

# La sélection de l'antécédent du pronom dans les systèmes de traitement automatique des langues naturelles

Frédérique Depain-Delmotte

Centre de Linguistique Lucien Tesnière,  
Université de Franche-Comté, 30. rue Megevand, 25030 Cedex, France  
frederique.delmotte@univ-fcomte.fr  
<http://tesniere.univ-fcomte.fr>

**Mots-clés :** pronom, relation anaphorique, traitement automatique des langues, classe d'objets, relation d'appropriation, relation d'appropriation à distance, saillance.

## Résumé

Les études récentes sur le problème de la sélection automatique de l'antécédent des pronoms montrent combien sont importantes les connaissances syntaxiques et sémantiques, mais aussi, dans quelle mesure sont requises les connaissances pragmatiques et encyclopédiques pour rendre compte du phénomène anaphorique. Nous présentons ici, une nouvelle approche du phénomène anaphorique développée autour de la notion de *relation d'appropriation* en vue de proposer un modèle linguistique pour la sélection automatique de l'antécédent du pronom. Tout d'abord, nous allons très brièvement dresser un bilan partiel des procédures de résolution traditionnellement utilisées dans la plupart des systèmes de résolution automatique. Puis, nous essaierons de montrer en quoi le choix d'un type particulier de description sémantique a une incidence sur la manière d'appréhender le phénomène anaphorique en traitement automatique des langues naturelles. La méthode proposée ici s'appuie sur la notion de *classes d'objets*, telle qu'elle est définie par Gaston Gross. Enfin, la notion de *relation d'appropriation 'à distance'* sera définie, et utilisée pour la sélection des antécédents des pronoms personnels. Nous envisagerons aussi le problème du verbe polysémique, véritable obstacle à la résolution de l'anaphore pronominale construite sur une représentation en termes de prédicat-arguments.

## 1. Introduction

Trouver l'antécédent des pronoms est depuis longtemps reconnu comme un problème difficile à résoudre en traitement automatique des langues naturelles. Les recherches développées sur cette question en linguistique et en psycholinguistique

tendent à montrer l'hétérogénéité des connaissances impliquées dans le mécanisme interprétatif de l'anaphore, d'où la grande diversité des approches proposées qui vont de l'analyse syntaxique, à l'analyse sémantique et pragmatique, en passant par l'analyse des facteurs psycholinguistiques, ou encore par le calcul statistique.

Nous présentons ici, une nouvelle approche du phénomène anaphorique développée autour de la notion d'*appropriation* en vue de proposer un modèle linguistique pour la sélection automatique de l'antécédent du pronom. Tout d'abord, nous exposerons très brièvement les procédures de résolution utilisées dans la plupart des systèmes de résolution automatique. L'objectif de cet article est de montrer en quoi le choix d'un type de description sémantique a une incidence sur la manière d'appréhender le phénomène anaphorique en traitement automatique des langues naturelles. Après avoir défini la notion de *relation d'appropriation à distance*, nous montrerons comment la relation anaphorique peut être décrite en termes de relation bilatérale, et nous expliquerons en quoi la notion des *classes d'objets* au sens de Gaston Gross<sup>1</sup>, est intéressante pour le traitement automatique de l'anaphore.

## 2. Procédures de résolution

La plupart des systèmes de résolution anaphorique accordent un intérêt particulier à l'anaphore pronominale ayant pour antécédent un syntagme nominal. Pourtant, force est de constater que le pronom admet des antécédents de nature très variée, allant du verbe à l'adjectif, de la proposition à un ensemble de plusieurs phrases, ou encore du paragraphe à plusieurs segments d'un discours. Envisager l'intégralité de ces possibilités reste problématique car la difficulté réside dans le choix des fonctions à exécuter par le système lorsqu'un pronom est repéré, et dans la définition des critères

<sup>1</sup> GROSS G., 1995, « Une sémantique nouvelle pour la traduction automatique : les classes d'objets », *La tribune des industries de la langue et de l'information électronique*, p.17-19, Paris.

qui permettent de décider de la nature de l'antécédent à rechercher. Georges Kleiber<sup>2</sup> critique la manière dont les chercheurs en intelligence artificielle et en psycholinguistique appréhendent le phénomène anaphorique. Selon lui, ils ont tort d'avoir pour souci principal de trouver le critère qui intervient dans le recrutement de l'antécédent (ces critères étant la proximité, la position syntaxique, le parallélisme, l'analyse casuelle ; nous y reviendrons un peu plus loin), et de focaliser leur attention sur l'anaphore pronominale ayant pour antécédent un syntagme nominal avec lequel il coréfère. Il est vrai qu'il s'agit là de ne considérer qu'un fonctionnement archétypal de l'anaphore, parmi l'ensemble fort complexe des phénomènes anaphoriques. En réalité, la raison de cet intérêt pour le pronom ayant pour antécédent un syntagme nominal est d'ordre technique. Tout d'abord, les pronoms et les syntagmes nominaux sont facilement repérables. On peut donc aisément constituer des listes de syntagmes nominaux. On obtient alors la liste des candidats antécédents auxquels des informations de différentes natures peuvent être attribuées, ce qui permet véritablement la résolution de l'anaphore.

Les systèmes de résolution anaphorique tente d'identifier le meilleur antécédent possible du pronom à partir de procédures qui reposent traditionnellement sur des stratégies telles que l'accord morpho-syntaxique, la contrainte de la c-commande et celle de Langacker (ces deux dernières étant des contraintes syntaxiques issues de la théorie chomskienne du liage), le parallélisme syntaxique, le parallélisme sémantique, le calcul de la saillance et la règle de récence. Décrivons et illustrons par des exemples, certaines de ces procédures (les items coréférentiels sont indiqués par des indices identiques).

#### • L'accord en genre et en nombre

Cette procédure repose sur l'accord en genre et en nombre du pronom avec son antécédent. Elle élimine de la liste des syntagmes nominaux ceux qui violent l'accord morpho-syntaxique.

(1) *Marie<sub>i</sub> a avoué à Paul<sub>j</sub> et à ses sœurs<sub>k</sub> qu'elle<sub>i</sub> avait eu tort.*

#### • Les contraintes syntaxiques

Issues de la théorie du gouvernement et du liage, de nombreuses restrictions ont été formulées qui permettent d'éliminer les candidats inacceptables pour les relations anaphoriques intraphrastiques, par exemple, la contrainte de la commande et la contrainte de Langacker. Cette dernière se définit

en ces termes : un pronom ne peut à la fois précéder et c-commander son antécédent.

(2) \* [<sub>P1</sub> Ils<sub>i</sub> pensent [<sub>P2</sub> que les Martiens<sub>i</sub> envahiront la planète.]]

#### • Le parallélisme syntaxique

Le parallélisme syntaxique permet de sélectionner les syntagmes nominaux qui ont la même fonction syntaxique que le pronom anaphorique.

(3) *L'enfant<sub>i</sub> reconnut le roi<sub>j</sub> alors qu'il<sub>i</sub> ne l'<sub>j</sub> avait jamais rencontré auparavant.*

#### • Le parallélisme sémantique

Il s'agit d'une procédure très puissante. Elle n'est intégrable que dans les architectures qui possèdent un module d'analyse sémantique utilisant une grammaire casuelle, autrement dit, qui soit capable d'identifier automatiquement les rôles casuels des arguments des prédicats verbaux dans les phrases. Seuls les syntagmes nominaux qui possèdent le même rôle casuel que le pronom sont sélectionnés comme étant de potentiels antécédents.

(4) *Le livre<sub>i</sub> a été envoyé à Luc<sub>j</sub> par Martin<sub>k</sub>. Il<sub>k</sub> le<sub>i</sub> lui<sub>j</sub> prête quelques jours*

#### • La contrainte sémantique du rôle casuel

Les rôles casuels imposent des contraintes sémantiques sur le pronom qui occupe une place dans la structure d'arguments du prédicat verbal. Pour que la contrainte soit satisfaite, les marques sémantiques assignées aux arguments doivent être identiques. Dans l'exemple suivant, la contrainte sémantique (+comestible) sur le rôle casuel objet du verbe 'manger' suffit pour trouver l'antécédent du pronom 'le'.

(5) *Le chien qui voit un sucre<sub>i</sub> monte sur le buffet. Il veut le<sub>i</sub> manger.* (Sabah, 1989:225)

#### • Autres stratégies mises en œuvre

Généralement, le calcul de la saillance s'effectue à partir de plusieurs critères tels que la récence, la fréquence des occurrences, la fonction grammaticale de l'antécédent, le repérage d'expressions focalisantes du type 'il y a ...', ou 'c'est...qui/que'. Issus principalement des recherches en psycho-linguistique et en intelligence artificielle, ces critères de résolution apparaissent comme des procédures qui permettent, lorsqu'il demeure une ambiguïté sur l'antécédent, de sélectionner parmi les candidats possibles, celui qui est le plus proche, ou celui qui se trouve être le plus fréquemment mentionné dans le texte, ou encore celui qui a la fonction grammaticale la plus saillante. Leur nécessité dans les systèmes d'analyse automatique paraît évidente, même si leur

<sup>2</sup> KLEIBER G., 1991. « L'anaphore : d'un problème à l'autre », In Journée SCICOGNEST, Publication du Groupe Anaphore et Deixis, N°4.

influence dans la résolution de l'anaphore est difficile à évaluer de manière systématique.

Toutes ces stratégies de résolution ont été mises en œuvre dans différentes approches : par exemple, dans l'approche en surface de Carter (1986), dans l'approche multi-stratégique de Carbonell et Brown (1988), dans les approches syntaxiques de Lappin et Leass (1994), et de Stuckardt (1996), dans l'algorithme de Kennedy et Boguraev (1996), ou encore dans le modèle intégré de Mitkov (1994 ; 1996). On constate une grande variabilité dans l'utilisation de ces procédures de résolution à l'intérieur de l'architecture de ces systèmes. En effet, certaines d'entre elles apparaissent comme étant des contraintes de sélection, - c'est-à-dire qu'elles filtrent la liste des candidats antécédents dans le but d'éliminer les syntagmes nominaux qui ne respectent pas les conditions de la règle -, ou alors comme règles de sélection préférentielle, - sélectionnant ou proposant à partir de la liste des potentiels antécédents (lorsque l'analyse par contraintes a échoué) tel syntagme nominal comme étant le meilleur antécédent possible -.

Si les grammaires de cas ont suscité un intérêt certain de la part des chercheurs qui se sont intéressés au problème de la sélection automatique de l'antécédent du pronom, c'est parce qu'elles reposent sur trois principes fondamentaux qui sont la structure de type prédicat-arguments, l'étiquetage casuel des arguments, et l'assignation de valeurs sémantiques à chacun des cas dans la structure prédictive. Il découle de la théorie des grammaires de cas, la possibilité d'assigner au pronom un certain nombre de valeurs à partir de la place qu'il occupe dans la structure prédictive, valeurs qui permettent la recherche de l'antécédent par filtrage de traits. Cependant, nous faisons ici trois remarques :

[1] Tout d'abord, ce ne sont pas tant les étiquettes casuelles dans la structure prédictive que les traits sémantiques qui leur sont assignés qui permettent la sélection de l'antécédent, comme l'illustre l'exemple (4). La contrainte sémantique du rôle casuel propose une discrimination plus forte.

[2] Ensuite vient la question de la description sémantique des langues. La contrainte sémantique du rôle casuel n'est possible qu'à partir d'un étiquetage sémantique qui souvent est réalisé de manière intuitive. Du codage sémantique du lexique dépend le succès de la résolution de l'anaphore. En effet, le mécanisme interprétatif du pronom s'effectue à partir de la vérification de l'adéquation des marques sémantiques attribuées au pronom et à son antécédent ; la cohérence dans l'étiquetage des unités lexicales d'une part, et des arguments des prédicats d'autre part, est le garant du succès de l'analyse. Il s'agit d'un véritable casse-tête pour celui qui s'attelle à la description sémantique d'une

langue en tentant de rendre compte d'une telle adéquation dans le codage en se fiant totalement à son intuition. Par exemple, doit-on coder le mot 'robot' comme étant un objet concret, animé, non humain ? A partir de quels critères peut-on décider d'attribuer le trait « NON(humain) », alors que ce robot dans un texte possèdera des caractéristiques humaines telles que se déplacer, agir, regarder, analyser, et en même temps, il pourra rouiller, s'oxyder, se briser. Il s'agit là d'un exemple extrême mais qui donne une bonne illustration de la difficulté à décrire sémantiquement une langue à partir de l'intuition.

[3] Même si les contraintes sémantiques sont bien codées et qu'une réelle adéquation des traits sémantiques est observée, certaines difficultés persistent, écrivent Jaime Carbonell et Ralf Brown<sup>3</sup>. Pour illustrer leur propos, les auteurs donnent les exemples suivants :

(6) *John took the cake from the table and ate [it].*

(7) *John took the cake from the table and washed [it].*

Dans cet article, les auteurs montrent que les contraintes sémantiques du rôle casuel suffisent à résoudre la phrase (6) puisque les substances susceptibles d'être ingérées sont de préférence des aliments comestibles. Au contraire, pour la phrase (7), il paraît plus difficile d'automatiser un système qui soit capable d'exclure la solution du cas objet instancié par 'cake' comme étant l'objet lavé, puisqu'on ne peut pas simplement écrire la restriction « NON(comestible) » sur le cas objet du verbe 'laver'. D'ailleurs, le fait que les légumes et les fruits soient susceptibles d'être nettoyés avant d'être mangés interdit l'utilisation d'un tel trait. Pour répondre à de telles difficultés, Carbonell et Brown suggèrent d'intégrer des connaissances pragmatiques, qui, manipulées par un mécanisme d'inférence, permettent d'extrapoler les conséquences de l'immersion d'un gâteau dans une baignoire d'eau, ou encore d'utiliser des inférences plus abstraites qui pourraient rendre compte du fait qu'après immersion dans l'eau, l'objet du verbe 'laver' doit resté inchangé dans sa constitution. Mais ces deux propositions sont très coûteuses, c'est pourquoi les auteurs préconisent une approche construite à partir de la combinaison de plusieurs stratégies de résolution (proches de celles présentées ci-dessus).

Ce que nous voulons montrer maintenant, c'est que la description sémantique en termes de *classes d'objets*, outre le fait qu'elle réponde au principe d'adéquation nécessaire à la résolution automatique

<sup>3</sup> CARBONELL J., BROWN R., 1988. « Anaphora Resolution : A Multi-strategy approach », In COLING'88, Vol.1, p.96-101, Budapest.

de l'anaphore, permet de résoudre ce genre de problème. Cette nouvelle approche repose sur la relation de *compatibilité* entre le prédicat verbal et les arguments qu'il sélectionne ; elle nous permet de formuler une explication originale du phénomène anaphorique. C'est ce dont il est question maintenant.

### 3. Résolution de l'anaphore et classes d'objets

Le type de description sémantique a une incidence sur la manière d'appréhender le phénomène anaphorique en traitement automatique des langues naturelles. Mais, en quoi la description sémantique en termes de *classes d'objets* est-elle intéressante pour la résolution de l'anaphore ? Avant de répondre à cette question, il nous faut définir les notions de *relation d'appropriation* et de *classes d'objets* selon l'approche de Gaston Gross.

#### 3.1 Terminologie de base

Selon un principe communément admis, le prédicat verbal sélectionne ses arguments, et dans un certain nombre de cas, une spécification sémantique précise peut leur être assignée. La description au moyen des classes d'objets permet une analyse très fine des restrictions de sélection : tout d'abord, chaque argument, dans le schéma syntaxique du verbe, est caractérisé par des traits syntactico-sémantiques (leur nombre est actuellement de 8) ; ils permettent une première discrimination. Ensuite, un '*typage*' des arguments est réalisé « sur la base des compatibilités entre opérateurs et arguments »<sup>4</sup>. Des classes lexicales sont constituées en répondant à la question suivante : « pour un prédicat (ou ensemble de prédicats) donné, quel type précis de nom peut se trouver en position d'argument »<sup>5</sup>. Ainsi, chaque classe d'objets est définie à partir des prédicats qui sélectionnent de façon appropriée des mots sémantiquement équivalents. Par exemple, les noms d'<aliments> et de <boissons> entretiennent une relation de nature privilégiée avec des prédicats verbaux comme *manger* et *boire*, mais aussi avec leurs variantes : *absorber*, *avalier*, *consommer*, etc. Reprenons les exemples (6) et (7) de tout à l'heure. Nous avons montré, à travers les propos de Carbonell et Brown, l'insuffisance du trait « NON(comestible) » assigné au cas objet du verbe '*laver*' pour résoudre l'anaphore, la raison étant que certains aliments, tels que les fruits et les légumes, peuvent être lavés avant d'être consommés. Grâce au principe selon lequel les noms entretiennent une relation d'appropriation avec un nombre limité de prédicats verbaux, les anaphores des exemples (6)

et (7) peuvent être résolues. En effet, le mot '*table*' sélectionne de manière privilégiée des verbes comme '*nettoyer*', '*essuyer*', '*laver*', alors que le mot '*gâteau*', qui appartient à la classe d'objets <aliment> et qui sélectionne de manière privilégiée des verbes comme '*manger*', n'entretient pas de relation privilégiée avec de tels verbes.

- (8a) *John lave la table*
- (8b) \* *John lave un gâteau*
- (8c) *John lave une pomme*

- (9a) \* *John mange la table*
- (9b) *John mange un gâteau*

On remarquera que certains noms d'aliments (appartenant à la classe des fruits, légumes, poissons, etc.) peuvent sélectionner de manière privilégiée des verbes comme '*laver*', '*nettoyer*', '*rincer*', etc., alors que d'autres noms tels que '*gâteau*', '*pizza*' ne sélectionnent pas de tels verbes. Une sous-classe de la classe <aliments> peut rendre compte de cette contrainte.

#### 3.2 Relation d'appropriation à distance

Ce sont les critères utilisés pour définir les schémas syntaxiques des prédicats verbaux qui sont à la base de l'adéquation exigée dans le codage sémantique pour la recherche de l'antécédent, car ils reposent sur la *compatibilité* entre le prédicat verbal et ses arguments. Ce que nous voulons montrer, c'est que la compatibilité entre opérateurs et arguments peut être utilisée pour la sélection de l'antécédent. On postulera alors que l'antécédent entretient une relation d'appropriation à distance avec le verbe, grâce à la relation anaphorique établie avec le pronom-substitut. Mais illustrons, par un exemple, notre propos en montrant comment la *relation d'appropriation à distance* permet la sélection de l'antécédent du pronom.

- (10) *Les passants qui traversaient le parc pouvaient apprécier la mélodie. Le violoniste reconnaissait parfaitement cet air musical. Son père le jouait souvent au piano.*

Imaginons que le verbe '*jouer*' ne soit pas un verbe polysémique et qu'il n'accepte que le sens *jouer un air*. Le phénomène d'accord en genre et en nombre entre l'antécédent et le pronom est un auxiliaire précieux à la résolution automatique de l'anaphore. Les contraintes morpho-syntaxiques doivent être satisfaites par les candidats antécédents. Le pronom anaphorique '*le*' possède les marques suivantes :

- (*Le*
- (POSITION ( pronom objet direct)),
- (GENRE (masculin)),
- (NOMBRE (pluriel))).

<sup>4</sup> LERAT P., 1996. « Terminologie et connaissances sémantiques », In JSLB'96 ; p.84.

<sup>5</sup>GROSS G., 1994. « Classes d'objets et description des verbes », *Langages 115*, Larousse, Paris, p.11.

A l'issue de l'analyse des marques morpho-syntaxique, la liste des syntagmes nominaux potentiellement antécédents est construite :

*le violoniste* : <humain>  
*le parc* : <locatif>  
*cet air musical* : <air>

La structure prédicative du verbe *jouer* est donnée comme suit (T pour trait syntactico-sémantique ; C pour classe d'objets) :

(prédicat verbal : jouer  
 (Arg0 : T (humain),  
 C([de préférence : instrumentiste])),  
 (Arg1 : T (inanimé concret), C (air)))

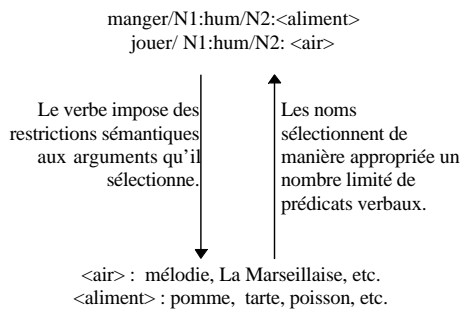
Le pronom *le* reçoit les informations de la structure prédicative du verbe *jouer*, elles vont permettre de déterminer lequel des potentiels antécédents est le meilleur antécédent. On obtient la structure suivante :

(prédicat verbal : 'jouer'  
 (Arg0 : 'son père',  
 (T (humain), C(relationnel))),  
 (Arg1 : 'le'  
 ((GENRE (masculin)),  
 (NOMBRE (pluriel))),  
 (T (inanimé concret),  
 C (air)))).  
 → a pour antécédent : '*cet air musical*'.

Ce sont ces contraintes de sélection qui permettent d'inférer la nature sémantique de l'antécédent du pronom, et de trouver le meilleur antécédent. En l'occurrence, dans la séquence (8), un seul candidat antécédent dans la liste proposée, possède les mêmes caractéristiques sémantiques que celles attribuées au pronom *le*. Il s'agit du syntagme '*cet air musical*'. Il est donc sélectionné comme étant le meilleur antécédent possible.

### 3.3 Mécanisme exploratoire bilatéral

Pour résumer notre pensée, la solution que nous proposons montre que le phénomène anaphorique peut être interprété et décrit comme relevant d'un mécanisme bilatéral : d'une part le pronom acquière des informations grâce aux contraintes sémantiques assignées aux arguments dans le schéma syntaxique du verbe, et d'autre part les noms eux-mêmes ont des relations particulières et privilégiées avec un nombre limité de prédicats verbaux. Les contraintes de sélection qui s'appliquent au pronom, s'appliquent aussi à son antécédent (voir le schéma ci-dessous). Ceci nous permet d'affirmer que la relation d'appropriation entre les noms et les verbes qu'ils sélectionnent peut s'établir à distance.



La prise en compte d'une double relation dans l'établissement des contraintes de sélection renforce la procédure de résolution en réduisant le champ des possibles dans la liste de candidats antécédents, laquelle se trouve considérablement réduite. Mais, qu'en est-il des cas où le verbe est polysémique ?

La polysémie des verbes constitue une caractéristique qui pose les problèmes les plus ardues à la description des langues. Le verbe *jouer* est un verbe polysémique : *jouer l'indifférence*, *jouer la Marseillaise*, *jouer du piano*, etc. Les différents emplois d'un verbe sont rendus explicites dans les schémas syntaxiques grâce aux classes d'objets :

*jouer* <jeu> : *jouer aux billes*  
*jouer* <sentiment> *jouer l'indifférence*  
*jouer* <air> *jouer la Marseillaise*  
*jouer* <instrument> *jouer de la mandoline*  
*jouer* <spectacle> *jouer le Cid*  
*jouer* <sport> *jouer au football*

Gaston Gross propose d'utiliser la valeur sémantique des syntagmes nominaux qui occupent la place d'un argument dans la structure prédicative du verbe pour décider de l'emploi du verbe. C'est donc le syntagme nominal qui lève l'ambiguïté du verbe :

1. Localiser le verbe ( : *prendre*)
  - 1.1. si le verbe est polysémique, rechercher l'argument à droite (pour l'interpréter)
  - 1.2. si le mot rencontré est *avion*, rechercher la classe d'objets de *avion* (la classe d'objets est <moyen de transport en commun>)
2. interpréter le verbe *prendre* comme *voyager*

Plusieurs questions s'imposent : comment trouver l'antécédent d'un pronom quand le verbe admet plusieurs structures prédicatives, et, comment déterminer parmi les différents emplois du verbe *jouer*, la signification du verbe sans connaître les caractéristiques des arguments utilisés. Reprenons l'exemple (10).

(10) *Les passants qui traversaient le parc pouvaient apprécier la mélodie. Le violoniste reconnaissait parfaitement cet air musical. Son père le jouait souvent au piano.*

[1] Tout d'abord, les occurrences d'un texte participent à la levée de l'ambiguïté sur le verbe. À partir de l'analyse des classes d'objets auxquelles appartiennent les syntagmes nominaux repérés, il est possible de faire des prédictions sur la 'bonne' structure prédicative du verbe polysémique à partir des classes d'objets déjà utilisées dans le texte. Une vérification de la correspondance entre les classes d'objets assignées aux syntagmes nominaux et celles attendues dans les structures prédicatives du verbe permettra de lever l'ambiguïté sur le verbe, et du même coup, de trouver l'antécédent du pronom. Dans certains cas, comme dans l'exemple (12), le degré de saillance d'une classe d'objets - calculable à partir de paramètres tels que la récence, la fréquence, la position syntaxique, la proximité - peut être un critère pertinent pour la désambiguïsation du verbe. Ici, la classe d'objets <air> est utilisée par deux fois : '*la mélodie*', et '*cet air musical*'<sup>6</sup> ; en outre, une des structures prédicatives admet comme argument cette même classe d'objets. On remarque aussi qu'aucune des autres classes d'objets acceptées par le verbe n'est utilisée dans le texte, ce qui est une condition importante qui facilite la levée de l'ambiguïté sur le verbe.

[2] Ensuite, si on élargit le schéma syntaxique du verbe, on se rend compte que la structure '*jouer <air>*' peut être étendue à '*jouer <air> à <instrument de musique>*', comme dans la phrase '*Son père le jouait souvent au piano*'. La présence du syntagme '*au piano*' participe très fortement à l'interprétation du verbe '*jouer*'.

[3] Et enfin, la nature du pronom est un indice important qui permet de lever l'ambiguïté sur le verbe et de trouver la nature sémantique du pronom. En effet, dans le cas du verbe *jouer*, elle peut aider à déterminer lequel des emplois admis du verbe est mis en œuvre, et par extension, apporter l'information concernant la classe d'objets de l'argument, et permettre la sélection de l'antécédent :

Le Cid <spectacle> → *il le joue*  
du piano <instrument> → *il en joue*  
aux cartes <jeu> → *il y joue*

<sup>6</sup> Une condition importante et qui permet de calculer le degré de saillance d'une classe d'objets, consiste à vérifier si une relation de co-référence est effectivement établie entre les syntagmes nominaux en question. Pour ce faire, par exemple, on vérifiera si la succession des articles indéfinis, définis, démonstratifs est respectée.

(11) ... *il en joue tous les jours.*

Si on analyse cette phrase volontairement incomplète, on constate qu'une attente très forte détermine notre interprétation quant à l'objet utilisé pour réaliser l'action de jouer, alors que la polysémie du verbe devrait laisser totalement équivoque cette phrase. C'est donc la nature du pronom qui guide la sélection de l'antécédent en introduisant l'information *instrument de musique*.

Nous voulions mettre l'accent sur le fait que le repérage et l'utilisation de tels indices dans un système d'analyse automatique permet de reconnaître lequel des emplois verbaux est utilisé dans un texte et, du même coup, permet de résoudre l'anaphore pronominale grâce aux informations récupérées sur la valeur sémantique des arguments dans la structure prédicative sélectionnée. Le phénomène de saillance apparaît essentiel dans les analyses qui viennent d'être faites. On a vu que, d'une part, la saillance pouvait être d'une aide très précieuse dans l'interprétation sémantique du verbe, et d'autre part, que les informations issues de la structure prédicative, couplées avec celles obtenues par le calcul de la saillance, permettent de résoudre l'anaphore.

## 4. Conclusion

Dans cet exposé, nous nous sommes attachée, dans un premier temps, à présenter les différentes stratégies de résolution traditionnellement mises en œuvre dans les systèmes d'analyse automatique pour la sélection de l'antécédent du pronom. Ensuite, nous avons porté notre attention sur la question de la description sémantique des langues en montrant, à partir d'un exemple problématique, quelle influence elle pouvait avoir sur la manière d'appréhender le phénomène anaphorique. Un de nos objectifs était de montrer dans quelle mesure la notion de classes d'objets, au sens de Gaston Gross, est intéressante pour la résolution automatique de l'anaphore ; c'est surtout autour de la notion de relation d'appropriation que se construit notre démonstration. Aussi, nous avons abordé le problème de la recherche de l'antécédent des pronoms dans le cas où le verbe est polysémique. Il ne s'agit certainement pas du chemin le plus simple pour réfléchir sur le problème déjà fort complexe de l'anaphore. Mais, cette réflexion a permis de soulever des interrogations sur deux points : d'une part, sur le rapport entre co-référence et classe d'objets, et d'autre part, sur les procédures d'évaluation de la saillance des occurrences, calculée non pas à partir des occurrences, mais à partir des classes d'objets qui leur sont assignées.

## Références

- BALTAZART, Didier, 1993. « Contribution au traitement des ambiguïtés polysémiques et anaphoriques », Thèse de Doctorat, Université de Technologie de Compiègne, Compiègne.
- BOGUREAV, Branimir & KENNEDY, Christopher, 1996. « Anaphora for everyone : pronominal anaphora resolution without a parser », In : Proceedings of the 16th International Conference on Computational Linguistics (COLING'96), Vol.1, p.113-118, Copenhagen.
- CARBONELL, Jaime & BROWN, Ralf, 1988. « Anaphora Resolution : A Multi-strategy approach », In : Proceedings of the 12th International Conference on Computational Linguistics (COLING'88), Vol.1, p.96-101, Budapest.
- CARTER, David, 1987. « Interpreting anaphors in natural language texts », Ellis Horwood limited, Chichester.
- GIRY-SCHNEIDER, Jacqueline, 1994. « Les compléments nominaux du verbe *lire*, une illustration de la notion de 'classe d'objets' », *Langages 115*, Larousse, Paris.
- GROSS, Gaston, Vivès Robert, 1986. « Les constructions nominales et l'élaboration d'un lexique-grammaire », *Langue française 69*, Larousse, Paris.
- GROSS, Gaston, 1992. « Forme d'un dictionnaire électronique », in Clas A., Safar H., *L'environnement traductionnel*, Presses de l'Université du Québec, Sillery (Canada).
- GROSS, Gaston, 1994. « Classes d'objets et description des verbes », *Langages 115*, Larousse, Paris.
- GROSS, Gaston, 1995. « Une sémantique nouvelle pour la traduction automatique : les classes d'objets », *La tribune des industries de la langue et de l'information électronique*, Paris.
- KLEIBER, Georges, 1991. « L'anaphore : d'un problème à l'autre », In Journée SCICOGNEST, Publication du Groupe Anaphore et Deixis, N°4.
- LAPPIN, Shalom & LEASS, Herbert, 1994. « An algorithm for pronominal anaphora resolution » In Computational Linguistics, vol.20, n°4, MIT Press, Cambridge, Massachusetts, USA.
- LE PESANT, Denis, 1996. « Principes d'organisation des données lexicales dans un dictionnaire électronique », *Sémiotiques N° 11*.
- LE PESANT, Denis, 1998. « Utilisation des propriétés des anaphores dans la définition des relations lexicales », *Langages 131*, Larousse, Paris.
- LE PESANT, Denis & MATHIEU-COLAS, Michel, 1998. « Introduction aux classes d'objets », *Langages 131*, Larousse, Paris.
- LERAT, Pierre, 1996. « Terminologie et connaissances sémantiques », In : Journées de Sémantique Lexicale Brestoise (JSLB'96).
- MITKOV, Ruslan, 1994. « An integrated model for anaphora resolution », In : Proceedings of the 15<sup>th</sup> International Conference on Computational Linguistics (COLING'94), Kyoto, Japan.
- MITKOV, Ruslan, 1996. « Anaphora resolution : a combinaison of Linguistics and historical approaches », in Proceedings of the Discours Anaphora and Anaphor resolution, (DAARC'96), Lancaster, UK.
- SABAH, Gérard, 1989. *L'intelligence artificielle et le langage*, Tome 2, Editions Hermès, Paris.
- STUCKARDT, Roland, 1996. « Anaphor Resolution and the Scope of Syntactic Constraints », In : Proceedings of the 16th International Conference on Computational Linguistics (COLING'96), Copenhagen.