U_i. Dans (S7), les formes pronominales *gli* et *glie* sont incompatibles avec le verbe *pensare* (**Eva gli pensa* (*Eva lui pense)). Les listes $U_{[sujet]}$ et $U_{[a\text{-}objet]}$ sont donc les suivantes :

$U_{[sujet]}$ = ([pro] : *lei* [EVA],
[eff] : ∅ [EVA, UGO],
[redes] : *la ragazza* [EVA])
$U_{[a\text{-}objet]}$ = ([pro] : *lui* [UGO],
[redes] : *il ragazzo* [UGO])

4.2.2 TOUTES LES VERSIONS POSSIBLES POUR UN SCHEMA DE PHRASE

Combinaison des listes Ui :

Après avoir calculé, pour chaque étiquette, l'ensemble Ui des formes (pro)nominales avec les antécédents morpho-sémantiques de chacune d'elles, on forme les listes Cj qui forment la combinaison, pour chaque constituant i, des éléments de Ui. Plus précisément, si le schéma de phrase à synthétiser a n constituants, et que :

$U_1 = (forme^1_1, \ldots, forme^1_k, \ldots, forme^1_{g_1})$, a g_1 éléments

...

$U_i = (forme^i_1, \ldots, forme^i_l, \ldots, forme^i_{g_i})$, a g_i éléments

...

$U_n = (forme^n_1, \ldots, forme^n_m, \ldots, forme^n_{g_n})$, a g_n éléments

alors il existe $g_1 \times \ldots \times g_i \times \ldots g_n$ combinaisons Cj, avec :

$Cj = (forme^1_k \ldots forme^i_l \ldots forme^n_m)$

Parmi ces combinaisons, on ne conserve que celles qui vérifient les contraintes établies pour chaque forme pronominale. Ainsi, les combinaisons Cj contenant à la fois :

- soit les ppv *lo* et *gli*,
- soit la ppv *glie* et une forme non atone pour l'objet direct

sont supprimées, puisqu'elles ne respectent pas les conditions liées à *gli* et *glie*. Les combinaisons obtenues pour l'exemple ci-dessus sont alors les suivantes, où le premier élément de Cj représente le sujet, le second, l'objet direct et le dernier, le [a-objet] :

C1 = (*lei* [EVA], *lo* [UGO, MAX], *lui* [UGO, MAX])
C2 = (*lei* [EVA], *lo* [UGO, MAX], *il ragazzo* [UGO])
C3 = (*lei* [EVA], *lo* [UGO, MAX], *glie* [UGO, EVA, MAX])
C4 = (*lei* [EVA], *il vecchietto* [MAX], *gli* [UGO, MAX])
C5 = (*lei* [EVA], *il vecchietto* [MAX], *lui* [UGO, MAX])
C6 = (*lei* [EVA], *il vecchietto* [MAX], *il ragazzo* [UGO])
C7 = (ø [EVA, UGO, MAX], *lo* [UGO, MAX], *lui* [UGO, MAX])
C8 = (ø [EVA, UGO, MAX], *lo* [UGO, MAX], *il ragazzo* [UGO])
C9 = (ø [EVA, UGO, MAX], *lo* [UGO, MAX], *glie* [UGO, EVA, MAX])
C10 = (ø [EVA, UGO, MAX], *il vecchietto* [MAX], *gli* [UGO, MAX])
C11 = (ø [EVA, UGO, MAX], *il vecchietto* [MAX], *lui* [UGO, MAX])
C12 = (ø [EVA, UGO, MAX], *il vecchietto* [MAX], *il ragazzo* [UGO])
C13 = (*la ragazza* [EVA], *lo* [UGO, MAX], *lui* [UGO, MAX])
C14 = (*la ragazza* [EVA], *lo* [UGO, MAX], *il ragazzo* [UGO])
C15 = (*la ragazza* [EVA], *lo* [UGO, MAX], *glie* [UGO, EVA, MAX])
C16 = (*la ragazza* [EVA], *il vecchietto* [MAX], *gli* [UGO, MAX])
C17 = (*la ragazza* [EVA], *il vecchietto* [MAX], *lui* [UGO, MAX])
C18 = (*la ragazza* [EVA], *il vecchietto* [MAX], *il ragazzo* [UGO])

Antécédents syntaxiques :

Pour chaque Cj, on effectue le calcul des antécédents syntaxiques des formes pronominales que Cj contient, en appliquant d'abord la contrainte qui détermine que deux objets sont non-coréférents, puis la contrainte de non coréférence entre le sujet et les objets. Dans le cas de la liste :

C8 = (ø [EVA, UGO, MAX], *lo* [UGO, MAX], *il ragazzo* [UGO])

la première règle permet de supprimer l'étiquette UGO de la liste des antécédents associés à la ppv *lo*, puisque le [a-objet] *il ragazzo* est défini de façon non ambiguë :

C'8 = (ø [EVA, UGO, MAX], *lo* [MAX] <cnc-objets : UGO>, *il ragazzo* [UGO])

L'information <cnc-objets : UGO>, placée derrière la liste des antécédents du pronom *lo*, indique que la contrainte de non coréférence entre les objets a été appliquée pour modifier cette liste, et que le résultat en a été l'élimination de l'étiquette UGO. Cette information est utilisée pour que la contrainte de non coréférence entre deux objets ne soit pas utilisée deux fois. Nous utiliserons de même des informations telles que <cnc-sujet-dirobjet : TOKi> ou <cnc-sujet-aobjet : TOKi> pour que la contrainte de non coréférence entre sujet et objet ne soit pas appliquée deux fois.

CONTRAINTES DE NON-COREFERENCE SUR LES Cj

Cl	CONTRAINTE OBJET-OBJET	CONTRAINTE SUJET-DIROBJET	CONTRAINTE SUJET-AOBJET	PHRASE OBTENUE
C1 = (lei[EVA], lo[UGO, MAX], lui[UGO,MAX])	C'1 = C1	C"1 = C'1	C""1 = C"1	(6 - 1) *Lei lo [presentare] a lui* (Elle le [présenter] à lui)
C2 = (lei[EVA], lo[UGO, MAX], il ragazzo[UGO])	C'2 = (lei[EVA], lo[MAX] <cnc-objets : UGO>, il ragazzo[UGO])	C"2 = C'2	C""2 = C"2	(6 - 2) *Lei lo [presentare] al ragazzo* (Elle le [présenter] au jeune homme)
C3 = (lei[EVA], lo[UGO, MAX], glie [UGO,MAX, EVA])	C'3 = C3	C"3 = C'3	C""3 = (lei[EVA], lo[UGO, MAX], glie [UGO,MAX]<cnc-suj-ao : EVA>)	(6 - 3) *Lei glielo [presentare]* (Elle le lui [présenter])
C4 = (lei[EVA], il vecchietto[MAX], gli[UGO, MAX])	C'4 = (lei[EVA], il vecchietto[MAX], gli[UGO] <cnc-objets : MAX>)	C"4 = C'4	C""4 = C"4	(6 - 4) *Lei gli [presentare] il vecchietto* (Elle lui [présenter] le vieillard)
C5 = (*lei[EVA], il vecchietto[MAX]*, lui[UGO,MAX])	C'5 = (*lei[EVA], il vecchietto[MAX]*, lui[UGO]<cnc-objets : MAX>)	C"5 = C'5	C""5 = C"5	(6 - 5) *Lei [presentare] il vecchietto a lui* (Elle [présenter] le vieillard à lui)
C6 = (lei[EVA], il vecchietto[MAX], il ragazzo[UGO])	C'6 = C6	C"6 = C'6	C""6 = C"6	(6 - 6) *Lei [presentare] il vecchietto al ragazzo* (Elle [présenter] le vieillard au jeune h.)
C7 = (Ø[EVA, MAX, UGO], lo[UGO, MAX], lui[UGO,MAX])	C'7 = C7	C"7 = C'7	C""7 = C"7	(6 - 7) *Lo [presentare] a lui*
C8 : cet exemple a été étudié au début de cette section				
C9 = (Ø[EVA, MAX, UGO], lo[UGO, MAX], glie[UGO,MAX, EVA])	C'9 = C9	C"9 = C'9	C""9 = C"9	(6 - 9) *Glielo [presentare]* ([Elle + il] le lui [présenter])
C10 = (Ø[EVA, MAX, UGO], il vecchietto[MAX], gli[UGO,MAX])	C'10 = (Ø[EVA, MAX, UGO], il vecchietto[MAX], gli[UGO]<cnc-objets : MAX>)	C"10 = (Ø[EVA, UGO]<cnc-suj-diro : MAX>, il vecchietto[MAX], gli[UGO]<cnc-objets : MAX>)	C""10 = (Ø[EVA]<cnc-suj-diro : MAX> <cnc-suj-ao : UGO>, il vecchietto[MAX], gli[UGO]<cnc-objets : MAX>)	(6 - 10) *Gli [presentare] il vecchietto* ([Elle] lui [présenter] le vieillard)
C11 = (Ø[EVA, MAX, UGO], il vecchietto[MAX], *lui[UGO,MAX]*)	C'11 = (Ø[EVA, MAX, UGO], il vecchietto[MAX], *lui[UGO]<cnc-objets : MAX>*)	C"11 = (Ø[EVA,UGO]<cnc-suj-diro : MAX>, il vecchietto[MAX], *lui[UGO] <cnc-objets : MAX>*)	C""11 = (Ø[EVA]<cnc-suj-diro : MAX> <cnc-suj-ao : UGO>, il vecchietto[MAX], lui[UGO] <cnc-objets : MAX>)	(6 - 11) *[Presentare] il vecchietto a lui* ([Elle] [présenter] le vieillard à lui)
C12 = (Ø[EVA, MAX, UGO], il vecchietto[MAX], il ragazzo[UGO])	C'12 = C12	C"12 = (Ø[EVA,UGO]<cnc-suj-diro : MAX>, il vecchietto[MAX], il ragazzo[UGO])	C""12 = (Ø[EVA]<cnc-suj-diro : MAX> <cnc-suj-ao : UGO>, il vecchietto[MAX], il ragazzo[UGO])	(6 - 12) *[Presentare] il vecchietto al ragazzo* ([Elle] [présenter] le vieillard au jeune h.)
C13 = (la ragazza[EVA], lo[UGO, MAX], lui[UGO,MAX])	C'13 = C13	C"13 = C'13	C""13 = C"13	(6 - 13) *La ragazza lo [presentare] a lui* (La jeune fille le [présenter] à lui)
C14 = (la ragazza[EVA], lo[UGO, MAX], il ragazzo[UGO])	C'14 = (la ragazza[EVA], lo[MAX]<cnc-objets : UGO>, il ragazzo[UGO])	C"14 = C'14	C""14 = C'14	(6 - 14) *La ragazza lo [presentare] al ragazzo* (La jeune f. le [présenter] au jeune h.)
C15 = (la ragazza[EVA], lo[UGO, MAX], glie[UGO,MAX, EVA])	C'15 = C15	C"15 = C'15	C""15 = (la ragazza[EVA], lo[UGO, MAX], glie[UGO,MAX] <cnc-suj-ao : EVA>)	(6 - 15) *La ragazza glielo [presentare]* (La jeune f. le lui [présenter])
C16 = (la ragazza[EVA], il vecchietto[MAX], gli[UGO,MAX])	C'16 = (la ragazza[EVA], il vecchietto[MAX], gli[UGO]<cnc-objets : MAX>)	C"16 = C'16	C""16 = C"16	(6 - 16) *La ragazza gli [presentare] il vecchietto* (La jeune f. lui [présenter] le vieillard)
C17 = (la ragazza[EVA], il vecchietto[MAX], *lui[UGO,MAX]*)	C'17 = (la ragazza[EVA], il vecchietto[MAX], *lui[UGO]<cnc-objets : MAX>*)	C"17 = C'17	C""17 = C"17	(6 - 17) *La ragazza [presentare] il vecchietto a lui* (La jeune f. [présenter] le vieillard à lui)
C18 = (la ragazza[EVA], il vecchietto[MAX], il ragazzo[UGO])	C'18 = C18	C"18 = C'18	C""18 = C"18	(6 - 18) *La ragazza [presentare] il vecchietto al ragazzo* (La jeune f. [présenter] le vieillard au jeune h.)

Maintenant, grâce à la contrainte de non coréférence entre le sujet et les objets, les étiquettes MAX et UGO sont éliminées de la liste des antécédents du sujet effacé : UGO, pour les raisons déjà évoquées ci-dessus :

C"8 = (ø [EVA, MAX] <cnc-sujet-aobjet : UGO>,
lo [MAX] <cnc-objets : UGO>, *il ragazzo* [UGO])

et MAX, parce que les raisons de l'unicité de son antécédent sont indépendantes du sujet. Après l'application de cette contrainte, la liste C"8 devient :

C'"8 = (ø [EVA] <cnc-sujet-aobjet : UGO> <cnc-sujet-dirobjet : MAX>,
lo [MAX] <cnc-objets : UGO>, *il ragazzo* [UGO])

Le verbe est synthétisé après les autres constituants de la phrase, comme on le verra plus loin dans ce chapitre. De ce fait, la phrase que l'on pourrait obtenir à partir de C'"8 est la suivante, si l'on ne tient pas compte de la conjugaison du verbe, dont la forme infinitive apparaît entre crochets :

(6-8) *Lo [presentare] al ragazzo*
([Elle] le *[présenter]* au jeune homme.)

Dans le tableau de la page précédente, nous présentons brièvement comment s'appliquent les contraintes de non coréférence dans le cas des autres combinaisons Cj, et sur quels textes les listes ainsi modifiées pourraient déboucher :

- la première colonne contient les combinaisons Cj de départ ; on applique à chacune la contrainte de non coréférence entre objets, et on obtient (colonne 2) des listes C'j ;
- on applique ensuite aux C'j la contrainte de non coréférence entre le sujet et l'objet direct (colonne 3) ;
- enfin, on applique aux listes C"j résultantes la contrainte de non coréférence entre le sujet et le [a-objet] (colonne 4) ;
- les combinaisons finales (C'"j) donneraient alors lieu aux phrases (6 - j), qui sont exposées dans la dernière colonne, en laissant de côté la synthèse du verbe.

L'égalité dans une colonne donnée entre deux listes de combinaisons (e.g. C'1 = C1) est due à la non application de l'une des contraintes de non coréférence, pour l'une des deux raisons :

- les entités comparées n'ont pas d'antécédents en commun (c'est le cas pour C'6 = C6),
- les entités comparées ont toutes deux plusieurs antécédents (c'est le cas pour C'1 = C1).

Règles de préférence :

Après le calcul des antécédents syntaxiques pour toutes les Cj, on est amené à effectuer un certain nombre de choix.

a) Règles de préférence liées au contexte :

Soulignons qu'il existe des règles de préférence basées sur la pragmatique, la distance et le parallélisme qui permettent éventuellement d'apporter des modifications à ces listes :

nous supposerons que ces règles de préférence ont été prises en compte au cours du mécanisme de synthèse. Les critères de parallélisme de structure et le calcul de la distance entre une forme pronominale et ses antécédents morphologiques, sémantiques et syntaxiques sont présentés au prochain chapitre, où nous expliquerons à quel moment dans l'algorithme ils doivent être appliqués.

b) Règles de préférence liées à l'ambiguïté :

Nous laissons pour l'instant de côté les Cj dont l'un des éléments a plus d'un antécédent ; c'est, par exemple, le cas de la liste C9 dans laquelle, après calcul des antécédents syntaxiques, chaque élément a encore au moins deux antécédents possibles :

C9 = (ϕ [EVA, MAX, UGO], *lo* [UGO, MAX], *glie* [UGO, MAX, EVA])

Dans un premier temps, on prendra donc pour hypothèse que le contexte n'intervient pas ; on supprime les Cj dont un élément au moins a plus d'un antécédent. Cette décision est alors correcte, car nous avons vu que la liste C9, par exemple, déboucherait à la phrase (6-9) qui est ambiguë. On remarquera que cette règle, effectuée dans l'approche globale, équivaut à la décision finale effectuée dans l'approche séquentielle, et qui consiste à synthétiser un pronom seulement si celui-ci n'a qu'un antécédent. Dans l'exemple de notre schéma de phrase (S6), après application de cette règle de préférence, il reste les liste Cj suivantes :

C'2, C'4, C'5, C6, C'"8, C'"10, C'"11, C'12, C'14, C'16, C'17 et C18.

c) Règles de préférence stylistiques :

On effectue une gradation par ordre de préférence des Cj restantes. Il est connu que dans un texte, on a tendance à éviter les répétitions et à rechercher le plus possibles les formes pronominales, pour la cohérence qu'elles donnent à ce texte : c'est pourquoi, plus une liste contient de formes pronominales préverbales (le sujet vide est compté comme ppv), plus la note attribuée à cette liste est élevée; une ppv a une note plus élevée qu'un pronom non préverbal, qui lui même est préféré à une redescription : la note la plus basse est ainsi réservée aux listes qui ne contiennent que des groupes nominaux[3]. Dans notre exemple, l'ordre des préférences est le suivant :

- les listes C'"8 et C'"10, qui contiennent un sujet effacé et une ppv
- les listes C'2, C'4, C'"11 qui contiennent une ppv (ou un sujet effacé, pour C'"11) et une forme tonique.
- la liste C'5 : deux formes toniques
- les listes C'12, C'14 et C'16, contenant chacune une ppv (ou le sujet effacé, pour C'12)
- les listes C6 et C'17, avec chacune une forme tonique.
- la liste C18, constituée de trois groupes nominaux.

3 Une ppv (ainsi que le sujet effacé) est notée 3, un pronom non préverbal, 2, un démonstratif, 1, et un groupe nominal 0 : pour connaître la note attribuée à une liste Cj, on additionne les notes obtenues pour chacun de ses éléments.

Cet ordre est justifié, puisque l'on préfère écrire :

Gli presenta il vecchietto.
([Elle] lui$_{masc}$ présente le vieillard.)

qui correspond à C'"10, à :

Lei gli presenta il vecchietto.
(Elle lui$_{masc}$ présente le vieillard.)

qui résulte de C'2, cette phrase étant elle même meilleure que la version provenant de C'5 :

Lei presenta a lui il vecchietto.
(Elle présente à lui le vieillard.)

Cette dernière phrase est, quant à elle, préférée à la suivante, produite à partir de C'12 :

Presenta il vecchietto al ragazzo.
([Elle] présente le vieillard au jeune homme.)

Enfin, le dernier choix que l'on puisse faire, quand aucune forme pronominale n'est possible, est la version provenant de C18, dernière liste dans l'ordre de préférence :

La ragazza presenta il vecchietto al ragazzo.
(La jeune fille présente le vieillard au jeune homme.)

Ce système de notation de l'approche globale peut être mis en parallèle avec le fait suivant dans l'approche séquentielle : quand on envisage de synthétiser une étiquette en pronom, on tente d'abord la synthèse d'une forme préverbale; en cas d'échec, la forme tonique, si elle existe, est envisagée. Enfin, si aucune forme pronominale n'est possible, on synthétise cette étiquette en un groupe nominal (i.e. en une redescription ou une répétition). Cette remarque sera développée au chapitre 5 où le fait suivant sera également mis en évidence : dans l'approche séquentielle, les Cj sont construites et examinées l'une après l'autre, jusqu'à obtention d'une Cj satisfaisante. Par contre, dans l'approche globale, toutes les Cj pouvant représenter une version du schéma de phrase du départ sont établies en même temps. Dans l'exemple de la synthèse de (S6), les listes qui sont conservées après l'application des règles de préférence stylistiques sont C'"8 et C'"10 :

C'"8 = (ϕ [EVA], *lo* [MAX], *il ragazzo* [UGO])
C'"10 = (ϕ [EVA], *il vecchietto* [MAX], *gli* [UGO])

La synthèse du verbe :

Le verbe est synthétisé après l'étape ci-dessus. En italien, pour respecter le fait que la synthèse de l'objet direct dépend de la conjugaison du verbe, il suffit en effet, quand on envisage une forme pronominale [dir-objet], de vérifier qu'elle est compatible avec les paramètres de conjugaisons du verbe, ainsi qu'avec la première lettre de la forme infinitive de celui-ci. Ces mêmes examens sont faits pour le calcul de la liste des antécédents

morphologiques du pronom. En français, la situation est plus complexe ; nous y reviendrons en détail au chapitre suivant.

Par exemple, si on envisage la synthèse du pronom [dir-objet] féminin-singulier de la 3ème personne, et que les informations liées au verbe indiquent que celui-ci doit être conjugué à un temps composé (avec *avere*), alors la ppv choisie est *l'*, et la liste des antécédents morphologiques est la même que celle qui serait calculée pour la ppv *la*.

Par contre, si le verbe doit être conjugué à un temps simple, alors l'analyse de la forme infinitive du verbe permet de décider quelle sera la ppv choisie (*la* ou *l'*) et comment calculer les antécédents en genre et nombre. Quand la (les) liste(s) finale(s) Cj a (ont) été sélectionnée(s), le verbe est conjugué et éventuellement accordé en genre et nombre avec l'objet direct. Si on suppose que verbe de (S6) est conjugué à un temps simple, on obtient : *presenta*. Finalement, les deux phrases résultantes de la synthèse de (S6) sont :

1) avec C'"8 = (ø [EVA], *lo* [MAX], *il ragazzo* [UGO]) : (6-8), qui est égale à la phrase (6a) obtenue par l'approche séquentielle, dans la section précédente :

(6-8) *Lo presenta al ragazzo.*
([Elle] le présente au jeune homme.)

2) avec C'"10 = (ø [EVA], *il vecchietto* [MAX], *gli* [UGO]) : (6-10), qui est la phrase (6b) que l'on ne pouvait pas obtenir par l'approche séquentielle :

(6-10) *Gli presenta il vecchietto.*
([Elle] lui présente le vieillard.)

4.3 CONCLUSION

Les calculs effectués sur les listes d'antécédents (recherche des antécédents morphologiques, sémantiques et syntaxiques d'une forme pronominales) ont pour résultat l'obtention de textes qui ne sont jamais ambigus. Cependant, l'absence d'ambiguïté n'est pas le seul critère d'acceptabilité d'un texte :

- dans certaines circonstances, la synthèse d'une forme pronominale (ou l'effacement de ce pronom) est possible, même quand le nombre d'antécédents est supérieur à un. Ainsi, dans :

(8) *Ugo incontra Eva. Egli l'abbraccia.*
(Ugo rencontre Eva. Il l'embrasse.)

le sujet *egli* de la deuxième phrase peut être effacé, même si l'élément vide a deux antécédents : l'interprétation préférentielle du sujet vide reste le sujet de la phrase précédente :

Ugo incontra Eva. L'abbraccia.
(Ugo rencontre Eva. [Il] l'embrasse.)

Une autre illustration est que dans :

(9) *Max si è rotto una gamba. Eva e Lia si sono occupate della casa. Ugo le ha aiutate. Sono molto grate a Ugo.*
(Max s'est cassé la jambe. Eva et Lia se sont occupées de la maison. Ugo les a aidées. [Elles] sont très reconnaissantes à Ugo.)

l'objet direct *Ugo* de la dernière phrase peut être pronominalisé en *gli*, bien que la ppv ait à priori deux antécédents : les étiquettes représentant *Max* et *Ugo*.

Max si è rotto una gamba. Eva e Lia si sono occupate della casa. Ugo le ha aiutate. Gli sono molto grate.
(Max s'est cassé la jambe. Eva et Lia se sont occupées de la maison. Ugo les a aidées. [Elles] lui sont très reconnaissantes.)

- à l'inverse, il est possible que, malgré l'unicité des antécédent d'une forme pronominale, la synthèse de celle-ci soit déconseillée, voire interdite ; ainsi, dans le texte (10) :

(10) *Eva ha incontrato Max. Egli l'ha vista ieri.*
(Eva a rencontré Max. Il l'a vue hier.)

L'effacement du pronom sujet *egli* de la deuxième phrase n'engendre pas d'ambiguïté, à cause de la contrainte de non coréférence (l'objet direct *l'*, dont le genre est marqué par le participe passé *vista*, ne peut référer qu'à une entité féminin-singulier). Pourtant, dans la langue écrite, on préfère conserver le pronom :

?Eva ha incontrato Max. L'ha vista ieri.
(Eva a rencontré Max. [Il] l'a vue hier.)

De façon analogue, on ne peut pas pronominaliser (ni effacer) le sujet *Max* de la dernière phrase du texte (11) :

(11) *Max si è rotto una gamba. Eva e Lia si sono occupate della casa. Hanno lavato i piatti. Eva ha ridipinto le pareti. Lia ha stirato le lenzuola. Max è stato trattato come un principe.*
(Max s'est cassé la jambe. Eva et Lia se sont occupées de la maison. [Elles] ont fait la vaisselle. Eva a repeint les murs. Lia a repassé les draps. Max a été traité comme un prince.)

bien que le pronom envisagé (*egli*) n'ait qu'un antécédent possible :

**Max si è rotto una gamba. Eva e Lia si sono occupate della casa. Hanno lavato i piatti. Eva ha ridipinto le pareti. Lia ha stirato le lenzuola. Egli è stato trattato come un principe.*
(*Max s'est cassé la jambe. Eva et Lia se sont occupées de la maison. [Elles] ont fait la vaisselle. Eva a repeint les murs. Lia a repassé les draps. [Il] a été traité comme un prince.)

Plusieurs phénomènes interviennent pour expliquer ces résultats sur la pronominalisation :

- les relations pragmatiques et la notion de focus ont déjà été évoquées au chapitre précédent. L'exemple (8) peut être interprété comme relevant de la notion de focus : le focus des deux phrases (i.e. le sujet) représentent le même individu, donc le sujet de la deuxième phrase peut être effacé ;

- mais la notion de "préférence d'interprétation" (illustrée par les textes (8) et (10)) peut être expliquée également en termes de syntaxe et de géométrie d'arbres : le parallélisme existant entre deux structures peut justifier non seulement la coréférence entre les sujets, mais aussi celle des objets. Dans le chapitre qui suit sont présentés certains résultats formels liés à ce phénomène. Nous y exposerons également des cas de coréférence obligatoire des sujets de deux phrases consécutives, en fonction de la valeur lexicale de ces sujets et nous essaierons de donner une explication à la différence d'acceptabilité qui existe entre le sujet vide et le pronom sujet *egli*, quand cette différence n'est pas perceptible en français, comme c'est le cas de l'exemple (10) ;

- enfin, comme le montrent les textes (9) et (11), la distance (dont nous reparlerons également au chapitre 5) intervient pour autoriser ou interdire la synthèse d'une forme pronominale. Rappelons que ce concept peut être soit calculé de façon linéaire (i.e. en comptant le nombre de mots ou de phrases qui séparent deux entités, ce qui est illustré dans nos exemples), soit en termes de profondeur, ce qui nous rapproche d'études théoriques (N. Chomsky (1980), T. Reinhard (1976), J. Guéron (1979)), qui ont comparé référent d'anaphore dans un arbre et géométrie de cet arbre (notion de domaine, de c-commande, etc ...). Soulignons cependant que ces recherches concernent essentiellement les relations anaphoriques à l'intérieur d'une même phrase.

Distance et parallélisme sont deux considérations interphrastiques dont les résultats peuvent être interprétés comme des règles de préférence. Contrairement aux notions de pragmatique et de focus, elles ont l'avantage de s'appuyer sur des calculs formels et indépendants du domaine (peut-on concevoir un système assez puissant pour y intégrer des relations pragmatiques universelles ?). Du fait de leur fiabilité moins grande que les contraintes morphologiques, sémantiques et syntaxiques, ces deux phénomènes ont donné lieu à des résultats qui ne peuvent être jugés ni définitifs ni absolus. Ces résultats sont décrits dans le chapitre suivant, ainsi que leur intégration parmi les autres opérations de synthèse : comme nous l'avons déjà annoncé, ils permettent de justifier le choix de l'approche globale.

5 CONSIDERATIONS INTERPHRASTIQUES DE PRONOMINALISATION : CHOIX D'UN ALGORITHME

Les deux chapitres précédents étaient consacrés à la recherche de contraintes morphologiques, sémantiques et syntaxiques de pronominalisation ; le contexte gauche n'y est pris en compte qu'en tant qu'ensemble non structuré d'étiquettes, pour déterminer la forme (pro)nominale à synthétiser ainsi que son genre. Ces contraintes mettent en jeu des opérations purement formelles dont les résultats ne sont pas discutables. Elles ont servi de base à l'élaboration de deux algorithmes. Dans ces deux chapitres, nous avons également présenté les règles de préférence stylistiques qui favorisent entre autres l'emploi d'une particule préverbale à celle d'une forme tonique.

Les règles que nous présentons ici sont également des règles de préférence. Elles reposent, elles, sur la structure du contexte gauche des pronoms dont la synthèse est envisagée, d'où leur nom. Contrairement aux contraintes morphologiques, sémantiques et syntaxiques, les résultats auxquelles elles aboutissent sont discutables : on ne peut pas juger en effet de la même façon une règle d'accord en genre et nombre et une règle de préférence entre deux formes pronominales. C'est pourquoi nous avons laissé la description des règles de préférence interphrastiques pour ce cinquième chapitre. Rappelons que ceci ne signifie pas que ces règles s'appliquent après les autres : à la suite de la présentation de ces règles, i.e. du parallélisme de structure et du calcul sur la distance, nous indiquerons à quel moment elles peuvent ou doivent être appliquées dans les deux algorithmes de synthèse. Cette description débouchera sur une discussion sur ces deux algorithmes, où nous donnerons nos arguments en faveur de l'approche globale. Les mérites de cette approche par rapport à l'approche séquentielle seront de plus confirmés dans le prochain chapitre, consacré aux complétives.

5.1 PARALLELISME DE STRUCTURE

Le parallélisme de structure est, nous l'avons annoncé au chapitre 3, une méthode pour la recherche de l'antécédent d'un pronom qui s'appuie sur la comparaison de structures ; ceci rend cette notion plus formalisable que celle de focus (cf. C. Sidner (1981) et B. Grosz (1982), B. Grosz et alii (1982)), qui repose sur des règles beaucoup plus intuitives. L. Danlos (1985, 1987) a mis en avant la notion de parallélisme de structure en comparant deux à deux des phrases plus ou moins complexes, construites autour de verbes identiques ou appartenant à la même table du lexique-grammaire. Dans ce qui suit, nous allons étudier les critères permettant de déterminer quand un couple (P1, P2) de phrases à structure simple (i.e. sans complétive, ni relative, ni subordonnée) sont parallèles. Nous allons utiliser les premiers résultats obtenus pour apporter des conditions de pronominalisation (ou d'effacement) du sujet et des objets de P2.

5.1.1 PRESENTATION DU PROBLEME

Idée de départ :

L'idée de rechercher un parallélisme existant entre deux structures vient de la constatation suivante : il est des cas où l'interprétation d'un pronom PRO est non ambiguë,

alors que PRO a plusieurs antécédents morphologiques, sémantiques et syntaxiques. Cette constatation est illustrée par le discours (1) :

(1) *Ieri, Max_i ha presentato Ugo_k a Bob_m. Oggi, Eva lo_k ha presentato a Lia.*[1].
(Hier, Max_i a présenté Ugo_k à Bob_m. Aujourd'hui, Eva l'_k a présenté à Lia.)

Dans cet exemple, le pronom *lo* de la deuxième phrase n'a qu'une interprétation possible, i.e. *Ugo*. Pourtant ce pronom a trois antécédents morphologiques, sémantiques et syntaxiques ; par ailleurs cet exemple ne relève ni de la notion de focus, ni de la pragmatique. Par contre, on remarque que les phrases qui constituent le texte (1) sont issues de structures isomorphes (i.e. identiques aux feuilles près), qui composent le schéma de texte (T1) :

(T1) *(MAX presentare UGO a BOB. EVA presentare UGO a LIA.)*
(MAX présenter UGO à BOB. EVA présenter UGO à LIA.)

Vu le parallélisme de structure entre ces deux schémas de phrases, la deuxième occurrence de l'étiquette UGO, qui est [dir-objet] dans les deux structures, peut être pronominalisée.

Conséquences en génération :

Ce résultat est important car il permet de dégager deux conclusions en génération, que nous développons dans la section 5.1.2 :

- si la structure de deux schémas de phrases qui se suivent est la même (i.e. elle est représenté par des arbres isomorphes), et si les feuilles qui occupent la même position syntaxique POS dans les deux arbres sont **identiques**, alors la feuille occupant la position POS dans le deuxième schéma de phrase peut être **effacée** (si POS = sujet) ou **pronominalisée** (si POS = [x-objet]).

- si la structure de deux schémas de phrases qui se suivent est la même (i.e. elle est représenté par des arbres isomorphes), et si les feuilles qui occupent la même position syntaxique POS dans les deux arbres sont **différentes**, alors la feuille occupant la position POS dans le deuxième schéma de phrase ne peut être **ni effacée** (si POS = sujet) **ni pronominalisée** (si POS = [x-objet]).

La section 5.1.2 sera donc consacrée à la description des critères de parallélisme qui ont été mis en évidence dans les travaux effectués jusqu'ici ; nous indiquerons ensuite en détail les conséquences que l'on peut tirer du parallélisme pour l'interprétation des pronoms.

Coréférence des sujets :

Un autre résultat présenté dans ce chapitre ne dépend pas directement du parallélisme entre deux structures, mais relève quand même de l'interprétation non ambiguë d'un pronom, sans que la syntaxe, la sémantique ou la pragmatique jouent un rôle : il s'agit de la coréférence des sujets de deux phrases, qui est dans certains cas obligatoire. Nous verrons dans la section 5.1.3 que cette propriété est liée à la forme qui synthétise le sujet dans les deux phrases. Ainsi, examinons le texte (2) :

1 Les indices indiquent les éléments coréférents.

(2) *Paul$_i$ est entré dans la pièce. Il$_i$ a vu Max$_k$. Il$_i$ lui$_k$ a serré la main...*

Les structures des deux dernières phrases ne sont pas parallèles, mais la coréférence de leurs sujets est néanmoins très probable, car le sujet de la deuxième phrase est pronominalisé. Nous verrons également dans la section 5.1.3 les conséquences de la coréférence des sujets sur la synthèse des autres éléments de la phrase. Dans la dernière phrase de (2), par exemple, elle implique la coréférence du pronom *lui* avec le nom propre *Max* de la phrase précédente.

5.1.2 PARALLELISME

5.1.2.1 CRITERES

Les discours que nous avons étudiés en français et en italien et que nous présentons ci-dessous sont constitués de deux phrases indépendantes, à arguments nominaux (i.e. sans complétives), juxtaposées ou coordonnées par *e* (et) ou *ma* (mais). La deuxième de ces phrases contient une ou plusieurs formes pronominales[2], et nous avons recherché l'antécédent le plus probable de cette (ces) forme(s) dans la première phrase.

Un cas trivial de parallélisme de structure a été fourni par l'exemple (1) :

(1) *Ieri, Max$_i$ ha presentato Ugo$_k$ a Bob$_m$. Oggi, Eva lo$_k$ ha presentato a Lia.*
(Hier, Max$_i$ a présenté Ugo$_k$ à Bob$_m$. Aujourd'hui, Eva l'$_k$ a présenté à Lia.)

constitué de deux phrases ayant le même verbe (utilisé dans le même emploi et appartenant par conséquent à la même table du lexique-grammaire), le même temps et la même voix. On remarque que quel que soit l'argument pronominalisé (ou effacé, s'il s'agit du sujet) dans la deuxième phrase, son interprétation préférentielle est l'étiquette qui, dans la première phrase, occupe la même position syntaxique que cette forme pronominale :

[sujet] ***Max$_i$*** *ha presentato Ugo$_k$ a Bob$_m$. Poi* ***Ø$_i$*** *ha presentato Eva a Lia.*[3]
(**Max$_i$** a présenté Ugo$_k$ à Bob$_m$. Puis **[il]$_i$** a présenté Eva à Lia.)

[dir-objet] *Max$_i$ ha presentato* ***Ugo$_k$*** *a Bob$_m$. Poi Eva* ***lo$_k$*** *ha presentato a Lia.*
(Max$_i$ a présenté **Ugo$_k$** à Bob$_m$. Puis Eva **l'$_k$** a présenté à Lia.)

[a-objet] *Max$_i$ ha presentato Ugo$_k$ a* ***Bob$_m$****. Poi Eva* ***gli$_m$*** *ha presentato Lia.*
(Max$_i$ a présenté Ugo$_k$ à **Bob$_m$**. Puis Eva **lui$_m$** a présenté Lia.)

De même, si deux arguments de la deuxième phrase sont pronominalisés, alors ces deux pronoms (ou ce pronom et le sujet effacé) sont non ambigus :

[sujet + dir-objet] ***Max$_i$*** *ha presentato* ***Ugo$_k$*** *a Bob$_m$. Poi* ***Ø$_i$ lo$_k$*** *ha presentato a Lia.*
(**Max$_i$** a présenté **Ugo$_k$** à Bob$_m$. Puis **[il]$_i$ l'$_k$** a présenté à Lia.)

---

2 Pour simplifier, le terme "forme pronominale" englobera dans ce qui suit à la fois les pronoms et le sujet vide italien.

3 Pour permettre l'indiçage d'un sujet effacé on le représente par le symbole Ø.

[sujet + a-objet] ***Max$_i$ ha presentato Ugo$_k$ a Bob$_m$. Poi Ø$_i$ gli$_m$ ha presentato Lia.***
(**Max$_i$** a présenté **Ugo$_k$** à **Bob$_m$**. Puis **[il]$_i$ lui$_m$** a présenté Lia.)

[dir-objet + a-objet] *Ieri, Max$_i$ ha presentato **Ugo$_k$ a Bob$_m$**. Oggi, Eva **glie$_m$lo$_k$** ha presentato di nuovo.*
(Hier, Max$_i$ a présenté **Ugo$_k$** à **Bob$_m$**. Aujourd'hui, Eva **le$_k$ lui$_m$** a présenté de nouveau.)

En conclusion, deux phrases issues d'arbres isomorphes sont totalement parallèles ; en cas d'égalité des feuilles des deux arbres, tous les arguments dans la deuxième phrase peuvent être pronominalisés (effacé) :

[sujet + dir-objet + a-objet] *Ieri, **Max$_i$** ha presentato **Ugo$_k$ a Bob$_m$**. **Ø$_i$ glie$_m$lo$_k$** ha presentato dopo cena.*
(Hier, **Max$_i$** a présenté **Ugo$_k$** à **Bob$_m$**. **[Il]$_i$ le$_k$ lui$_m$** a présenté après dîner.)

Le lecteur pourra se convaincre que le parallélisme que nous venons de mettre en évidence dans deux phrases ayant pour verbe *presentare* se retrouve dans n'importe quel couple de phrases ayant des verbes à mêmes paramètres (donc même classe ; même temps ; même voix ; même nombre d'objets), ce qui peut être schématisé ainsi :

(S) Si P1 = (SUJET $verbe_1$ [TABLE$_1$, TEMPS$_1$, VOIX$_1$] (N$_{OBJET=i}$))
et
Si P2 = (SUJET $verbe_1$ [TABLE$_1$, TEMPS$_1$, VOIX$_1$] (N$_{OBJET=i}$)) avec $0 \leq i \leq 2$
alors P1 et P2 sont parallèles

Afin d'élargir ces critères, nous allons ci-dessous tester l'interprétation des formes pronominales de la deuxième phrase, après changement de chacun des paramètres :

- le temps du verbe de l'une des deux phrases (TEMPS$_1$)
- sa voix (VOIX$_1$)
- sa valeur ($verbe_1$)
- la table à laquelle il appartient (TABLE$_1$)

dans un discours constitué de deux phrases où les éléments de la première phrase sont synthétisés en groupes nominaux.

Nous verrons dans quelle mesure le parallélisme de structure est conservé, en fonction de la modification de chacun de ces paramètres. Nous étudierons ensuite les conséquences du parallélisme de structure sur l'interprétation des formes pronominales exprimant une étiquette TOK, en fonction de la synthèse de l'occurrence précédente de TOK ; ainsi, dans la section 5.1.2.2, nous verrons le rôle joué par la catégorie lexicale (nom, pronom, élément vide) des arguments de la première phrase dans l'interprétation des pronoms de la seconde, et la différence entre les résultats observés pour les objets, et ceux observés pour les sujets.

Changement du temps :

Si le temps des deux verbes est modifié en même temps et dans le même sens (i.e. les deux verbes sont au présent, ou à l'imparfait, ou au futur ...) alors le parallélisme est conservé. Il en va de même, si la modification n'affecte que l'un des deux verbes, comme l'illustrent les textes suivants :

(3a) [sujet] *Max_i ha sempre aiutato Ugo_k. $Ø_i$ aiuterà anche Bob.*
(Max_i a toujours aidé Ugo_k. $[Il]_i$ aidera aussi Bob.)

(3b) [dir-objet] *Max_i ha aiutato Ugo_k. Eva lo_k aiuterà la prossima volta.*
(Max_i a aidé Ugo_k. Eva l'_k aidera la prochaine fois.)

(3c) [sujet+dir-objet] *Max_i ha sempre aiutato Ugo_k. $Ø_i$ lo_k aiuterà ancora questa volta.*
(Max_i a toujours aidé Ugo_k. $[Il]_i$ l'_k aidera encore cette fois-ci.)

Changement de la voix :

Si les deux verbes sont au passif (tous deux avec ou sans agent), alors le parallélisme est conservé : une forme pronominale de la deuxième phrase est coréférente au groupe nominal de la première phrase qui occupe la même position syntaxique qu'elle. Ce résultat est illustré par les exemples (4a), (4b) et (4c), où les arguments de *amare* sont synthétisés en formes pronominales l'un après l'autre, puis tous en même temps :

(4a) [sujet] *Ugo_i è molto amato da Bob_m. $Ø_i$ è amato anche da Eva.*
(Ugo_i est très aimé par Bob_m. $[Il]_i$ est aimé aussi par Eva.)

(4b) [prep-objet] *Ugo_i è molto amato da Bob_m. Anche Eva è amata da lui_m.*
(Ugo_i est très aimé par Bob_m. Eva aussi est aimée par lui_m.)

(4c) [sujet+prep-objet] *Ugo_i è molto amato da Bob_m. $Ø_i$ è amato da lui_m da anni.*
(Ugo_i est très aimé par Bob_m. $[Il]_i$ est aimé par lui_m depuis des années.)

L'inacceptabilité éventuelle d'un discours peut être due à ce que la pronominalisation d'un complément d'agent est une opération parfois douteuse du point de vue stylistique :

(4d) **Ugo_i è stato presentato a Max_k da Bob_m. $Ø_i$ gli_k è stato presentato da lui_m l'altra sera.*
(*Ugo_i a été présenté à Max_k par Bob_m. $[Il]_i$ lui_k a été présenté par lui_m l'autre soir.)

Considérons maintenant un discours de deux phrases issues de la structure profonde suivante :

(SUJET presentare DIR-OBJET a A-OBJET.)
(SUJET présenter DIR-OBJET à A-OBJET.)

Supposons que l'on conjugue les deux verbes à des voix différentes. Les deux phrases résultantes ont alors comme catégories syntaxiques communes le sujet et le [a-objet], puisque

l'objet direct est transformé dans la phrase devenue passive. On constate que la coréférence des sujets est conservée, contrairement à celle des [a-objets] :

Phrase 1 au passif avec ou sans agent ; Phrase 2 à l'actif :

(4e) [sujet] *Ugo$_i$ è stato presentato a Max$_k$ (Ø + da Bob$_m$). Poi Ø$_i$ ha presentato Eva a Lia.*
(Ugo$_i$ a été présenté à Max$_k$ (Ø + par Bob$_m$). Puis [Il]$_i$ a présenté Eva à Lia.)

(4f) [a-objet] **Ugo$_i$ è stato presentato a Max$_k$ (Ø + da Bob$_m$). Poi Eva gli$_{i+k}$ ha presentato Lia.*
(*Ugo$_i$ a été présenté à Max$_k$ (Ø + par Bob$_m$). Puis Eva lui$_{i+k}$ a présenté Lia.)

Phrase 1 à l'actif ; Phrase 2 au passif avec ou sans agent :

(4g) [sujet] *Ugo$_i$ ha presentato Bob$_m$ a Max$_k$. Poi Ø$_i$ è stato presentato (Ø + da Eva) a Lia.*
(Ugo$_i$ a présenté Bob$_m$ à Max$_k$. Puis [Il]$_i$ a été présenté (Ø + par Eva) à Lia.)

(4h) [a-objet] **Ugo$_i$ ha presentato Bob$_m$ a Max$_k$. Poi Eva gli$_{i+k}$ è stata presentata (Ø + da Lia).*
(*Ugo$_i$ a présenté Bob$_m$ à Max$_k$. Puis Eva lui$_{i+k}$ a été présentée (Ø + par Lia).)

Ces résultats montrent que la notion de parallélisme a un sens différent pour les sujets et pour les objets. Ils nous amènent aux conclusions suivantes :

1) on ne peut pas considérer deux phrases à des voix différentes comme parallèles,

2) si deux phrases ayant la même structure profonde sont conjuguées à des voix différentes, alors l'antécédent du sujet effacé de la deuxième phrase est le sujet de la première phrase : cette conclusion sera reprise dans la section 5.1.3, où seront étudiés les cas de coréférence obligatoire des sujets.

Changement de la valeur du verbe :

Si les deux phrases ont deux verbes différents mais appartenant à la même table, alors les tests effectués dans les différentes tables ont permis de certifier que les résultats ci-dessus restent vrais, ce qui est illustré par l'ensemble de discours suivants, dont les verbes (*amare / adorare*) appartiennent tous les deux à la table 43 de l'italien (12 du français) :

(5a) actif-actif ; mêmes temps

[sujet] *Ugo$_i$ ama Max$_k$. Ø$_i$ Adora Eva.*
(Ugo$_i$ aime Max$_k$. [Il]$_i$ adore Eva.)
[dir-objet] *Ugo$_i$ ama Max$_k$. Eva lo$_k$ adora.*
(Ugo$_i$ aime Max$_k$. Eva l'$_k$ adore.)
[sujet+dir-objet] *Ugo$_i$ ama Max$_k$. Ø$_i$ lo$_k$ adora da anni.*
(Ugo$_i$ aime Max$_k$. [Il]$_i$ l'$_k$ adore depuis des années.)

(5b) actif-actif ; temps différents

[sujet] *Ugo$_i$ ama Max$_k$. Ø$_i$ adorava Eva.*
(Ugo$_i$ aime Max$_k$. [Il]$_i$ adorait Eva.)
[dir-objet] *Ugo$_i$ ama Max$_k$. Eva lo$_k$ adorava.*
(Ugo$_i$ aime Max$_k$. Eva l'$_k$ adorait.)
[sujet+dir-objet] *Ugo$_i$ ama Max$_k$. Anni fa Ø$_i$ lo$_k$ adorava.*
(Ugo$_i$ aime Max$_k$. Autrefois [il]$_i$ l'$_k$ adorait.)

(5c) actif-passif; mêmes temps

[sujet] *Ugo$_i$ ama Max$_k$. Ø$_i$ è adorato da Eva.*
(Ugo$_i$ aime Max$_k$. [Il]$_i$ est adoré par Eva.)

(5d) actif-passif; temps différents

[sujet] *Ugo$_i$ ama Max$_k$. Ø$_i$ era adorato da Eva.*
(Ugo$_i$ aime Max$_k$. [Il]$_i$ était adoré par Eva.)

(5e) passif-actif ; mêmes temps

[sujet] *Ugo$_i$ è amato da Max$_k$. Ø$_i$ adora Eva.*
(Ugo$_i$ est aimé par Max$_k$. [Il]$_i$ adore Eva.)

(5f) actif-passif ; temps différents

[sujet] *Ugo$_i$ è amato da Max$_k$. Ø$_i$ ha sempre adorato Eva.*
(Ugo$_i$ est aimé par Max$_k$. [Il]$_i$ a toujours adoré Eva.)

(5g) passif-passif ; mêmes temps

[sujet] *Ugo$_i$ è amato da Max$_k$.*
(Ugo$_i$ est aimé par Max$_k$.)
Ø$_i$ è adorato da Eva.
([Il]$_i$ est adoré par Eva.)
[prep-objet] *Ugo$_i$ è amato da Max$_k$.*
(Ugo$_i$ est aimé par Max$_k$.)
Anche Eva è adorata da lui$_k$.
(Eva aussi est adorée par lui$_k$.)
[sujet+prep-objet] *Ugo$_i$ è amato da Max$_k$. Ø$_i$ è adorato da lui$_k$ da anni.*
(Ugo$_i$ est aimé par Max$_k$. [Il]$_i$ est adoré par lui$_k$ depuis des années.)

(5h)passif-passif ; temps différents

[sujet] *Ugo$_i$ è amato da Max$_k$.*
(Ugo$_i$ est aimé par Max$_k$.)
Ø$_i$ era adorato da Eva.
([Il]$_i$ était adoré par Eva.)
[prep-objet] *Ugo$_i$ è amato da Max$_k$.*
(Ugo$_i$ est aimé par Max$_k$.)
Anche Eva era adorata da lui$_k$.
(Eva aussi était adorée par lui$_k$.)

[sujet+prep-objet] *Ugo$_i$ è amato da Max$_k$. Ø$_i$ sarà adorato da lui$_k$ per anni.*
(Ugo$_i$ est aimé par Max$_k$. [Il]$_i$ sera adoré par lui$_k$ pendant des années.)

Changement de table :

On ne peut plus parler de parallélisme quand on met en présence deux verbes qui appartiennent à des tables différentes, et qui, par conséquent, se construisent avec des arguments dont le nombre, la nature sémantique et la catégorie syntaxique peuvent être différents. Comparons par exemple les textes (6) et (7), construits autour des verbes *incontrare* (rencontrer) et *dare appuntamento* (donner rendez-vous), dont les structures respectives sont :

(SUJET *incontrare* DIR-OBJET.)
et (SUJET *dare appuntamento a* A-OBJET prep$_{loc}$ LOC-OBJET.)

(6) *Ieri, Ugo$_i$ ha incontrato Max$_k$. Ø$_i$ gli$_k$ ha dato appuntamento da Eva.*
(Hier, Ugo$_i$ a rencontré Max$_k$. [Il]$_i$ lui$_k$ a donné rendez-vous chez Eva.)

(7) **Ieri, Ugo$_i$ ha incontrato Max$_k$. Oggi, Eva gli$_{i+k}$ ha dato appuntamento da Lia.*
(Hier, Ugo$_i$ a rencontré Max$_k$. Aujourd'hui, Eva lui$_{i+k}$ a donné rendez-vous chez Lia.)

Dans (6), l'antécédent du sujet vide de la deuxième phrase est le sujet *Ugo* de la première. De même, les objets *gli* et *Max* sont coréférents : il semble donc que le couple de verbes (*incontrare / dare appuntamento*), qui ont des structures différentes, sont pourtant parallèles. Mais cette impression est contredite par le texte (7), où seul le [a-objet] est pronominalisé dans la deuxième phrase : ce pronom est ambigu. Les textes (6) et (7) illustrent donc un autre cas de coréférence des sujets (comme les exemples (4e) et (4g) actif-passif et passif-actif), entraînant celle des objets, dans des structures non parallèles, et qui seront vus dans la section 5.1.3.

L'association de deux verbes issus de tables différentes risque, de plus, d'introduire une sémantique de discours spécifique, ce qui fausse les conclusions, comme le montre le texte (8) :

(8) *Ugo$_i$ ha licenziato Max$_k$. Ø$_k$ lavorava troppo poco.*
(Ugo$_i$ a licencié Max$_k$. [Il]$_k$ travaillait trop peu.)

La relation causale qui se dégage de ce texte et qui fait que le sujet effacé de *lavorare* est coréférent à *Max* est due au couple de verbes mis en jeu (*licenziare / lavorare*) ; cette sémantique est accentuée par l'opposition des temps des deux verbes, et par l'ambiguïté d'interprétation du connecteur sous-entendu.

Les résultats obtenus pour le parallélisme de structure nous permettent de modifier le schéma (S) proposé au début de la section 5.1.2.1 :

(S') Si P1 = (SUJET *verbe*$_{i_1}$ [TABLE$_{j_1}$, TEMPS$_{k_1}$, VOIX$_{m_1}$] (N$_{OBJET=i}$))
et
Si P2 = (SUJET *verbe*$_{i_2}$ [TABLE$_{j_1}$, TEMPS$_{k_2}$, VOIX$_{m_1}$] (N$_{OBJET=i}$)), avec $0 \leq i \leq 2$
alors P1 et P2 sont parallèles

UNIVERSITE DU MAINE - BIBLIOTHEQUE

En d'autres termes, P1 et P2 ont des structures parallèles si leurs verbes (qui peuvent être différents, et peuvent avoir des temps différents) appartiennent à la même table et sont conjugués à la même voix. Dans ces conditions, nous allons voir, dans la section ci-dessous, quelles sont les conséquences pour les sujets et les objets des phrases issues de P1 et P2. En particulier, la coréférence des objets nécessite des contraintes supplémentaires, que nous décrivons ci-dessous.

5.1.2.2 CONSEQUENCES DU PARALLELISME DE STRUCTURE

Dans cette section, nous supposerons que le parallélisme des exemples que nous présentons a été établi, i.e. que ces exemples proviennent de structures obéissant à (S'). Voyons alors dans quelle mesure le parallélisme de structure implique le parallélisme de surface : étudions le rôle de la valeur lexicale des étiquettes de P1 dans l'interprétation des formes pronominales de la deuxième phrase.

a) Nous avons pris pour hypothèse, dans la section précédente, que toutes les étiquettes de P1 étaient synthétisées en groupes nominaux : le parallélisme de structure se concrétise alors en parallélisme total. En d'autres termes, si l'étiquette TOK de P2 est sujet, et si l'étiquette TOK de P1 est elle aussi sujet, alors l'occurrence de TOK dans la deuxième phrase est synthétisée en élément vide ; si d'autre part l'étiquette TOK de P2 est [x-objet], et si l'étiquette TOK de P1 est elle aussi [x-objet], alors l'occurrence de TOK dans la deuxième phrase est pronominalisée.

b) Supposons maintenant que la première phrase, issue de P1, contienne un pronom personnel, ou un sujet effacé. Nous avons dressé la combinatoire des cas réunissant une forme pronominale dans la première phrase et une autre dans la deuxième phrase. Dans chaque cas, examinons l'interprétation de la forme pronominale de la deuxième phrase :

(9a)
[P1 : sujet effacé[4], P2 : sujet effacé]
$\emptyset_i$ parla di Bob$_k$ con Max$_m$. $\emptyset_i$ discute di Lia con Eva.
(**[Il]**$_i$ parle de Bob$_k$ avec Max$_m$. **[Il]** $_i$ discute de Lia avec Eva.)

(9b)
[P1 : sujet effacé, P2 : x-objet pro.]
*$\emptyset_i$ parla di Bob$_k$ con **Max$_m$**. Lia discute di Eva con **lui$_{i+m}$**.*
(*[Il]$_i$ parle de Bob$_k$ avec **Max$_m$**. Lia discute de Eva avec **lui$_{i+m}$**.)

(9c)
[P1 : x-objet pro., P2 : sujet effacé]
*Ugo$_i$ parla di **lui$_k$** con Max$_m$. $\emptyset_i$ discute di Lia con Eva.*
(**Ugo$_i$** parle de **lui$_k$** avec Max$_m$. **[Il]** $_i$ discute de Lia avec Eva.)

(9d)
[P1 : x-objet pro., P2 : x-objet pro.]
*Ugo$_i$ parla di **lui$_k$** con Max$_m$. Lia discute di **lui$_k$** con Eva.*
(Ugo$_i$ parle de **lui$_k$** avec Max$_m$. Lia discute de **lui$_k$** avec Eva.)

(9e)
[P1 : y-objet pro., P2 : x-objet pro.]
Ugo$_i$ parla di **lui$_k$ con **Max$_m$**. Lia discute di Eva con **lui$_{i+m}$**.*
(Ugo$_i$ parle de **lui$_k$** avec **Max$_m$**. Lia discute de Eva avec **lui$_{i+m}$**.)

4 On suppose pour simplifier que le sujet effacé dans la première phrase ne peut référer qu'à un humain masculin.

Les exemples ci-dessus montrent que

- l'étiquette sujet de P2, quand elle est effacée, est coréférente au sujet de P1, quelle que soit la forme (nominale, pronominale) du sujet et des objets de la première phrase (exemples (9a) et (9c)). On remarque, de plus, une impression plus "forte" de coréférence entre deux sujets effacés (exemple (9a)). Ce détail sera repris dans la section 5.1.3.

- le [x-objet] de P2, quand il est pronominalisé, est coréférent au [x-objet] de P1 quand celui-ci est pronominalisé (exemple (9d)). Cette interprétation n'est par contre plus vérifiée quand la première phrase contient un sujet effacé (exemple (9b)), ou un pronom dont la catégorie syntaxique est différente de [x-objet] (exemple (9e)).

c) Enfin, si la première phrase contient deux formes pronominales, alors les résultats sont les mêmes qu'en b), comme l'indiquent les exemples ci-dessous :

(10a)
[P1 : sujet effacé, x-objet pro., P2 : sujet effacé]
$\emptyset_i$ parla di ***lui***$_k$ *con* Max_m. *$\emptyset_i$ discute di Lia con Eva.*
([**II**]$_i$ parle de **lui**$_k$ avec Max_m. [**Il**] $_i$ discute de Lia avec Eva.)

(10b)
[P1 : sujet effacé, x-objet pro., P2 : x-objet pro.]
$\emptyset_i$ parla di* Bob_k *con* ***lui*$_m$. *Lia discute di Eva con* ***lui***$_{i+m}$.
(*[**Il**]$_i$ parle de Bob_k avec **lui**$_m$. Lia discute de Eva avec **lui**$_{i+m}$.)

(10c)
[P1 : sujet effacé, y-objet pro., P2 : x-objet pro.]
$\emptyset_i$ parla di* ***lui*$_k$ *con* Max_m. *Lia discute di Eva con* ***lui***$_{i+m}$.
(*[**Il**]$_i$ parle de **lui**$_k$ avec Max_m. Lia discute de Eva avec **lui**$_{i+m}$.)

(10d)
[P1 : x-objet pro., y-objet pro., P2 : sujet effacé]
Ugo$_i$ parla di ***lui***$_k$ *con* Max_m. *$\emptyset_i$ discute di Lia con Eva.*
(**Ugo**$_i$ parle de **lui**$_k$ avec Max_m. [**Il**] $_i$ discute de Lia avec Eva.)

(10e)
[P1 : x-objet pro., : y-objet pro.] P2 : x-objet pro.]
Ugo$_i$ ne$_k$ parla con ***lui***$_m$. *Lia discute di Eva con* ***lui***$_{i+m}$.
(Ugo$_i$ **en**$_k$ parle avec **lui**$_m$. Lia discute de Eva avec **lui**$_{i+m}$.)

Des points a), b) et c) on peut conclure que :

- étant données deux structures parallèles P1 et P2 qui contiennent toutes deux l'étiquette TOK à la position syntaxique POS

- si les entités de P1 sont synthétisées en groupes nominaux, alors l'occurrence de TOK de P2 est effacée, si POS = sujet, ou pronominalisée, si POS= [x-objet].

- si l'une ou plusieurs entités de P1 est une forme pronominale, alors

  - si POS = sujet, l'occurrence de TOK dans P2 est effacée ,

    - si la seule entité pronominalisée de P1 est TOK, en position POS, alors l'occurrence de TOK dans P2 est effacée ou pronominalisée en fonction de POS.

- sinon, on ne peut rien dire sur l'interprétation de la forme pronominale qui synthétise l'occurrence de TOK dans P2.

Voyons maintenant en détail les conséquences du parallélisme sur la synthèse des étiquettes de P2. Rappelons que les critères de parallélisme pour deux structures P1 et P2 sont schématisés dans (S') :

(S') Si P1 = (SUJET $verbe_{i_1}$ [$TABLE_{j_1}$, $TEMPS_{k_1}$, $VOIX_{m_1}$] ($N_{OBJET=i}$))
et
Si P2 = (SUJET $verbe_{i_2}$ [$TABLE_{j_1}$, $TEMPS_{k_2}$, $VOIX_{m_1}$] ($N_{OBJET=i}$)), avec $0 \leq i \leq 2$
alors P1 et P2 sont parallèles

1) Conséquences pour les [x-objet] de P2 :

A) Si l'un des critères énumérés dans (S') n'est pas vérifié, on ne peut rien prévoir pour l'interprétation des [x-objet] de P2. Si P1 et P2 sont parallèles, mais que P1 contient une forme pronominale qui n'est pas [x-objet], alors on ne peut rien dire non plus.

B) Si P1 et P2 sont parallèles et que les étiquettes de P1 sont synthétisées en groupes nominaux (sauf éventuellement le [x-objet]), alors on peut tirer des conséquences pour le [x-objet] de P2.

a) si les feuilles [x-objet] de P1 et P2 sont les mêmes, alors la feuille [x-objet] de P2 est pronominalisée : la seule interprétation de ce pronom est le [x-objet] de P1. Le parallélisme s'applique à toutes les catégories syntaxiques d'objets :

[dir-objet] *Ugo_i odia Max_k. Paolo lo_k disprezza.*
(Ugo_i déteste Max_k. Paolo le_k méprise.)

[di-objet] *Ugo_i ride di Max_k con $Paolo_m$ e Piero parla di lui_k con Eva.*
(Ugo_i rit de Max_k avec $Paolo_m$ et Piero parle de lui_k avec Eva.)

[a-objet] *Ugo_i presenta $Paolo_k$ a Max_m. Piero gli_m presenta Eva.*
(Ugo_i présente $Paolo_k$ à Max_m. Piero lui_m présente Eva.)

[da-objet] *Ugo_i libera $Piero_k$ da Max_m e Eva allontana Pia da lui_m.*
(Ugo_i libère $Piero_k$ de Max_m et Eva éloigne Pia de lui_m.)

[prep-objet] *Ugo_i si lamenta di Max_k con Bob_m e Piero si lagna con lui_m di Eva*
(Ugo_i se lamente de Max_k auprès de Bob_m et Piero se plaint auprès de lui_m de Eva.)

b) si les feuilles [x-objet] de P1 et P2 sont différentes, par contre, comme dans le schéma de texte (T11) :

(T11) *(UGO presentare PAOLO a MAX . PIERO presentare MAX a EVA.)*
(UGO présenter PAOLO à MAX . PIERO présenter MAX à EVA.)

alors le [x-objet] de P2 est synthétisé en un groupe nominal :

(11a) *Ugo presenta Paolo a Max. Piero presenta Max a Eva.*
(Ugo présente Paolo à Max. Piero présente Max à Eva.)

En effet, la synthèse d'une ppv est interdite :

(11b) **Ugo$_i$ presenta Paolo a Max. Piero lo$_i$ presenta a Eva.*
(*Ugo$_i$ présente Paolo à Max. Piero le$_i$ présente à Eva.)

1) Conséquences pour le sujet de P2 :

A) Si l'un des critères de (S') n'est pas vérifié, alors on ne peut rien affirmer sur la coréférence des sujets : on ne peut rien conclure sur l'antécédent du sujet effacé de P2, que P1 et P2 aient la même étiquette en position sujet ou pas.

B) Si tous ces critères sont respectés, alors P1 et P2 sont parallèles. On peut en dégager les conséquences suivantes, qui sont légèrement différentes en français et en italien :

- Si l'étiquette sujet de P2 est identique à l'étiquette sujet de P1, ce qui est le cas dans le schéma de texte (T12) :

(T12) *(MAX criticare sempre UGO. MAX odiare UGO.)*
(MAX critiquer toujours UGO. MAX détester UGO.)

le sujet de P2 est effacé en italien, ce qui est illustré par l'exemple (12) :

(12) *Max$_i$ critica sempre Ugo$_k$. Ø$_i$ lo$_k$ odia.*
(Max$_i$ critique toujours Ugo$_k$. [Il]$_i$ le$_k$ déteste.)

et en français, la deuxième occurrence de MAX est pronominalisée :

(12') *Max$_i$ critique toujours Ugo$_k$. Il$_i$ le$_k$ déteste.*

- Si, par contre, l'étiquette sujet de P2 est différente de l'étiquette sujet de P1, ce qui est le cas dans le schéma de texte (T13a) :

(T13a) *(MAX criticare sempre UGO. UGO odiare MAX.)*
(MAX critiquer toujours UGO. UGO détester MAX.)

le sujet de P2 est synthétisé en un pronom personnel, ou en un pronom démonstratif en italien, ce qui est illustré par les exemples (13a) et (13'a) :

(13a) *Max$_i$ critica sempre Ugo$_k$. Egli$_k$ lo$_i$ odia.*
(Max$_i$ critique toujours Ugo$_k$. ([litt : il] + Celui-ci)$_k$ le$_i$ déteste.)

(13'a) *Max$_i$ critica sempre Ugo$_k$. Questi$_k$ lo$_i$ odia.*
(Max$_i$ critique toujours Ugo$_k$. ([litt : il] + Celui-ci)$_k$ le$_i$ déteste.)

ou encore en un groupe nominal, ce qui est illustré par l'exemple (13b), issu du schéma de texte (T13b) :

(T13b) *(MAX presentare UGO a BOB. UGO presentare EVA a LIA.)*
(MAX présenter UGO à BOB. UGO présenter EVA à LIA.)

(13b) *Max_i ha presentato Ugo_k a Bob_m. (*$Ø_i$ + *$egli_{k+m}$ + *$questi_m$ + Ugo) ha presentato Eva a Lia.*
(13'b) (Max_ia présenté Ugo_k à Bob_m. (*[il]$_i$ + *([litt : il] celui-ci)$_{k+m}$ + *celui-ci$_m$ + Ugo) a présenté Eva à Lia.)

et en français, la deuxième occurrence de MAX est synthétisée en démonstratif, comme le montre l'exemple (13"a), issu de (T13a) :

(13"a) *Max_i critique toujours Ugo_k. Celui-ci$_k$ le$_i$ déteste.*

ou en un groupe nominal, comme le montre l'exemple (13'b) ci-dessus, issu de (T13b).

La comparaison des textes (13a) et (13'a) indique que le pronom personnel sujet *egli* peut jouer le rôle d'un démonstratif, et par conséquent, référer à la dernière entité mentionnée. Ceci n'est cependant vrai que si P1 ne contient au plus que deux entités du même genre, l'une sujet et l'autre objet :

**Ieri, ho presentato Max_i a $Jean_k$ e oggi, $egli_{i+k}$ ha presentato Eva a Pia.*
(*Hier, j'ai présenté Max_i à $Jean_k$ et aujourd'hui, il_{i+k} a présenté Eva à Pia.)

L'exemple ci-dessus, où *egli* est ambigu, indique que sa valeur démonstrative n'a pas la puissance du pronom démonstratif *questi* (celui-ci) qui, lui, réfère toujours à la dernière entité masculine mentionnée :

Ieri, ho presentato Max_i a $Jean_k$ e oggi, $questi_k$ ha presentato Eva a Pia.
(Hier, j'ai présenté Max_i à $Jean_k$ et aujourd'hui, celui-ci$_k$ a présenté Eva à Pia.)

Nous pouvons, pour résumer, faire les remarques suivantes, étant données deux phrases P1 et P2 à structures parallèles :

- si l'étiquette sujet de P1 est égale à l'étiquette sujet de P2, alors l'étiquette sujet de P2 est synthétisée en élément vide en italien et en pronom personnel en français ;

- sinon, supposons qu'une occurrence de TOKi apparaisse en position sujet dans P2, et qu'une autre occurrence de TOKi soit mentionnée dans P1. En fonction de la position de TOKi dans la structure de surface de P1, on synthétise en français un pronom démonstratif (exemple (13"a)) ou un groupe nominal (exemple (13'b)). En italien, on regarde également le nombre d'étiquettes, mentionnées dans P1, qui sont de même genre de TOKi : on synthétise soit un pronom personnel (exemple (13a)), soit un pronom démonstratif (exemple (13'a)), soit un groupe nominal (exemple (13b)).

5.1.3 COREFERENCE DES SUJETS

Nous avons vu, dans les exemples (4e) et (4g) de la section 5.1.2.1 d'une part, et (9a), (10a) de la section 5.1.2.2 d'autre part, des cas de coréférence des sujets, dans des structures reconnues comme non parallèles : étant données deux phrases issues de structures

quelconques, la coréférence des sujets semble être parfois l'interprétation la plus naturelle. Dans cette section, nous allons rechercher les critères qui rendent, en génération, la coréférence des sujets obligatoire, quels que soient la structure et la sémantique du discours, ainsi que les considérations pragmatiques mises en jeu : ces critères sont basés uniquement sur les valeurs de la catégorie lexicale des sujets étudiés. En fonction de ces valeurs, nous allons déterminer si la coréférence des sujets est obligatoire, interdite, si une autre interprétation est possible, ou encore s'il faut appliquer d'autres critères pour parvenir à une conclusion, ou enfin si la réponse est impossible à donner.

5.1.3.1 ETUDE DES DIFFERENTS CAS

a) Si, en français, deux phrases qui se suivent ont pour sujet un pronom personnel de même genre et nombre, alors les deux sujets sont obligatoirement coréférents :

$Paul_i$ est entré dans la maison. Il_i a salué Bob_k. Il_{i+*k} a pris une bière dans le frigidaire ...

Cette affirmation est confirmée par le contraste entre les deux couples de phrases suivants :

**Eva_i ne verra plus Lia_k. $Elle_{i+k}$ est partie hier.*
(...Eva_i est triste.) $Elle_i$ ne verra plus Lia_k. $Elle_{i+*k}$ est partie hier ...

Dans le premier discours, le sujet de la première phrase n'est pas pronominalisé et le pronom sujet de la deuxième est ambigu. Par contre, dans le deuxième discours, le sujet de la deuxième phrase est le pronom *elle* (coréférent à *Eva*). L'interprétation du pronom sujet *elle* de la deuxième phrase est *Eva* : le pronom sujet n'est pas ambigu.

Considérons le discours emprunté à Pitrat (1985), où la coréférence entre le pronom sujet *il* et le groupe nominal *l'élève* est dictée par la pragmatique et le temps du verbe de la 2ème phrase :

(14) *Le professeur$_i$ a puni l'élève$_k$. Il$_k$ lui$_i$ avait lancé des boulettes.*

Remarquons tout d'abord qu'en changeant le temps de la dernière phrase, on change l'interprétation des pronoms :

(14') *Le professeur$_i$ a puni l'élève$_k$. Il$_i$ lui$_k$ a lancé des boulettes.*

Maintenant, supposons que le sujet de la première phrase soit un pronom. Ici encore, l'interprétation des pronoms est modifiée :

(15) *Le professeur$_i$ est furieux. Il$_i$ a puni l'élève$_k$. Il$_i$ lui$_k$ avait lancé des boulettes.*

Bien entendu, (15) est bizarre : il faut un univers de discours particulier et supposer que le discours est prolongé. Mais ceci montre que la coréférence des sujets peut être plus forte que des règles de pragmatique.

b) Si, en italien, deux phrases qui se suivent ont pour sujet l'élément vide, alors la coréférence des deux sujets dépend de la morphologie, i.e. de la présence de marque permettant de distinguer le genre des deux éléments vides. Ainsi, dans le texte (16a), où le genre féminin du sujet effacé de la dernière phrase est indiqué par le participe passé *partita*, les sujets ne sont pas coréférents :

(16a) *Max$_i$ è triste. Ø$_i$ non vedrà più Lia$_k$. Ø$_{*i+k}$ è partita ieri ...*
(Max$_i$ est triste. [Il]$_i$ ne verra plus Lia$_k$. [Elle + *il]$_{*i+k}$ est partie hier ...)

Par contre, la coréférence des sujets est obligatoire dans le texte (16b), où le genre du sujet vide de la dernière phrase n'est pas indiqué :

(16b) *Max$_i$ è triste. Ø$_i$ non vedrà più Lia$_k$. Ø$_{i+*k}$ parte domani ...*
(Max$_i$ est triste. [Il]$_i$ ne verra plus Lia$_k$. [*Elle + il]$_{i+*k}$ part demain ...)

c) On ne peut rien affirmer pour la coréférence des sujets, quand le premier est un groupe nominal (ou un nom propre). En effet, dans certains cas elle constitue l'interprétation la plus probable, comme dans les exemples (4e) et (4g) :

(4e) [sujet] *Ugo$_i$ è stato presentato a Max$_k$ (Ø + da Bob$_m$). Poi Ø$_i$ ha presentato Eva a Lia.*
(Ugo$_i$ a été présenté à Max$_k$ (Ø + par Bob$_m$). Puis [Il]$_i$ a présenté Eva à Lia.)

(4g) [sujet] *Ugo$_i$ ha presentato Bob$_m$ a Max$_k$. Poi Ø$_i$ è stato presentato (Ø + da Eva) a Lia.*
(Ugo$_i$ a présenté Bob$_m$ à Max$_k$. Puis [Il]$_i$ a été présenté (Ø + par Eva) à Lia.)

ou dans l'exemple suivant, dans lequel on associe deux verbes définis par des structures différentes :

(17) *Max$_i$ ha incontrato Bob$_k$. Ø$_i$ gli$_k$ ha reso le chiavi.*
(Max$_i$ a rencontré Bob$_k$. [Il]$_i$ lui$_k$ a rendu les clés.)

Cependant cette coréférence peut être remise en cause quand certains paramètres modifient la sémantique du discours. Considérons par exemple le texte (18a), où les verbes sont définis par la structure SUJET verbe DIR-OBJET. Les sujets sont coréférents :

(18a) *Le professeur$_i$ a sermonné l'élève$_k$. Il$_i$ l'$_k$ a giflé.*

Si le verbe *gifler* est conjugué au passif sans agent, on perd la coréférence des sujets, et on obtient le texte ambigu (18b) :

(18b) **Le professeur$_i$ a sermonné l'élève$_k$. Il$_{i+k}$ a été giflé.*

De même, si *gifler* est conjugué à l'actif, mais à un temps antérieur à celui de la première phrase, l'interprétation du pronom sujet est ambiguë, avec une préférence pour *l'élève*. Le changement de temps modifie en effet l'interprétation du connecteur sous-entendu, qui introduit dans le texte résultant (18c) (stylistiquement douteux, et qu'on cherche par conséquent à éviter en génération) une relation causale :

(18c) *?Le professeur$_i$ a sermonné l'élève$_k$. Il$_{*i\ +\ ?k}$ l'avait giflé.*

Des point a), b) et c) nous pouvons conclure les faits suivants :

- Deux pronoms sujets ayant le même genre, ou deux sujets effacés dont le genre n'est pas indiqué par un autre élément de la phrase, sont coréférents dans deux phrases qui se suivent.

- La coréférence entre un groupe nominal sujet et un pronom sujet (ou un sujet effacé, pour l'italien) situé dans la phrase suivante, ne peut pas être établie : en effet, le temps des verbes, le changement de la voix et la valeur des connecteurs peuvent modifier la sémantique du discours et par conséquent l'interprétation de la forme pronominale sujet de la deuxième phrase.

5.1.3.2 CONSEQUENCES

En analyse, le point c) soulève des problèmes, puisqu'il y a antagonisme entre structure et pragmatique. Ainsi, G. Hirst (1981) utilise la pragmatique pour justifier l'interprétation du pronom *elle* dans les deux textes suivantes, identiques au verbe principal près :

(19a) *Quand Eva$_i$ va dîner chez Lia$_k$, elle$_i$ mange du cassoulet.*

(19b) *Quand Eva$_i$ va dîner chez Lia$_k$, elle$_k$ sert du cassoulet.*

En génération, il est préférable d'éviter la synthèse de (19b) :

(19'b) *Quand Eva$_i$ va dîner chez Lia$_k$, **celle-ci$_k$** sert du cassoulet.*

Plus généralement, on évitera dans un premier temps la synthèse de formes pronominales sujet non coréférentes au groupe nominal sujet de la phrase précédente. Si cela est possible, le composant stratégique génère un schéma de texte permettant d'éviter une telle situation. On aura alors, à la place de (19'b), le texte (19c) où les phrases ont le même sujet :

(19c) *Quand Lia$_i$ reçoit Eva$_k$ à dîner, elle$_i$ sert du cassoulet.*

Quant aux conséquences, en génération, des points a) et b), elles sont similaires aux conséquences du parallélisme pour les sujets qui sont présentées dans la section 5.1.2.2 :

- si P1 et P2 ont pour sujet l'étiquette TOK, et que TOK est synthétisée en une forme pronominale dans P1 (i.e. effacée en italien), alors l'occurrence de TOK dans P2 peut être pronominalisée en français, ou effacée en italien.

- si P1 a pour sujet l'étiquette TOK, synthétisée en un pronom (ou effacée), si P2 a pour sujet l'étiquette TOK', et si TOK et TOK' sont de même genre et nombre, alors la pronominalisation en français, et l'effacement en italien, de TOK' est interdite. On envisage, en français, la synthèse d'un groupe nominal ou d'un pronom démonstratif, et en italien, d'un pronom personnel sujet, ou d'un démonstratif ou d'un groupe nominal (cf. section 5.1.2.2).

- si P1 a pour sujet l'étiquette TOK, synthétisée en un pronom (ou effacée), si P2 a pour sujet l'étiquette TOK', et si TOK et TOK' sont de genre différent, alors on ne peut rien prévoir sur la synthèse de TOK'.

Enfin, la coréférence obligatoire des sujets a une conséquence directe sur le calcul des antécédents syntaxiques des objets. En effet, considérons les textes (20) et (21) respectivement en italien et en français, où la coréférence des sujets est établie, pour l'italien, par la présence de sujets vides dont le genre n'est pas indiqué, et pour le français, par celle de pronoms sujets de même genre :

(20) *...$\emptyset_i$ ha visto Ugo$_k$. $\emptyset_i$ lo$_k$ ha salutato ...*
(... [Il + elle]$_i$ a vu Ugo$_k$. [Il + elle]$_i$ l'$_k$ a salué ...)

(21) *Il$_i$ a vu Ugo$_k$. Il$_i$ l'$_k$ a salué.*

Les critères de coréférence des sujets permettant d'établir que le sujet effacé, dans la deuxième phrase de l'exemple (20), est coréférent au sujet effacé de la première phrase, nous pouvons appliquer alors la contrainte de non coréférence entre le sujet et un objet pour affirmer que le pronom objet direct *lo* en italien n'a qu'un antécédent possible : *Ugo*. De même, le pronom *il*, dans la deuxième phrase de l'exemple (21) est obligatoirement coréférent au sujet de la première phrase ; par conséquent la contrainte de non coréférence entre sujet et objet établit que le pronom [dir-objet] *l'* est obligatoirement coréférent à *Ugo*. La relation entre coréférence obligatoire des sujets et contrainte de non coréférence implique que les critères que nous venons de voir doivent être appliqués dans l'algorithme avant le calcul des antécédents syntaxiques du pronom. Cette remarque sera développée dans la section 5.3.

5.1.4 ORIENTATION DES RECHERCHES

Nous venons de proposer un certain nombre de critères permettant de donner une préférence pour l'interprétation des pronoms. Ces critères sont calculés en fonction de la structure des phrases mises en jeu et des valeurs lexicales des étiquettes de ces structures. Ils concernent un couple de phrases, coordonnées ou juxtaposées. Il s'agit d'un côté du parallélisme, et, de l'autre, des conditions de coréférence obligatoire (ou interdite) des sujets. Les recherches que nous menons actuellement concernent l'élargissement des critères de parallélisme. Ce travail suit deux directions : l'élargissement des recherches, dans le cas d'un couple (P1, P2) de structures juxtaposées ou coordonnées, à des verbes n'appartenant pas à la même table, et l'étude de structures ni juxtaposées, ni coordonnées.

<u>- Etude des couples P1, P2 dont les verbes appartiennent à des tables différentes :</u>

Le problème auquel on se heurte est la combinatoire quasi incontrôlable des couples de phrases à tester. Ainsi, on peut décider de se restreindre à la recherche du parallélisme pour les couples de verbes ayant le même nombre d'arguments. Prenons le cas des verbes transitifs, se construisant avec un objet : on s'aperçoit qu'il existe une opposition entre un verbe appartenant à la table 4 du français (41 de l'italien), et un verbe transitif d'une autre table. En effet, les verbes de la table 4 (41), qui sont définis par la structure Que P V N1, sont sémantiquement homogènes ; ils sont caractérisés par des sujet qui "déclenchent" un sentiment "éprouvé" par l'objet (cf. M. Gross (1975)) :

[table 41] *(Ugo + Il modo di fare di Ugo) diverte Max.*
[table 4] ((Ugo + Le comportement de Ugo) amuse Max.)

cette sémantique s'oppose à celle des verbes des autres tables, où le sujet exerce généralement une action consciente sur l'(les) objet(s).

Si on juxtapose ou si on coordonne une phrase à verbe de la table 4 (41) avec une phrase dont le verbe appartient à une autre table, quel que soit l'ordre d'apparition de ces deux phrases, l'antécédent d'un pronom dans la deuxième phrase est ambigu. De plus, le résultat est toujours stylistiquement désagréable, même en l'absence de problèmes de coréférence : une telle juxtaposition / coordination peut être en effet comparée à la réunion de deux phrases dont l'une serait à l'actif et l'autre au passif, puisque l'agent d'un verbe (à l'actif) de la table 4 (41) est exprimé non pas par le sujet, mais par l'objet. L'ambiguïté du sujet effacé (du pronom sujet) de P2 peut être illustrée par le texte (22) :

(22) **Paolo$_i$ odia Ugo$_k$ ma Ø$_{i+k}$ diverte Max.*
(*Paolo$_i$ déteste Ugo$_k$ mais [il]$_{i+k}$ amuse Max.)

Des résultats analogues sont observés quand on cherche à juxtaposer ou coordonner un verbe de la table 5 du français (42 de l'italien), qui est définie par la structure Que P V Prép N1, et qui a des propriétés sémantiquement et syntaxiquement proches à celles d'un verbe de la table 4 (41) :

[table 42] *(Ugo + Il modo di fare di Ugo) piace a Max.*
[table 5] ((Ugo + Le comportement de Ugo) plaît à Max.)

et un verbe intransitif à un complément, introduit par la même préposition, et appartenant à une autre table :

(23) **Paolo$_i$ piace a Ugo$_k$ ma Ø$_{i+k}$ pensa a Max.*
(*Paolo$_i$ plaît à Ugo$_k$ mais [il]$_{i+k}$ pense à Max.)

Par contre, la juxtaposition ou la coordination de phrases construites toutes les deux autour de verbes des tables 4 (41) et 5 (42) ne pose pas de problèmes stylistiques. Un tel couple de phrases est parallèle. Le sujet effacé de P2 est alors interprété comme coréférent au sujet de P1 :

(24) *Ugo$_i$ diverte Max$_k$. Ø$_i$ piace a Piero.*
(Ugo$_i$ amuse Max$_k$. [Il]$_i$ plaît à Piero.)

(25) *Ugo$_i$ piace a Max$_k$ e Ø$_i$ diverte Piero.*
(Ugo$_i$ plaît à Max$_k$ et [Il]$_i$ amuse Piero.)

De même, les objets sont préférentiellement coréférents :

(24') *Ugo$_i$ diverte Max$_k$ e Bob gli$_k$ piace.*
(Ugo$_i$ amuse Max$_k$ et Bob lui$_k$ plaît.)

(25') *Ugo$_i$ piace a Max$_k$ e Bob lo$_k$ diverte.*
(Ugo$_i$ plaît à Max$_k$ et Bob l'$_k$ amuse.)

Le parallélisme est total :

(24') *Ugo$_i$ diverte Max$_k$. Ø$_i$ gli$_k$ piace molto.*
(Ugo$_i$ amuse Max$_k$. [Il]$_i$ lui$_k$ plaît beaucoup.)

(25') *Ugo$_i$ piace a Max$_k$. Ø$_i$ lo$_k$ diverte.*
(Ugo$_i$ plaît à Max$_k$. [Il]$_i$ l'$_k$ amuse.)

- Parallélisme dans les structures complexes :

La comparaison d'arbres représentant des phrases à complétives, à relatives ou à compléments du nom, par exemple, peut aboutir à des conditions de coréférence pour les sujets ou les objets de ces structures enchâssées ; ainsi, dans l'exemple (26), proposé par K. McKeown (1983) pour ses règles sur le focus, et repris par L. Danlos (1985) pour illustrer le parallélisme :

(26) *Pensiamo che Max$_i$ prenderà il treno per Roma con Ugo$_k$, e speriamo che Ø$_i$ prenderà l'autobus per Ostia.*
([Nous] pensons que Max$_i$ prendra le train pour Rome avec Ugo$_k$, et [nous] espérons qu'[il]$_i$ prendra l'autobus pour Ostia.)

le sujet vide de la complétive enchâssée dans la deuxième coordonnée est interprété comme coréférent à *Max* et non à *Ugo*. Les questions que l'on peut se poser à propos du parallélisme dans les complétives concernent les principales, et en particulier les propriétés des verbes qui contrôlent ces complétives, comme l'illustrent les exemples (26') et (26") :

(26') **Piero$_i$ pensa che Max$_k$ prenderà il treno per Roma con Ugo$_m$, e Ø$_i$ spera che Ø$_{i+k}$ prenderà l'autobus per Ostia.*
(*Piero$_i$ pense que Max$_k$ prendra le train pour Rome avec Ugo$_m$, et [il]$_i$ espère qu'[il]$_{i+k}$ prendra l'autobus pour Ostia.)

Dans (26'), l'antécédent du sujet effacé dans la deuxième complétive est ambigu : ce sujet représente soit l'étiquette MAX, soit, plus probablement, PIERO. Cette étiquette est en effet synthétisée en sujet effacé du verbe *sperare*. En effet, d'une part, les deux principales sont parallèles. Elles sont construites autour des verbes *pensare* (penser) et *sperare* (espérer), tous deux de la table 43 de l'italien, définie par N0 V che F, et dont les traductions appartiennent à la table 6 du français, définie par N0 V que P. Par conséquent, le sujet effacé, dans la deuxième principale, est coréférent à *Piero*, sujet de la première principale. D'autre part, les propriétés du verbe *sperare* permettent à son sujet d'être coréférent au sujet effacé de la complétive que *sperare* contrôle. Ceci n'est pas le cas, par contre, de verbes comme *volere* (vouloir) :

Ø$_i$ vuole che Ø$_{k,\ k\neq i}$ vada a Parigi.
([Il + elle]$_i$ veut qu'[il + elle]$_{k,\ k\neq i}$ aille à Paris.)

Par conséquent, les deux sujets effacés, dans (26'), peuvent représenter la même entité, i.e. *Piero*, c'est pourquoi le sujet effacé de la deuxième complétive est ambigu.

(26") *Piero$_i$ desidera che Max$_k$ prenda il treno per Roma con Ugo$_m$, e Ø$_i$ vuole che Ø$_k$ prenda l'autobus per Ostia.*
(Piero$_i$ désire que Max$_k$ prenne le train pour Rome avec Ugo$_m$, et [il]$_i$ veut qu'[il]$_k$ prenne l'autobus pour Ostia.)

A l'inverse de l'exemple (26'), la coréférence est interdite entre le sujet de *volere* et le sujet de la complétive objet de ce verbe dans (26")[5]. Par conséquent on peut parler ici de parallélisme dans

5 La coréférence entre le sujet effacé de *volere* et *Piero* est assurée par l'appartenance des deux verbes à la même table (6 du français, 43 de l'italien) : les deux principales sont, encore ici, parallèles.

les complétives : le sujet vide de *prendere l'autobus* est coréférent à *Max*, sujet de *prendere il treno* dans la complétive objet de *desiderare*.

Le principe de la coréférence dans des structures complexes amène à une autre considération interphrastique de la pronominalisation : la notion de distance. Les exemples (26) et (26") montrent que l'antécédent préférentiel d'une forme pronominale peut être une entité assez éloignée quand les deux éléments se retrouvent dans des structures parallèles. D'autres exemples montrent au contraire que l'antécédent le plus probable d'une forme pronominale est l'entité la plus proche ; la section ci-dessous est consacrée à la présentation des deux points de vue.

5.2 DISTANCE

La notion de distance entre un pronom et une entité a été abordé par le biais du parallélisme : dans deux structures parallèles, l'antécédent le plus probable d'une forme pronominale n'est pas forcément la plus proche. La distance devient un paramètre intéressant pour la synthèse des pronoms si elle peut être décomposée en deux questions :

question 1 : étant données deux occurrences d'une même étiquette, la distance entre les deux occurrences permet-elle d'envisager de synthétiser la deuxième en une forme pronominale ? En d'autres termes, la distance intervient-elle dans la décision d'engager le mécanisme de synthèse d'un pronom ?

question 2 : étant donnée une forme pronominale dont on envisage la synthèse et dont on a calculé les antécédents possibles, la distance séparant ce pronom et ses antécédents permet-elle de réduire le nombre de ceux-ci (i.e. de réduire les ambiguïtés) ?

Dans ce qui suit, nous présentons deux aspects de la distance entre deux termes : la distance linéaire, tout d'abord, est la première qui vient à l'esprit : à combien de mots, ou de phrases, remonte-t-on dans le contexte gauche pour la recherche des antécédents ? Nous allons voir que formaliser cette notion correspond à une décision arbitraire, parce que la distance ne peut pas être formulée uniquement de façon linéaire. L'autre aspect de la distance qui sera abordé est celui qui résulte des applications de critère de parallélisme : des exemples illustreront le fait que si deux phrases sont parallèles, alors l'antécédent le plus probable d'une forme pronominale n'est pas le plus proche.

5.2.1 DISTANCE LINEAIRE

En général, lorsque l'on parle de distance, on pense à la notion de proximité linéaire entre deux mots. Dans ces conditions, la distance répond aux questions 1 et 2. Examinons en effet le texte (26) que nous avions introduit dans la conclusion du chapitre 4 :

(26) **?Max$_i$ si è rotto una gamba. Eva e Lia si sono occupate della casa. Hanno lavato i piatti. Eva ha ridipinto le pareti. Lia ha stirato le lenzuola. Ø$_i$ è stato trattato come un principe ...*
(*?Max$_i$ s'est cassé la jambe. Eva et Lia se sont occupées de la maison. [Elles] ont fait la vaisselle. Eva a repeint les murs. Lia a repassé les draps. [Il]$_i$ a été traité comme un prince ...)

les deux occurrences de *Max* sont trop éloignées l'une de l'autre pour que la deuxième puisse être effacée ou pronominalisée. On synthétise alors la deuxième occurrence de MAX en un groupe nominal :

(26') *Max_i si è rotto una gamba. Eva e Lia si sono occupate della casa. Hanno lavato i piatti. Eva ha ridipinto le pareti. Lia ha stirato le lenzuola. Il $ragazzo_i$ è stato trattato come un principe ...*
(**Max_i** s'est cassé la jambe. Eva et Lia se sont occupées de la maison. [Elles] ont fait la vaisselle. Eva a repeint les murs. Lia a repassé les draps. **Le jeune $homme_i$** a été traité comme un prince ...)

Dans l'exemple (27) :

(27) *Max_i ha passato l'inverno in campagna. Difatti, la città è inquinata. Il virus dell'influenza ha colpito vecchi e bambini. Perfino Ugo_k si è ammalato. Eva gli_k ha portato dei regali ...*
(Max_i a passé l'hiver à la campagne. En effet la ville est polluée. Le virus de la grippe a touché les vieux et les enfants. Même **Ugo_k** est tombé malade. Eva **lui_k** a porté des cadeaux ...)

on envisage de synthétiser la deuxième occurrence de *Ugo* (dans la dernière phrase) en un pronom personnel [dir-objet]. Cette synthèse est envisageable puisque *Ugo* est mentionné dans la phrase précédente (question 1). La liste des antécédents morphologiques, sémantiques et syntaxiques du pronom *gli* envisagé contient l'étiquette représentant *Ugo* et celle représentant *Max*. *Max*, qui n'apparaît que dans la première phrase de (27), est cependant un antécédent très peu probable, étant donné son éloignement par rapport au pronom (question 2). La présence d'un seul antécédent probable (i.e. *Ugo*) rend par conséquent la synthèse du pronom possible.

Il semble donc que la notion de distance se calcule seulement en nombre de mots ou de phrases ; le problème consisterait donc uniquement à déterminer à partir de combien de mots, par exemple, la distance entre deux termes est jugée assez grande pour rendre impossible la coréférence entre ces deux termes. Posé ainsi, le problème n'est déjà pas simple à résoudre, à moins de définir une distance arbitraire : par exemple, on peut poser arbitrairement qu'en italien, la distance entre deux occurrences d'une même entité ne doit pas dépasser 15 mots. Cette règle est vérifiée pour l'exemple suivant, où la deuxième occurrence de l'étiquette représentant l'entité *bomba* n'est ni effaçable, ni pronominalisable :

(28) *Ronald Reagan è uscito illeso da un attentato, ieri a New Haven. Dei terroristi hanno fatto esplodere **una bomba a orologeria$_i$** mentre l'ex-presidente degli Stati Uniti partecipava a un comizio politico, a Yale. (*$Ø_i$ + *$essa_i$ + la bomba) conteneva quattro chili di dinamite.*
(*Ronald Reagan est sorti indemne d'un attentat, hier à New Haven. Des terroristes ont fait exploser **une bombe à retardement$_i$** pendant que l'ex-président des Etats Unis participait à un meeting, à Yale. (**[Elle]$_i$** + **elle$_i$** + **la bombe**) contenait quatre kilos de dynamite.)

Mais cette règle est contredite par l'exemple suivant :

(29) *Eva_i è uscita a fare una passegiata nel bosco. E' primavera, il sole riscalda i prati. Il vento **le_i** accarezza il viso ...*

(**Eva**$_i$ est sortie faire une promenade dans le bois. C'est le printemps, le soleil réchauffe les prés. Le vent **lui**$_i$ caresse le visage ...)

où la synthèse de la deuxième occurrence de *Eva* en pronom personnel *le* est possible, bien que la distance entre le pronom et la première occurrence de *Eva* soit supérieure à 15 mots et bien qu'il n'y ait pas de parallélisme de structure en jeu.

La différence entre les exemples (28) et (29) réside dans la structure des phrases mises en jeu : le dernier texte est constitué de phrases simples juxtaposées[6], alors que le premier contient un certain nombre de modifieurs, parmi lesquels une subordonnée. L'étude de la distance entre deux termes nécessite donc celle de la forme des arbres dont sont issues les phrases contenant ces termes. L'intervention de la géométrie des arbres dans le calcul de la distance devient manifeste quand elle met en évidence des structures parallèles : dans ce cas, la distance linéaire ne joue plus aucun rôle.

5.2.2 DISTANCE ET PARALLELISME

Dans cette dernière section consacrée à la distance entre une forme pronominale et son (ses) antécédent(s), nous présentons plusieurs exemples où l'antécédent le plus probable d'une forme pronominale n'est pas le plus proche : c'est le cas quand ces deux termes appartiennent à deux structures P1 et P2 parallèles.

Si P1 et P2 ont des structures parallèles, alors l'antécédent le plus probable d'un pronom situé dans P2 est en effet l'entité occupant la même position syntaxique que lui dans P1. Ainsi, l'exemple (30), que nous avons présenté dans la section 5.1.5 à propos du parallélisme dans les complétives

(30) *Pensiamo che Max$_i$ prenderà il treno per Roma con Ugo$_k$, e speriamo che Ø$_i$ prenderà l'autobus per Ostia.*
([Nous] pensons que Max$_i$ prendra le train pour Rome avec Ugo$_k$, et [nous] espérons qu'[il]$_i$ prendra l'autobus pour Ostia.)

répond à la question 1 (la distance entre les deux occurrences de MAX permet-elle d'envisager l'effacement du pronom ?) : les deux complétives sont parallèles, donc la synthèse de la deuxième occurrence de *Max* en un élément vide est envisagée. L'exemple (30) répond également à la question 2 (la distance entre le sujet effacé et ses antécédents permet-elle de supprimer certains d'entre eux ?) : dans deux structures parallèles, le sujet vide de la deuxième phrase est obligatoirement coréférent au sujet de la première. Donc le seul antécédent possible du sujet vide de la deuxième complétive est le sujet de la première complétive.

Nous retrouvons ces mêmes résultats dans d'autres types de structures complexes : ainsi, dans l'exemple (31)

(31) *Max, che è andato con Pia$_i$ a Los Angeles, ha mandato a Lea$_k$ una cartolina. Ugo, che è andato con lei$_i$ a San Francisco, ci ha mandato una lettera.*
(Max, qui est allé avec Pia$_i$ à Los Angeles, a envoyé à Lea$_k$ une carte postale. Ugo, qui est allé avec elle$_i$ à San Francisco, nous a envoyé une lettre.)

6 L'infinitive : *fare una passeggiata* est argument du verbe *uscire* : les termes "phrase simple" indiquent ici l'absence de modifieurs.

le pronom [prep-objet] *lei*, dans la relative de la deuxième phrase a pour antécédent préférentiel *Pia*, malgré leur éloignement, et bien que *Pia* ne soit pas la dernière entité compatible mentionnée. De même, dans l'exemple (32) :

(32) *La sorella$_i$ di Ugo$_k$ ha invitato Max$_m$ a una festa e suo$_k$ fratello ha invitato Eva al cinema.*
(La soeur$_i$ de Ugo$_k$ a invité Max$_m$ à une fête et son$_k$ frère a invité Eva au cinéma.)

l'antécédent du possessif *suo* (son) est *Ugo*.

Pour conclure sur la description des règles de préférence interphrastiques que sont le parallélisme et la distance, nous pouvons dire ceci :

- le parallélisme est un calcul sur la géométrie des arbres, qui dépend aussi de la valeur des feuilles de ces arbres (i.e. de la synthèse des éléments des phrases issues de ces arbres) ;

- la distance entre une forme pronominale et son antécédent le plus probable est le nombre de mots ou de phrases séparant ces termes. Cette définition reste encore assez arbitraire et doit être combinée à une étude de la structure des phrases mises en jeu. Cette distance linéaire n'est plus à prendre en compte quand les structures à comparer sont parallèles.

- en regard des règles de parallélisme et de distance qui ont pu être établies jusqu'à maintenant, et des deux algorithmes de synthèse d'une phrase qui ont été proposés au chapitre 4, nous allons déterminer maintenant comment intégrer ces règles dans les opérations qui ont été décrites dans les deux chapitres précédents, et nous allons dégager de ces règles les arguments en faveur de l'un des deux algorithmes de synthèse d'une phrase : l'approche globale.

5.3 INTEGRATION DANS L'ALGORITHME DES CONDITIONS INTERPHRASTIQUES.

Les règles que nous venons de décrire ont pour rôle d'effectuer un choix entre des formes lexicales, qui sont calculées par les règles intra et interphrastiques de morphologie, sémantique et syntaxe. On aurait pu croire que ces règles, dont le rôle est d'apporter une préférence pour une forme (pro)nominale, et qui sont fondées sur des critères moins fiables que les opérations morphologiques, sémantiques et syntaxiques, doivent être appliquées après ces opérations ; or, nous allons voir que le parallélisme (i.e. le contexte) peut prendre le pas sur la morpho-syntaxe. Nous allons voir également que l'établissement de l'ordre d'application des différentes règles de pronominalisation examinées jusqu'ici permet aussi parfois de minimiser les opérations de synthèse d'un pronom. C'est pourquoi cette section est consacrée à la description de l'intégration des règles de préférence interphrastiques dans les deux algorithmes, en fonction de considérations théoriques linguistiques d'une part, et informatiques (i.e. liées aux économies de calculs) d'autre part.

5.3.1 QUAND APPLIQUER LES REGLES DE PREFERENCE INTERPHRASTIQUES ?

Application des règles de parallélisme :

Il serait logique de penser que les règles de parallélisme (qui sont des règles de préférence) s'appliquent sur un ensemble de formes (pro)nominales sélectionnées par les règles de calcul morphologiques, sémantiques et syntaxiques. Cependant, le contexte peut se révéler plus important que la morpho-syntaxe, ce qui remet cet ordre en question. Examinons tout d'abord le schéma de texte (T33) :

(T33) *(UGO presentare ieri MAX a EVA. BOB presentare oggi EVA a PIERO.)*
(UGO présenter hier MAX à EVA. BOB présenter aujourd'hui EVA à PIERO.)

Supposons que l'on applique d'abord les règles morphologiques, sémantiques et syntaxiques de pronominalisation. Celles-ci indiquent que la deuxième occurrence de *Eva* peut être synthétisée en *l'*, ce qui n'engendre pas d'ambiguïté. Cependant, le texte ainsi obtenu :

(33) **Ieri, Ugo$_i$ ha presentato Max$_k$ a Eva$_m$. Oggi, Bob l'$_m$ ha presentata a Piero.*
(*Hier, Ugo$_i$ a présenté Max$_k$ à Eva$_m$. Aujourd'hui, Bob l'$_m$ a présentée à Piero.)

est inacceptable bien que non ambigu. L'inacceptabilité de (33) s'explique par le fait que les schémas de phrases de (T33) sont parallèles, et les feuilles en position [dir-objet] sont différentes : la pronominalisation de EVA dans la deuxième phrase est donc interdite. Dans le même ordre d'idées, considérons le schéma de texte (T34) :

(T34) *(UGO parlare ieri a EVA di SARDEGNA. EVA parlare oggi a UGO di SICILIA.)*
(UGO parler hier à EVA de SARDEGNA. EVA parler aujourd'hui à UGO de SICILIA.)

En appliquant les règles morphologiques, sémantiques et syntaxiques, on aboutit à l'effacement du sujet de la deuxième phrase, grâce à la contrainte de non coréférence entre le sujet et le [a-objet] qui est synthétisé en un pronom personnel masculin singulier (*gli*). Le texte résultant (34)

(34) **Ieri, Ugo$_i$ ha parlato a Eva$_k$ della Sardegna. Oggi, Ø$_k$ gli$_i$ ha parlato della Sicilia.*
(*Hier, Ugo$_i$ a parlé à Eva$_k$ de la Sardaigne. Aujourd'hui, [elle]$_k$ lui$_{masc,i}$ a parlé de la Sicile.)

est cependant inacceptable : les deux phrases sont parallèles et leurs sujets sont différents, l'effacement du deuxième sujet est donc interdit. Les exemples (33) et (34) montrent que les critères de parallélisme peuvent être plus puissants que ceux de la morpho-syntaxe. Dans un autre ordre d'idées, examinons maintenant avec le schéma de texte (T35), dans lequel on supposera que la première occurrence de MAX est synthétisée en un pronom personnel, un nouvel exemple en faveur de l'application des règles de parallélisme avant les règles syntaxiques de contrainte de non coréférence :

(T35) *(... MAX rencontrer UGO. MAX serrer la main à UGO.)*

Si la deuxième occurrence de MAX est pronominalisée, la coréférence des deux pronoms sujet est obligatoire (cf. section 5.1.2.2). Supposons que cette opération (qui est, comme le parallélisme, une règle de préférence interphrastique) soit effectuée après les règles morphologiques, sémantiques et syntaxiques : quel que soit l'algorithme choisi, on ne peut appliquer, dans un premier temps, ni la contrainte de non coréférence entre le sujet et l'objet, ni la contrainte de non coréférence entre l'objet et le sujet. En effet, les étiquettes MAX et UGO ont chacune deux antécédents morpho-syntaxiques : MAX et UGO. Par contre, si on applique la règle de coréférence des sujets avant les contraintes syntaxiques, il est possible de faire jouer la contrainte de non coréférence entre sujet et objet, puisque l'unicité de l'antécédent du pronom *il* ne dépend pas de critères syntaxiques : la contrainte de non coréférence peut alors être employée pour la synthèse du pronom [a-objet]. On obtient le texte non ambigu suivant :

(35) *Il a rencontré Ugo. Il lui a serré la main.*

Cet exemple montre qu'on a intérêt à appliquer le parallélisme avant les règles syntaxiques pour pouvoir appliquer les contraintes de non coréférence.

Application de la règle de coréférence des sujets :

La règle de coréférence des sujets peut être résumée ainsi :

Si les sujets de Pi et de Pi+1 sont identiques, et si le sujet de Pi est effacé alors le sujet de Pi+1 est effacé aussi.

Si leurs sujets TOK_i (qui est effacé) et TOK_{i+1} sont différents, alors la décision d'interdire l'effacement du sujet de Pi+1 dépend du genre du sujet TOK_{i+1} de Pi+1. Si le genre de TOK_{i+1} est identique à celui de TOK_i, ou si le genre du sujet effacé de Pi+1 n'est pas indiqué dans Pi+1, alors l'effacement de TOK_{i+1} est interdit. Si, par contre, le genre de TOK_{i+1} est marqué par un autre élément dans Pi+1, et qu'il est différent du genre de TOK_i, alors on ne peut rien conclure.

En d'autres termes, cette décision dépend du résultat des opérations morphologiques : l'application des règles de coréférence des sujets dépend de celle des règles morphologiques.

Application des règles de distance :

Rappelons que nous avons distingué essentiellement deux types de calculs de la distance : le premier est lié au parallélisme ; l'application de telles règles suit par conséquent les remarques ci-dessus. Le deuxième, qui nous intéresse ici, est lié au nombre de mots ou de phrases[7].

Contrairement aux règles de parallélisme, les règles de distance linéaire sont des règles de préférence qui peuvent toujours, au niveau théorique, s'appliquer après les règles morphologiques, sémantiques et syntaxiques. En pratique cependant, les règles de distance qui répondent à la question 1[8] doivent être appliquées en début d'algorithme dans l'approche

7 Nous avons présenté un troisième type de calcul de la distance, à la fois linéaire et dépendant de la structure de l'arbre. Mais les recherches sur ce type de distance n'ont pas encore donné de résultats implémentables.

8 Rappelons que la question 1, présentée en introduction à la section 5.2, est la suivante : étant donnée la distance entre deux occurrences d'une entité, peut-on envisager la synthèse de la seconde en un pronom ?

séquentielle, pour des raisons d'économie de calcul : en effet il est inutile d'effectuer tous les calculs aboutissant à la conclusion qu'un pronom envisagé est synthétisable si ensuite les règles de distance interdisent une telle synthèse. Mais voyons plus en détail l'intégration des règles de préférence interphrastiques dans l'approche séquentielle d'abord, dans l'approche globale ensuite.

5.3.2 INTEGRATION DES REGLES DE PREFERENCE INTERPHRASTIQUES DANS LES DEUX APPROCHES

Approche séquentielle :

L'algorithme de synthèse d'un schéma de phrase Pi en approche séquentielle peut être envisagé en deux étapes principales :

1) Application sur Pi des règles de parallélisme,

2) Synthèse de chaque élément de Pi, dans l'ordre défini par la pronominalisation, et suivant les opérations décrites au chapitre précédent ; de plus, la forme envisagée pour une étiquette donnée est calculée en fonction de la distance séparant cette occurrence de l'étiquette avec l'occurrence précédente. Enfin, les règles de distance répondant à la question 2, i.e. permettant de réduire les ambiguïtés restant après le calcul des antécédents syntaxiques, sont appliquées après le calcul de la liste L3 ou L'3 (rappelons, cf. chapitre 3, que les antécédents syntaxiques d'un pronom constituent une liste notée L3, et que les antécédents syntaxiques du sujet effacé forment la liste L'3).

Les différentes opérations résultant de l'intégration des règles de préférence interphrastiques dans l'approche séquentielle sont résumées ci-dessous pour la synthèse d'un schéma de phrase Pi, pour $i>1$[9] :

1) Application du parallélisme sur Pi et Pi-1 :
   a) Pi et Pi-1 sont parallèles :
      - si $[\text{étiquette-sujet}]_{Pi} = [\text{étiquette-sujet}]_{Pi-1}$
        - alors le sujet de Pi est effacé (on n'appelle pas la fonction de synthèse du sujet)
        - sinon, on indique que l'effacement du sujet est interdit.
      - si $[\text{étiquette-x-objet}]_{Pi} = [\text{étiquette-x-objet}]_{Pi-1}$, et si Pi-1 ne contient pas d'argument pronominalisé, différent de [x-objet]
        - alors le [x-objet] de Pi est synthétisé en un pronom personnel (la seule opération à effectuer est dans ce cas le calcul de la forme du pronom : on n'appelle pas la fonction de synthèse du [x-objet])
        - sinon on indique que la synthèse du [x-objet] en un pronom personnel est interdite (i.e. on synthétise un groupe nominal)

   b) le sujet de Pi-1 est effacé (en français : pronominalisé)
      - si $[\text{étiquette-sujet}]_{Pi} = [\text{étiquette-sujet}]_{Pi-1}$
        - alors le sujet de Pi est effacé (on n'appelle pas la fonction de synthèse du sujet)
        - sinon l'interdiction d'effacer le sujet de Pi dépend de la morphologie

---

9 Le problème ne se pose pas pour P1 puisque nous ne tenons pas compte de la pronominalisation à gauche.

c) sinon, aucune indication ne peut être apportée sur la pronominalisation / effacement

2) Synthèse des TOKi de Pi

- choix d'une forme à synthétiser, en fonction des résultats de l'étape 1. Si on envisage la synthèse d'un pronom :

- calcul des antécédents morphologiques du pronom et de ceux du pronom effacé, si TOKi est sujet, et qu'il n'y a pas de marque qui en interdise l'effacement (cf. calcul sur la distance et la coréférence des sujets)

- calcul des antécédents morphologiques et sémantiques du pronom (et de ceux du pronom effacé, si TOKi est sujet, et qu'il n'y a pas de marque qui en interdise l'effacement)

- calculs des antécédents syntaxiques du pronom

- application des règles de distance

--> le pronom est ou n'est pas synthétisé

De plus, si TOKi est sujet et qu'on en envisage l'effacement, les opérations suivantes sont effectuées après la synthèse des objets :

- calculs des antécédents syntaxiques du sujet vide

- application des règles de distance

--> le pronom sujet est ou n'est pas effacé

Approche globale :

Les caractéristiques de l'approche globale permettent de faire intervenir les règles de préférence interphrastiques "le plus tard possible". Ainsi, les règles de parallélisme et les règles de coréférence des sujets peuvent être appliquées sur les listes de combinaisons Cj (cf. chapitre 4), juste avant les contraintes de non coréférence. Par exemple, si Pi et Pi-1 sont parallèles, et que leurs sujets TOK_{sujet} sont les mêmes, les règles de parallélisme permettent de supprimer les Cj où l'on envisage la synthèse du sujet de Pi en :

- un pronom personnel,

- un démonstratif,

- un groupe nominal.

On ne conserve donc que les Cj où le sujet est effacé, et les antécédents de cet élément vide sont réduits à l'étiquette TOK_{sujet}. Quant aux règles de distance, elles sont utilisées :

- pour supprimer certaines listes Cj (s'il s'agit de règles répondant à la question 1),

- pour réduire le nombre des antécédents des formes pronominales de Cj (s'il s'agit de règles répondant à la question 2).

Elles peuvent donc être appliquées en dernier dans l'algorithme. En particulier, on peut les faire intervenir après les règles de préférence stylistiques (qui notent les Cj en fonction des formes pronominales contenues, cf. chapitre 4). Les différentes opérations résultant de l'intégration des

règles de préférence interphrastiques dans l'approche globale sont résumées ci-dessous pour la synthèse d'un schéma de phrase Pi, où i>1 :

1) Pour tous les TOKi :

- constitution des listes Ui des formes possibles
- calcul pour chaque élément de Ui des antécédents morphologiques et sémantiques

2) Formation des combinaisons Cj

- calculs sur le parallélisme et sur la coréférence des sujets qui débouchent sur la suppression éventuelle de certaines Cj
- calcul des antécédents syntaxiques pour chaque Cj restant
- règles de préférence stylistiques (ppv > pronom > nom)
- calculs sur la distance
  - suppression de certaines Cj (distance trop grande pour pronominaliser)
  - réduction des ambiguïtés (distance trop grande pour certains antécédents)
- choix final d'une version, en fonction du contexte gauche ou du contexte droit (cf. section 5.4)

5.4 CONCLUSION : AVANTAGES DE L'APPROCHE GLOBALE

Rappelons que les principales différences entre l'approche globale et l'approche séquentielle consistent en ces trois points :

- l'approche globale envisage en même temps toutes les possibilités de synthèse pour une entité donnée, alors que l'approche séquentielle effectue un choix unique au départ (et peut éventuellement revenir sur sa décision) ;
- l'approche globale permet de visionner toutes les possibilités de synthèse d'une phrase, par contre l'approche séquentielle donne, l'une après l'autre, les synthèses définitives des éléments de la phrase, dans un ordre donné ;
- l'approche globale aboutit à un choix entre plusieurs versions d'un même texte, ce qui n'est pas le cas de l'approche séquentielle, qui accorde la priorité à une version.

Ces trois points constituent-ils un avantage pour l'approche globale, qui par ailleurs, nécessite des calculs beaucoup plus lourds que l'approche séquentielle, à cause de la gestion de listes complexe qu'elle met en jeu ? En d'autres termes, choisir l'approche globale revient à justifier les affirmations suivantes :

affirmation 1 : Utilité d'avoir en même temps toutes les possibilités de synthèse pour une entité donnée,

affirmation 2 : Utilité de visionner une phrase dans son ensemble,

affirmation 3 : Utilité d'avoir toutes les versions possibles d'une même phrase.

Nous allons décrire les raisons de notre choix, qui peuvent être séparées en trois catégories :

- raisons linguistiques,
- raisons algorithmiques,
- adéquation entre l'ordre des opérations de synthèse, leur rôle et leur fiabilité ;

et chaque type de raison citée sera illustré par des exemples qui justifient une ou plusieurs des affirmations ci-dessus.

1) Raisons linguistiques :

L'approche globale fournit des résultats auquels on ne peut pas accéder avec l'approche séquentielle, pour le traitement des verbes symétriques, la résolution des dépendances croisées du français, et enfin l'utilisation du contexte droit :

a) Traitement des constructions symétriques :

Le problème posé par la dépendance croisée entre la synthèse d'un [dir-objet] et celle d'un [prep-objet] a été évoquée à plusieurs reprises au chapitre 3. Elle est illustrée par l'exemple (S36). Le verbe utilisé dans ce schéma de phrase se construit avec un objet direct et un [prep-objet] introduit par *con* (avec) :

(S36) *(UGO confondere sempre MAX con BOB.)*
(UGO confondre toujours MAX avec BOB.)

Si l'on envisage la synthèse de l'objet direct en un pronom personnel, on aboutit à la phrase (36a) :

(36a) [MAX --> lo] *Ugo **lo** confonde sempre con Bob.*
(Ugo le confond toujours avec Bob.)

Par contre, la pronominalisation de BOB en pronom tonique *lui* produit le résultat (36b), dont l'acceptabilité est douteuse :

(36b) [BOB --> lui] *?Ugo confonde sempre Max con **lui**.*
(?Ugo confond toujours Max avec lui.)

Maintenant, supposons que les occurrences de MAX et BOB de (S36) doivent être synthétisées en pronoms, par exemple supposons que (S36) soit inclus dans le schéma de texte (T36) :

(T36) *(MAX e BOB essere gemelli. UGO confondere sempre MAX con BOB.)*
(MAX et BOB être jumeaux. UGO confondre toujours MAX avec BOB.)

La synthèse conjointe de MAX et BOB en pronoms personnels aboutit au texte (36c) :

(36c) *?*Max e Bob sono gemelli. Ugo lo confonde sempre con lui.*
(?*Max et Bob sont jumeaux. Ugo le confond toujours avec lui.)

qui est inacceptable.

Etant donné que le verbe *confondere* se construit avec des objets symétriques (A. Elia et alii (1981)) :

(a) *Ugo confonde Max con Bob = Ugo confonde Bob con Max.*
(Ugo confond Max avec Bob = Ugo confond Bob avec Max.)

il est possible d'intervertir les deux objets, et d'obtenir ainsi la phrase (36b'), analogue à (36b) et stylistiquement plus acceptable :

(36b') [BOB -> dir-objet et MAX -> prep-objet
BOB --> *lo*] *Ugo lo confonde sempre con Max.*[10].
(Ugo le confond toujours avec Max.)

Considérons maintenant les implications de la symétrie pour le texte (36c). Il est équivalent, dans (36c), de considérer *lo* comme coréférent à *Max* et *lui* comme coréférent à *Bob* ou l'inverse. Cette propriété de symétrie, qui est une information liée au verbe, est contenue dans le lexique-grammaire (qui a été présenté au chapitre 2). On profite alors de cette propriété du verbe *confondere* , qui permet une construction à objets coordonnés :

(b) *Ugo confonde Max con Bob = Ugo confonde Bob con Max.*
<=> ***Ugo confonde Max e Bob.***
(Ugo confond Max avec Bob = Ugo confond Bob avec Max
<=> Ugo confond Max et Bob.)

pour synthétiser (36d) :

(36d) *Max e Bob sono gemelli. Ugo **li** confonde sempre.*
(Max et Bob sont jumeaux. Ugo les confond toujours.)

On remarque que la structure de la phrase construite autour de *confondere* dépend de la pronominalisation :

10 Cette possibilité d'inverser les objets rapproche d'une certainé manière les verbes comme *confondere* aux verbes à constructions dites "standard/croisé", qui ont été étudiées par J.P. Boons, A. Guillet et Ch. Leclère (1976a - 1976b), A. Guillet et Ch. Leclère (1981), M. Salkoff (1983) et J.P. Boons (1984) :
(a) *Max a chargé les oranges sur le camion.*
(b) *Max a chargé le camion d'oranges.*
Notons, cependant, qu'il existe une symétrie absolue entre les deux phrases (c) et (d) ci-dessous :
(c) *Ugo confonde Max con Luca.*
(Ugo confond Max avec Luc.)
(d) *Ugo confonde Luca con Max*
(Ugo confond Luc avec Max.)
contrairement à ce qui se passe pour la paire (a), (b), qui ont un sens légèrement différent.

- le verbe se construit avec deux objets si l'un d'eux est synthétisé en un pronom :

CONSTRUCTION A 2 OBJETS	CONSTRUCTION A COORDINATION D'OBJETS
Ugo lo confonde con Bob. (Ugo le confond avec Bob	*??Ugo confonde lui e Bob.* ??Ugo confond lui et Bob)
?Ugo confonde Max con lui. *-->Ugo lo confonde con Max.* (Ugo le confond avec Max	*??Ugo confonde lui e Max.* ??Ugo confond lui et Max)

- Il se construit avec un [dir-objet] fait d'une coordination d'objets si ceux-ci sont synthétisés en pronom personnel :

**Ugo lo confonde con lui.* (*Ugo le confond avec lui	*Ugo li confonde.* Ugo les confond)

- Sinon la structure de la phrase est indifféremment l'une ou l'autre des constructions ci-après :

Ugo confonde Max con Bob. (Ugo confond Max avec Bob	*Ugo confonde Max e Bob.* Ugo confond Max et Bob)

Il est impossible de décider quelle structure de discours sera choisie, avant de savoir comment seront synthétisés les deux objets. Cette conclusion entre en conflit avec le fait que le composant stratégique (qui choisit la structure de discours) opère avant le composant syntaxique (qui décide de la synthèse des constituants de cette structure de discours en pronoms). Cet exemple montre que :

- la construction syntaxique dépend de la pronominalisation,
- la synthèse de l'objet direct dépend de celle du [prep-objet], comme on le voit en comparant les phrases (36a) et (36d),
- la synthèse de l'objet prépositionnel dépend de celle du [dir-objet], ce qui est illustré par les phrases (36b) et (36d).

Pour résoudre ces dépendances, on peut imaginer la séquence d'opérations suivantes :

- le composant stratégique choisit une construction syntaxique par défaut, e.g. la construction à deux objets ;
- en fonction du résultat des pronominalisations, le composant syntaxique décide soit de conserver la structure choisie par le composant stratégique, soit d'inverser la catégorie des deux objets, soit de transformer la construction à deux objets en construction à objets coordonnés.

On s'aperçoit que l'approche séquentielle est incapable de prendre de telles décisions, puisqu'elles nécessitent de connaître le résultat de la synthèse des deux objets simultanément. Par contre, l'approche globale est tout à fait adaptée :

1) si le [dir-objet] seul doit être pronominalisé, on procède comme d'habitude ;

2) s'il s'agit du [prep-objet], on procède comme d'habitude, après avoir interverti les deux objets ;

3) si les deux objets peuvent être pronominalisés, on envisage en plus des pronoms correspondant à chaque objet la synthèse d'un pronom [dir-objet] pluriel.

b) Résolution du schéma des dépendances en français :

Le composant syntaxique est conçu pour la génération de textes dans les langues romanes. Il a été testé pour l'italien, mais les problèmes soulevés par la pronominalisation ont été étudiés également pour le français. Ces recherches, résumées ci-dessous, mettent en évidence une série complexe de dépendances non locales et croisées pour la synthèse de l'objet direct.

Les premiers résultats sont communs au français et à l'italien :

1) la synthèse de l'objet direct et celle du verbe dépendent l'une de l'autre. En effet, rappelons que la synthèse du verbe est conditionnée par la présence d'un pronom réflexif, et par conséquent, par la synthèse de l'objet direct et du [à-objet] :

*Eva **a** lavé Paul.*
*Eva **s'est** lavée.*

D'autre part, la synthèse de l'objet direct dépend de la conjugaison du verbe ; en effet, en fonction de la lettre initiale du verbe conjugué, la forme de la ppv [dir-objet] 3ème personne du singulier est soit *le / la* soit *l'*. Ainsi, la synthèse d'une ppv [dir-objet] est possible dans :

*Eva arriva. Paul **l'accueillit** avec plaisir.*

ou dans :

*Eva et Luc sont arrivés. Paul **l'a accueillie** avec plaisir.*

mais pas dans :

Eva et Luc arrivèrent. Paul **l'accueillit avec plaisir.*

où le pronom *l'* est ambigu.

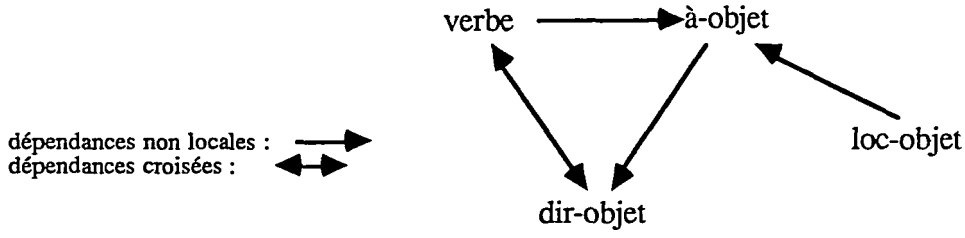
2) La synthèse du [à-objet] dépend de celle du [dir-objet]. Comme on l'a vu pour l'italien, la synthèse d'une ppv [à-objet] est interdite en présence d'une ppv [dir-objet] de la 1ère et 2ème personne, et la phrase ci-dessous est agrammaticale :

**Je me lui présente.*

3) Enfin, le dernier résultat commun au français et à l'italien qui nous intéresse est que la synthèse du [loc-objet] dépend de celle du [à-objet], comme l'illustre l'exemple ci-dessous :

**Je retourne au bureau$_i$. J'y$_i$ penserai (Ø + à ce problème).*

Ces dépendances conduisent au schéma intermédiaire suivant :



Il est possible de les résoudre par l'approche séquentielle, en adoptant l'ordre de synthèse suivant :

a - on détermine si le [dir-objet] ou le [à-objet] doivent être synthétisés en pronoms réflexifs,

b - on effectue la synthèse du verbe, en fonction des résultats de l'étape a, et sans tenir compte d'un éventuel accord avec l'objet direct,

c - on synthétise l'objet direct en fonction de b,

d - on synthétise le [à-objet],

e - on synthétise le [loc-objet],

f - on achève la synthèse du verbe, s'il a un participe passé à accorder avec le [dir-objet].

Cependant, les règles de pronominalisation mettent en jeu, en français, d'autres dépendances. En effet, l'élision d'un pronom [dir-objet] français ne dépend pas uniquement de la première lettre du verbe : elle dépend aussi de l'éventuelle synthèse des ppv : *lui*, *leur*, *y*, *en*[11]. Ainsi, la synthèse de l'objet direct dépend de celle du [à-objet]. En effet :

- la ppv [dir-objet] est placée en français à gauche de la ppv [à-objet],

- l'élision d'une ppv [dir-objet] de la 3ème personne dépend, rappelons-le, de la première lettre par laquelle commence le mot suivant ce pronom.

Ces deux remarques ont des conséquences sur la synthèse du [dir-objet], que l'on peut illustrer avec le schéma de phrase suivant :

(MOI offrir CADEAU à MAX.)

En supposant que CADEAU a déjà été mentionné, on n'envisage pas de la même façon la synthèse de cette étiquette selon que MAX apparaisse ici pour la première fois :

Je l'offre à Max.

11 Je remercie Pollet Samvilian qui m'a rappelé cet aspect du problème.

ou que la synthèse de MAX en une ppv soit possible :

*Je **le lui** offre.*

les antécédents morphologiques de la forme envisagée (*l'*) dans le premier cas sont les étiquettes féminin-singulier et masculin-singulier, contrairement à ce qui se passe dans le second cas, où les antécédents en genre et nombre de la ppv *le* sont les étiquettes masculin-singulier uniquement. En d'autres termes, la synthèse du [à-objet] conditionne la forme de la ppv objet direct envisagée, et par conséquent ses conditions de synthèse : la synthèse du [dir-objet] dépend de celle du [à-objet].

Un résultat similaire s'observe avec les synthèses du [loc-objet] de destination et celle du [dir-objet] : pour les mêmes raisons que ci-dessus, la synthèse du [dir-objet] dépend de celle du [loc-objet]. Examinons par exemple le schéma de texte :

(MOI mettre CADEAU dans ARMOIRE.)

où CADEAU a déjà été mentionné. On envisage sa synthèse

- soit en ppv *le* :

*Je **le** mets **dans l'armoire.***

- soit en ppv *l'* :

*Je **l'y** mets.*

en fonction de la synthèse de l'étiquette ARMOIRE.

Les mêmes types de dépendances sont mises en jeu avec la synthèse d'un [loc-objet] de provenance et celle d'un [dir-objet] ; la synthèse de :

(MOI sortir CADEAU de ARMOIRE.)

débouche sur :

Je le sors de l'armoire.

ou sur :

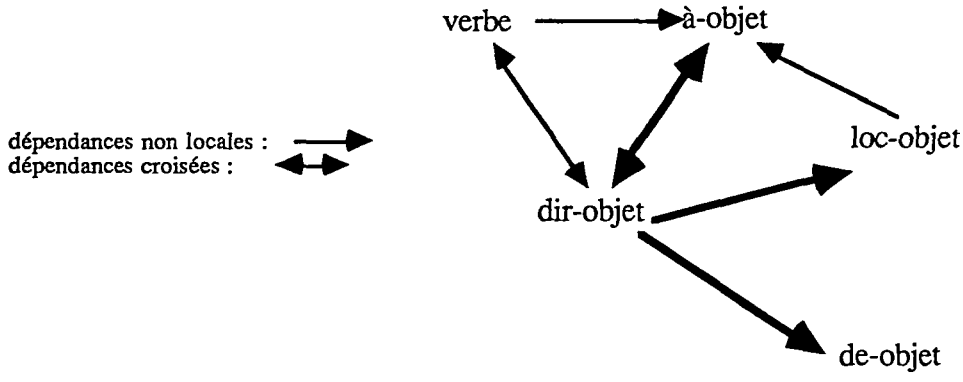
Je l'en sors.

en fonction de la synthèse de ARMOIRE.

Enfin , signalons que les mêmes problèmes sont soulevés avec la synthèse des [de-objet], comme le montre le contraste entre ces deux exemples :

*Je **le** dispense de cette tâche.*
*Je **l'en** dispense.*

On aboutit, avec ces résultats propres au français, à un nouveau schéma des dépendances :



L'ordre de synthèse présenté ci-dessus pour l'approche séquentielle ne peut plus être appliqué : en particulier, il est impossible d'ordonner les points c, d et e, c'est à dire la synthèse des trois objets. En d'autres termes la synthèse d'une phrase du français est quasiment impossible par l'approche séquentielle sans l'intervention de nombreux retours-arrière. Par contre, il est clair que l'approche globale (où l'on envisage toutes les formes de synthèse d'un [dir-objet] de la 3ème personne du singulier) est tout à fait adéquate pour résoudre cet ensemble de dépendances croisées. Rappelons en effet qu'à chaque forme envisagée est associée une série de conditions, appliquées dès que possible :

[ppv]:*l'* < première-lettre-verbe = voyelle,
a-objet ≠ *lui,leur*
ou
loc-objet = *y,en*
a-objet ≠ *lui,leur*
ou
a-objet ≠ *lui,leur*
de-objet = *en*
>

[ppv]:*le* (*la*) < première-lettre-verbe ≠ voyelle,
loc-objet ≠ *y,en*
de-objet ≠ *en*
ou
a-objet = *lui,leur*
>

c) Utilisation du contexte droit pour le choix d'une version :

L'exemple ci-dessous met en évidence la véracité de l'affirmation 3 : l'utilité d'avoir plusieurs versions d'une phrase pour construire un texte, cohérent du point de vue de la pronominalisation, non seulement en fonction du contexte gauche de Pi, mais également en fonction de son contexte droit, i.e. du schéma de phrase représentant la phrase Pi+1. Examinons en effet le schéma de phrase (S6) du chapitre 4 :

(EVA presentare MAX a UGO.)
(EVA présenter MAX à UGO.)

Rappelons que MAX peut être défini en "un vieillard" et UGO en "un jeune homme". Supposons que ce schéma appartiennent au schéma de texte (T37) :

(T37) *(MAX e UGO fare amicizia. EVA presentare MAX a UGO.)*
(MAX et UGO faire connaissance. EVA présenter MAX à UGO.)

Avec l'approche globale, on obtient deux synthèses pour le deuxième schéma de phrase de (T37), et donc deux textes possibles, (37a) et (37b) :

(37a) *Max e Ugo hanno fatto amicizia. Eva lo ha presentato al giovane.*
(Max et Ugo ont fait connaissance. Eva l'a présenté au jeune homme.)

(37b) *Max e Ugo hanno fatto amicizia. Eva gli ha presentato il vecchietto.*
(Max et Ugo ont fait connaissance. Eva lui a présenté le vieillard.)

Maintenant, considérons deux façons possibles d'élargir (T37) dans son contexte droit :

(T37-38a) *(MAX e UGO fare amicizia. EVA presentare MAX a UGO. LIA presentare poi MAX a FAMIGLIA.)*
(MAX et UGO faire connaissance. EVA présenter MAX à UGO. LIA présenter ensuite MAX à FAMIGLIA.)

(T37-38b) *(MAX e UGO fare amicizia. EVA presentare MAX a UGO. LIA presentare poi FAMIGLIA a UGO.)*
(MAX et UGO faire connaissance. EVA présenter MAX à UGO. LIA présenter ensuite FAMIGLIA à UGO.)

La synthèse de (T37-38a) doit conduire à :

(37a-38a) *Max_i e Ugo_k hanno fatto amicizia. Eva lo_i ha presentato al $giovane_k$.Poi, Lia lo_i ha presentato al resto della famiglia.*
(Max_i et Ugo_k ont fait connaissance. Eva l'_i a présenté au jeune $homme_k$. Ensuite, Lia l'_i a présenté au reste de la famille.)

alors que celle de (T37-38b) ne peut aboutir qu'à :

(37b-38b) *Max_i e Ugo_k hanno fatto amicizia. Eva gli_k ha presentato il $vecchietto_i$. Poi, Lia gli_k ha presentato il resto della famiglia.*
(Max_i et Ugo_k ont fait connaissance. Eva lui_k a présenté le $vieillard_i$. Ensuite, Lia lui_k a présenté le reste de la famille.)

On remarque que les deux derniers schémas de phrases de (T37-38a) et (T37-38b) ont des structures parallèles. Leurs sujets (EVA / LIA) sont différents. Par contre, les critères de parallélisme peuvent être appliqués pour la synthèse de leurs objets. Examinons tout d'abord la synthèse de (T37-38a). En choisissant la version (37a) pour la synthèse du deuxième schéma de phrase, on rend possible la synthèse de MAX en une ppv (*lo*) dans le dernier schéma de phrase :

(37a) *Eva **lo** ha presentato al giovane.*
(Eva **l'**a présenté au jeune homme.)
(38a) *Poi, Lia **lo** ha presentato al resto della famiglia.*
(Ensuite, Lia **l'**a présenté au reste de la famille.)

En effet, les critères de parallélisme pour les [dir-objets] sont réunis :

- parallélisme des structures des deux schémas de phrases,
- égalité des [dir-objets], i.e. MAX,
- synthèse des arguments de *presentare* (sauf l'objet direct) dans (49a), en groupes nominaux.

A l'inverse, le choix de (37b) a pour conséquence une synthèse du dernier schéma de phrase en une phrase stylistiquement maladroite :

> *Max_i e Ugo_k hanno fatto amicizia. Eva gli_k ha presentato il $vecchietto_i$. Poi, Lia ha presentato Max al resto della famiglia.*
> (Max_i et Ugo_k ont fait connaissance. Eva lui_k a présenté au $vieillard_i$. Ensuite, Lia a présenté Max au reste de la famille.)

puisque la pronominalisation de MAX est interdite (le texte résultant est ambigu) :

> **Max_i e Ugo_k hanno fatto amicizia. Eva gli_k ha presentato il $vecchietto_i$. Poi, Lia lo_i ha presentato al resto della famiglia.*
> (*Max_i et Ugo_k ont fait connaissance. Eva lui_k a présenté au $vieillard_i$. Ensuite, Lia l'_i a présenté au reste de la famille.)

Un raisonnement analogue conduit au choix de (37b) lors de la synthèse de (T37-38b) :

(37b) *Eva **gli** ha presentato il vecchietto.*
(Eva **lui** a présenté le vieillard.)
(38b) *Poi, Lia **gli** ha presentato il resto della famiglia.*
(Ensuite, Lia **lui** a présenté le reste de la famille.)

puisqu'il permet l'application des critères de parallélisme pour la synthèse de UGO en une ppv dans (38b). Cette pronominalisation est interdite avec le choix de la version (37a), qui impose la synthèse du texte stylistiquement désagréable suivant :

> *Max_i e Ugo_k hanno fatto amicizia. Eva lo_i ha presentato al $giovane_k$. Poi, Lia ha presentato il resto della famiglia a Ugo.*
> (Max_i et Ugo_k ont fait connaissance. Eva l'_ia présenté au jeune $homme_k$. Ensuite, Lia a présenté le reste de la famille à Ugo.)

Les exemples (T37-38a) et (T37-38b) illustrent l'utilisation du contexte droit d'une version d'une phrase, et par là même dans l'avantage d'avoir plusieurs versions d'une même phrase : ce résultat privilégie donc l'approche globale par rapport à l'approche séquentielle, qui ne permet pas de disposer d'un tel choix.

2) Raisons algorithmiques :

Avec l'approche globale, on évite les retours en arrière et on supprime certaines hypothèses de calcul complexes inhérentes à l'approche séquentielle, notamment en ce qui concerne la choix entre deux formes pronominales, et celui d'une version en fonction du contexte gauche.

a) Choix entre une forme préverbale et un pronom personnel tonique :

Si la synthèse d'une forme pronominale est interdite, parce qu'elle engendre des ambiguïtés, on peut être amené, si cela est possible, à envisager la synthèse d'une autre forme pronominale, contenant plus d'information (sur le genre ou la nature sémantique) concernant l'étiquette à synthétiser, ce qui implique une duplication des calculs avec l'approche séquentielle. Ainsi, considérons le schéma de texte (T39) :

(T39) *(UGO e MAX paragonare CASA a PALAZZO. UGO e MAX parlare spesso di CASA.)*
(UGO et MAX comparer CASA à PALAZZO. UGO et MAX parler souvent di CASA.)

Si la seconde occurrence de CASA était synthétisée en ppv *ne*, le texte (39a) résultant serait ambigu :

(39a) **Ugo e Max paragonano questa catapecchia a questo palazzo. Ne parlano spesso.*
(*Ugo et Max comparent cette bicoque à ce palais. [Ils] en parlent souvent.)

puisque *ne* réfère aussi bien à CASA qu'à PALAZZO. On envisage alors la synthèse de la forme tonique *essa*, qui n'a qu'un antécédent morpho-sémantique : CASA, et dont la synthèse n'engendre, elle, pas d'ambiguïté :

(39b) *Ugo e Max paragonano questa catapecchia a questo palazzo. Parlano spesso di essa.*
(Ugo et Max comparent cette bicoque à ce palais. [Ils] parlent souvent d'elle$_{non\text{-}hum.}$)

En appliquant l'approche séquentielle, on est amené à recommencer deux fois les mêmes calculs (listes L1[di-objet], L2[di-objet] et L3[di-objet]) pour la ppv *ne* d'abord, pour la forme tonique *essa* ensuite : on est obligé de recourir à un retour-arrière qui est évité dans l'approche globale, où les deux formes pronominales sont envisagées en même temps.

b) Choix d'une version en fonction du contexte gauche : rôle de la distance.

Considérons le schéma de texte :

(T40) *(EVA e LIA litigare sempre. MAX dar ragione a EVA. EVA odiare LIA.)*
(EVA et LIA se disputer toujours. MAX donner raison à EVA. EVA détester LIA.)

On y trouve trois schémas de phrase dont les structures sont deux à deux non-parallèles. La synthèse des deux premiers schémas de phrases donnent lieu aux phrases :

Eva e Lia litigano sempre. Max dà ragione a Eva
(Eva et Lia se disputent toujours. Max donne raison à Eva.)

A priori, la synthèse du dernier schéma de phrase aboutit à deux versions, étant donné que le contexte gauche de ce schéma contient les étiquettes EVA et LIA :

(a) *Lei odia Lia.*
(Elle déteste Lia.)
(b) *Eva la odia.*
(Eva la déteste.)

Le fait que EVA apparaisse dans la deuxième phrase, contrairement à LIA, impose le choix de la version (a), dans la synthèse de (T40) :

(40a) *Eva e Lia litigano sempre. Max dà ragione a Eva. Lei odia Lia ...*
(Eva et Lia se disputent toujours. Max donne raison à Eva. Elle déteste Lia ...)

En effet la synthèse de la phrase (b) conduit à un texte inacceptable, bien que non ambigu :

(40b) **Eva e Lia litigano sempre. Max dà ragione a Eva. Eva la odia ...*
(*Eva et Lia se disputent toujours. Max donne raison à Eva. Eva la déteste ...)

L'inacceptabilité est due à la distance entre la ppv *la* et son antécédent LIA, comparée à la proximité des deux occurrences de EVA : il faut tenir compte des critères de distance linéaire dans le choix d'un pronom, et donc d'une version.

Examinons maintenant maintenant le schéma de texte (T40'), identique à (T40) excepté pour la valeur de l'objet dans le deuxième schéma de phrase, qui est ici égal à LIA :

(T40') *(EVA e LIA litigare sempre. MAX dar ragione a LIA. EVA odiare LIA.)*
(EVA et LIA se disputer toujours. MAX donner raison à LIA. EVA détester LIA.)

Dans ce schéma de texte, l'étiquette LIA apparaît dans les deux derniers schémas de phrases, contrairement à EVA : les critères de distance linéaire conduisent ici à choisir la version (b) dans la synthèse de (T40') :

(40'b) *Eva e Lia litigano sempre. Max dà ragione a Lia. Eva la odia ...*
(Eva et Lia se disputent toujours. Max donne raison à Lia. Eva la déteste ...)

En effet, le choix de (a) débouche sur un texte incorrect :

(40'a) **Eva e Lia litigano sempre. Max dà ragione a Lia. Lei odia Lia ...*
(*Eva et Lia se disputent toujours. Max donne raison à Lia. Elle déteste Lia ...)

Pour réaliser ces choix, en fonction de la distance, les calculs à effectuer dans l'approche globale, où l'on a les deux versions en présence, sont plus simples que dans l'approche séquentielle, où il faut formuler des hypothèses tenant compte :

- de la distance entre les étiquettes EVA,
- de la distance entre les étiquettes LIA,
- de l'absence de parallélisme et de coréférence des sujets,

pour décider des formes (pro)nominales à envisager pour chaque entité mentionnée dans (S40).

3) Raisons liées au niveau de certitudes des différents types de règles qui interviennent dans la synthèse d'un pronom :

Il y a dans l'approche globale une meilleure adéquation entre le niveau de fiabilité des règles, leur rôle et l'ordre dans lequel elles interviennent. Comparons en effet les règles dont le rôle est de sélectionner une forme (pro)nominale par le calcul de ses antécédents, et les règles dont le rôle est d'opérer un choix entre plusieurs formes données. Nous appellerons les premières règles "règles de calcul", les secondes "règles de préférence". Contrairement aux règles de calcul, les règles de préférence sont plus ou moins discutables, car plus ou moins subjectives. De plus, bien que les règles de préférence soient parfois plus puissantes que les règles de calcul (cf. le parallélisme dans les exemples (33) et (34)), on peut en général concevoir que les règles de préférences s'appliquent sur des données déjà existantes, donc après les autres. En conclusion, il semble y avoir une certaine corrélation entre le rôle des règles (une règle de préférence prend en entrée le résultat d'une règle de calcul), leur fiabilité (une règle de calcul est plus fiable qu'une règle de préférence) et l'ordre dans lequel elles doivent être appliquées (règle de calcul < règle de préférence) : or, dans l'approche globale, cet ordre est dans l'ensemble respecté ; par contre, dans l'approche séquentielle, les règles de préférences sont appliquées brutalement en début d'algorithme.

Ces différentes raisons (logiques, linguistiques et formelles), basées sur l'étude des phrases simples font que l'approche globale présente des avantages énormes par rapport à l'approche séquentielle. Le chapitre suivant, où il est question de la synthèse des phrases complexes à complétives renforcera encore cette affirmation.

6 ARGUMENTS COMPLEXES DE LA PHRASE : SYNTHESE DES COMPLETIVES

6.1 INTRODUCTION

L'approche globale permet de synthétiser un schéma de phrase dont les étiquettes représentent des entités. Nous avons vu que la synthèse d'une entité donne lieu à un groupe nominal, un nom propre, un pronom démonstratif, un pronom personnel ou un élément vide. Il n'a été question, jusqu'ici, que de la synthèse des groupes nominaux. Cependant, une étiquette peut représenter non seulement une entité mais également un prédicat, comme nous l'avons vu : à partir d'une représentation sémantique exprimant un prédicat et instanciant un champ d'un autre prédicat, on génère un schéma de phrase enchâssée, qui occupe la position sujet ou [x-objet] d'un verbe. Ce schéma de phrase est alors synthétisé en complétive sujet ou [x-objet]. Ainsi, rappelons le schéma de texte présenté au chapitre 2, contenant le schéma d'une complétive [dir-objet] :

(T1) (:cl (:table 47)
(:sujet HOMME1)
(:verbe *ordinare*)
(:dir-objet
(:cl (:table 23)
(:sujet FEMME1)
(:verbe *vendere*)
(:dir-objet OBJET1)
(:a-objet HOMME2)))
(:a-objet FEMME1))

et dont la synthèse permet d'obtenir :

(1) *Max ha dato ordine a Eva di vendere la casa a Ugo.*
(Max a donné l'ordre à Eva de vendre la maison à Ugo.)

Un argument phrastique, ou complétive, peut s'exprimer sous la forme récursive f (x_1, x_2, ... x_i ... x_n), où f est une fonction à n arguments, mettant en relation x_1, x_2 ... x_n ; x_i est une étiquette pouvant représenter soit une entité, soit une complétive (dans ce cas, xi est elle-même dérivée en g (y_1, y_2, ... y_j ... y_m)). Ainsi, le schéma (T1) peut se réécrire :

f (HOMME1, TRANSACTION, FEMME1)

les paramètres HOMME1 et FEMME1 sont des entités, TRANSACTION peut, lui, être dérivé en :

g (FEMME1, OBJET1, HOMME2)

dont les trois arguments sont des entités. La fonction f se dérive donc récursivement en :

f (HOMME1, g (FEMME1, OBJET1, HOMME2), FEMME1)

La difficulté de la synthèse des complétives, qui apparaît dans l'exemple ci-dessus, est la présence, à plusieurs niveau d'enchâssement, d'une même entité, i.e. FEMME1. L'exemple (1) illustre un cas particulier : l'égalité entre le sujet de la complétive et ce qui est généralement appelé son contrôleur ; cette égalité peut imposer la réduction de la complétive à l'infinitive. Cette propriété a été largement étudiée dans les écoles linguistiques ; citons notamment Manzini (1983) pour la théorie du Government and Binding, et Gazdar et alii (1985) pour les travaux effectués à ce sujet dans le cadre de la théorie GPSG ; d'autre part, l'équipe du L.A.D.L. dirigée par M. Gross (1975), et A. Elia et alii (1981) ont établi la liste des verbes à complétive (respectivement pour le français et l'italien) dont ils ont décrit les propriétés.

Les conséquences de la notion de "contrôle" en génération seront développées dans ce chapitre. Plus généralement, nous étudierons les problèmes posés par la synthèse d'une étiquette située à la fois dans la principale et dans la complétive. En d'autres termes, nous verrons que l'application des contraintes de non coréférence dans la principale peuvent se répercuter sur la synthèse des pronoms dans la complétive, et réciproquement. Nous indiquerons alors une méthode d'intégration de la synthèse des complétives dans l'approche globale, et nous proposerons une solution pour la gestion de l'historique des étiquettes. Etant donné, par exemple, le schéma de phrase :

(EVA *annunciare* (MAX *tradire* PIERRE) *a* PIERRE.)
(EVA annoncer (MAX trahir PIERRE) à PIERRE.)

où EVA, PIERRE et MAX sont mentionnés pour la première fois, nous verrons comment parvenir à synthétiser :

Eva ha annunciato a Pierre che Max lo ha tradito.
(Eva a annoncé à Pierre que Max l'a trahi.)

tout en évitant les combinaisons suivantes :

**Eva gli ha annunciato che Max ha tradito Pierre.*
(*Eva lui a annoncé que Max a trahi Pierre.)

**Eva gli ha annunciato che Max lo ha tradito.*
(*Eva lui a annoncé que Max l'a trahi.)

**Eva ha annunciato a Pierre che Max ha tradito Pierre.*
(*Eva a annoncé à Pierre que Max a trahi Pierre.)

**Eva ha annunciato che Max ha tradito Pierre a questi*
(*Eva a annoncé que Max a trahi Pierre à celui-ci.)

Mais tout d'abord, nous présentons ci-dessous une description linguistique des arguments phrastiques.

6.2 DESCRIPTION D'UNE COMPLETIVE

Une complétive est une phrase qui occupe la position sujet ou objet d'un verbe V. La classification des verbes en fonction de leur comportement vis-à-vis des complétives a été décrite pour l'italien dans A. Elia et alii (1981), A. Elia (1982,1984). En effet c'est le verbe V qui définit un certain nombre de propriétés des complétives :

6.2.1 NOMBRE DE COMPLETIVES

Un verbe V donné se construit :

- soit sans complétive, comme le verbe *dormire* :

(2) *(Paul + *il fatto che Paul venga domani) dorme.*
((Paul + *le fait que Paul vienne demain) dort.)

- soit avec une complétive occupant une position syntaxique précise, comme le verbe *volere*, qui accepte une complétive en position objet direct :

(2a) [dir-objet] *Paul vuole **che Eva parta**.*
(Paul veut **que Eva parte**.)

(2b) [sujet] **Che Eva parta vuole Paul.*
(*Que Eva parte veut Paul.)

- soit avec deux complétives (encore une fois leur position syntaxique est fixe). Ainsi le verbe *accorgersi* (s'apercevoir) accepte une complétive en position objet direct et une autre en position [da-objet] :

(3) *Eva si accorge **che è tardi*** $_{\text{dir-objet}}$ ***dal fatto che i negozi chiudono*** $_{\text{da-objet}}$.
(Eva s'aperçoit **qu'il est tard** $_{\text{dir-objet}}$ **du fait que les magasins ferment**$_{\text{da-objet}}$.)

- soit avec trois complétives, i.e. chaque argument de V peut être phrastique ; ce cas est assez rare, car le résultat est stylistiquement peu naturel, comme le montre l'exemple emprunté à A. Elia et alii (1981) :

(4) *Il fatto che tutto vada in rovina amalgama il fatto che nessuno va più al cinema con il fatto che i criminali scorazzano per la città.*
(Le fait que tout se détériore amalgame le fait que personne ne va plus au cinéma avec le fait que les criminels courent à travers la ville.)

6.2.2 MODE DE LA COMPLETIVE ET SUBSTANTIFS QUI LA REGISSENT

Mode de la complétive

Le mode de la complétive (indicatif/subjonctif) dépend du verbe principal. Parfois, tous les verbes d'une même table du lexique-grammaire induisent le même mode à leur complétive.

C'est le cas par exemple des verbes de la table 49, qui se construisent avec une complétive [a-objet], dont le verbe est toujours au subjonctif :

(5) [table 49] *Eva tiene a che Max (resti + *resta) a casa.*
(Eva tient à ce que Max reste à la maison.)

Dans d'autres tables, la valeur du mode de la complétive change d'un verbe à l'autre. Ainsi, *volere* (vouloir) exige le subjonctif, alors qu'une complétive dépendant de *credere* (croire) doit être à l'indicatif :

(6) [table 43] *Eva vuole che Max se ne (**vada** + *va).*
(Eva veut que Max s'en (**aille** + *va).)

*Eva crede che Max se ne (**va** + *vada).*
(Eva croit que Max s'en (**va**+ *aille).)

Pour certains verbes, comme *credere* (croire), la négation dans la principale peut induire le subjonctif dans la complétive :

Ugo non crede che Max se ne (vada + va).
(Ugo ne croit pas que Max s'en (aille + *va).)

Substantif qui régit la complétive

Une complétive est introduite par la conjonction *che* (que). La plupart du temps, elle peut être précédée par le groupe nominal *il fatto* (le fait), sans qu'il y ait modification du sens de la phrase :

(7) *Eva concepisce (Ø + il fatto) che Max voglia partire.*
(Eva conçoit (Ø + le fait) que Max veuille partir.)

Mais certains verbes n'ont pas cette double possibilité. Ainsi, pour les verbes de la table 43, le groupe nominal *il fatto* ne peut pas être adjoint à la complétive [dir-objet] :

(8) *Eva vuole (*il fatto + Ø) che Max se ne vada.*
(Eva veut (*le fait + Ø) que Max s'en aille.)

Par contre, les complétives [da-objet] des verbes de la table 46 doivent obligatoirement être introduite par *il fatto* :

(9) *Questa situazione risulta (dal fatto + *Ø + *da) che Max se n'è andato.*
(Cette situation résulte (du fait + ?*Ø + de ce) que Max est parti.)

Enfin pour certains verbes, le groupe nominal régissant la complétive **peut** être différent de *il fatto* :

(10a) [table 49] *Max abitua Eva (al fatto + all'idea) che Ugo se ne va.*
(Max habitue Eva (au fait + à l'idée) que Ugo s'en va.)

ou **doit** être différent de *il fatto* :

(10b) [table 47] *Max verifica (*il fatto + la possibilità) che Ugo abbia detto la verità.*
(Max vérifie (*le fait + la possibilité) que Ugo ait dit la vérité.)

Une complétive se construisant au subjonctif change parfois son mode de conjugaison quand elle est introduite par *il fatto*. Ainsi, on contrastera l'exemple (5) que nous rappellerons :

(5) *Eva tiene a che Max (resti + *resta) a casa.*
(Eva tient à ce que Max reste à la maison.)

avec l'exemple (11) :

(11) *Eva tiene al fatto che Max (resta + resti) a casa.*
(Eva tient au fait que Max reste$_{\text{subjonctif + indicatif}}$ à la maison.)

Ces propriétés, notées sous forme de paramètres dans le lexique-grammaire, ne posent pas de problème de synthèse : il s'agit de transformations que l'on peut effectuer à la dernière étape d'un système de génération.

6.2.3 REDUCTION DES COMPLETIVES A L'INFINITIVE

La principale caractéristique des complétives est de pouvoir être réduites à l'infinitive ou en groupes nominaux prédicatifs. Contrairement aux précédentes, cette propriété est cruciale pour l'algorithme de génération.Rappelons qu'à partir du concept PROMENADE, le composant stratégique décide de l'une des transformations suivantes :

[verbe simple] *NO passeggiare.*
(N0 se promener.)
[verbe support] *N0 fare una passeggiata.*
(N0 faire une promenade.)
[g.n. prédicatif] *La passeggiata di N0.*
(La promenade de N0.)
[nominalisation de l'infinitive] *Il passeggiare di N0.*
([litt :] Le promener de N0.)

La nominalisation des verbes à l'infinitif est un phénomène propre à l'italien que nous avons laissé de côté dans nos recherches actuelles. Les autres transformations, et en particulier la mise sous forme de groupe nominal prédicatif, dépasse le cadre de la synthèse des complétives : elles concernent celui de la synthèse des groupes nominaux complexes à complément du nom.

La réduction des complétives à l'infinitive est une opération "préparée" par le composant stratégique (cf. chapitre 7) et effectuée par le composant syntaxique. C'est cette dernière propriété linguistique des complétives que nous présentons ici : nous en faisons ci-dessous la description ; nous en verrons ensuite les conséquences sur la pronominalisation dans les complétives et sur la façon d'intégrer leur synthèse dans l'approche globale.

Dans certaines conditions, une complétive peut (ou doit, en fonction du verbe principal) être réduite à l'infinitive. Cette réduction a lieu en cas d'égalité entre l'un des arguments du verbe, appelé **contrôleur**, et le **sujet de la complétive**. La position syntaxique du contrôleur est déterminée par le verbe principal : il s'agit d'un autre paramètre inscrit dans le lexique-grammaire.

(12) [contrôleur=sujet] *Max$_i$ vuole Ø$_i$ andarsene.*[1]
(Max$_i$ veut Ø$_i$ s'en aller.)

(13) [contrôleur=dir-objet] *Max$_i$ obliga Eva$_k$ ad Ø$_k$ andarsene.*
(Max$_i$ oblige Eva$_k$ à Ø$_k$ s'en aller.)

(14) [contrôleur=a-objet] *Max$_i$ suggerisce a Eva$_k$ di Ø$_k$ andarsene.*
(Max$_i$ suggère à Eva$_k$ de Ø$_k$ s'en aller.)

(15) [contrôleur=prep-objet] *La situazione$_i$ equivale per Max$_k$ di Ø$_k$ andarsene.*
(La situation$_i$ équivaut pour Max$_k$ de Ø$_k$ s'en aller.)[2]

Notons encore que certains verbes ont plusieurs contrôleurs possible pour la réduction à l'infinitif de leur complétive. C'est le cas par exemple, du verbe *proporre* (proposer) qui se construit avec une complétive objet direct :

(16) *Ugo propone ad Eva che Bob se ne vada.*
(Ugo propose à Eva que Bob s'en aille.)

Le contrôleur de la complétive peut être soit le sujet :

(17a) ***Ø$_i$** Ho proposto ad Eva di Ø$_i$ andarmene.*
(**J'$_i$** ai proposé à Eva de Ø$_i$ **m'**en aller.)

soit le [a-objet] :

(17b) *Eva **mi$_i$** ha proposto di Ø$_i$ andarmene.*
(Eva **m'$_i$** a proposé de Ø$_i$ **m'**en aller.)

soit encore le résultat de la coordination des deux entités :

(17c) ***Eva$_i$ mi$_j$** ha proposto di Ø$_{(i\ et\ j)}$ andarcene.*
(**Eva$_i$ m'$_j$** a proposé de Ø$_{(i\ et\ j)}$ **nous** en aller.)

Ce dernier exemple illustre un cas qui n'a pas été implémenté. Quant aux textes (17a) et (17b), ils montrent que la possibilité d'avoir deux contrôleurs pose un problème pour l'interprétation du sujet effacé de la complétive, quand la personne de celui-ci n'est pas indiquée dans le reste de la phrase :

---

1 Rappelons (cf. chapitre 2) que les ppv objets d'un verbe à l'infinitif sont accolées à celui-ci, et que le *e* final du verbe tombe. Ainsi :
se + ne + andare -> andarsene
Rappelons en outre que la notation Ø, symbolisant le sujet vide, permet de visualiser l'interprétation coréférentielle indiquée par l'indice i, entre le sujet vide de l'infinitive et *Max*, sujet de la principale.

2 Une certaine catégorie de verbes, dits verbes de perception, se construisent avec des complétives dont la réduction à l'infinitive n'est pas régie par un contrôleur, mais par le sujet de l'infinitive :

?Max ascolta che la pioggia cade -> *Max ascolta cadere la pioggia*
(?Max écoute que la pluie tombe -> Max écoute tomber la pluie)

(17'b) *Eva$_i$ mi$_j$ ha proposto di* $\emptyset_{i+j+(i\ e\ j)}$ *partire.*
(Eva$_i$ m'$_j$ a proposé de $\emptyset_{i+j+(i\ et\ j)}$ partir.)

ou quand les deux contrôleurs possibles sont de la même personne :

(17"b) *Eva$_i$ ha proposto a Max$_j$ di* $\emptyset_{i+j+(i\ e\ j)}$ *partire.*
(Eva$_i$ a proposé à Max$_j$ de $\emptyset_{i+j+(i\ et\ j)}$ partir.)

(17'"b) *Eva$_i$ ha proposto a Max$_j$ di* $\emptyset_{i+j}$ *andarsene.*
(Eva$_i$ a proposé à Max$_j$ de $\emptyset_{i+j}$ s'en aller.)

les textes (17'b), (17"b) et (17'"b) sont intrinsèquement ambigus. Signalons au passage que l'infinitive objet direct apparaît parfois directement :

(18) *Ugo detesta parlare con Eva.*
(Ugo déteste parler avec Eva.)

D'autres fois, l'infinitive est introduite par *di* (de), ce qui a été illustré par les exemples (17a) à (17'"b). Enfin elle peut être précédée par *il fatto di* (le fait de) :

(19) *Ugo detesta il fatto di parlare con Eva.*
(Ugo déteste le fait de parler avec Eva.)

Ces informations sont stockées dans le lexique-grammaire, comme c'est le cas des informations décrites dans les sections 6.2.1 et 6.2.2.

Pour en finir avec la description linguistique de la réduction des complétives à l'infinitive, rappelons un point que nous avons évoqué au chapitre précédent : quand le sujet d'une complétive dépendant d'un verbe V est égal au contrôleur, alors la nécessité de réduire la complétive à l'infinitif est une propriété liée à V. Par exemple, la réduction est obligatoire si V=*volere* (vouloir) :

(20)	(a)	*Ugo vuole che Eva parta.*	(Ugo veut que Eva parte.)
	(b)	**Ugo$_i$ vuole che Ugo$_i$ parta.*	(*Ugo$_i$ veut que Ugo$_i$ parte.)
	(c)	**Ugo$_i$ vuole che* $\emptyset_i$ *parta.*	(*Ugo$_i$ veut que [il]$_i$ parte.)
	(d)	*Ugo$_i$ vuole* $\emptyset_i$ *partire.*	(Ugo$_i$ veut $\emptyset_i$ partir.)

Elle est facultative si V=*credere* (croire) :

(21)	(a)	*Ugo crede che Eva parte domani.*	(Ugo croit que Eva part demain.)
	(b)	**Ugo$_i$ crede che Ugo$_i$ parte domani.*	(*Ugo$_i$ croit que Ugo$_i$ part demain.)
	(c)	*Ugo$_i$ crede che* $\emptyset_i$ *parte domani.*	(Ugo$_i$ croit qu' [il]$_i$ part demain.)
	(d)	*Ugo$_i$ crede di* $\emptyset_i$ *partire domani.*	(Ugo$_i$ croit $\emptyset_i$ partir demain.)

Dans la section ci-dessous, nous allons voir l'utilisation de cette dernière propriété dans le processus de pronominalisation pour les phrases à complétives.

6.3 CONTRAINTES DE NON COREFERENCE DANS LES PHRASES A COMPLETIVES

Au chapitre 4, nous avons présenté les contraintes de non coréférence comme un phénomène intraphrastique permettant de calculer les antécédents syntaxiques d'une forme pronominale. Dans ce qui suit, nous allons voir que ces contraintes sont applicables entre une principale et sa complétive, et ce, à cause des propriétés du contrôleur de la principale.

6.3.1 COMPLETIVES NON REDUITES A L'INFINITIVE DONT LE SUJET EST UNE FORME PRONOMINALE

Considérons les deux phrases suivantes construites respectivement avec les verbes *volere* et *credere*, dont la complétive objet a pour contrôleur le sujet de la principale, la réduction étant obligatoire avec *volere*, facultative avec *credere* :

(22) *Ugo$_i$ vuole che Ø$_j$ parta domani.*
i≠j (Ugo$_i$ veut qu' [il]$_j$ parte demain.)

(23) *Ugo$_i$ crede che Ø$_i$ parte domani.*
(Ugo$_i$ croit qu' [il]$_i$ part demain.)

Dans (22), *Ugo* est syntaxiquement incompatible avec le sujet effacé du verbe conjugué de la complétive, car la réduction est obligatoire en cas de coréférence :

(22') **Ugo$_i$ vuole che Ø$_i$ parta domani.* => *Ugo$_i$ vuole Ø$_i$ partire domani.*
(*Ugo$_i$ veut qu' [il]$_i$ parte demain.) => (Ugo$_i$ veut Ø$_i$ partir demain.)

En d'autres termes, le contrôleur d'une complétive à réduction infinitive obligatoire est syntaxiquement incompatible avec le sujet de cette complétive, sauf quand elle réduite à l'infinitive.

Par contre, dans (23) *Ugo* est syntaxiquement compatible avec le sujet effacé du verbe conjugué de la complétive car la réduction est facultative. (23) est alors, sous certaines conditions, équivalente à :

(23') *Ugo$_i$ crede di Ø$_i$ partire domani.*
(Ugo$_i$ croit Ø$_i$ partir demain.)

Pour exprimer que le sujet de la complétive est syntaxiquement incompatible avec le sujet du verbe *credere*, il faut conserver le pronom :

(23'') *Ugo$_i$ crede che egli$_j$ parte domani.*
i≠j (Ugo$_i$ croit qu' il$_j$ part demain.)

Ce pronom est également syntaxiquement incompatible avec le sujet du verbe *volere*:

(22'') *Ugo$_i$ vuole che egli$_j$ parta domani.*
i≠j (Ugo$_i$ veut qu' il$_j$ parte demain.)

La contrainte de non coréférence entre le sujet d'une complétive, noté $N_{sujet\text{-}complétive}$ et le contrôleur de celle-ci, noté $N_{contrôleur\text{-}complétive}$ peut être résumée comme suit :

- si $N_{contrôleur\text{-}complétive}$ et $N_{sujet\text{-}complétive}$ sont représentés par les mêmes étiquettes, alors

  si la complétive est réduite à l'infinitive, $N_{contrôleur\text{-}complétive}$ est le seul antécédent syntaxique du sujet effacé $N_{sujet\text{-}complétive}$:

  Ugo$_i$ (vuole + crede di) Ø$_i$ partire domani.
  (Ugo$_i$ (veut +croit) Ø$_i$ partir demain.)

  sinon $N_{contrôleur\text{-}complétive}$ est l'un des antécédents syntaxiques du sujet effacé $N_{sujet\text{-}complétive}$:

  *Ugo$_i$ (*vuole + crede) che Ø$_i$ parte domani.*
  (Ugo$_i$ (*veut +croit) qu'[il]$_i$ part demain.)

- si $N_{contrôleur\text{-}complétive}$ et $N_{sujet\text{-}complétive}$ sont différents, et si $N_{sujet\text{-}complétive}$ est pronominalisé ou synthétisé en groupe nominal, alors $N_{contrôleur\text{-}complétive}$ et $N_{sujet\text{-}complétive}$ sont non coréférents :

  Ugo$_i$ (vuole + crede) che egli$_{k\neq i}$ parte domani.
  (Ugo$_i$ (veut + croit) qu'il$_{k\neq i}$ part(e) demain.)

- si $N_{contrôleur\text{-}complétive}$ et $N_{sujet\text{-}complétive}$ sont différents, et si $N_{sujet\text{-}complétive}$ est effacé et que la réduction à l'infinitive de la complétive est obligatoire, alors $N_{contrôleur\text{-}complétive}$ est un antécédent interdit pour le sujet effacé $N_{sujet\text{-}complétive}$:

  Ugo$_i$ vuole che Ø$_{k\neq i}$ parte domani.
  (Ugo$_i$ veut qu' [il]$_{k\neq i}$ parte demain.)

- si $N_{contrôleur\text{-}complétive}$ et $N_{sujet\text{-}complétive}$ sont différents, et si $N_{sujet\text{-}complétive}$ est effacé, et que la réduction est facultative, alors $N_{contrôleur\text{-}complétive}$ est un antécédent syntaxique du sujet effacé $N_{sujet\text{-}complétive}$:

  Ugo$_i$ crede che Ø$_{k+i}$ parte domani.
  (Ugo$_i$ croit qu' [il]$_{k+i}$ part demain.)

6.3.2 COMPLETIVES REDUITES A L'INFINITIVE

Nous venons de voir comment la contrainte de non coréférence entre le contrôleur de la principale et le sujet de la complétive pouvait s'appliquer pour calculer les antécédents syntaxiques du sujet de cette complétive. L'exemple ci-dessous illustre, à l'inverse, le cas où les formes pronominales d'une complétive permettent de déterminer les antécédents syntaxiques des éléments de la principale. Examinons en effet (24) :

(24) *Lei$_i$ vuole Ø$_i$ ucciderlo$_j$.*
(Elle$_i$ veut Ø$_i$ le$_j$ tuer.)

Supposons que le contexte gauche de (24) contienne les étiquettes MAX et EVA. On envisage d'effacer dans (24) le pronom sujet *lei* de la principale. L'examen de la forme :

(24a) *Lei vuole* <phrase>
(Elle veut <phrase>)

suggère que l'effacement du pronom personnel sujet rend le texte ambigu. Dans :

(24b) *Ø vuole* <phrase>
([Elle + il] veut <phrase>)

le sujet effacé de *volere* peut référer aussi bien à MAX qu'à EVA.

A présent, examinons non pas uniquement le niveau de la principale, qui contient le sujet dont on envisage l'effacement, mais la phrase toute entière, i.e. avec la complétive objet direct. Comme l'indiquent les indices, dans (24), le sujet de cette complétive réduite à l'infinitive est obligatoirement coréférent au contrôleur, i.e. le sujet de *volere*. Par ailleurs, la contrainte de non coréférence entre sujet et objet direct implique que le sujet effacé de *uccidere* et l'objet direct masculin-singulier (*lo*) sont syntaxiquement incompatibles. Le seul antécédent possible de ce pronom personnel masculin est MAX, donc (contrainte de non coréférence) le seul antécédent possible du sujet effacé de *uccidere* est EVA, seul autre humain présent dans le contexte. Comme le sujet effacé d'une infinitive est obligatoirement coréférent au contrôleur de celle-ci, le seul antécédent possible du sujet de *volere* , même s'il est effacé, est EVA. Donc le pronom sujet *lei* peut être effacé :

(24') $Ø_{EVA}$ *vuole* $Ø_{EVA}$ *ucciderlo*$_{MAX}$.
([Elle]$_{EVA}$ veut $Ø_{EVA}$ le$_{MAX}$ tuer.)

La contrainte de non coréférence dans une complétive peut ainsi avoir des répercussions sur le contrôleur de cette complétive.

Une autre utilisation des contraintes de non coréférence est donnée par la phrase :

(25) *Elle veut lui parler.*

dont le contexte gauche est encore une fois constitué des étiquettes EVA et MAX. A priori, il est impossible d'appliquer les contraintes de non coréférence dans l'infinitive puisque les antécédents possibles du sujet vide sont EVA et MAX, ainsi que ceux de la ppv [à-objet] *lui*. Mais cette affirmation est invalidée par la coréférence obligatoire entre le sujet de l'infinitive et le contrôleur : celui-ci, i.e. *elle* est un pronom féminin-singulier qui ne peut référer qu'à EVA. De ce fait, le seul antécédent du sujet vide de l'infinitive est EVA, et par conséquent on peut appliquer la contrainte de non coréférence entre sujet et objets pour supprimer EVA de la liste des antécédents de la ppv *lui* : ainsi, la phrase (25) est non ambiguë. La contrainte de non coréférence entre sujet et objets peut être appliquée même si le sujet en question est effacé.

Enfin, la phrase :

(26) *Je lui ai ordonné de la tuer.*

dont le contexte gauche contient uniquement EVA et MAX, est une dernière illustration de la répercussion des contraintes de non coréférence. La synthèse de la ppv *lui* de la principale est *a priori* ambiguë, puisqu'elle réfère aussi bien à MAX qu'à EVA. Cependant, en appliquant la contrainte de non coréférence entre sujet et objets dans l'infinitive, on peut conclure que le sujet effacé ne peut référer qu'à MAX, puisque l'objet direct féminin-singulier *la* n'a qu'un antécédent possible, i.e. EVA. Etant donnée la coréférence obligatoire entre le sujet effacé de

l'infinitive et le [à-objet] *lui*, contrôleur de la complétive, cette ppv ne peut référer qu'à MAX, et n'est donc pas ambiguë.

Ces résultats, ainsi que ceux qui ont été vus dans la section 6.3.1, justifient l'algorithme de synthèse d'une complétive que nous présentons ci-dessous. L'algorithme permettra en outre de résoudre d'autres problèmes que nous évoquerons, comme la mise à jour de l'historique des étiquettes, ou la synthèse des complétives sujet.

6.4 SYNTHESE D'UNE COMPLETIVE

La synthèse des complétives sont une justification de l'approche globale, dont nous avons déjà discuté les avantages au chapitre précédent. Les propriétés que nous venons de voir, à propos de la répercussion des contraintes de non coréférence, ont pour conséquence des résultats dont l'obtention est impossible par l'approche séquentielle (où, rappelons-le, chaque constituant est synthétisé à tour de rôle). Par contre, l'approche globale permet, par une application récursive, de faire jouer les conséquences des contraintes de non coréférence à tous les niveaux de l'arbre représentant la phrase à complétive à synthétiser.

6.4.1 APPLICATION RECURSIVE DE L'APPROCHE GLOBALE

Considérons les quatre schémas de phrase, dont le contexte gauche contient les étiquettes EVA et MAX :

(S27) *(EVA volere (EVA uccidere MAX).)*
(EVA vouloir (EVA tuer MAX).)

(S28) *(PIERRE volere (MAX uccidere EVA).)*
(PIERRE vouloir (MAX tuer EVA).)

(S29) *(PIERRE credere (MAX conoscere EVA).)*
(PIERRE croire (MAX connaître EVA).)

(S30) *(PIERRE credere (PIERRE conoscere EVA).)*
(PIERRE croire (PIERRE connaître EVA).)

Ils symbolisent respectivement les quatre situations dans lesquelles on peut se retrouver avec une construction à complétive :

- la réduction à l'infinitive est obligatoire, le sujet de la complétive est coréférent au contrôleur,
- la réduction à l'infinitive est obligatoire, le sujet de la complétive est non coréférent au contrôleur,
- la réduction à l'infinitive est facultative, le sujet de la complétive est non coréférent au contrôleur,
- la réduction à l'infinitive est facultative, le sujet de la complétive est coréférent au contrôleur.

Voici tout d'abord la description détaillée des opérations effectuées pour la synthèse de (S27).

1) Calcul des listes Ui des formes (pro)nominales pour chaque étiquette, ainsi que des antécédents morpho-sémantiques de ces formes :

U[SUJET] = (Ø [EVA, MAX], *lei* [EVA], *Eva* [EVA])

Complétive	U[SUJET] = (Ø [EVA, MAX]) U[OBJET] = (*lo* [MAX], *Max* [MAX])

La liste U[SUJET] associée au sujet de la complétive est réduite à la forme Ø, étant donné que les conditions sont réunies ici pour la réduction obligatoire de la complétive à l'infinitive : le verbe de la principale est *volere*, et le sujet de la complétive est coréférent au contrôleur (EVA).

2) Construction des combinaisons sujet-objet, notées CCi, dans la complétive :

CC1 = (Ø [EVA, MAX], *lo* [MAX])
CC2 = (Ø [EVA, MAX], *Max* [MAX])

Application des règles syntaxiques :

CC1' = (Ø [EVA <cnc - sujet-objet : MAX>], *lo* [MAX])
CC2' = (Ø [EVA <cnc - sujet-objet : MAX>], *Max* [MAX])

Application des règles de préférence stylistiques, en se basant uniquement sur la valeur des objets dans chaque combinaison : on préfère la ppv *lo* au groupe nominal *Max*, donc on ne retient que la combinaison CC1'.

3) Construction des combinaisons sujet-objet, notées Cj, dans la principale :

C1 = (Ø [EVA, MAX], (Ø [EVA], *lo* [MAX]))
C2 = (*lei* [EVA], (Ø [EVA], *lo* [MAX]))
C3 = (*Eva* [EVA], (Ø [EVA], *lo* [MAX]))

Application du fait que le sujet de la principale et le sujet de la complétive dont la réduction est obligatoire sont coréférents : on ne conserve que **l'intersection** de leurs antécédents. On obtient alors :

C1' = (Ø [EVA <MAX n'appartient pas à l'intersection>], (Ø [EVA], *lo* [MAX]))
C2' = C2 = (*lei* [EVA], (Ø [EVA], *lo* [MAX]))
C3' = C3 = (*Eva* [EVA], (Ø [EVA], *lo* [MAX]))

Application des règles de préférence stylistique, qui privilégient C1', ce qui conduit à :

(27) *Vuole ucciderlo.*
([Elle] veut le tuer.)

En d'autres termes, on procède à deux séries d'opérations syntaxiques, qui montrent bien la récursivité de l'approche globale :

- On calcule les combinaisons CCi de la (ou des) complétive(s) en prenant en compte les phénomènes de réduction à l'infinitive : dans l'exemple (S27) ils concernent la réduction

obligatoire et la coréférence entre le contrôleur et le sujet de la complétive ; on ramène une ou plusieurs CCi.

- On calcule les combinaisons Cj de la principale en intégrant les CCi. Il y a réduction à l'infinitive, donc on fait intervenir le fait qu'il y a coréférence entre le contrôleur et le sujet de l'infinitive en prenant l'intersection de leurs antécédents ; on ramène une ou plusieurs Cj.

Examinons la synthèse de (S28), que nous rappelons :

(S28) *(PIERRE volere (MAX uccidere EVA).)*
(PIERRE vouloir (MAX tuer EVA).)

1) Calcul des listes Ui des formes (pro)nominales :

$U_{[SUJET]}$ = (*Pierre*[PIERRE])

Complétive | $U_{[SUJET]}$ = (Ø [EVA,MAX,PIERRE],*egli* [MAX,PIERRE],*Max* [MAX])
$U_{[OBJET]}$ = (*la* [EVA],*Eva* [EVA])

La liste $U_{[SUJET]}$ associée au sujet de la complétive contient les formes Ø, *egli* et *Max*, car les conditions présentes ici interdisent la réduction de la complétive à l'infinitive : le verbe de la principale (*volere*) impose la réduction obligatoire, mais le sujet de la complétive (MAX) et le contrôleur (PIERRE) sont non coréférents.

2) Construction des combinaisons sujet-objet, notées CCi, dans la complétive :

CC1 = (Ø [EVA, MAX, PIERRE], *la* [EVA])
CC2 = (Ø [EVA, MAX, PIERRE], *Eva* [EVA])
CC3 = (*egli* [MAX, PIERRE], *la* [EVA])
CC4 = (*egli* [MAX, PIERRE], *Eva* [EVA])
CC5 = (*Max* [MAX], *la* [EVA])
CC6 = (*Max* [MAX], *Eva* [EVA])

Application des règles syntaxiques :

CC1' = (Ø [MAX, PIERRE <cnc - sujet-objet : EVA>], *la* [EVA])
CC2' = (Ø [MAX, PIERRE <cnc - sujet-objet : EVA>], *Eva* [EVA])
CC3' = CC3 = (*egli* [MAX, PIERRE], *la* [EVA])
CC4' = CC4 = (*egli* [MAX, PIERRE], *Eva* [EVA])
CC5' = CC5 = (*Max* [MAX], *la* [EVA])
CC6' = CC6 = (*Max* [MAX], *Eva* [EVA])

Application des règles de préférence stylistique, en se basant uniquement sur la valeur des objets dans chaque combinaison : on retient donc les combinaisons CC1', CC3' et CC5', dont l'objet est représenté par la ppv *la*.

3) Construction des combinaisons sujet-objet, notées Cj, dans la principale

On intègre les combinaisons CC1', CC3' et CC5' :

C1 = (*Pierre* [PIERRE], (*Ø* [MAX, PIERRE], *la* [EVA]))
C2 = (*Pierre* [PIERRE], (*egli* [MAX, PIERRE], *la* [EVA]))
C3 = (*Pierre* [PIERRE], (*Max* [MAX], *la* [EVA]))

Application du fait que le sujet de la principale et le sujet de la complétive, dont la réduction est obligatoire, sont non coréférents : on ne conserve, dans chaque combinaison, que **le complémentaire de l'intersection** de leurs antécédents. On obtient alors :

C1' = (*Pierre* [PIERRE], (*Ø* [MAX <PIERRE appartient à l'intersection>], *la* [EVA]))
C2' = (*Pierre* [PIERRE], (*egli* [MAX <PIERRE appartient à l'intersection>], *la* [EVA]))
C3' = (*Pierre* [PIERRE], (*Max* [MAX], *la* [EVA]))

Application des règles de préférence stylistique, qui privilégient C1', ce qui conduit à :

(28) *Pierre vuole che la uccida.*
(Pierre veut qu'[il] la tue.)

Considérons maintenant la synthèse de (S29), dont le verbe principal *credere* (croire), contrairement à *volere* (vouloir), n'impose pas la réduction obligatoire à sa complétive objet :

(S29) *(PIERRE credere (MAX conoscere EVA).)*
(PIERRE croire (MAX connaître EVA).)

1) Calcul des listes Ui des formes (pro)nominales :

$U_{[SUJET]}$ = (*Pierre*[PIERRE])

Complétive	$U_{[SUJET]}$ = (*Ø* [EVA,MAX,PIERRE],*egli* [MAX,PIERRE],*Max* [MAX])
	$U_{[OBJET]}$ = (*la* [EVA], *Eva* [EVA])

La liste $U_{[SUJET]}$ associée au sujet de la complétive contient les formes *Ø*, *egli* et *Max*, car les conditions présentes ici interdisent la réduction de la complétive à l'infinitive : le verbe de la principale (*credere*) n'impose pas la réduction obligatoire, et le sujet de la complétive (MAX) et le contrôleur (PIERRE) sont non coréférents.

2) Construction des combinaisons sujet-objet, notées CCi, dans la complétive :

Les combinaisons, ainsi que les résultats des contraintes de non coréférence, sont identiques à celles obtenues dans l'exemple précédent (S28). On retient donc les combinaisons CC1', CC3' et CC5' :

CC1' = (*Ø* [MAX, PIERRE <cnc - sujet-objet : EVA>], *la* [EVA])
CC3' = CC3 = (*egli* [MAX, PIERRE], *la* [EVA])
CC5' = CC5 = (*Max* [MAX], *la* [EVA])

3) Construction des combinaisons sujet-objet, notées Cj, dans la principale

On intègre les combinaisons CC1', CC3' et CC5' :

C1 = (*Pierre* [PIERRE], (*Ø* [MAX, PIERRE], *la* [EVA]))
C2 = (*Pierre* [PIERRE], (*egli* [MAX, PIERRE], *la* [EVA]))
C3 = (*Pierre* [PIERRE], (*Max* [MAX], *la* [EVA]))

Application du fait que le sujet de la principale et le sujet de la complétive dont la réduction est facultative sont non coréférents : On ne peut pas modifier C1, où le sujet de la complétive est l'élément vide ; par contre, on ne conserve dans C2 que **le complémentaire de l'intersection** des antécédents de *Pierre* et *egli*. On obtient alors :

C1' = C1 = (*Pierre* [PIERRE], (*Ø* [MAX, PIERRE], *la* [EVA]))
C2' = (*Pierre* [PIERRE], (*egli* [MAX <PIERRE appartient à l'intersection>], *la* [EVA]))
C3' = C3 = (*Pierre* [PIERRE], (*Max* [MAX], *la* [EVA]))

Application des règles de préférence stylistique, qui privilégient C2', ce qui conduit à :

(29) *Pierre crede che egli la conosce.*
(Pierre$_i$ croit qu'il$_{k\neq i}$ la connait.)

Finalement, la synthèse du schéma de phrase (S30), qui combine réduction facultative de la complétive à l'infinitive, et coréférence entre le contrôleur et le sujet de la complétive, i.e. PIERRE, et que nous rappelons :

(S30) *(PIERRE credere (PIERRE conoscere EVA).)*
(PIERRE croire (PIERRE connaître EVA).)

conduit soit à :

(30a) *Pierre crede che la conosce.*
(Pierre croit qu'il la connait.)

soit à :

(30b) *Pierre crede di conoscerla.*
(Pierre croit la connaître.)

Dans notre système de génération, nous avons décidé d'effectuer systématiquement la réduction à l'infinitive dès qu'elle est possible, pour éviter des complications inutiles. Ainsi, nous générons (30b) et non (30a), et on peut se rapporter, pour la synthèse de (S30), à la liste des opérations nécessaires à la synthèse de (S27).

Les quatre exemples que nous venons d'étudier représentent les différents cas possibles de réduction à l'infinitive d'une complétive objet. L'approche globale appliquée récursivement permet d'obtenir des résultats concernant la pronominalisation en faisant intervenir uniquement la syntaxe, i.e. les propriétés liées à la réduction des complétives à l'infinitive. Les sections suivantes sont consacrées aux autres problèmes liés à la synthèse des complétives, et aux solutions apportées par l'approche globale.

6.4.2 GESTION DE L'HISTORIQUE DES ETIQUETTES

Chaque fois qu'une étiquette est mentionnée, on la mémorise dans une liste, ainsi que d'autres informations, comme par exemple le genre et le nombre du groupe nominal qui est synthétisé. Les exemples de la section précédente mettent en jeu des entités déjà mentionnées dans le contexte gauche. Maintenant supposons que l'on ait à synthétiser un schéma de phrase contenant la même étiquette dans la principale et la complétive, et que cette étiquette n'apparaisse pas dans le contexte gauche. Par exemple, considérons le schéma de phrase (S31) :

(S31) (UGO *annunciare* (MAX *odiare* EVA) *a* EVA.)
(UGO annoncer (MAX détester EVA) à EVA.)

où chaque étiquette est mentionnée pour la première fois et où le mécanisme de réduction de la complétive à l'infinitive n'intervient pas, puisque le sujet de la complétive, i.e. MAX, est différent du contrôleur, i.e. EVA. Le problème qui se pose ici concerne la synthèse des deux occurrences de EVA. A priori, (S31) pourrait être synthétisé en :

(31a) **Ugo le annuncia che Max odia Eva.*
(*Ugo **lui**$_{fem}$ annonce que Max déteste **Eva**.)

(31b) **Ugo annuncia che Max odia Eva a lei.*
(*Ugo annonce que Max déteste **Eva à elle**.)

(31c) **Ugo annuncia a Eva che Max odia Eva.*
(*Ugo annonce **à Eva** que Max déteste **Eva**.)

(31d) **Ugo le annuncia che Max la odia.*
(Ugo **lui**$_{fem}$ annonce que Max **la** déteste.)

(31e) *?*Ugo annuncia che Max la odia a Eva.*
(?*Ugo annonce que Max **la** déteste **à Eva**.)

(31f) *Ugo annuncia a Eva che Max la odia.*
(Ugo annonce **à Eva** que Max **la** déteste.)

Chaque phrase obtenue correspond à un ordre de synthèse différent des deux occurrences de EVA :

- (31a), qui est inacceptable du point de vue de l'interprétation, présuppose que l'occurrence de EVA dans la complétive est connue lors de la synthèse du [à-objet] ;

- (31b) est agrammaticale, et résulte du même raisonnement que (31a) ;

- (31c), qui est également agrammaticale, est obtenue en supposant que les deux occurrences de EVA sont mentionnées en même temps pour la première fois ;

- (31d), qui est sémantiquement incorrecte, résulte d'un raisonnement analogue, où la synthèse simultanée des deux occurrences de EVA impose la double pronominalisation ;

- (31e) est ambiguë ; on l'obtient en supposant que l'ordre d'apparition des occurrences de l'étiquette EVA est le suivant : [à-objet] de la principale < [dir-objet] de la complétive, et en plaçant la complétive avant le groupe nominal dans la phrase synthétisée ;

- enfin, en gardant les mêmes hypothèses quant à l'historique des occurrences de EVA, mais en inversant la position des deux objets dans la phrase synthétisée, on obtient (31f) qui est la seule version parfaitement acceptable à partir de (S31).

Cet ensemble de résultats a été largement commenté et analysé : R. Langacker (1969) introduit la notion de c-commande pour décrire formellement ces contraintes de pronominalisation ; signalons que les règles qu'il a établies acceptent la version (31e), que nous refusons. J. Ross (1969) avance, quant à lui, l'hypothèse que le pronom répond à une règle cyclique obligatoire. Ces théories, ainsi que leurs applications, sont indissociables d'un certain formalisme : pour cette raison, il est impossible de les adapter à notre système.

Les phrases (31a-f) montrent qu'il faut effectuer une gestion interne de l'historique des étiquettes, quand celles-ci sont mentionnées à plusieurs reprises et pour la première fois dans le schéma de phrase en cours de synthèse. La synthèse des étiquettes de la complétive dépend de celle des étiquettes de la principale ; cette constatation est d'une certaine manière en accord avec les principes de c-commande énoncés par les générativistes (même si, répétons-le, nous ne pouvons pas nous servir de leur résultats, faute de formalisme adéquate). La solution de l'approche globale consiste donc à ordonner le calcul des listes Ui : les Ui des étiquettes attachées à la principale sont calculées d'abord ; on remet à jour l'historique des étiquettes en ajoutant les étiquettes mentionnées, ainsi que le genre et le nombre ; puis on répète ces opérations pour les Ux des étiquettes attachées à la complétive [x-objet][3].

Ainsi, l'ensemble des opérations nécessaires à la synthèse de (S31), que nous rappellons :

(S31) (UGO *annunciare* (MAX *odiare* EVA) *a* EVA.)
(UGO annoncer (MAX détester EVA) à EVA.)

sont les suivantes :

1) Calcul des listes Ui des formes (pro)nominales :

Etiquettes de la principale :

$U_{[SUJET]}$ = (*Ugo* [UGO])
$U_{[A\text{-}OBJET]}$ = (*Eva* [EVA])

Remise à jour de la liste gérant l'historique des étiquettes de la principale : les étiquettes EVA et MAX, ainsi que leur genre et nombre respectifs y sont ajoutés.

Etiquettes de la complétive :

$U_{[SUJET]}$ = (*Max* [MAX])
$U_{[DIR\text{-}OBJET]}$= (*la* [EVA], *Eva* [EVA])

La liste $U_{[DIR\text{-}OBJET]}$ associée à l'objet direct de la complétive contient la forme pronominale *la*, à cause du contenu de la liste des étiquettes déjà mentionnées, dans laquelle apparaît EVA.

3 S'il y a plusieurs complétives en position [xi-objet] dans la phrase, l'ordre de calcul des Uxi et de la gestion des étiquettes est arbitraire, et il est égal à l'ordre d'apparition des complétives dans la phrase synthétisée.

2) Construction des combinaisons sujet-objet, notées CCi, dans la complétive :

CC1 = (*Max* [MAX], *la* [EVA])
CC2 = (*Max* [MAX], *Eva* [EVA])

Application des règles de préférence stylistiques : on ne retient que la combinaison CC1.

3) Construction des combinaisons sujet-aobjet-dirobjet, notées Cj, dans la principale :

C1 = (*Ugo*[UGO], *Eva* [EVA], (*Max* [MAX], *la* [EVA]))

On obtient finalement (31f) :

(31f) *Ugo annuncia a Eva che Max la odia.*
(Ugo annonce à Eva que Max la déteste.)

6.4.3 AUTRES CONSEQUENCES DE LA SYNTHESE DES COMPLETIVES

Dans cette dernière section, nous évoquerons deux derniers résultats liés à la synthèse des complétives, et compatibles avec l'approche globale récursive.

1) Il s'agit tout d'abord de la synthèse du verbe de la complétive : la synthèse du verbe d'une complétive dépend de celle de son contrôleur. Examinons en effet le schéma de phrase (S32) :

(S32) (TABLE *aver bisogno di* (TABLE *essere riparato*).)
(TABLE avoir besoin de (TABLE être réparé).)

L'expression verbale *aver bisogno* (avoir besoin) se construit avec une complétive en position [di-objet], dont le contrôleur est le sujet. Dans (S32), le contrôleur (TABLE) est égal au sujet de la complétive, ce qui entraîne la réduction obligatoire de celle-ci à l'infinitive, étant données les propriétés de *aver bisogno*. Rappelons (cf. chapitre 3) qu'en italien le genre est lexical et non sémantique. L'étiquette TABLE peut être synthétisée en *il tavolo*, groupe nominal masculin-singulier, ou en *la tavola*, groupe nominal féminin-singulier. Comme le verbe de l'infinitive s'accorde en genre avec son sujet, même si celui-ci est effacé :

(32a) ***Il tavolo** ha bisogno di essere (riparato + *riparata).* ou :
(32b) ***La tavola** ha bisogno di essere (riparata + *riparato) .*
(La table a besoin d'être réparée.)

il est nécessaire de connaître le genre du contrôleur afin de générer la complétive. En appliquant, avec l'approche globale, le principe de la gestion ordonnée de l'historique des étiquettes, on calcule tout d'abord la liste U_{SUJET} de l'étiquette TABLE de la principale. Puisque l'on mémorise le genre du groupe nominal qui a été ainsi synthétisé, le verbe de l'infinitive peut s'accorder en genre et nombre avec son sujet effacé, i.e. au masculin-singulier dans le cas de (32a) et au féminin-singulier dans le cas de (32b).

2) Enfin, signalons que l'approche globale récursive rend possible la synthèse des complétives sujet. La synthèse des complétives sujet soulève le problème inverse de celui posé par la synthèse des complétives objet : la synthèse des entités en position objet dans une phrase peut dépendre de la synthèse d'une complétive sujet. Examinons en effet le schéma de phrase (S33) :

(S33) ((EVA *salvare* UGO) *legare per la vita* UGO *a* EVA.)
((EVA sauver UGO) lier pour la vie UGO à EVA.)

où l'on suppose que les étiquettes UGO et EVA sont mentionnées pour la première fois. Si on applique le principe : calcul des Ui de la principale < calcul des Ui de la complétive, on obtient la phrase

(33a) *Il fatto che lo abbia salvato lega Ugo a Eva per la vita.*
(Le fait qu'[elle] l'ait sauvé lie Ugo à Eva pour la vie.)

qui est acceptée par certains locuteurs, mais dont la compréhension est moins évidente que celle de la phrase :

(33b) *Il fatto che Eva abbia salvato Ugo lo lega a lei per la vita.*
(Le fait que Eva ait sauvé Ugo le lie à elle pour la vie.)

qui est obtenue, elle, en appliquant le principe suivant :

calcul des Ui de la complétive < calcul des Ui de la principale.

6.5 CONCLUSION : ADEQUATION DE L'APPROCHE GLOBALE RECURSIVE

Le schéma ci-dessous résume les opérations qui doivent être effectuées pour synthétiser une phrase pouvant être construite avec N complétives, i.e. une complétive sujet et une ou deux complétives objet :

Calcul des Ui de la principale et des N complétives

Calcul des Ui de **la complétive sujet**
Mémorisation des étiquettes rencontrées, ainsi que leur genre et nombre.

Calcul des Ui de **la principale**
Mémorisation des étiquettes rencontrées, ainsi que leur genre et nombre.

Calcul des Ui de **la complétive objet 1**
Mémorisation des étiquettes rencontrées, ainsi que leur genre et nombre.

Calcul des Ui de **la complétive objet 2**
Mémorisation des étiquettes rencontrées, ainsi que leur genre et nombre.

Elaboration des listes de combinaisons dans les N complétives

Application des règles syntaxiques en prenant en compte les phénomènes de réduction à l'infinitive :

- coréférence entre le contrôleur et le sujet de la complétive,
- réduction obligatoire ou facultative, puis application des règles de préférence stylistiques, sauf celles qui concernent le sujet.

Elaboration des listes de combinaisons dans la principale

Intégration des listes de combinaisons sélectionnées dans l'étape précédente.

Application des règles syntaxiques qui aboutissent (entre autres) au calcul de l'intersection ou du complémentaire de l'intersection des antécédents des étiquettes sujet de la complétive et contrôleur, en fonction du mode de réduction de la complétive et de l'égalité des deux étiquettes.

Application des règles de préférence stylistiques.

L'approche globale récursive est donc une réponse satisfaisante aux problèmes complexes posés par la synthèse des complétives. Ces problèmes, qui ont été illustrés par des exemples en italien, mais dont l'illustration par des exemples français donnent des résultats tout aussi intéressants, sont les suivants :

- réduction à l'infinitive,
- antécédents syntaxiques des pronoms dans la principale et dans les complétives,
- synthèse du verbe d'une complétive réduite à l'infinitive,
- synthèse des complétives sujet.

Notons en outre que les solutions apportées par cet algorithme ne font intervenir que la syntaxe, dont on vient de montrer une autre de ses applications en génération de textes, et qui permet d'autres utilisations, exploitables dans de prochaines recherches : ainsi, on peut comparer la synthèse des complétives à celle des relatives et des groupes nominaux à complément(s), autres constituants complexes de la phrase. Comme une complétive, une relative ou un groupe nominal à complément(s) peut s'écrire sous la forme :

$$f\ (x_1, x_2 \ldots x_i \ldots x_n)$$

Les problèmes que soulèvent ces constituants ont été cités, entre autres, par R. Langacker (1969) et R. Kayne (1971) . Ils y ont apporté des solutions théoriques, en ce qui concerne la pronominalisation (cf. chapitre 3). Comme pour la pronominalisation dans les complétives, ces solutions, nécessitant un formalisme particulier, ne sont pas utilisables par notre générateur. Par contre, les propriétés de l'approche globale récursive, qui ont été prouvées avec la synthèse des complétives, doivent pouvoir être employées pour permettre, à partir de :

(S34) (MOI *parlare di* MADRE (*di* EVA) *a* EVA.)
(MOI parler de MADRE (de EVA) à EVA.)

la synthèse de :

(34a) *Parlo a Eva di sua madre.*
([Je] parle à Eva de sa mère.)

tout en évitant les phrases :

(34b) **Le parlo della madre di Eva.*
(*[Je] lui$_{fem}$ parle de la mère d'Eva.)

(34c) *?Parlo di sua madre a Eva.*
(?[Je]parle de sa mère à Eva.)

7 PROGRAMME

7.1 INTRODUCTION

Le travail que nous avons présenté dans les chapitres précédents concerne essentiellement le composant syntaxique pour un système de génération de texte. Notre effort a surtout porté sur l'étude des pronoms, dont le résultat est l'élaboration d'un algorithme de synthèse. Finalement, nous avons, au chapitre précédent, abordé la synthèse des complétives et proposé des méthodes d'intégration de ces constituants enchâssés dans notre algorithme.

Nos recherches en génération nous ont amenée également à amorcer une relation entre pronominalisation et composant stratégique, et plus généralement, entre composant stratégique et composant syntaxique. Cette constatation, tout comme l'intégration des arguments complexes, est encore à l'état d'ébauche : nous avons dégagé les principales opérations à effectuer pour obtenir un schéma de texte optimal pour la coréférence des sujets ou la réduction des complétives, comme nous allons le voir dans la section 7.3. Cependant, d'autres recherches sont en cours pour mettre au point des règles établissant des relations dans le but de faciliter la synthèse des pronoms, entre composant stratégique et composant syntaxique, tout en conservant pour chacun leur identité propre.

L'ensemble des résultats auxquels nous avons abouti à propos du composant syntaxique ont été testés dans un programme écrit en Common Lisp sur station Sun, et dont le domaine d'application consiste à résumer des opéras de Giuseppe Verdi. Dans le présent chapitre, nous exposerons (cf. section 7.3) les heuristiques employées pour l'élaboration de schémas de textes adaptés au domaine indiqué ci-dessus. Nous présenterons ensuite (section 7.4) trois textes pour lesquels nous décrirons les faits linguistiques pertinents (e.g. nous expliquerons à chaque fois que cela sera nécessaire les raisons motivant ou prohibant une pronominalisation).

Nous n'avons pas cherché à faire une étude sur la structure, le contenu ou la stylistique des résumés, ni sur ceux des opéras italiens du 19ème siècle : notre choix de domaine a pour buts d'introduire les phrases à complétives et de tester les résultats de la pronominalisation sur des textes longs. Les textes présentés sont tirés des traductions de résumés d'opéras qui ont été trouvés dans Télé 7 jours (**Don Carlos**), le Guide de l'opéra (H. Rosenthal et alii (1986)) (**I vespri siciliani**), ou reconstitués à partir de synopsis (**Un ballo in maschera** (EMI 1986)).

Nous allons commencer la présentation de notre programme en décrivant la représentation sémantique[1] que nous avons adaptée au domaine (e.g. personnages de rang noble, concepts de meurtre, d'embuscade, de trahison, ...).

1 Rappelons (cf. chapitre 2) que nos recherches se sont limitées à utiliser une représentation sémantique déjà existante : il nous a semblé hors du cadre de notre travail d'effectuer des études approfondies dans le domaine de la représentation sémantique des connaissances.

7.2 REPRESENTATION SEMANTIQUE NECESSAIRE AU DOMAINE

Parmi les opéras que G. Verdi a écrit, nous nous sommes intéressée à ceux qui partagent un certain nombre de thèmes, ce pour des raisons de lexique ainsi que pour des raisons de représentation des connaissances. On se limite donc à un type donné d'opéras dont les concepts principaux sont :

- l'amour, l'amitié
- la jalousie, la confiance
- la traîtrise
- le meurtre
- la tyrannie d'un dictateur sur un peuple opprimé
- la machination (contre un tyran, un rival)

Ainsi, nous avons généré une version des résumés de **Don Carlos**, **I vespri siciliani** et **Un ballo in maschera**. Les opéras dont l'intrigue repose sur une guerre ont été écartés pour des raisons de lexique (**Aida** ou **Nabucco**). De même, les opéras à intrigue trop complexe (**Ernani** ou **La Forza del destino**) ont été évités. Enfin, nous n'avons pris en compte que les opéras dont les protagonistes étaient définis par des rôles constants et des relations relativement simples : ainsi, nous ne nous sommes pas attaquée au résumé de **Il trovatore**, où le héros Manrico croit être le fils d'une bohémienne et découvre finalement qu'il est le frère d'un Comte. Un tel changement dans les rôles et les relations demanderait d'intégrer dans la représentation sémantique des phénomènes relevant de la logique non monotone, ce qui est hors du cadre de notre travail.

La description de la représentation sémantique nécessaire à la génération de résumés d'opéras que nous avons retenus nous a conduit à distinguer :

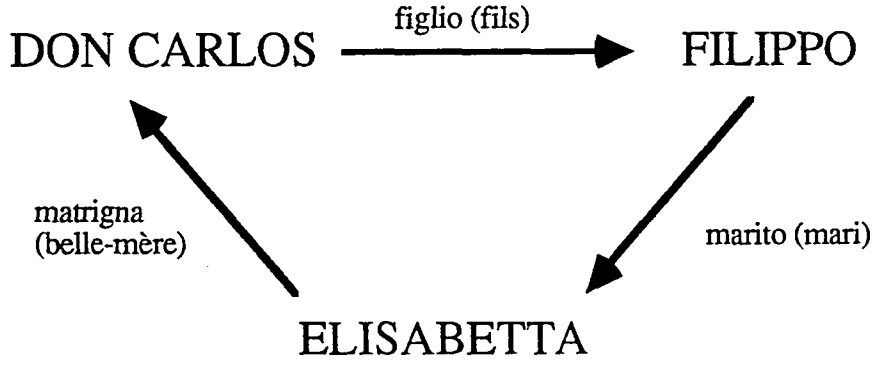
- la représentation sémantique du récit,
- la représentation sémantique des personnages (et autres entités) indispensables au récit.

Les entités présentes dans les opéras de Verdi qui nous intéressent sont les personnages et les choeurs, ainsi que certains endroits et objets clés. Les personnages sont caractérisés par leur nom, leur sexe, leur(s) rôle(s) politique et leur(s) relation(s) de parenté ou d'amitié avec les autres personnages. Notons que ces relations sont unilatérales : par exemple, l'étiquette ELENA de **I vespri siciliani** est définie comme étant *la sorella di* FEDERICO (la soeur de FEDERICO) :

ELENA := **PERSONNAGE**
NAME = "Elena"
ROLEI = "la sorella" (FEDERICO)
REDESCRIPTION = "la ragazza"
SEX = FEM

mais l'étiquette FEDERICO n'est pas définie comme étant *il fratello di* ELENA (le frère de ELENA), ce qui entraînerait la redondance d'une information déjà liée à ELENA et

éventuellement des boucles infinies. De même, le trio DON-CARLOS, FILIPPO, ELISABETTA (de **Don Carlos**) est défini de telle façon que chacun ait une relation a sens unique avec l'un des autres personnages, suivant le schéma :



DON-CARLOS := **PERSONNAGE**
NAME = "Don Carlos"
ROLE1 = "l' Infante di Spagna"
ROLE2 = "il figlio" (FILIPPO)
SEX = MASC

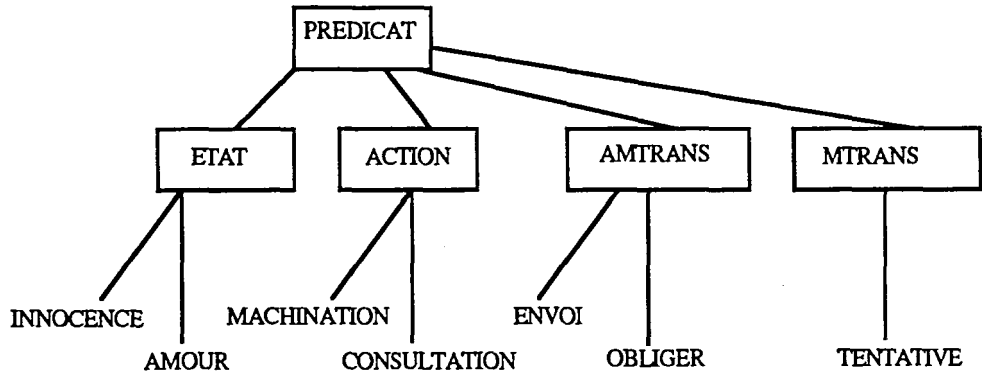
FILIPPO := **PERSONNAGE**
NAME = "Filippo II"
ROLE1 = "il re di Spagna"
ROLE2 = "il marito"(ELISABETTA)
SEX = MASC

ELISABETTA := **PERSONNAGE**
NAME = "Elisabetta"
ROLE1 = "la principessa di Valois"
ROLE2 = "la matrigna" (DON-CARLOS)
REDESCRIPTION = "la regina"
SEX = FEM

Dans le texte synthétisé, un personnage est décrit la première fois par son nom, son titre et son lien de parenté avec les autres personnages. Ensuite, lorsque la redescription est nécessaire, on utilisera à tour de rôle le nom et le titre, à moins que le champ REDESCRIPTION soit non vide, ce qui est le cas pour ELISABETTA et ELENA. Dans ce cas, on alterne le nom et la redescription du personnage.

La représentation sémantique du récit est constituée d'une suite d'évènements qui en forment les étapes principales. Cette suite d'évènements est fournie au générateur par l'utilisateur au moyen d'un menu, qui lui demande dans quel ordre les différents thèmes proposés apparaissent dans l'histoire qu'il invente. Nous ne décrirons pas ce menu dont la raison d'être n'est que de faciliter l'entrée des données nécessaires au générateur.

Chaque évènement E_i est dérivé à partir de la classe PREDICAT (cf. chapitre 2). La figure ci-dessous donne un extrait de la hiérarchie utilisée :



Les sous-classes ETAT, ACTION, MTRANS et AMTRANS sont définies ainsi :

- ETAT : un concept de la sous-classe ETAT fait intervenir un AGENT humain (E22) et éventuellement, un OBJET, humain lui aussi (E1) :

E22 := **ETAT**
CONCEPT = INNOCENCE
AGENT = DON-CARLOS

E1 := **ETAT**
CONCEPT = AMOUR
AGENT = DON-CARLOS
OBJET = ELISABETTA

L'ETAT est soit une situation psychologique que subit l'AGENT, soit la manifestation d'un sentiment de sa part.

- ACTION : dans cette classe, l'agent agit consciemment, contrairement à son comportement dans la classe ETAT. Un concept de type ACTION peut avoir un (E18) ou deux champs du type OBJET, i.e. OBJET et OBJET1 (E33), humains ou non :

E18 := **ACTION**
CONCEPT = CONSULTATION
AGENT = AMELIA
OBJET = ULRICA

E23 := **ACTION**
CONCEPT = ENVOI
AGENT = RODRIGO
OBJET = LETTRE
OBJET1 = FILIPPO

- MTRANS : cette classe regroupe les concepts où l'AGENT (humain) est en relation avec un PROCES, qui est instancié par un évènement :

E17 := **MTRANS**
CONCEPT = TENTATIVE
AGENT = RODRIGO
PROCES = E18

- AMTRANS : cette classe, contient les concepts dont l'AGENT induit l'objet à se mettre en relation avec un PROCES. L'AGENT et l'OBJET sont humains ; le champ PROCES est instancié par un prédicat :

E7 := **AMTRANS**
CONCEPT = OBLIGER
AGENT = FILIPPO
OBJET = DON-CARLOS
PROCES = E8

Nous rappelons (cf. chapitre 2) que les termes MTRANS et AMTRANS sont empruntés à R. Schank (1975).

Les concepts peuvent inclure des informations sur le lieu, le temps ou la manière qui sont enregistrées respectivement sous les champs LIEU, TEMPS, MANIERE, comme illustré dans l'exemple ci-dessous :

E7 := **ACTION**
CONCEPT = CONSULTATION
AGENT = AMELIA
OBJET = ULRICA
MANIERE = "di nascosto"

Ils peuvent aussi inclure des informations sur la cause, le but, le moyen ou la conséquence. Les champs BUT, CAUSE, CONSEQUENCE et MOYEN ont pour valeur une instance de la classe prédicat :

E9 := **ACTION**
CONCEPT = MACHINATION
AGENT = EBOLI
CAUSE = E10
BUT = E12

E20 := **MTRANS**
CONCEPT = TENTATIVE
AGENT = RODRIGO
PROCES = E21
MOYEN = E23

Les évènements E_i principaux, i.e. ceux qui n'occupent ni un champ PROCES, ni un des champs subordonné à un autre évènement E_j, sont dits **indépendants**. Ils constituent les étapes successives du récit, qui est lui-même représenté par le symbole *opera*. Un exemple de représentation sémantique d'un récit est donné ci-dessous avec un extrait de la représentation de **Don Carlos**, qui décrit les quatre premiers évènements indépendants :

opera = (E1 E2 E6 E7 ...)

E1 := **ETAT**
CONCEPT = AMOUR
AGENT = DON-CARLOS
OBJET = ELISABETTA

E2 := **AMTRANS**
CONCEPT = INCITER
AGENT = RODRIGO
OBJET = DON-CARLOS
PROCES = E3

E3 := **ACTION**
CONCEPT = DEPLACEMENT-ACTION
AGENT = DON-CARLOS
OBJET = FIANDRE
BUT = E4

E4 := **ACTION**
CONCEPT = DEFENDRE
AGENT = DON-CARLOS
OBJET = FIAMMINGHI
CAUSE = E5

E5 := **ACTION**
CONCEPT = OPRESSION
AGENT = FILIPPO
OBJET = FIAMMINGHI

E6 := **AMTRANS**
CONCEPT = ENCOURAGER
AGENT = ELISABETTA
OBJET = DON-CARLOS
PROCES = E3
BUT = E6BIS

E6BIS := **MTRANS**
CONCEPT = OUBLI
AGENT = DON-CARLOS
PROCES = E1

E7 := **AMTRANS**
CONCEPT = OBLIGER
AGENT = FILIPPO
OBJET = DON-CARLOS
PROCES = E8

E8 := **MTRANS**
CONCEPT = RENONCEMENT
AGENT = DON-CARLOS
PROCES = E3

7.3 OBTENTION D'UN SCHEMA DE TEXTE

La formation d'un schéma de texte est le résultat d'un certain nombre d'**opérations** qui sont effectuées dans un certain **ordre** sur des **données** propres au domaine. En d'autres termes, le schéma de texte est obtenu en choisissant pour chaque concept une phrase simple, et en articulant (par juxtaposition, coordination, subordination ou relativation) les phrases simples entre elles. Plus précisément, les données sont réunies :

1) en un lexique,

2) en une grammaire de discours.

Le lexique contient toutes les phrases simples pouvant exprimer un concept donné. Ces phrases diffèrent soit par leur matériel lexical, soit par des transformations syntaxiques. Ainsi, au concept AMOUR sont associées les phrases simples suivantes :

N0 amare N1	(N0 aimer N1)
N1 essere amato da N0	(N1 être aimé par N0)
N0 essere innamorato di N1	(N0 être amoureux de N1)
l'amore di N0 per N1	(l'amour de N0 pour N1)
N0 voler bene a N1	(N0 aimer $N1_{hum}$)

De même, le concept AMITIE correspond aux phrases simples :

N0 essere l'amico di N1	(N0 être l'ami de N1)
N1 ottenere l'amicizia di N0	(N1 obtenir l'amitié de N0)
l'amicizia di N0 per N1	(l'amitié de N0 pour N1)
N0 essere legato da amicizia a N1	(N0 être lié d'amitié à N1)

La grammaire de discours établit l'ensemble des combinaisons autorisées du point de vue de la syntaxe et de la sémantique, pour articuler entre elles les phrases simples choisies pour chaque concept.

Le choix d'une phrase simple, ainsi que celui de l'articulation de deux phrases simples est le résultat des opérations suivantes :

- recherche du matériel lexical exprimant un concept ayant la précision sémantique la plus grande,

- calcul de la transformation syntaxique la plus élégante pour les phrases à complétives,

- calcul des caractéristiques de la subordonnée (exprime-t-elle la cause, le but ...) dans le but d'effectuer le choix le plus adéquate du mode d'articulation de deux phrases,

- établissement des conditions les plus favorables à la pronominalisation entre deux schémas de phrases juxtaposés : recherche de constructions parallèles, ou à sujets coréférents.

Chacune des opérations décrites ci-dessus peut entrer en conflit avec les autres. En d'autres termes, les décisions concernant l'organisation du texte, celles concernant l'articulation de la cause, le but, ... autour d'une "principale", et celles concernant le choix d'une phrase simple dépendent les unes des autres. Etant donnée la longueur des textes que nous générons, nous avons privilégié **le texte**, suivi de la phrase, i.e. de la construction d'un schéma de phrase à partir de l'un des concepts qui constituent les étapes du récit ; nous emploierons le terme d'**évènement indépendant**. Enfin, les décisions laissées pour la fin concernent ce qu'on appellera l'**évènement de base**, c'est à dire le choix de la phrase simple et, éventuellement, de la (les) complétive(s) qui en dépend(ent).

Donc, le composant stratégique, qui intègre les deux bases de données linguistiques (i.e. le lexique et la grammaire de discours) aboutit à un schéma de texte à partir d'une réflexion à trois niveaux : le texte, l'évènement indépendant et l'évènement de base.

- La représentation sémantique d'un **texte** est constitué d'une suite chronologique d'évènements indépendants. Les questions que l'on doit résoudre concernent donc la construction syntaxique de chaque évènement, et la façon dont ils sont ordonnés et articulés les uns aux autres.

- La représentation sémantique d'un **évènement indépendant** E_i est formée d'un **noyau** exprimant la $i^{ème}$ étape du récit, complétée d'un certain nombre d'évènements satellites représentant la cause, la conséquence, le but, etc ... de E_i. La grammaire de discours et le lexique sont utilisés, ici encore, pour déterminer le nombre, l'ordre des phrases qui constituent l'évènement principal E_i et la construction syntaxique de chacune d'elle.

- Enfin, pour un **évènement de base** (qui peut contenir un champ PROCES, et donc un autre évènement de base) il faut effectuer les choix lexicaux (en fonction du concept) et les transformations syntaxiques éventuelles.

Nous avons montré au chapitre 2, que les décisions, à chaque niveau, dépendent les unes des autres. Nous allons donc, pour chaque division présentée ci-dessus, décrire nos décisions, expliquer les raisons pour lesquelles elles ont été prises et l'ordre dans lequel ces choix ont été effectués.

7.3.1 TEXTE

L'ordre des informations contenues dans un texte dépend de la nature de ce texte (e.g. récit policier, résumé d'histoire). Ainsi, L. Pemberton (1989) indique l'ordre dans lequel doivent apparaître les informations relatives à un récit policier : l'ordre chronologique, qui place le meurtre en première position, ne peut être adopté, à moins qu'on ne veuille faire savoir au lecteur dès le départ qui est le criminel. Dans le domaine des résumés d'opéras qui nous intéressent, nous conserverons l'ordre chronologique, c'est à dire l'ordre dans lequel l'auteur de l'opéra a présenté les différentes phases de son oeuvre.

Tels qu'ils sont conçus dans nos représentations sémantiques, les évènements d'un récit sont indépendants. Leur mode d'articulation devrait donc être la juxtaposition ou la coordination. Les conditions de coordination de deux phrases ont été décrites par L. Danlos et

appliquées pour réaliser son générateur. Nous avons décidé, dans cette première version de notre système où l'accent a été mis surtout sur le composant syntaxique, de ne pas nous occuper des contraintes de coordination dans un texte long : nous nous contentons donc de relier les différents évènements du texte par juxtaposition.

Les opérations à effectuer au niveau du texte ont trois motivations :

1) Elles visent à obtenir des schémas de phrase dont les principales sont soit parallèles, soit à sujets coréférents : on essaie d'obtenir des couples de structures ayant le même sujet ou le même [x-objet]. Ainsi, examinons les représentations sémantiques E_{23} et E_{24}, extraites de **Un ballo in maschera** :

E23 := **ACTION**	E24 := **ACTION**
CONCEPT = DEPLACEMENT-ACTION	**CONCEPT** = RENCONTRE
AGENT = AMELIA	**AGENT** = RICCARDO
OBJET = POTENCE	**OBJET** = AMELIA
	OBJET1 = POTENCE

Le schéma de phrase associé à E_{23} a pour sujet AMELIA. On a le choix, pour le concept RENCONTRE, entre les phrases simples :

	RICCARDO incontrare AMELIA	(RICCARDO rencontrer AMELIA)
et	*AMELIA incontrare RICCARDO*	(AMELIA rencontrer RICCARDO)

en tenant compte de la propriété de symétrie du verbe *incontrare*. Pour obtenir une structure dont le sujet soit coréférent à celui du schéma de phrase précédent, le composant stratégique choisit la deuxième phrase simple. Les schémas de phrase juxtaposés exprimant E_{23} et E_{24} sont alors les suivants :

```
(:ps (V2 "recarsi")
        (:sujet AMELIA)
        (:verbe "recarsi")
        (:loc-objet POTENCE :prep "a"))
 "."
(:ps (VAZIO2 "incontrare")
        (:sujet AMELIA)
        (:verbe "incontrare")
        (:objet RICCARDO)
        (:loc-objet POTENCE))
```

2) Elles permettent de pallier à l'absence de l'agent : le composant stratégique choisit une phrase simple qui est la transformation passive de l'un des items lexicaux exprimant le concept concerné, comme le montre le passage de la représentation sémantique E_{27} de **Don Carlos** :

E27 := **ACTION**
CONCEPT = LIBERATION
OBJET = DON-CARLOS

en schéma de phrase :

```
(:ps (VAZIO2 "liberare")
        (:sujet DON-CARLOS)
        (:verbe "liberare" :voix :passif))
```

3) Enfin, s'il n'y a pas contre-indication avec les points 1) et 2), les opérations sur le texte privilégient les constructions à l'actif pour des raisons stylistiques de simplicité.

En conclusion, l'ordre de priorité choisi pour la construction des schémas de phrases principales sacrifie le choix lexical (ou plus précisément, l'acuité sémantique de l'item lexical) au profit de l'organisation générale du texte.

7.3.2 EVENEMENT INDEPENDANT

L'évènement est constitué d'un **noyau** E_i, qui sera synthétisé en "principale" et dont les transformations viennent d'être décrites ci-dessus, et d'un ensemble de prédicats représentant la cause, la conséquence, le but, ... de E_i. Le choix du nombre de phrases pour décrire les différentes composantes d'un évènement est une décision stylistique. Nous avons choisi d'articuler ces différentes composantes dans une unique phrase complexe pour deux raisons :

- préserver la cohésion des différentes parties d'un évènement indépendant,
- assurer une meilleure précision sémantique dans une relation entre deux concepts, ce que l'on obtient de façon plus satisfaisante par relativation ou subordination que par juxtaposition.

L'inconvénient de cette solution est le risque d'obtenir des phrases lourdes, si elles contiennent trop de relatives ou de subordonnées. C'est pourquoi le nombre de ces prédicats enchâssés est limité au moment même de l'élaboration de la représentation sémantique de l'évènement indépendant.

Le mode d'articulation des schémas de phrases exprimant le moyen, la conséquence, la cause et le but au schéma de phrase principal dépend de la nature (CAUSE, BUT, MOYEN, ...) de ces concepts ainsi que de la valeur de leurs champ AGENT, OBJET et OBJET1.

7.3.2.1 MOYEN

On exprime le moyen par une phrase subordonnée, dont le verbe est au participe présent. Le sujet (effacé) de ce verbe et celui de la principale doivent être coréférents.

> *Rodrigo$_i$ tenta di convincere Filippo$_j$ dell'innocenza di Don Carlos$_k$ Ø$_i$ spedendo al re una lettera.*
> (Rodrigo$_i$ tente de convaincre Philippe$_j$ de l'innocence de Don Carlos$_k$ en Ø$_i$ envoyant au roi une lettre.)

Un prédicat exprimant le moyen est donc placé par le composant stratégique en position subordonnée gérondive du prédicat exprimant la principale si les champs AGENT des deux concepts ont la même valeur, comme dans cet extrait de la représentation sémantique de **Don Carlos** :

E20 := **MTRANS**	E23 := **ACTION**
CONCEPT = TENTATIVE	**CONCEPT** = ENVOI
AGENT = RODRIGO	**AGENT** = RODRIGO
PROCES = E21	**OBJET** = LETTRE
MOYEN = E23	**OBJET1** = FILIPPO

qui conduit au schéma de texte :

```
(:ps (V56 "tentare")
        (:sujet RODRIGO)
        (:verbe "tentare")
        (:objet
                (:ps (V50 "convincere")
                        (:sujet RODRIGO)
                        (:verbe "convincere")
                        (:objet FILIPPO )
                        (:di-objet "l'innocenza" (DON-CARLOS)))
        (:sub (:conj-sub "")
                (:ps (VAZIO2 "spedire")
                        (:sujet RODRIGO)
                        (:verbe "spedire" :mode :participe-present)
                        (:objet LETTRE))))
```

Le moyen est également représenté par une subordonnée au gérondif si la valeur de l'AGENT du concept exprimant la principale est égale à la valeur de l'OBJET du concept exprimant le moyen. Dans ce cas, la formation d'une subordonnée-gérondive dépend de la possibilité, pour le concept MOYEN, de s'exprimer par une phrase simple résultant, par exemple, de la transformation passive, qui place le contenu de son champ OBJET en position sujet : nous dirons pour simplifier que le concept est passivable.

7.3.2.2 CONSEQUENCE

La conséquence s'exprime au moyen d'une relative qui est introduite par le pronom relatif sujet et qui est attachée à l'une des entités de la principale : cela suppose l'égalité entre le contenu de l'un des champs AGENT, OBJET, OBJET1 du concept exprimant la principale, et :

- soit le contenu du champ AGENT du concept exprimant la conséquence,
- soit le contenu du champ OBJET de ce concept, qui doit dans ce cas être passivable.

Par exemple, examinons les représentations sémantiques suivantes, extraites de **I Vespri Siciliani** :

E16 := **ACTION**	E17 := **ACTION**
CONCEPT = GRACE	**CONCEPT** = LIBERATION
AGENT = MONTFORT	**OBJET** = CONGIURATI
OBJET = CONGIURATI	
CONSEQUENCE = E17	

Parmi les phrases simples associées au concept LIBERATION :

(a)	*N0 liberare N1*	(N0 libérer N1)
(b)	*N1 essere liberato (da N1 + Ø)*	(N1 être libéré (par N1 + Ø))

(c) *la libertà di N1* (la liberté de N1)

...

le composant stratégique choisit la forme (b), ce qui permet d'obtenir le schéma de phrase :

```
(:PS (VAZIO1 "accordare la grazia a")
        (:sujet MONTFORT)
        (:verbe "accordare")
        (:objet "la grazia")
        (:a-objet CONGIURATI
                :relative
                (:ps (VAZIO2 "liberare")
                        (:sujet CONGIURATI)
                        (:verbe "liberare" :voix :passif))))
```

qui est synthétisé en :

> *" ... Montfort accorda la grazia ai congiurati che sono liberati ... "*
> (" ... Montfort accorde la grâce aux conjurés qui sont libérés ...")

7.3.2.3 CAUSE

Une phrase exprimant la cause peut être rattachée à la principale soit par relativation, soit par subordination. La construction en relative, qui est plus élégante, doit être introduite par le pronom relatif sujet (L. Danlos (1985)). Ainsi, la phrase :

> *Don Carlos prende le difese dei Fiamminghi, che sono oppressi da Filippo II.*
> (Don Carlos prend la défense de Flamands, qui sont opprimés par Philippe II.)

exprime une relation causale de façon beaucoup plus marquée que :

> *Don Carlos prende le difese dei Fiamminghi, che Filippo II opprime.*
> (Don Carlos prend la défense de Flamands, que Philippe II opprime.)

et dans un style plus élégant que :

> *Don Carlos prende le difese dei Fiamminghi, perchè (Filippo II li opprime + sono oppressi da Filippo II).*
> (Don Carlos prend la défense de Flamands, parce que (Philippe II les opprime + ils sont opprimés par Philippe II).)

Le composant stratégique essaie donc tout d'abord de représenter par une relative le prédicat exprimant la cause, en effectuant les mêmes opérations que celles décrites en 7.3.2.2 pour la conséquence. Par exemple, la représentation sémantique :

E4 := **ACTION**
CONCEPT = DEFENDRE
AGENT = DON-CARLOS
OBJET = FIAMMINGHI
CAUSE = E5

E5 := **ACTION**
CONCEPT = OPRESSION
AGENT = FILIPPO
OBJET = FIAMMINGHI

conduit au schéma de phrase :

```
(:ps (VAZIO1 "prendere le difese di")
        (:sujet DON-CARLOS)
        (:verbe "prendere")
        (:objet "le difese"
                (FIAMMINGHI :relative
                        (:ps (VAZIO2 "opprimere")
                                (:sujet FIAMMINGHI)
                                (:verbe "opprimere" :voix :passif)
                                (:prep-objet :prep "da" FILIPPO)))))
```

grâce au choix de la phrase simple exprimant la transformation passive du verbe *opprimere*. Ce schéma de phrase peut alors être synthétisé en :

> *Don Carlos prende le difese dei Fiamminghi, che sono oppressi da Filippo II.*
> (Don Carlos prend la défense de Flamands, qui sont opprimés par Philippe II.)

Quand le prédicat exprimant la CAUSE ne peut pas être représenté par une relative, le composant stratégique produit une subordonnée à l'indicatif introduite par *perchè* (parce que).

7.3.2.4 BUT

Un prédicat exprimant le but est une subordonnée au subjonctif introduite par *perchè* (pour que) ou *affinchè* (afin que) :

> *Elisabetta incoraggia Don Carlos a compiere questa impresa affinchè dimentichi il proprio amore per lei.*
> (Elisabeth encourage Don Carlos à mener cette entreprise afin qu'[il] oublie son [litt : propre] amour pour elle.)

Cette subordonnée peut être réduite à l'infinitive, quand son sujet est égal au sujet de la principale. Quand il en a la possibilité, le composant stratégique choisit la phrase simple exprimant le but qui permet une telle réduction. A titre d'exemple, examinons la représentation sémantique extraite de **I vespri siciliani** :

```
E5 := ACTION                    E6 := AMTRANS
CONCEPT = ALLIANCE              CONCEPT = VENGEANCE
AGENT = ELENA                   AGENT = ELENA
OBJET = PROCIDA                 OBJET = FEDERICO
BUT = E6
```

Elle donne lieu au schéma de phrase :

```
(:ps (VAZIO2 "allearsi")
        (:sujet ELENA)
        (:verbe "allearsi")
        (:a-objet PROCIDA)
        (:sub (:conj-sub "perche'")
                (:ps (V50 "vendicare")
                        (:sujet ELENA)
                        (:verbe"vendicare")
                        (:objet FEDERICO))))
```

qui est synthétisé en :

Elena si allea a Procida per vendicare Federico.
(Elena s'allie à Procida pour venger Federico.)

7.3.3 EVENEMENT DE BASE

L'évènement de base E_i est constitué d'un concept (que nous noterons "principale" pour simplifier) reliant des entités et/ou un prédicat (le PROCES). L'ensemble des opérations nécessaires au passage de la représentation sémantique de l'évènement de base au schéma de phrase dépend de contraintes externes à l'évènement (ainsi, la transformation passive peut dépendre de l'évènement précédent si E_i est un évènement indépendant, cf. section 7.3.1, ou de sa nature, si E_i exprime le but, la conséquence, la cause ou le moyen d'un autre évènement, cf. section 7.3.2). Il dépend aussi d'un certain nombre de décisions prises dans l'ordre suivant :

1) La précision sémantique

La recherche d'une précision sémantique maximale est le choix que l'on fait en priorité. Ainsi si le concept AMOUR est mentionné pour la première fois, et concerne un OBJET humain, comme dans E_1 (extrait de **Don Carlos**) :

E1 := **ETAT**
CONCEPT = AMOUR
AGENT = DON-CARLOS
OBJET = ELISABETTA

alors il sera exprimé par la phrase simple

AGENT essere innamorato di OBJET
(AGENT être amoureux de OBJET)

plutôt que par :

AGENT amare OBJET
(AGENT aimer OBJET)

ce qui permet d'obtenir :

```
(:ps (EADN "innamorato")
     (:sujet DON-CARLOS)
     (:verbe "essere")
     (:attribut "innamorato")
     (:di-objet ELISABETTA :prep "di"))
```

Il est logique de considérer ce choix comme prioritaire puisqu'il revient à faire du concept AMOUR plusieurs sous-concepts sémantiquement plus précis.

2) La construction syntaxique de la principale

Le choix, parmi un même matériel lexical, d'une construction syntaxique donnée (e.g. actif, passif avec agent ...) doit tenir compte du PROCES (qui sera synthétisé en complétive) car on essaie de respecter deux contraintes stylistiques :

- éviter les complétives sujet qui alourdissent la phrase,
- faire en sorte de réduire les complétives à l'infinitive ou en groupes nominaux prédicatifs, pour rendre la phrase plus fluide.

Pour éviter la complétive sujet, on choisit pour la principale une phrase simple acceptant une complétive en position objet. Ainsi la représentation sémantique E_{10} de **Don Carlos** :

E10 := **MTRANS**
CONCEPT = DECEPTION
PROCES = E11
OBJET = EBOLI

dont le concept peut s'exprimer soit par la phrase simple :

Ch F deludere N1 (Qu Ph décevoir N1),

soit par :

N1 essere deluso dal fatto ch F (N1 être déçu par le fait qu Ph)

et dont le champ PROCES vaut :

E11 := **ACTION**
CONCEPT = REPOUSSER
AGENT = DON-CARLOS
OBJET = EBOLI

est donc représenté par le schéma de phrase :

```
(:ps (V41 "deludere")
        (:sujet EBOLI )
        (:verbe "deludere":voix :passif)
        (:prep-objet :prep "da"
                :completive
                ... (représentation de E11) ... )
```

3) La construction syntaxique du PROCES

CAS 1 : Le PROCES est l'instance du champ AGENT du concept principal, et la seule phrase simple exprimant cette principale n'est pas passivable : le composant stratégique recherche alors pour le concept exprimant le PROCES une construction nominale prédicative. Par exemple, examinons la représentation sémantique :

E := AMTRANS
CONCEPT = CONVENIR
PROCES = E'
OBJET = PAUL

Le concept CONVENIR s'exprime par la phrase simple :

Il fatto ch F conviene a N1 (Le fait que Ph convient à N1)

Il n'y a pas de phrase simple passivable pour exprimer CONVENIR. Par contre, le lexique associe au concept RETOUR les phrases simples suivantes :

N0 ritorna (Ø + prep LOC)	(N0 revient (Ø + prép LOC))
N0 far ritorno (Ø + prep LOC)	(?N0 faire un retour (Ø + prép LOC))
il ritorno di N0 (Ø + prep LOC)	(le retour de N0 (Ø + prép LOC))

parmi lesquelles le composant stratégique choisit le groupe nominal prédicatif *il ritorno di N0 (Ø + prep LOC)*, plutôt que l'une des constructions verbales, afin d'obtenir :

Il ritorno di N0 conviene a Paul.
(Le retour de N0 convient à Paul.)

jugé plus élégant que :

Il fatto che N0 (ritorni + faccia ritorno) conviene a Paul.
(Le fait que N0 (revienne + ?fasse un retour) convient à Paul.)

CAS 2 : Le PROCES occupe la position objet du concept principal : le choix de la phrase simple qui l'exprime dépend alors :

- de la classe (ETAT, ACTION ...) à laquelle le PROCES appartient,
- de la position syntaxique qu'occupe le PROCES dans le schéma de phrase principal.

En effet, si le PROCES appartient à la classe ETAT et qu'il doit occuper la position [di-objet], [da-objet] ou [dir-objet], alors le composant stratégique choisit en priorité, pour ce concept, une construction nominale prédicative. Cette décision est d'ordre stylistique. Par exemple, on préfère synthétiser :

(1) *Rodrigo convince Filippo dell'innocenza di Don Carlos.*
(Rodrigue convainc Philippe de l'innocence de Don Carlos.)

plutôt que :

(1') *Rodrigo convince Filippo che Don Carlos è innocente.*
(Rodrigue convainc Philippe que Don Carlos est innocent.)

Ainsi, à partir de :

E21 := **AMTRANS**
CONCEPT = PERSUADER
AGENT = RODRIGO
OBJET = FILIPPO
PROCES = E22

E22 := **ETAT**
CONCEPT = INNOCENCE
AGENT = DON-CARLOS

le composant stratégique produit le schéma de phrase :

```
(:ps (V50 "convincere")
        (:sujet RODRIGO)
        (:verbe "convincere")
        (:objet FILIPPO )
        (:di-objet "l'innocenza"(DON-CARLOS))
```

synthétisé en (1). De même, la représentation sémantique :

E18 := **AMTRANS**
CONCEPT = COMMUNIQUER
AGENT = RODRIGO
OBJET = FILIPPO
PROCES = E19

E19 := **ETAT**
CONCEPT = AMOUR
AGENT = RODRIGO
OBJET = "la liberta'"

conduit au schéma de phrase :

```
(:ps (V47 "comunicare")
        (:sujet RODRIGO)
        (:verbe "comunicare")
        (:objet"l'amore per la liberta'" (RODRIGO))
        (:a-objet FILIPPO))
```

où le concept AMOUR est représenté par le groupe nominal prédicatif *l'amore di N0 per N1* (l'amour de N0 pour N1), ce qui permet la synthèse de :

(2) *Rodrigo comunica a Filippo il proprio amore per la libertà.*
(Rodrigue communique à Philippe son [propre] amour pour la liberté.)

jugé meilleur que :

(2') *?Rodrigo comunica a Filippo che ama la libertà.*
(?Rodrigue communique à Philippe qu'il aime la liberté.)

ou :

(2'') *?Rodrigo comunica a Filippo di amare la libertà.*
(?Rodrigue communique à Philippe d'aimer la liberté.)

CAS 3 : On ne peut pas, ou on n'est pas obligé de choisir un groupe nominal prédicatif pour exprimer un concept PROCES : le composant stratégique choisit alors dans la mesure du possible une phrase simple permettant la réduction de la complétive à l'infinitive. Par exemple, examinons le couple de représentations sémantiques (E_{10},E_{11}) :

E10 := **MTRANS**	E11 := **ACTION**
CONCEPT = DECEPTION	**CONCEPT** = REPOUSSER
PROCES = E11	**AGENT** = DON-CARLOS
OBJET = EBOLI	**OBJET** = EBOLI

nous avons vu que la principale (E_{10}) est mise au passif pour éviter d'avoir une complétive en position sujet. La phrase simple exprimant E_{10} est construite autour du verbe *deludere* (décevoir). Ce verbe appartient à la table 41 : quand la phrase est au passif le contrôleur de la complétive est le sujet. Or celui-ci est représenté ici par EBOLI :

```
(:ps (V41 "deludere")
        (:sujet EBOLI)
        (:verbe "deludere" :voix :passif)
        (:prep-objet :prep "da"
        :completive
        ... (représentation de E11) ... )
```

Considérons maintenant E_{11} dans lequel le champ OBJET a pour valeur l'étiquette EBOLI. Le concept REPOUSSER peut être représenté entre autres par :

	(a)	*N0 respingere N1*	(N0 repousser N1)
ou par	(b)	*N1 essere respinto da N0*	(N1 être repoussé par N0)

Le choix de la phrase simple (b) permet d'avoir l'égalité entre le contrôleur de la complétive et le sujet de celle-ci (i.e. N1=EBOLI). On obtient alors le schéma de phrase :

```
(:ps (V41 "deludere")
        (:sujet EBOLI )
        (:verbe "deludere" :voix :passif)
        (:prep-objet :prep "da"
         :completive
                (:ps (VAZIO2 "respingere")
                        (:sujet EBOLI)
                        (:verbe "respingere" :voix :passif)
                        (:prep-objet DON-CARLOS :prep "da"))))
```

qui permettra la synthèse d'une infinitive :

> " ... *La Principessa d'Eboli$_i$ è delusa di Ø$_i$ essere respinta da Don Carlos* ..."
> (" ... La Princesse d'Eboli$_i$ est déçue d' Ø$_i$ être repoussée par Don Carlos ...")

Ainsi le composant stratégique effectue des opérations en vue de la réalisation de certains phénomènes apparemment réservés à l'action du composant syntaxique, comme la réduction des complétives à l'infinitive, ou (cf. section 7.3.1) l'effacement des sujets coréférents et la pronominalisation des objets identiques dans les structures parallèles. Voyons un dernier exemple à propos de la réduction des complétive à l'infinitive. Considérons les représentations sémantiques :

E15 := **MTRANS**
CONCEPT = REUSSIR
AGENT = RODRIGO
PROCES = E16

E16 := **ETAT**
CONCEPT = AMITIE
AGENT = FILIPPO
OBJET = RODRIGO

Le concept REUSSIR de E_{15} est représenté par la phrase simple *N0 riuscire a che F* (N0 réussir à ce que Ph), avec pour contrôleur (N0) RODRIGO. Quant au concept AMITIE, on a vu qu'il était associé, entre autres, aux phrases simples suivantes :

N0 essere l'amico di N1	(N0 être l'ami de N1)
N1 ottenere l'amicizia di N0	(N1 obtenir l'amitié de N0)

Le composant stratégique choisit la deuxième phrase simple, à cause de la valeur du sujet (N1 = RODRIGO), qui permet la réduction à l'infinitive de la complétive. Le schéma de phrase obtenu est alors :

```
(:ps (V44 "riuscire")
        (:sujet RODRIGO)
        (:verbe "riuscire")
        (:a-objet :prep "a"
        :completive
                (:ps (VSTATO1 "ottenere l amicizia di")
                        (:sujet RODRIGO)
                        (:verbe "ottenere")
                        (:objet"l'amicizia" (FILIPPO)))))
```

qui est synthétisée en :

" ... *Rodrigo riesce ad ottenere l'amicizia del re* ... "
(... Rodrigue réussit à obtenir l'amitié du roi ...)

7.3.4 CONCLUSION : EXEMPLE DE SCHEMA DE TEXTE

Pour conclure sur la formation du schéma de texte, nous présentons ci-dessous, à partir de l'extrait de la représentation sémantique de **Don Carlos** proposé à la fin de la section 7.2 :

opera = (E1 E2 E6 E7 ...)

E1 := **ETAT**
CONCEPT = AMOUR
AGENT = DON-CARLOS
OBJET = ELISABETTA

E2 := **AMTRANS**
CONCEPT = INCITER
AGENT = RODRIGO
OBJET = DON-CARLOS
PROCES = E3

E3 := **ACTION**
CONCEPT = DEPLACEMENT-ACTION
AGENT = DON-CARLOS
OBJET = FIANDRE
BUT = E4

E4 := **ACTION**
CONCEPT = DEFENDRE
AGENT = DON-CARLOS
OBJET = FIAMMINGHI
CAUSE = E5

E5 := **ACTION**
CONCEPT = OPRESSION
AGENT = FILIPPO
OBJET = FIAMMINGHI

E6 := **AMTRANS**
CONCEPT = ENCOURAGER
AGENT = ELISABETTA
OBJET = DON-CARLOS
PROCES = E3
BUT = E6BIS

E6BIS := **MTRANS**
CONCEPT = OUBLI
AGENT = DON-CARLOS
PROCES = E1

E7 := **AMTRANS**
CONCEPT = OBLIGER
AGENT = FILIPPO
OBJET = DON-CARLOS
PROCES = E8

E8 := **MTRANS**
CONCEPT = RENONCEMENT
AGENT = DON-CARLOS
PROCES = E3

le schéma de texte correspondant :

```
(E1)  ((:ps (EADN "innamorato")
            (:sujet  DON-CARLOS)
            (:verbe "essere")
            (:attribut "innamorato")
            (:di-objet ELISABETTA :prep "di"))
       ".")
(E2)  (:ps (V49 "incitare")
            (:sujet  RODRIGO)
            (:verbe "incitare")
            (:objet DON-CARLOS)
            (:a-objet
            :completive
                (:ps (V2 "recarsi")
                    (:sujet DON-CARLOS)
                    (:verbe "recarsi")
                    (:loc-objet "nelle Fiandre")
                    (:sub (:conj-sub "perche'")
                        (:ps (VAZIO1 "prendere le difese di")
                            (:sujet DON-CARLOS)
                            (:verbe"prendere")
                            (:objet "le difese" (FIAMMINGHI)
                                :relative
                                (:ps (VAZIO2 "opprimere")
                                    (:sujet FIAMMINGHI)
                                    (:verbe "opprimere" :voix :passif)
                                    (:prep-objet :prep "da" FILIPPO))))))))
       ".")
(E6)  (:ps (V49 "incoraggiare")
            (:sujet ELISABETTA)
            (:verbe "incoraggiare")
            (:objet DON-CARLOS)
            (:a-objet
             :completive
                (:ps (VAZIO2 "compiere")
                    (:sujet DON-CARLOS)
                    (:verbe "compiere")
                    (:objet  "questa impresa")))
            (:sub (:conj-sub "affinche'")
                (:ps (V43 "dimenticare")
                    (:sujet DON-CARLOS)
                    (:verbe "dimenticare")
                    (:objet "l'amore"(DON-CARLOS))
                    (:adverbe ELISABETTA :prep "per"))))
       ".")
```

```
(E7)  (:ps (V49 "costringere")
           (:sujet FILIPPO)
           (:verbe "costringere")
           (:objet DON-CARLOS)
           (:a-objet
            :completive
               (:ps (V44 "rinunciare")
                    (:sujet DON-CARLOS)
                    (:verbe"rinunciare")
                    (:a-objet "questo viaggio"))))
      )
```

7.4 EXEMPLES DE TEXTES GENERES

Dans cette dernière section, nous présentons trois textes obtenus à partir des schémas de texte dont l'élaboration, décrite dans la section ci-dessus, est due à l'algorithme qui a été présenté dans les chapitres 3,4,5 et 6. Pour chaque texte présenté, nous expliquerons les raisons pour lesquelles une forme pronominale est synthétisée ou au contraire est manquante, à chaque fois que ce résultat est pertinent, et correspond à l'un des points soulevés dans les chapitres précédents.

Rappelons qu'en règle générale, notre choix a été d'éviter la génération de pronoms dès qu'il y avait ambiguïté. Ce choix peut être contestable, puisque les récits d'origine présentent parfois des ambiguïtés : elles sont souvent gommées par les connaissances pragmatiques du lecteur. Dans notre système, qui ne fait pas intervenir de relations pragmatiques, nous avons jugé que le choix d'éviter les ambiguïtés était meilleur que le choix inverse, qui aurait eu pour conséquence des textes absolument incompréhensibles.

7.4.1 DON CARLOS

"Don Carlos, Infante di Spagna e figlio di Filippo II, re di Spagna, e' innamorato della propria matrigna Elisabetta, un tempo principessa di Valois. Rodrigo, Marchese di Posa e amico dell' Infante, incita Don Carlos a recarsi nelle Fiandre per prendere le difese dei Fiamminghi, che sono oppressi da Filippo II. Elisabetta lo incoraggia a compiere questa impresa affinche' dimentichi il proprio amore per lei. Il re lo costringe a rinunciare a questo viaggio. La Principessa D'Eboli, che e' delusa dal fatto di essere respinta da Don Carlos, ordisce una trama in modo che Filippo II sia persuaso che egli e' l' amante della regina. L' Infante e' condannato a morte. Rodrigo riesce a ottenere l' amicizia del re. Cerca di comunicargli il proprio amore per la liberta'. Tenta di convincerlo dell' innocenza di Don Carlos spedendo a Filippo II una lettera in cui il Marchese si accusa di amare Elisabetta. Il re lo fa ammazzare per ordine del grande Inquisitore. L' Infante e' liberato. Va a fare gli addii alla regina quando Filippo II sopraggiunge. Don Carlos scappa mentre Carlo Quinto, l' Imperatore defunto, protegge la fuga di lui."

(Don Carlos, Infant d'Espagne et fils de Philippe II, roi d'Espagne, est amoureux de sa propre belle-mère Elisabeth, jadis princesse de Valois. Rodrigo, Marquis de Posa et ami de l'Infant, incite Don Carlos à se rendre dans les Flandres pour prendre la défense des Flamands, qui sont opprimés par Philippe II. Elisabeth l'encourage à mener cette entreprise afin qu'il oublie son [propre] amour pour elle. Le roi l'oblige à renoncer à ce voyage. La princesse d'Eboli, qui est déçue d'être repoussée par Don Carlos, ourdit une machination de façon à ce que Philippe II soit persuadé qu'il$_{\text{Don Carlos}}$ est l'amant de la reine. L'Infant est condamné à mort. Rodrigo réussit à gagner l'amitié du roi. Il essaie de lui communiquer son [propre] amour pour la liberté. Il tente de le convaincre de l'innocence de Don Carlos en envoyant à Philippe II une lettre dans laquelle le Marquis s'accuse d'aimer Elisabeth. Le roi le fait tuer sur l'ordre du grand Inquisiteur. L'Infant est libéré. Il va faire [litt : les] ses adieux à la reine quand Philippe II

surgit. Don Carlos s'enfuit pendant que Charles Quint, l'Empereur défunt, protège la fuite de celui-ci)

Le texte d'origine, tiré de Télé-7-Jours, est le suivant :

Don Carlos, Infant d'Espagne, est amoureux de sa belle-mère Elisabeth. Le marquis de Posa l'incite à se rendre dans les Flandres et à prendre la défense des Flamands persécutés par l'inquisition instaurée par Philippe II, son père. La reine elle-même pousse le jeune homme à mener cette entreprise afin qu'il oublie son amour pour elle. Philippe II le contraint à renoncer à cette expédition. La princesse d'Eboli, déçue d'avoir été repoussée par Don Carlos, complote une sombre machination afin de persuader Philippe que son fils est l'amant d'Elisabeth. Le marquis de Posa parvient à s'attirer l'amitié de Philippe. Il essaie de lui communiquer sa flamme pour son idéal de liberté et tente de le convaincre de l'innocence de son fils en lui envoyant une lettre où il se déclare lui-même coupable d'aimer Elisabeth. Sur les instances du Grand Inquisiteur, Philippe fait abattre Posa. Libéré, Carlos va faire ses adieux à Elisabeth lorsque son père surgit. Carlos s'échappe, protégé dans sa fuite par l'ombre de l'empereur défunt Charles Quint.

Phrase 1

> "Don Carlos, Infante di Spagna e figlio di Filippo II, re di Spagna, e' innamorato della **propria** matrigna Elisabetta, un tempo principessa di Valois."
> (Don Carlos, Infant d'Espagne et fils de Philippe II, roi d'Espagne, est amoureux de sa propre belle-mère Elisabeth, jadis princesse de Valois.)

Utilisation du possessif *proprio* à la place de *suo* dans le groupe nominal *della propria matrigna... :*

Les trois formes pronominales possibles pour un complément du nom sont :

- le possessif :

  suo, sua, suoi, sue, loro
  (son, sa, ses_{masc}, ses_{fem}, leur(s))

- le possessif propre :

  proprio, propria, propri, proprie
  ($propre_{masc}$, $propre_{fem}$,$propres_{masc}$, $propres_{fem}$)

- la forme tonique :

  di lui, di lei, di loro
  (de celui-ci, de celle-ci, de (ceux-ci + celles-ci))

Le possessif propre s'emploie pour référer au sujet :

> *Max_i ha parlato a Ugo_k della $propria_i$ madre.*
> (Max_i a parlé à Ugo_k de sa $propre_i$ mère.)

La forme tonique ne peut pas, par contre, référer au sujet de la phrase :

> *Max_i ha parlato a Ugo_k della madre di lui_k.*

(Max$_i$ a parlé à Ugo$_k$ de la mère de celui-ci$_k$.)

Quant à *suo,* il s'utilise quelle que soit la position syntaxique de son antécédent :

Max$_i$ ha parlato a Ugo$_k$ di sua$_{i+k}$ madre.
(Max$_i$ a parlé à Ugo$_k$ de sa$_{i+k}$ mère.)

Ainsi, pour éviter toute ambiguïté, le possessif synthétisé dans la phrase 1 est *propria* et non *sua* qui pourrait référer à DON-CARLOS ou à FILIPPO :

"Don Carlos, Infante di Spagna e figlio di Filippo II, re di Spagna, e' innamorato della sua matrigna Elisabetta, un tempo principessa di Valois."
(Don Carlos, Infant d'Espagne et fils de Philippe II, roi d'Espagne, est amoureux de sa belle-mère Elisabeth, jadis princesse de Valois.)

Phrases 2 et 3

[Rodrigo, Marchese di Posa e amico dell' Infante, incita Don Carlos a recarsi nelle Fiandre per prendere le difese dei Fiamminghi, che sono oppressi da Filippo II]
"Elisabetta lo incoraggia a compiere questa impresa affinche' Ø dimentichi il proprio amore per lei."
([Rodrigo, Marquis de Posa et ami de l'Infant, incite Don Carlos à se rendre dans les Flandres pour prendre la défense des Flamands, qui sont opprimés par Philippe II.]
Elisabeth l'encourage à mener cette entreprise afin qu'[il] oublie son [propre] amour pour elle.)

Parallélisme des principales des phrases 2 et 3 :

Les verbes *incitare* (inciter) et *incoraggiare* (encourager) appartiennent à la même table (i.e. la table 49) ; donc les phrases 2 et 3 ont des structures parallèles. Les arguments non phrastiques de *incitare* sont synthétisés en groupes nominaux et les objets directs des deux phrases représentent la même entité, i.e. DON-CARLOS. C'est pourquoi le [dir-objet] de la phrase 3 peut être pronominalisé : il ne peut référer qu'à DON-CARLOS, objet direct de la phrase précédente.

Effacement du sujet dans la circonstancielle de but de la phrase 3 :

"... affinchè Ø dimentichi il proprio amore per lei."
(... afin qu'[il] oublie son [litt : propre] amour pour elle.)

La contrainte de non coréférence entre le sujet et le [prep-objet] a pour conséquence que l'antécédent du sujet effacé est différent de l'antécédent de *lei*, i.e. de ELISABETTA. Il s'agit donc de l'un des autres humains présents dans le contexte : DON-CARLOS, FILIPPO ou RODRIGO. FILIPPO et RODRIGO sont trop éloignés, donc le sujet effacé n'a qu'un antécédent possible (le plus probable), DON-CARLOS, et n'est donc pas ambigu.

Phrase 4

"Il re lo costringe a rinunciare a questo viaggio."
(Le roi l'oblige à renoncer à ce voyage.)

Parallélisme des principales des phrases 3 et 4 :

La situation est la même qu'entre les principales des phrases 2 et 3 : le pronom *lo* de la phrase 4 ne peut référer qu'à DON-CARLOS. L'impression de coréférence entre les objets directs est accru du fait que le [dir-objet] de la principale de la phrase 3 est pronominalisé.

Phrase 5

> "La Principessa D'Eboli, che e' delusa dal fatto di essere respinta da Don Carlos, ordisce una trama in modo che Filippo II sia persuaso che **egli** e' l' amante della regina."
> (La Princesse d'Eboli, qui est déçue du fait d'être repoussée par Don Carlos, ourdit une machination de façon à ce que Philippe II soit persuadé qu'il est l'amant de la reine.)

Antécédent du pronom *egli* :

Dans la circonstancielle exprimant le but :

> " ... *in modo che Filippo II sia persuaso che egli è l'amante della regina* ..."
> (... de façon à ce que Philippe II soit persuadé qu'il est l'amant de la reine ...)

FILIPPO est contrôleur de la complétive. La réduction de cette complétive à l'infinitive est facultative. Aussi, si le sujet de celle-ci est vide, il peut être coréférent à FILIPPO ; par contre cette étiquette est un antécédent interdit pour un pronom personnel sujet : les antécédents de *egli* sont alors DON-CARLOS et RODRIGO. Comme la précédente occurrence de RODRIGO est très éloignée, surtout par rapport à celle de DON-CARLOS, DON-CARLOS est le candidat le plus probable : le pronom *egli* est donc non ambigu.

Phrases 7 et 8

> [Rodrigo riesce a ottenere l' amicizia del re.]
> "**Ø** Cerca di comunicar**gli** il **proprio** amore per la liberta'."
> ([Rodrigo réussit à obtenir l'amitié du roi.]
> [Il] essaie de lui communiquer son [propre] amour pour la liberté.)

Coréférence des sujets des phrases 7 et 8 :

La possibilité de coréférence des sujets est, d'après la discussion au chapitre 5, réservée aux couples de sujets où le premier est pronominalisé ou effacé. Dans le cas présent les verbes se construisent avec une seule entité, i.e. le sujet. Ils sont tous les deux définis par :

N0 V (prep) Ch F.

Dans ces conditions, les sujets sont coréférents même si le sujet de la phrase 7 n'est pas pronominalisé : le sujet de la deuxième phrase est effacé puisque les deux étiquettes sujets sont égales à RODRIGO.

Conséquence pour la phrase 8 : contrainte de non coréférence

Le sujet effacé du verbe *comunicare* (communiquer) dans l'infinitive [dir-objet] de *cercare* (essayer) ne peut avoir qu'un seul antécédent, i.e. RODRIGO. Par application de la contrainte de non coréférence entre le sujet et le [a-objet], la seule interprétation possible pour la ppv *gli* est donc l'autre humain masculin-singulier présent dans le contexte "proche", i.e. FILIPPO.

Possessif :

Etant donné que la sémantique de *comunicare* évoque le passage d'une information détenue par N0 vers N1, on s'attend à la phrase :

> " ... *comunicargli il* ***suo*** *amore per la libertà* ..."
> (... communiquer son amour pour la liberté ...)

où l'antécédent de *suo* est le sujet effacé. Mais comme les connaissances pragmatiques n'ont pas été implémentées, *suo* reste *a priori* ambigu, ce qui explique le choix du possessif propre :

> " ... *comunicargli il* ***proprio*** *amore per la libertà* ..."
> (... communiquer son [propre] amour pour la liberté ...)

Phrase 9

> "**Ø** Tenta di convincer**lo** dell' innocenza di Don Carlos spedendo a **Filippo II** una lettera in cui **il Marchese** si accusa di amare Elisabetta."
> ([Il] tente de le convaincre de l'innocence de Don Carlos en envoyant à Philippe II une lettre dans laquelle le Marquis s'accuse d'aimer Elisabeth.)

Parallélisme entre les principales des phrases 8 et 9 :

Puisque les verbes *cercare* et *tentare* appartiennent à la même table (i.e. la table 56), et que leurs étiquettes sujets sont les mêmes (i.e. RODRIGO), alors le sujet de *tentare* est effacé. Il en résulte la même conséquence que dans la phrase 8 : par application de la contrainte de non coréférence entre le sujet et le [dir-objet], FILIPPO est pronominalisé en ppv [dir-objet] *lo*.

Synthèse des entités dans la subordonnée de moyen :

Dans la subordonnée au gérondif, on aurait pu observer :

- une pronominalisation qui n'est pas effectuée : le [a-objet] FILIPPO en ppv *gli*,
- l'effacement du sujet RODRIGO dans la relative, qui n'a pas lieu (le sujet est en effet synthétisé en un groupe nominal, i.e. *il Marchese*),

ce qui correspond à la phrase suivante :

> " ... *spedendo****gli*** *una lettera in cui Ø si accusa di amare Elisabetta* ..."
> (... en lui envoyant une lettre dans laquelle [il] s'accuse d'aimer Elisabeth ...)

L'interdiction de pronominaliser FILIPPO vient de ce que l'un des antécédents possibles de la ppv est DON-CARLOS, complément du nom *innocenza*. Quand à la redescription du sujet de *accusarsi* (s'accuser), elle est due au fait que la coréférence implicite du sujet effacé à

RODRIGO est d'ordre pragmatique (la relative décrit en effet le contenu d'une lettre qu'il envoie : c'est donc probablement lui qui l'a écrite).

Phrase 10

"Il re **lo** fa ammazzare per ordine del grande Inquisitore."
(Le roi le fait tuer sur l'ordre du grand Inquisiteur.)

Pronominalisation du [dir-objet] :

Elle est rendue possible par :

- la contrainte de non coréférence : l'antécédent de *lo* est soit RODRIGO, soit DON-CARLOS,
- la préférence pour l'antécédent le plus probable, i.e. RODIGRO. En effet, DON-CARLOS apparaît une seule fois, en position de complément de nom ; RODRIGO est mentionné deux fois, comme argument de verbe, et ses occurrences sont toutes les deux plus proches du pronom envisagé que celle de DON-CARLOS.

Phrases 12 et 13

[Va a fare gli addii alla regina quando Filippo II sopraggiunge.]
" **Don Carlos** scappa mentre Carlo Quinto, l'Imperatore defunto, protegge la fuga **di lui.**"
([[Il] va faire ses adieux à la reine quand Philippe II surgit.]
Don Carlos s'échappe pendant que Charles Quint, l'Empereur défunt, protège la fuite de celui-ci.)

Coréférence des sujets :

En comparant les phrases 12 et 13, on pourrait s'attendre à l'effacement du sujet de la principale de la phrase 13, étant donné que le sujet de la phrase 12 est effacé (*"....Ø va a fare ..."*) et que les étiquettes sujet des phrases 12 et 13 sont les mêmes, i.e. DON-CARLOS. Mais on obtiendrait alors un texte ambigu :

*$Ø_i$ *va a fare gli addii alla regina quando Filippo* II_k *sopraggiunge.* $Ø_{i+k}$ *scappa mentre Carlo Quinto, l'Imperatore defunto, protegge la fuga di lui."*
(*$[Il]_i$ va faire ses adieux à la reine quand Philippe II_k surgit. $[Il]_{i+k}$ s'échappe pendant que Charles Quint, l'Empereur défunt, protège la fuite de celui-ci.)

Il faut en effet tenir compte de la présence du sujet de la circonstancielle de temps, i.e. FILIPPO, qui empêche l'application de la règle de coréférence des sujets : le sujet de la phrase 13 est donc synthétisé en un groupe nominal.

Synthèse d'un possessif "tonique" :

Dans la subordonnée, on s'attend à la synthèse du complément du nom du [dir-objet] en possessif *sua* :

"Don Carlos$_i$ scappa mentre Carlo Quinto$_k$,, l'Imperatore defunto protegge la ***sua$_i$*** *fuga.""*
(Don Carlos$_i$ s'échappe pendant que Charles Quint$_k$, l'Empereur défunt, protège **sa$_i$** fuite.)

Cependant l'unicité de l'interprétation du possessif *sua* est due à des raisons pragmatiques (on parle de **la fuite de N0** pour un N0 qui s'**échappe** de quelque part, ce qui est le cas pour Don Carlos, mais pas pour Charles Quint) car dans une structure analogue où la sémantique n'intervient pas, l'antécédent du possessif *suo* est ambigu :

"Don Carlos$_i$ scappa mentre Carlo Quinto$_k$, l'Imperatore defunto protegge la ***sua$_{i+k}$*** *amica.""*
(Don Carlos$_i$ s'échappe pendant que Charles Quint$_k$, l'Empereur défunt, protège **son$_{i+k}$** amie.)

En l'absence de relation pragmatique, la synthèse du possessif tonique *di lui*, obligatoirement différent du sujet CARLO5, est le seul moyen d'éviter l'ambiguïté.

7.4.2 UN BALLO IN MASCHERA

"Riccardo, Conte di Warwick e Governatore di Boston, e' innamorato di Amelia, la moglie di Renato, consigliere fidato. Degli avversari politici ordiscono una trama per ammazzare Riccardo. Renato cerca di avvisarlo. Amelia, che vuole dimenticare il suo amore per il Conte, va a consultare Ulrica, una veggente, di nascosto. Ulrica le ordina di cogliere un' erba magica ai piedi della forca, a mezzanotte, per salvarsi da questa passione. Riccardo, che sente la conversazione, decide di seguirla. Finge di interrogare la fattucchiera sul proprio avvenire. Lei gli predice che il consigliere lo uccidera'. Questi rifiuta di crederle. Amelia si reca ai piedi della forca. Vi incontra il Conte. Confessa di amarlo. Renato, che e' persuaso che Riccardo e' l' amante di Amelia, decide di allearsi agli avversari politici per ucciderlo. Amelia tenta di avvertire il Conte del pericolo. Il consigliere lo ammazza, durante un ballo in maschera."

(Riccardo, Comte de Warwick et gouverneur de Boston, est amoureux d'Amelia, la femme de Renato, conseiller de confiance. Des adversaires politiques ourdissent une machination pour tuer Riccardo. Renato essaie de le prévenir. Amelia, qui veut oublier son amour pour le Comte, va consulter Ulrica, une voyante, en cachette. Ulrica lui ordonne de cueillir une herbe magique au pied de la potence, à minuit, pour échapper à cette passion. Riccardo, qui entend la conversation, décide de la suivre. Il feint d'interroger la sorcière sur son propre avenir. Elle lui prédit que le conseiller le tuera. [litt : Celui-ci] Il refuse de la croire. Amelia se rend au pied de la potence. Elle y rencontre le Comte. Elle confesse [litt : de l'aimer] qu'elle l'aime. Renato, qui est persuadé que Riccardo est l'amant d'Amelia, décide de s'allier aux adversaires politiques pour le tuer. Amelia tente d'avertir le Comte du danger. Le conseiller le tue, pendant un bal masqué.)

Phrases 7 et 8

[Finge di interrogare la fattucchiera sul proprio avvenire.]
"**Lei** gli predice che il consigliere lo uccidera'."
([Il feint d'interroger la sorcière sur son propre avenir.]
Elle lui prédit que le conseiller le tuera.)

Présence du pronom sujet dans la principale de la phrase 8 :

La présence du pronom personnel *lei* semble superflue puisque (par la contrainte de non coréférence) le sujet effacé ne peut être coréférent qu'à ULRICA. Cependant, la phrase 7 a un

sujet effacé : étant donnée la règle de coréférence des sujets, le sujet effacé dans la phrase 8 doit avoir le même antécédent que le sujet effacé de la phrase 7. Ce n'est pas le cas puisque l'étiquette sujet de la phrase 7 est RICCARDO. Donc, il faut conserver le pronom sujet *lei* dans la principale de la phrase 8.

Phrase 9

"**Questi** rifiuta di crederle."
(Celui-ci refuse de la croire.)

Synthèse d'un démonstratif en position sujet :

Le démonstratif *questi* réfère à la dernière entité humaine masculin-singulier mentionnée : l'objet direct de la phrase 8 pronominalisé en *lo* , i.e. RICCARDO. La synthèse d'un pronom personnel ou d'un sujet effacé aurait donné lieu à un texte ambigu :

*"...*Lei gli predice che il consigliere$_i$ lo$_k$ ucciderà. (Ø + egli)$_{i+j}$ rifiuta ..."*
(" ... *Elle lui prédit que le conseiller$_i$ le$_k$ tuera. ([Il] + Il)$_{i+j}$ refuse ...")

Phrases 11 et 12

[Vi incontra il Conte.]
"Ø confessa di amarlo."
([[Elle] y rencontre le Comte.]
[Elle] confesse [litt : de l'aimer] qu'elle l'aime.)

Effacement du sujet de la phrase 12 :

L'effacement du pronom sujet *lei* est explicable par la règle de coréférence des sujets. Cependant, on peut aussi y parvenir en appliquant la répercussion sur le contrôleur de la contrainte de non coréférence entre sujet et objet direct dans l'infinitive : puisque *lo* ne peut référer qu'à RICCARDO, par application de la contrainte de non coréférence, le sujet effacé du verbe à l'infinitif *amare* n'a qu'un antécédent (en raison de la proximité) i.e. AMELIA. Etant donné que ce sujet et le contrôleur de la principale (i.e. le sujet de *confessare*) sont obligatoirement coréférents, ce dernier n'a également qu'un antécédent possible, i.e. AMELIA : le sujet de *confessare* (confesser) peut être effacé.

7.4.3 I VESPRI SICILIANI

Voici un dernier exemple de résumé produit par le générateur. Les commentaires sur les formes pronominales synthétisées sont ici inutiles car ils répèteraient ceux des deux sections précédentes.

"Arrigo, giovane patriota e , senza saperlo, figlio di Montfort, Governatore di Sicilia, e' innamorato di Elena, la sorella di Federico, Duca d'Austria. Procida, un cospiratore Siciliano, ordisce una trama per ammazzare Montfort, che opprime la Sicilia. Elena si allea a Procida per vendicare Federico, che e' stato ucciso dal Governatore. Questi scopre di essere il padre di Arrigo. Rivela la verita' al giovane che fa cessare la congiura. La ragazza e il cospiratore sono condannati a morte. Arrigo si reca in prigione per giustificarsi con Elena che gli concede il perdono. Montfort accorda la grazia ai congiurati che sono liberati. Autorizza il giovane a sposare la ragazza. Procida decide di dare inizio a un' insurrezione nel momento in cui le campane delle nozze cominciano a suonare. Elena, che scopre il complotto, vuole rinunciare al matrimonio. Il Governatore rifiuta questa situazione. Fa suonare i vespri."

(Arrigo, jeune patriote et, sans le savoir, fils de Montfort, gouverneur de Sicile, est amoureux de Elena, la soeur de Federico, Duc d'Autriche. Procida, un conspirateur sicilien, ourdit une machination pour tuer Montfort, qui oppresse la Sicile. Elena s'allie à Procida pour venger Federico, qui a été tué par le gouverneur. Celui-ci découvre [litt : d'être] qu'il est le père de Arrigo. Il révèle la vérité au jeune homme qui fait cesser la conjuration. La jeune fille et le conspirateur sont condamnés à mort. Arrigo se rend en prison pour se justifier auprès de Elena qui lui concède son pardon. Montfort accorde la grâce aux conjurés qui sont libérés. Il autorise le jeune homme à épouser la jeune fille. Procida décide de commencer l'insurrection au moment où les cloches des noces commencent à sonner. Elena, qui découvre le complot, veut renoncer au mariage. Le gouverneur refuse cette situation. Il fait sonner les vêpres.)

La version générée en italien est inspirée du texte ci-dessous, tiré de H. Rosenthal et alii (1986) :

Le gouverneur français Montfort, qui tient la Sicile opprimée, a fait exécuter le Duc Frédéric d'Autriche, dont la soeur Hélène aime le jeune patriote Arrigo. Le patriote Procida, revenant d'exil, arme la conjuration qui doit tuer Montfort pendant un bal ; mais Montfort, qui a reconnu dans Arrigo son propre fils, lui révèle la vérité. Celui-ci refuse les honneurs qui lui sont offerts mais arrête le bras des conjurés qui allaient frapper son père. Hélène et Procida sont en prison attendant la condamnation à mort. Arrigo est plein de remords , il vient se justifier auprès d'Hélène qui l'aime encore malgré sa trahison. Montfort pardonne aux conjurés et ordonne les noces d'Hélène et Arrigo. Procida arme de nouveau la révolte, dont les cloches sonneront le ralliement. Hélène apprend que ses noces donnent le signal du massacre des français et veut renoncer à la main d'Arrigo. Montfort passe outre et donne l'ordre de sonner les vêpres.[2]

7.5 CONCLUSION

Nous venons de présenter un programme de génération de textes qui produit des résumés d'opéras en Italien, à partir de représentations sémantiques pouvant faire intervenir jusqu'à 60 concepts. Ce programme a été écrit en Kyoto Common Lisp sur SUN 3-160, il est constitué de 8320 lignes de code. Pour générer **Don Carlos**, qui comporte 13 phrases complexes, le temps CPU nécessaire est de 6 minutes.

Ce programme a permis en outre d'adapter à l'italien le générateur de L. Danlos, qui produit des récits d'attentats en français. En effet, nous avons pu coupler facilement son composant stratégique à notre composant syntaxique, étant donnée la similitude des grammaires de discours pour les deux langues. C'est ainsi qu'à partir de la représentation sémantique :

story := **ATTENTAT**
CIBLE = GANDHI
ACTEUR = HUM1
ACTE = ACT3
RESULTAT = RES1
TEMPS = "ieri"
LIEU = "New Dehli"

ACT3 := **COUP-DE-FEU**
ACTEUR = HUM1
MVMT-CIBLE = PTRANS3
CIBLE = GANDHI
LIEU-ATTENTAT = LOC3

RES1 := **MORT**
OBJET = GANDHI

PTRANS3 := **PTRANS**
ACTEUR = GANDHI
OBJET = GANDHI
DEST = LOC2
PROV = LOC3
LOCOMOTION = A_PIED

2 Dans l'esprit de Verdi, mais pas forcément dans celui du lecteur, les vêpres sont assimilées aux cloches des noces, d'où leur importance, le titre de l'oeuvre, et cette dernière phrase apparemment sans rapport avec le reste ...

où :

GANDHI := **CIBLE-CONNUE**
NAME = "Indira Gandhi"
ROLE = "il primo ministro indiano"
SEX = FEM

HUM1 := **TERRORISTE**
NB = 2
DEF-TYPE = INDEFINI
LEANINGS = "estremisti Sikh"
REDESCRIPTION = "terroristi"

LOC2 := **LIEU**
LIEU = "ufficio"

LOC3 := **LIEU**
LIEU = "casa sua"

nous engendrons le texte :

> *Due estremisti Sikh hanno assassinato Indira Gandhi , ieri a New Dehli, sparando al primo ministro indiano mentre questa partiva da casa sua, a piedi, per recarsi in ufficio.*
> (Deux extrémistes Sikhs ont assassiné Indira Gandhi, hier à New Dehli, en tirant sur le premier ministre indien alors que celle-ci partait de chez elle, à pied, pour se rendre à son bureau.)

Ces résultats confirment la robustesse du composant syntaxique que nous avons élaboré. Nous donnons dans ce qui suit encore quelques exemples de récits d'attentats générés :

Dei membri dell' IRA hanno ferito Margaret Thatcher, oggi a Londra. Hanno aperto il fuoco sulla macchina in cui il primo ministro Inglese arrivava in ristorante, venendo dal ministero degli affari esteri.
(Des membres de l'IRA ont blessé Margaret Thatcher, aujourd'hui à Londres. Ils ont ouvert le feu sur la voiture dans laquelle le premier ministre anglais arrivait au restaurant, en venant du ministère des affaires étrangères.)

Margaret Thatcher è rimasta ferita, oggi a Londra. Dei membri dell'IRA hanno aperto il fuoco sulla macchina in cui il primo ministro Inglese arrivava in ristorante, venendo dal ministero degli affari esteri.
(Margaret Thatcher a été blessée, aujourd'hui à Londres. Des membres de l'IRA ont ouvert le feu sur la voiture dans laquelle le premier ministre anglais arrivait au restaurant, en venant du ministère des affaires étrangères.)

Jacques Chirac e François Léotard sono stati feriti, giovedì, da due ecologisti che hanno fatto esplodere una bomba telecomandata nell'aereo in cui l'ex primo ministro francese e l'ex ministro della cultura arrivavano a Nizza, venendo da Parigi. La bomba conteneva un chilo di dinamite.
(Jacques Chirac et François Léotard ont été blessés, jeudi, par deux écologistes qui ont fait exploser une bombe télécommandée dans l'avion dans lequel l'ex-premier ministre français et l'ex-ministre de la culture arrivaient à Nice, en venant de Paris. La bombe contenait un kilo de dynamite.)

CONCLUSION GENERALE

Il a toujours été admis, en stylistique comme en rhétorique, que tant les pronoms personnels que le sujet effacé en italien, en espagnol, en portugais et en roumain jouent un rôle essentiel pour la cohérence, la clarté et la fluidité des textes. Cette constatation explique la fréquence élevée de ces phénomènes de pronominalisation dans tous les types de textes (romanciers, journalistiques, techniques) et l'intérêt qu'on leur porte en traitement automatique du langage.

Parmi les principales études consacrées à ce sujet, on peut distinguer deux écoles. La première concerne l'analyse de la phrase isolée : il s'agit d'établir les conditions de coréférence de deux constituants d'une phrase d'après leur position dans la structure de l'arbre représentant cette phrase. La seconde école concerne l'analyse ou la génération de textes où les considérations basées sur la notion de focus et de pragmatique sont utilisées pour calculer les conditions de coréférence de deux constituants.

La première école présente l'inconvénient de reposer uniquement sur la géométrie d'arbres syntaxiques : la structure de ces arbres est étroitement liée à la théorie linguistique dont ils relèvent ; ces arbres sont donc profondément différents d'une théorie à l'autre, et les résultats obtenus ici sont par conséquent inapplicables là. A la rigidité de cette première approche, on peut opposer la seconde, qui présente en un certain sens l'inconvénient inverse : se situant dans le cadre de l'intelligence artificielle, elle suppose que la représentation des connaissances est une question résolue ou facile à résoudre, et que la formalisation de notions intuitives comme la pragmatique ou le focus est acquise.

L'approche que nous avons dans son étude sur la pronominalisation se distingue de ces deux écoles , par les deux points ci-dessous :

1) La pragmatique et le focus présentent une importance intuitive certaine, mais le rôle qu'ils jouent dans les conditions de coréférence sont mal définis, c'est pourquoi nous avons essayé de n'employer que des critères formels : ainsi, nous avons utilisé des critères de parallélisme, qui prennent en compte le matériel lexical des phrases concernées, leur structure, ainsi que d'autres informations morphologiques, sémantiques et syntaxiques. Par exemple, rappelons la fragilité de l'interprétation préférentielle du texte (1), dont nous jugeons le style douteux[1] :

(1) *Le professeur$_i$ a puni l'élève$_k$. Il$_k$ lui$_i$ avait lancé des boulettes.*

qui est remise en question dès qu'un paramètre morpho-syntaxique, comme le temps dans la deuxième phrase, est modifié :

(2) *Le professeur$_i$ a puni l'élève$_k$. Il$_i$ lui$_k$ a lancé des boulettes.*

Notons que les critères formels que nous employons ont l'avantage d'être indépendants du domaine, contrairement aux règles basées sur la pragmatique. Ces résultats, positifs mais encore insuffisants, nous amènent à élargir et approfondir les travaux sur le parallélisme.

---

1 Nous préférons à (1), que nous n'engendrons pas, l'un des textes suivants, dont le sens est le même :

Le professeur$_i$ a puni l'élève$_k$. Celui-ci$_k$ lui$_i$ avait lancé des boulettes.

L'élève$_i$ a été puni par le professeur$_k$. Il$_i$ lui$_k$ avait lancé des boulettes.

2) Les critères morpho-syntaxiques que nous avons utilisés à l'intérieur d'une phrase, qui prennent en compte le fait qu'une phrase n'est qu'un élément dans un texte[2] , ne reposent pas sur une géométrie rigide de l'arbre représentant la phrase : il suffit par exemple de distinguer les notions de sujet et d'objet (ce qui est possible dans tous les formalismes) pour appliquer la contrainte de non coréférence entre sujet et objet. C'est pourquoi notre approche peut être utilisée dans des systèmes de génération de textes utilisant un formalisme différent du nôtre, e.g. une syntaxe différente pour les schémas de phrase.

Les règles morpho-syntaxiques que nous avons utilisées ont été largement décrites dans les grammaires traditionnelles ainsi qu'en linguistique descriptive. C'est le cas, entre autres, pour le phénomène d'élision des pronoms *le, la* (*je **le** donne, je **l'**ai donné, je **le lui** ai donné.*), pour les conditions de cooccurrence de certaines ppv (*Ugo **le** presenta Bob* [Ugo lui_{fem} présente Bob], **Ugo **mi le** presenta* [*Ugo me lui_{fem} présente], *Ugo **glielo** presenta* [Ugo le $lui_{fem\ +\ masc}$ présente]) ou pour la contrainte de non coréférence entre sujet et objet (**Elle$_i$ la$_i$ peigne*, vs *Elle$_i$ la$_k$ peigne*). Nous avons mis en évidence dans nos travaux l'importance capitale des conséquences, souvent mal explorées, de ces critères en génération automatique :

- l'utilisation de ces critères conduit à des résultats fiables dans des cas complexes, sans recours à la pragmatique ou à la notion de focus. Ainsi, nous engendrons (cf. chapitre 6) :

(3) *Ho presentato Piero$_i$ a Eva$_k$. Ø$_i$ vuole rivederla$_k$.*
([J']ai présenté Piero$_i$ à Eva$_k$. [Il$_i$ + *elle] veut la$_k$ revoir.)

grâce à des règles purement morpho-syntaxiques.

- leur application systématique a des conséquences sur la conception même d'un générateur de textes. Rappelons que la plupart des systèmes de génération sont basés sur le principe de la modularité, ce qui conduit les chercheurs à concevoir leur système comme une suite d'opérations locales et ordonnées. Déjà, L. Danlos avait prouvé dans ses travaux que certains aspects de ce découpage devaient être remis en question. Quant à nos recherches consacrées à la pronominalisation, elles nous amènent à reconsidérer dans son intégralité le bien fondé de la modularité. Rappelons en effet que :

1) Les calculs morphologiques ne doivent pas être relégués à la dernière étape d'un système de génération automatique puisqu'ils interviennent dans le mécanisme de synthèse des pronoms (cf. chapitre 3) :

(4a) *Pierre$_i$ est arrivé avec Eva$_k$. Je l'$_i$ ai invité à ma fête.*
(4b) *Pierre$_i$ est arrivé avec Eva$_k$. Je l'$_k$ ai invitée à ma fête.*
(4c) **Pierre$_i$ est arrivé avec Eva$_k$. Je l'$_{i\ +\ k}$ inviterai à ma fête.*

(5a) *Pierre$_i$ est arrivato con Eva$_k$. Ø$_i$ è simpatico.*
(Pierre est arrivé avec Eva. [Il + *elle] est sympathique.)
(5b) *Pierre$_i$ est arrivato con Eva$_k$. Ø$_k$ è simpatica.*
(Pierre est arrivé avec Eva. [*Il + elle] est sympathique.)

---

2 Ceci a pour conséquence, entre autres, que les phénomènes de pronominalisation gauche sont rares dans un texte :

Max$_i$ ha una brutta cera. Quando Ø$_{i\ +\ k}$ è ammalato, Ugo$_k$ è di cattivo umore.
(Max$_i$ a mauvaise mine. Quand il$_{i\ +\ k}$ est malade, Ugo$_k$ est de mauvaise humeur.)

(5c) **Pierre$_i$ est arrivato con Eva$_k$. Ø$_{i+k}$ è gentile.*
(Pierre est arrivé avec Eva. [Il + elle] est (gentil + gentille).)

2) La synthèse des pronoms personnels met en jeu de nombreuses dépendances non locales : la synthèse d'un [loc-objet] dépend de celle d'un [dir-objet] pour éviter la cooccurrence de deux pronoms *ci* (**Ci ci ho pensato* [*J'y y ai pensé], **Ci ho pensato a quella cosa* [*J'y ai pensé à cette chose], *Ci ho pensato in quel luogo* [J'y ai pensé dans ce lieu]) ; elle implique aussi de nombreuses dépendances croisées :

- la synthèse du [à-objet] dépend de celle du [dir-objet] pour éviter la cooccurrence de certaines ppv (**Ugo me t'a présenté*, **Ugo t'a présenté moi*, *Ugo m'a présenté à toi*),
- la synthèse du [dir-objet] dépend de celle du [à-objet] pour l'élision des ppv *le* et *la* (*je l'enverrai à Ugo*, *je le lui enverrai*).

Les dépendances non locales et croisées excluent définitivement un algorithme de synthèse séquentiel ;

3) Les critères permettant de décider si un pronom personnel doit être synthétisé ou si un pronom sujet doit être effacé ne reposent pas uniquement sur des considérations locales : ils nécessitent une vision globale de la phrase et même du texte entier. Supposons que *Max* est *un vieillard* et que *Ugo* est *un jeune homme*. Nous rappelons (cf. chapitre 5) que :

- la synthèse du pronom objet *l'* dans

(6) *Max$_i$ est arrivé avec Ugo$_k$. Bob l'$_i$ a présenté au jeune homme$_k$.*

dépend de la synthèse du [à-objet] en un groupe nominal non ambigu,

- de même, la synthèse du pronom [à-objet] *lui* dans

(7) *Max$_i$ est arrivé avec Ugo$_k$. Bob lui$_k$ a présenté le vieillard$_i$.*

dépend de la synthèse du [dir-objet] en un groupe nominal non ambigu,

- le choix entre (6) et (7) dépend de la suite du texte :

(6') *Max$_i$ est arrivé avec Ugo$_k$. Bob l'$_i$ a présenté au jeune homme$_k$. Eva l'$_i$ a ensuite présenté au reste de la famille.*
(7') *Max$_i$ est arrivé avec Ugo$_k$. Bob lui$_k$ a présenté le vieillard$_i$. Eva lui$_k$ a ensuite présenté le reste de la famille.*

Les trois points que nous venons de mentionner nous ont amenée à concevoir un algorithme de synthèse basé sur une "approche globale" (cf. chapitre 4). Dans cette approche, **toutes** les formes (pro)nominales possibles pour la synthèse d'un constituant d'une phrase sont envisagées simultanément, et ce pour **tous** les constituants de la phrase. Les choix définitifs pour la synthèse des constituants sont effectués globalement sur la phrase, en tenant éventuellement compte de son contexte droit. Cet algorithme peut être appliqué de façon récursive (cf. chapitre 6). Sa gestion est assez complexe, mais en revanche il règle les problèmes posés aux points 1 (le rôle de la morphologie dans la synthèse des pronoms personnels) et 2 (les dépendances non locales et croisées). De plus, il permet d'appliquer les

critères de pronominalisation dans un ordre souple qui reflète le niveau de fiabilité de ces critères, leur mode d'application et leur rôle.

Nous n'avons étudié que certains phénomènes de pronominalisation, i.e. la synthèse des pronoms personnels ou l'effacement des sujets singulier, définis et référant à une entité. Mais il semble que l'approche globale soit applicable à d'autres phénomènes de pronominalisation, e.g. la synthèse de pronoms personnels pluriel, indéfinis, référant à un prédicat, ou encore la synthèse de formes pronominales comme *celui-ci*. Par exemple, les conditions nécessaires à la synthèse de *celui-ci*, *celle-ci*, etc ..., peuvent être résumées ainsi : on engendre une telle forme pour synthétiser une étiquette TOKi lorsque TOKi ne peut pas être synthétisée en un pronom personnel, et que cette étiquette est la dernière mentionnée. La synthèse du schéma de texte (ST8) illustre ces conditions :

(ST8) *(UGO rencontrer MAX. MAX être très malade.)*

Si la deuxième occurrence de MAX est synthétisée en un pronom personnel, alors l'interprétation de ce pronom est ambiguë, voire préférentiellement égale à UGO :

(8a) *$*Ugo_i$ a rencontré Max_k. $Il_{?i+k}$ était très malade.*

Par contre cette occurrence peut être synthétisée en *celui-ci :*

(8b) *Ugo_i a rencontré Max_k. Celui-ci_k était très malade.*

Cependant, ces conditions ne sont pas suffisantes : si on les applique au schéma de texte (ST9) :

(ST9) *(UGO rencontrer MAX. MAX renvoyer BOB (alors que BOB être très malade).)*

elles aboutissent au texte suivant :

(9a) *?*Ugo_i a rencontré Max_k. Celui-ci_k a renvoyé Bob_m alors que celui-ci_m était très malade.*

qui n'est pas ambigu, mais stylistiquement inacceptable. En d'autres termes, si les conditions de formation de *celui-ci* sont correctes (cf. exemple (8b)), l'application de ces conditions deux fois dans la même phrase doit elle, par contre, être remise en question. L'approche globale est tout à fait désignée pour éviter cette double application : elle peut facilement rejeter la synthèse de (ST9) en (9a) et proposer les textes suivants, si MAX et BOB sont respectivement définis comme *le chef de service* et *le jeune homme* :

(9b) *Ugo_i a rencontré Max_k. Celui-ci_k a renvoyé Bob alors que le jeune homme était très malade.*

(9c) *Ugo_i a rencontré Max_k. Le chef de service a renvoyé Bob_m alors que celui-ci_m était très malade.*

La remise en cause de la modularité n'est pas limitée au composant syntaxique (qui transforme un schéma de texte en un texte) : nous avons en effet abordé le problème de l'interaction entre le composant stratégique (qui fournit le schéma de texte) et le composant syntaxique. Là aussi, les questions de pronominalisation jouent un rôle essentiel : idéalement,

un schéma de texte devrait être conçu pour permettre la synthèse du plus grand nombre possible de pronoms personnels ou de sujets effacés, ce afin d'éviter des répétitions, qui sont responsables de textes non ambigus mais stylistiquement lourds. Les principes stylistiques utilisés pour effectuer les calculs de pronominalisation dans le composant syntaxique doivent aussi servir à la formation du schéma de texte. Nous avons commencé des recherches sur l'intégration de ces principes à l'élaboration du composant stratégique (cf. chapitres 2 et 7). Ces recherches doivent être poursuivies, dans le but d'inclure au calcul d'un schéma de texte d'autres conditions de synthèse des pronoms, et de mettre ainsi en évidence l'interaction des composants syntaxiques et stratégiques.

BIBLIOGRAPHIE

Boons, J.P., 1984, *Introduction au cours sur le lexique-grammaire des verbes locatifs* , Langue Française 62, Larousse, Paris.

Boons, J.P., Guillet, A., Leclère, Ch., 1976a, *La structure des phrases simples en français. Constructions intransitives* , Droz, Genève.

Boons, J.P., Guillet, A., Leclère, Ch., 1976b, *La structure des phrases simples en français. Classes de constructions transitives* , Rapport de recherche de l'Université Paris 8, no 6.

Brunet, J., 1985, *Grammaire Critique de l'Italien - vol 8* , Presses Universitaires de Vincennes, Saint Denis.

Chomsky, N., 1980, 'On Binding', *Linguistic Inquiry* , 11.

Chomsky, N., 1981, *Lectures on Government and Binding* , Foris, Dordrecht.

Chomsky, N., 1982, *Some Concepts and Consequences of the Theory of Government and Binding* , MIT Press, Cambridge, Mass.

Corblin, F., 1980, *Mémoire du texte et Interprétation des syntagmes nominaux*, Thèse de 3ème cycle, Université de Caen.

D'Agostino, E., 1983, *Lessico e sintassi delle costruzioni locative* , Liguori, Naples.

Danlos, L., 1985, *Génération automatique de textes en langues naturelles* , Masson, Paris.

Danlos, L., 1987, *The linguistic basis of text generation* , Cambridge University Press, Cambridge.

Danlos, L., 1988a, 'Some pronominalization issues in generation of texts in romance languages', *Electronics dictionaries and automata in computational linguistics* , D. Perrin ed., Springler-Verlag publications, Berlin.

Danlos, L., 1988b, 'Les phrases à verbe support être Prép', *Language*, 90, Larousse, Paris.

Danlos, L., Emerard, F., Sorin, C., 1985, 'Une structure prosodique du français', *Actes du GALF*, Journées d'Etude sur la Parole, Paris.

Danlos, L., Emerard, F., Laporte, E., 1986, 'Synthesis of spoken messages from semantic representations (semantic-representation-to-speech-system)', in *Proceedings of Coling 86*, Bonn, RFA.

Danlos, L., Namer, F., 1988, 'Morphology and cross-dependencies in the synthesis of personal pronouns in Romance languages', in *Proceedings of the 12th Coling*, Budapest.

Elia, A., 1982, *Lessico - Grammatica dei verbi italiani a completiva* , Liguori, Naples.

Elia, A., 1984, *Le verbe italien* , Schena-Nizet, Paris.

Elia, A.,Martinelli M.,D'Agostino E., 1981, *Lessico e Strutture sintattiche* , Liguori, Naples.

Eschenbach, C., Habel, Ch., Herweg, M., Rehkämper, K., 1989, 'Remarks on plural anaphora', in *Proceedings of the fourth european conference of the Association for Computational Linguistics*, Manchester.

Fauconnier, G., 1974, *La coréférence : syntaxe ou sémantique ?* , Seuil, Paris.

Gazdar, G., Klein, E., Pullum, G.K., Sag, I., 1985, *Generalized Phrase Structure Grammar* , Harvard University Press, Cambridge, Mass.

Giry-Schneider, J., 1978, *Les nominalisations en français. L'opérateur faire dans le lexique*, Droz, Genève.

Giry-Schneider, J., 1987, *Les prédicats nominaux en français : les phrases simples à verbe support* , Droz, Genève.

Gross, G., 1989, *Etudes syntaxiques de constructions converses* , Droz, Genève.

Gross, M., 1969, *Grammaire transformationnelle du Français : syntaxe du verbe* , Larousse, Paris.

Gross, M., 1975, *Méthodes en syntaxe* , Hermann, Paris.

Gross, M., 1982, 'Une classification des phrases "figées" du français', *Revue Québécoise de linguistique,* Vol. 11, No 2, Presses de l'Université du Québec, Montréal.

Grosz, B., 1982, 'Focusing and Description in Natural Language Dialogues', in *Elements of discourse understanding* , A. Joshi *et al.*, ed., Cambridge University Press.

Grosz, B., Joshi A., Weinstein S., 1982, 'Providing a Unified account of Definite Noun Phrases In Discourse', in *Proceedings of the 21st Annual Meeting of the Association for Computational Linguistics.*

Guéron, J., 1979, *Relations de coréférence dans la phrase et dans le discours*, Langue Française 44, Larousse, Paris.

Guillet, A., Leclère, Ch., 1981, 'Restructuration du groupe nominal', *Langages*, 63, Larousse, Paris.

Harris, Z.S., 1968, *Mathematical Structures of Language* , Wiley-interscience, New York.

Hirst, G., 1981, 'Discourse oriented Anaphora resolution : a review', American Journal of Computational Linguistics, Vol. 7, no 2.

Kayne, R., 1971, A Pronominalization Paradox in French, *Linguistic Inquiry* II,2.

Lamiroy, B., 1988, 'Binding properties of French en', in *To honour Yuki Kuroda*, eds C. Georgopoulos & R. Ishihara, Foris, Dordrecht.

Langacker, R., 1969, 'Pronominalization and the chain of command', *dans* Reibel et Shane.

Lo Cascio, V., 1969, *Strutture pronominali e verbali italiane* , Zanichelli, Bologne.

Manzini, M.R., 1983, 'On control and Control Theory', *Linguistic Inquiry* 14.

McDonald, D., 1983, 'Natural Language Generation as a Computational Problem : an introduction', in *Computational Models of Discourse* , Brady and Berwick, ed., MIT Press, Cambridge, Mass.

McKeown, K., 1983, 'Focus constraints on language generation', in *Proceedings of the Eighth International Joint Conference on Artificial Intelligence* , Karlsruhe, RFA.

McKeown, K., 1985, *Text generation,* Cambridge University Press, Cambridge.

Meunier, A., 1980, *Nominalisation d'adjectifs par verbes supports.*, Thèse de 3ème cycle, Université Paris 7, LADL, Paris..

Milner, J.C., 1982, *Ordres et Raisons de Langue*, Seuil, Paris.

Namer, F., 1989, 'Subject erasing and pronominalization in italian texte generation', in *Proceedings of the fourth european conference of the Association for Computational Linguistics*, Manchester.

Pemberton, L., 1989, 'A modular approach to story generation', in *Proceedings of the fourth european conference of the Association for Computational Linguistics.*, Manchester

Pitrat, J., 1985, *Textes, ordinateurs et compréhension*, Eyrolles, Paris.

Pollard, C., 1984, *Generalized Phrase Structure Grammars, Head Grammars, and Natural Language* , Doctoral dissertation, Stanford University.

Pollard, C., 1985, 'Phrase structure Grammar Without Metarules', in *Proceedings of WCCFL 4* , Golberg J., Mackaye S. Westcoast M. eds.

Reinhard, T., 1976, *The syntactic domain of Anaphora* , Doctoral dissertation, MIT, Cambridge, Mass.

Reinhard, T., 1983, *Anaphora and semantic interpretation* , Croom Helm, London.

Rizzi, L., 1982, *Issues in Italian Syntax* , Foris, Dordrecht.

Rosenthal, H., Warrack J., 1986, *Guide de l'Opéra*, Fayard, Poitiers.

Ross, J., 1969, 'On the cyclic Nature of Pronominalization', *dans* Reibel et Shane.

Ruwet, N., 1990, '*En* et *y* deux clitiques pronominaux antilogophoriques', Language 97, Larousse, Paris.

Sabah, G., 1988, *L'intelligence artificielle et le langage,* Hermès, Paris.

Salkoff, M., 1983, 'Bees are swarming in the garden', *Language* , vol. 59, no 2.

Schank, R., 1975, Conceptual Information Processing , North Holland, Amsterdam.

Sensini, M., 1988, *Grammatica della lingua italiana. Guida alla conoscenza e all'uso dell'italiano scritto e parlato* , Mondadori, Turin.

Serianni, L., 1988, *Grammatica italiana. Suoni Forme Costrutti* ,UTET, Turin.

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

Zingarelli, N., 1986, *Il nuovo Zingarelli* , Zanichelli, Bologna.

Zock, M., 1987, 'Propositions pour une simulation en temps réel de la génération de phrases en privilégiant ou le choix des mots ou celui des structures syntaxiques', *Machines, Réseaux Intelligents* , Paris.

Zock, M., 1988, 'Natural Languages are flexible tools, that's what makes them hard to explain, to learn and to use', *in Advances in Natural Language Generation : an interdisciplinary perspective* , Zock M. & Sabah G. eds., Ablex, New Jersey & Pinter, London.

Zock, M., Francopoulo G., Laroui A., 1988, 'Language Learning As Problem Solving : Modelling logical aspects of inductive learning to generate sentences in French by man and machine', in *Proceedings of the 12th Coling*, Budapest.

Zock, M., Sabah G., Alviset C., 1986, 'From Structure to Process : Computer-assisted teaching of various strategies for generating pronouns constructions in French', in proceedings of the 11th Coling, Bonn.

UNIVERSITE de MARNE-la-VALLEE
BIBLIOTHEQUE