

Université Paris 7 Denis Diderot  
UFR de Linguistique

Master de Linguistique informatique  
Juin 2006

# Un algorithme pour la résolution d'anaphores événementielles

André Bittar

Sous la direction de Pascal Amsili

## **Remerciements**

Je remercie particulièrement Pascal Amsili de m'avoir guidé dans ce travail et de m'avoir aidé à suivre une démarche scientifique.

Je remercie également tous ceux qui ont participé aux exercices, notamment les étudiants du Master 2 de linguistique informatique et de phonétique, qui m'ont servi de cobayes linguistiques. Qu'ils n'entendent plus jamais la phrase « c'est arrivé hier ».

Merci à Stéphanie d'avoir relu mon travail et corrigé mes fautes et tournures de franglais.

Enfin, merci à mes proches de m'avoir encouragé et soutenu.

## Table des matières

<b>1. Introduction.....</b>	<b>5</b>
<b>2. Description théorique.....</b>	<b>6</b>
2.1. Anaphore et coréférence.....	6
2.1.1. La référence.....	6
2.1.2. La coréférence.....	8
2.1.3. L'anaphore.....	9
2.2. Entités abstraites.....	10
2.2.1. Les origines : Davidson et les événements.....	11
2.2.2. Une typologie des entités abstraites : généralisation.....	12
<b>3. Anaphore événementielle – étude linguistique.....</b>	<b>15</b>
3.1. Formes du matériau de reprise.....	15
3.1.1. Les pronoms cela, ça, ce, le, y, en.....	15
3.1.2. Conteneurs des expressions anaphoriques événementielles.....	16
3.2. Forme de l'antécédent.....	18
3.2.1. Les phrases événementielles.....	19
3.2.2. Les contrefactuelles.....	21
3.2.3. Les propositions subordonnées complétives.....	24
3.2.4. Les propositions subordonnées relatives.....	29
3.2.5. Événements coordonnés.....	30
3.2.6. Les temps verbaux.....	32
3.2.7. La modalité.....	34
3.3. Formes non anaphoriques.....	36
3.3.1. Le ça impersonnel.....	36
3.3.2. Le ce non anaphorique.....	38
<b>4. Algorithmes pour la résolution d'anaphores.....</b>	<b>41</b>
4.1. Résolution d'anaphores nominales.....	41

4.1.1. Hobbs.....	41
4.1.2. Lappin & Leass.....	42
4.1.3. Mitkov.....	44
4.2. Résolution d'anaphores abstraites.....	46
4.2.1. Eckert & Strube.....	46
4.3. Un algorithme pour la résolution d'anaphores événementielles.....	48
4.3.1. Algorithme.....	48
4.3.2. Description sommaire.....	57
<b>5. Résultats préliminaires.....</b>	<b>59</b>
<b>6. Conclusion et travail futur éventuel.....</b>	<b>62</b>
<b>7. Références.....</b>	<b>64</b>
<b>8. Annexe.....</b>	<b>67</b>
8.1. Les temps verbaux.....	67
8.2. Quantification temporelle.....	67

## 1. Introduction

La résolution des anaphores est une des branches de recherche les plus actives du domaine du Traitement automatique des langues (TAL). Le phénomène de l'anaphore est très répandu dans les langues naturelles, c'est pour cela qu'un module de résolution est nécessaire dans presque toutes les applications du TAL. Des travaux très variés ont été effectués dans ce domaine selon des approches théoriques différentes.

La résolution consiste simplement à trouver la référence d'un groupe nominal qui doit être interprété par rapport à un élément apparaissant avant lui dans le discours. Ce groupe nominal peut être un nom avec un déterminant (groupe nominal plein), un pronom personnel, un pronom démonstratif ou réfléchi. Le problème de la résolution comprend deux parties. La première est la résolution de la *coréférence*, qui consiste à établir une relation entre deux groupes nominaux pleins qui font référence à un même élément dans le discours. La deuxième consiste en la recherche de la référence d'un pronom, élément qui doit obligatoirement être interprété selon son contexte. Cette deuxième opération est ce que l'on appelle la résolution d'*anaphore*.

Les pronoms anaphoriques peuvent reprendre des éléments linguistiques de nature catégorielle diverse – des noms, des groupes nominaux (anaphore *individuelle*), mais aussi des verbes, ou des phrases (anaphore *abstraite*).

Le but de l'étude qui suit est d'examiner le phénomène de l'anaphore abstraite en français et, à partir de données empiriques, de formuler un algorithme pour la résolution d'un cas particulier de ce type d'anaphore. Le cas particulier que l'on envisage de traiter est celui de l'anaphore événementielle, où un pronom fait référence à un événement qui a été précédemment évoqué dans le discours.

Avant d'entamer l'étude linguistique proprement dite on adoptera un point de vue théorique en se basant sur des travaux préexistants dans les domaines philosophiques et linguistiques.

## 2. Description théorique

Cette partie constitue une introduction aux notions théoriques centrales à cette étude. On commencera d'abord par une description de la notion de référence, selon Jean-Claude Milner, qui donnera les bases pour l'étude des relations de coréférence et d'anaphore. Ensuite, on considérera la théorie des événements de Donald Davidson dans le but d'adopter une ontologie particulière. Finalement, on procédera à une étude des entités abstraites selon la typologie de Nicolas Asher.

### 2.1. Anaphore et coréférence

Afin de comprendre le phénomène que l'on souhaite traiter on commencera par un exposé des principes du fonctionnement de l'anaphore et de la coréférence. On exposera également les notions et principes qui les gouvernent. Comprendre la nature de la référence est essentiel pour la considération de ces phénomènes.

#### 2.1.1. La référence

Décrire la façon dont les mots de la langue font référence au monde qui nous entoure, que ce soit le monde que nous percevons au moyen de nos sens ou celui des idées et concepts abstraits<sup>1</sup>, constitue une grande question de philosophie. Comment définir la relation qui entre en jeu lorsque nous parlons d'un objet concret ou d'une idée abstraite, et comment décrire la façon dont elle relie nos mots à la réalité ? Le but est ici de définir la notion de *référence*.

Les mots, ou plus précisément les unités lexicales, des langues naturelles sont classifiés selon leur catégorie syntaxique. Les unités lexicales se combinent (en fonction de leurs catégories) pour former des groupes syntaxiques (ou syntagmes). Les groupes de types différents ne font pas référence au monde de la même façon. Par exemple, les groupes verbaux ne représentent pas le même type d'entités que les groupes nominaux. Ces deux sortes de référence sont distinctes.

Jean-Claude Milner (1982) s'intéresse à une description de la référence nominale, c'est-à-dire à une description de la façon dont les groupes nominaux désignent des entités dans le monde. On explorera principalement dans cette section cette théorie de la référence nominale.

Le référent d'une unité lexicale est la partie du monde ou l'entité à laquelle cette unité est associée. Celui-ci peut être de nature abstraite ou concrète. Les unités lexicales de la langue ne peuvent avoir pour référent n'importe quelle entité du

---

1 Par le terme *monde* on entendra la réalité en général – le monde des objets concrets, celui des idées, des concepts et tout ce qui peut exister entre les deux.

monde : les groupes nominaux *un homme* et *une femme*, par exemple, désignent des entités différentes. La langue impose des contraintes sur la référence de ses unités lexicales et les distingue en spécifiant le **type** d'entité que chacune peut désigner. Un type correspond à l'ensemble des propriétés qu'une entité doit représenter pour être le référent d'une unité lexicale donnée. Par exemple, pour être le référent de l'unité lexicale *homme*, une entité doit posséder un ensemble de propriétés, parmi lesquelles se trouvent *être humain* et *de sexe masculin*. Le type d'une unité lexicale reflète donc la réalité, mais ne réfère pas directement au monde.

Milner distingue deux types de référence : la *référence virtuelle* correspond à l'ensemble des propriétés qu'une unité lexicale possède tandis que la *référence actuelle* correspond à l'entité du monde associée à cette unité.

La référence virtuelle correspond donc à ce que l'on trouve dans une définition de dictionnaire : elle consiste en la description des propriétés d'une unité lexicale en fonction du monde réel. Toute unité lexicale a une référence virtuelle, mais pour avoir une référence actuelle, elle doit être mise en contexte. C'est dans l'emploi qu'une unité lexicale peut désigner une entité dans le monde : isolé, le mot *lion* ne sert qu'à délimiter un type d'entité. Les entités du monde ne sont donc pas associées à des unités individuelles, mais à des combinaisons d'unités lexicales : par exemple à des séquences qui composent des groupes nominaux ou verbaux. Pour illustrer ce point, prenons le mot *lion*, qui n'est pourvu que d'une référence virtuelle. Celle-ci représente l'ensemble des propriétés telles que *animal*, *carnivore*, *mammifère*, *mâle portant une crinière* etc. Seul, ce mot ne désigne aucune entité particulière. En revanche, lorsqu'il s'agit d'un groupe nominal (avec l'ajout d'un déterminant), comme *ce lion*, il y a référence actuelle. Dans ce cas, le groupe nominal fait référence à une entité dans le monde qui possède l'ensemble des propriétés d'un *lion*.

Quand on emploie plusieurs unités en séquence, leurs références virtuelles se combinent pour contraindre une référence actuelle possible. La référence actuelle en question correspond à la combinaison de ces contraintes, et non pas à chacune des unités de la séquence. Par exemple, le groupe nominal *le chasseur de lions* consiste en la combinaison des références virtuelles de *chasseur* et de *lion* (cela ne tient pas compte de la relation entre les deux qui est établie par la préposition). Le déterminant donne au groupe nominal sa référence actuelle.

On peut décrire la référence des verbes de la même manière que celle des noms. Parsons (1989), propose que les verbes eux-mêmes représentent des classes d'actions ou d'états et non pas des instances particulières. Selon les termes de Milner, les verbes isolés ne sont pourvu que d'une référence virtuelle. Ce n'est que lorsqu'un verbe a un temps et qu'il est accompagné d'arguments (au moins un

sujet *syntactique*), qu'il acquiert une référence actuelle. Pour les verbes, comme pour les noms, la référence actuelle se rapporte à la combinaison d'unités qui forment un groupe verbal. Le verbe *marcher*, par exemple, désigne un type d'action, un déplacement d'un endroit à un autre par un mouvement particulier des pieds. Mais *Jean marche*, désigne plutôt une instance particulière de ce type d'action, située temporellement dans le présent avec un actant qui s'appelle *Jean*. C'est un événement particulier et il y a référence au monde réel.

Dans le groupe verbal, les groupes nominaux (qui constituent les actants du verbe) et le verbe lui-même (éventuellement doté de propriétés de temps, d'aspect ou de mode) ont chacun leur propre référence virtuelle. En plus de ces propriétés, les verbes imposent des contraintes sur le type de leurs arguments. Par exemple, dans son interprétation habituelle, le verbe *accoucher* attend un sujet humain et féminin. Ainsi, la référence virtuelle d'un verbe impose des contraintes sur les références virtuelles de ses arguments. Le groupe entier représente donc la combinaison de plusieurs références virtuelles et sa référence actuelle consiste en la désignation d'une partie du monde, de façon analogue à celles des séquences nominales.

Avec cette définition de la notion de référence à l'esprit on procédera ensuite à une considération de deux relations qui en dépendent : la **coréférence** et l'**anaphore**.

### 2.1.2. La coréférence

En langue, il est possible que deux unités lexicales ou séquences d'unités fassent référence à la même chose. La **coréférence** est une relation linguistique qui s'établit entre deux unités lexicales qui ont la même référence. De la même manière qu'on distingue deux types de référence, on distingue *coréférence virtuelle* et *coréférence actuelle*.

Quand deux unités lexicales partagent un même référent, il y a **coréférence actuelle**. Dans ce cas elles désignent une même entité dans le monde. La coréférence actuelle est une relation symétrique entre deux unités et implique l'identité matérielle absolue des référents, mais pas forcément l'identité des unités elles-mêmes :

(1) Marie<sub>i</sub> est tombée dans la baignoire. [Cette fille]<sub>i</sub> est très maladroite.

Dans cet exemple, il y a coréférence actuelle entre les groupes nominaux *Marie* et *cette fille* car ils réfèrent à la même entité, même si les unités lexicales ne sont pas identiques.

Lorsque deux unités lexicales distinctes ont les mêmes propriétés, elles sont en relation de **coréférence virtuelle**. Cependant deux unités lexicales ne peuvent



avoir exactement le même ensemble de propriétés à moins d'être identiques. Comme le dit Milner, on peut douter de l'existence de la synonymie lexicale absolue dans les langues naturelles. Cela voudrait dire que la relation de coréférence virtuelle ne peut exister qu'entre une unité ayant des propriétés lexicales spécifiées et une autre qui n'en a pas. Ce sont les pronoms de troisième personne qui sont dépourvus de référence virtuelle. Il n'est pas possible d'attribuer au pronom *en*, hors contexte, un ensemble de propriétés qui permettrait de lui associer un type d'entité dans le monde. En revanche, en contexte il devient interprétable :

(2) J'ai tué dix des lions et toi tu en as tué six.

Dans cet exemple on a un groupe nominal, *des lions*, pourvu de références virtuelle et actuelle. Le pronom *en* entretient une relation de coréférence virtuelle avec le groupe nominal. Le pronom possède donc toutes les propriétés lexicales du groupe nominal *des lions* (coréférence virtuelle). Il désigne également les mêmes entités du monde (coréférence actuelle).

On a introduit deux types de relation de coréférence qui dépendent de la notion de référence. On procède maintenant à une discussion de la deuxième relation qui nous intéresse, l'anaphore.

### 2.1.3. L'anaphore

Dans un discours, à l'oral ou à l'écrit, on fait souvent référence à un même objet, fait, action ou événement de façon répétitive. Mais on ne l'évoque pas toujours de la même façon. Si on parle de *la voiture rouge de Monsieur Dupond*, on utilisera un pronom à chaque fois qu'on voudra en dire quelque chose : *elle a coûté cher*. Ceci évite la répétition inutile d'informations et assure la cohérence de notre discours. L'usage de pronoms pour la reprise d'éléments mentionnés précédemment dans un discours met en jeu la relation d'**anaphore**.

L'anaphore existe entre deux unités lexicales quand l'interprétation de l'une nécessite la présence de l'autre. Milner considère que les pronoms sont dépourvus de référence virtuelle propre, qu'ils ne sont pas *référentiellement autonomes*. Cela veut dire qu'ils ne sont pas capables de déterminer leur propre référent. Ils ne désignent pas de façon indépendante des entités du monde. On a vu qu'un pronom hors-contexte n'est associé à aucun ensemble de propriétés lexicales, mais qu'il peut être interprété en contexte. Il doit obtenir sa référence virtuelle à partir d'un autre élément dans le discours.

La relation d'*anaphore pronominale*<sup>2</sup> combine la coréférence avec une relation

---

2 Il existe d'autres types de relation anaphorique, tels que l'anaphore nominale où l'antécédent et l'anaphore sont des groupes nominaux, mais nous n'examinerons que l'anaphore pronominale.

asymétrique de reprise qui existe entre deux termes hétérogènes (groupe nominal – pronom ou nom – pronom), l'un étant référentiellement autonome, l'autre pas.

L'anaphore repose toujours sur la coréférence virtuelle ; parfois elle repose également sur la coréférence actuelle :

(3) Paul s'est levé puis il a pris son petit-déjeuner.

Le pronom *il* est ici en relation de coréférence actuelle avec le nom propre *Paul*, qui a une référence actuelle. Les deux références actuelles sont identiques. Le nom donne une référence virtuelle au pronom, qui en était dépourvu.

(4) J'ai mangé trois gâteaux et toi tu en as mangé cinq.

Dans ce cas, le pronom partage la référence virtuelle du nom *gâteaux*. En revanche, il ne partage pas la référence actuelle du groupe nominal *trois gâteaux*, car il ne s'agit pas des mêmes gâteaux. Il y a donc une relation de coréférence virtuelle entre le pronom et le nom qu'il reprend, mais pas de coréférence actuelle.

Ces deux exemples comportent une unité dépourvue de référence virtuelle, l'*anaphore*, et une unité lexicalement spécifiée qui la précède et qui a une référence virtuelle : l'*antécédent*.

Seule une unité lexicalement spécifiée peut jouer le rôle d'antécédent. Les pronoms, étant dépourvus de référence virtuelle propre, ne peuvent pas en fournir une, et ne peuvent donc pas être l'antécédent dans une relation anaphorique. Cependant, dans les chaînes de référence, où plusieurs pronoms anaphoriques ont le même antécédent, ils entretiennent une relation de coréférence.

Il est important de ne pas oublier la distinction entre la relation d'anaphore et celle de coréférence. Contrairement à la coréférence, l'anaphore est une relation asymétrique qui existe entre un élément anaphorisé (l'antécédent) et un deuxième élément anaphorisant (l'anaphore). Les deux relations peuvent coexister, et tel est le cas quand il y a anaphore pronominale, ce qui implique toujours une coréférence virtuelle. C'est justement grâce à la coréférence virtuelle qu'on arrive à interpréter le pronom dans une relation anaphorique.

## **2.2. Entités abstraites**

Comme on vient de voir, la référence porte sur la désignation d'entités du monde par des mots de la langue. Ces entités peuvent être de nature abstraite aussi bien que concrète. On appellera désormais *entité concrète* tout « objet » du monde pouvant être désigné et perçu au moyen de nos sens. Par contraste, les *entités abstraites* sont des concepts, des idées et des propositions, que l'on ne perçoit que grâce à notre faculté de raisonnement, à notre esprit. Mais on peut constater que cette division est peu fine, et qu'il existe des entités qui ne sont ni complètement concrètes, ni complètement abstraites. Prenons l'exemple des

situations, des états, ou des faits, qui dépendent du monde concret mais qui ne sont pas tangibles de la même façon que les entités concrètes. De même, les actions et les événements mettent en jeu des entités concrètes, et portent sur le monde que l'on perçoit. La question qui se pose concerne la façon de formuler une ontologie qui contienne toutes ces entités en les distinguant les une des autres.

### 2.2.1. Les origines : Davidson et les événements

Selon Donald Davidson (1967), comprendre les phrases d'action revient à évaluer les valeurs de vérité de la forme logique qu'elles représentent. Dans l'ontologie de Davidson les actions et les événements<sup>3</sup> sont des entités qui se rapprochent des objets. Les événements ont lieu à un certain moment dans le temps et ils ont une localisation spatio-temporelle. De plus, ils sont dénombrables et quantifiables. Ils se prêtent aussi à la « redescription », c'est-à-dire qu'en langue naturelle on peut employer divers modificateurs pour les qualifier ou pour décrire la façon dont ils se produisent. Un événement peut se passer *dans* un endroit particulier, à un moment spécifique. Une action peut être effectuée *avec* ou *par* un objet et *d'une* manière ou une autre. La forme logique des phrases qui décrivent des événements doit rendre compte de ces propriétés. Davidson adopte le principe de **réification** sur les événements et montre qu'un argument événementiel est présent dans la forme logique des phrases d'action. Les événements sont alors représentés en forme logique par une variable quantifiable.

La théorie davidsonienne diffère d'autres points de vue philosophiques, qui interprètent les verbes comme des prédicats logiques à  $n$  places. Pour illustrer la différence, prenons la phrase *Marie a frappé Jean dans la cuisine*. Intuitivement on pourrait admettre que sa forme logique est  $F(a,b,c)$ , où  $a$  est « Marie »,  $b$  est « Jean »,  $c$  est « la cuisine », et  $F$  est le prédicat à trois places « ...frapper...dans... ».

Pour Davidson, cette phrase équivaut logiquement à « il existe un événement  $e$  tel que  $e$  est une frappe de Jean par Marie et  $e$  a eu lieu dans la cuisine ». Sa forme logique serait plutôt :

$$\exists e(\text{frappe}(e, \text{Marie}, \text{Jean}) \wedge \text{dans\_la\_cuisine}(e) \wedge \text{passé}(e))$$

Ici l'événement lui-même est une **entité** (un individu) représentée par une variable sous la portée d'un quantificateur existentiel. Des propriétés de localisation physique et temporelle lui sont attribuées grâce à des prédicats supplémentaires.

---

3 Davidson ne fait pas de distinction entre une action et un événement. Le fait qu'un acte soit intentionnel ou pas n'a pas d'importance pour sa théorie. On parlera désormais d'événement pour regrouper les deux et d'action lorsqu'on parle d'un événement intentionnel.

Davidson montre que l'introduction d'un argument événementiel dans la forme logique évite de multiples difficultés rencontrées par d'autres théories philosophiques ou linguistiques. Depuis la parution de cet article fondateur, cette théorie a été largement adoptée par les communautés philosophiques et linguistiques.

### 2.2.2. Une typologie des entités abstraites : généralisation

Nicolas Asher (1993) adopte le principe davidsonien de réification des événements et le généralise pour d'autres types d'entités. Il considère non seulement les **événements** comme des individus, mais aussi les **états**, les **situations**, les **faits** et les **propositions**. Il élabore une typologie (voir Figure 1) des entités et les place sur une échelle d'abstraction selon leur niveau d'ancrage dans le monde (« scale of world immanence »). Il examine aussi les différentes façons dont on fait référence aux entités abstraites en langue naturelle.

Sur l'échelle, les entités concrètes sont à un extrême et les entités purement abstraites à l'autre. Entre les deux extrêmes se trouvent, entre autres, des entités semi-abstraites telles que les états, les événements, les situations et les faits. Afin d'aboutir à un tel classement, Asher a considéré les propriétés métaphysiques qui indiquent le niveau d'ancrage dans le monde d'une entité – en particulier ses propriétés causales et spatio-temporelles.

Les entités concrètes sont localisées dans le monde et ont des propriétés spatiales délimitées<sup>4</sup>. Elles sont représentées en langue par un groupe nominal (N<sup>n</sup>) : *sa voiture, cette pomme, l'homme, dix lions* ou encore *Paul, le maire de Paris, Jean-Claude van Damme*.

D'après Bach (1986) on regroupe sous le terme *éventualité* les événements et les états, qui sont des entités semi-abstraites ayant une portée sur le monde réel. Cependant, les **états** s'opposent aux événements car ce sont des entités homogènes qui ont une durée, mais dont on ne peut pas demander combien de temps ils ont pris, ou s'ils se sont terminés. De plus, les états sont des entités qui ne représentent pas un changement dans le monde. Les phrases *Jean se repose, Max connaît le chemin* et *la voiture est rouge* sont toutes des descriptions d'états.

Les **situations** consistent en la description d'un événement, d'un état des choses. Ce sont des entités qui font l'objet de la perception – on peut les percevoir au moyen de nos sens : les voir, les entendre, les sentir etc. Sur le plan linguistique ce type d'entité est défini par les types de complément que peuvent prendre les verbes de perception. Amsili et al. (2005) observent qu'en français, les

4 Les propriétés temporelles des entités concrètes ne sont pas aussi facilement délimitées, on ne repère pas facilement le début ou la fin de leur existence, mais ce n'est pas dans nos préoccupations ici.

situations sont représentées par des propositions infinitives comme dans *Jean a vu les gens démanteler le mur de Berlin*, mais aussi par des complétives comprenant un participe présent, *Jean a vu les gens démantelant le mur de Berlin*, et des relatives comme *Jean a vu les gens qui démantelaient le mur de Berlin*.

Les **propositions** sont des entités purement abstraites qui dénotent un ensemble de mondes possibles. Une proposition ne représente pas un énoncé dont on suppose la véracité, mais la valeur de vérité d'une proposition dépend du modèle du monde dans lequel on se place. D'un point de vue linguistique, on peut considérer que les propositions sont généralement des subordonnées complétives de verbes d'attitude propositionnelle (*croire, penser, prétendre* etc). La subordonnée dans la phrase *Jean croit que Marie est malade* est une proposition. D'autres exemples sont *Il est faux que le mur de Berlin est tombé* et *Léa pense que Max est beau*.

Les **faits** sont aussi des entités ayant une portée sur le monde concret (ou plutôt qui en dépendent), mais qui consistent cette fois en une affirmation de la véracité de certaines conditions du monde. *Le mur de Berlin est tombé en 1989* est un fait. Il s'agit donc de propositions qui sont considérées vraies. Toute phrase affirmative peut être considérée comme un fait.

Enfin, les **événements** sont des entités abstraites qui ont des propriétés spatio-temporelles et des effets causaux. Un événement est donc localisable dans le monde du fait qu'il a des participants – un agent/sujet qui le déclenche à un instant donné et éventuellement un thème/objet qui le subit, etc.

Afin de définir les événements comme entités abstraites, Asher a procédé à une étude des contextes d'apparition et leurs emplois en anglais. Amsili et al. (2005) ont ébauché un travail similaire pour le français. Dans les deux cas, la notion de *conteneur* de Zeno Vendler (1957) a été empruntée. Cette notion consiste en la délimitation des contextes linguistiques et des constructions qui demandent la présence d'un événement. Il s'agit de délimiter les expressions verbales qui sous-catégorisent un argument événementiel. Pour les événements, en français, ce sont les verbes *arriver, se produire, avoir lieu, se passer*, etc. qui prennent tous un événement, obligatoirement sous forme nominale, en position sujet.

(5) La chute du mur de Berlin est arrivée/s'est produite/a eu lieu hier.

Les verbes tels que *assister à, être témoin de, et manquer* attendent un objet événementiel, aussi sous forme nominale :

(6) Jean a assisté à la chute du mur de Berlin.

Les constructions clivées (7), pseudo-clivées (8) et impersonnelles avec des verbes comme *arriver, se produire* etc. (9) sont aussi des conteneurs qui requièrent l'intervention d'un événement, mais dans ces cas, l'argument événementiel est sous la forme d'une complétive, introduite par la conjonction de subordination *que*.

- (7) Ce qui s'est passé, c'est que le mur de Berlin est tombé.
- (8) C'est en 1989 que le mur de Berlin est tombé.
- (9) Il s'est passé que le mur de Berlin est tombé.

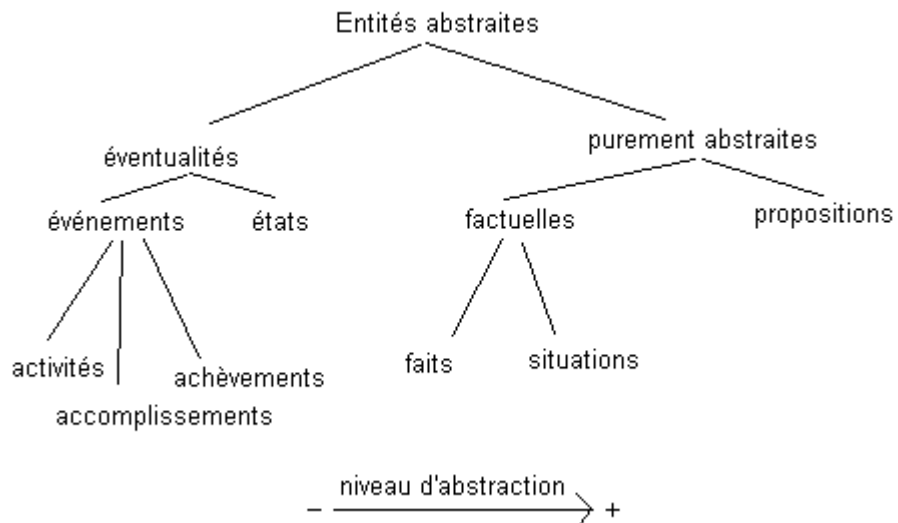


Figure 1 : L'échelle d'abstraction d'Asher (1993).

On a maintenant une idée de l'ontologie des différentes entités, et en particulier des différents types d'entités abstraites. La distinction entre les types d'entités a été établie. On a également repéré certains contextes linguistiques qui exigent l'intervention d'un argument événementiel.

### 3. Anaphore événementielle – étude linguistique

L'anaphore événementielle consiste en la reprise, par une expression pronominale, d'un événement qui apparaît précédemment dans le discours. L'antécédent dans cette relation consiste en la description d'un événement, qui peut être de forme nominale, *la relégation du PSG*, ou phrastique, *(que) le PSG a été relégué*. Dans le cadre de cette étude on se concentrera sur les antécédents phrastiques puisqu'il existe déjà des travaux portant sur les anaphores à antécédents nominaux.

#### 3.1. Formes du matériau de reprise

Dans cette partie, on examinera les formes de pronoms qui peuvent faire référence à un événement sous forme phrastique. On reviendra plus en détail sur la notion de conteneur événementiel, qui s'avérera vitale pour la sélection d'un antécédent de type événementiel par un pronom.

##### 3.1.1. Les pronoms *cela, ça, ce, le, y, en*

Un pronom, comme son nom l'indique, est l'unité lexicale qui peut prendre la place d'un nom dans le discours. Cependant, les pronoms ne fonctionnent que rarement comme l'équivalent du nom isolé qu'ils remplacent. Le plus souvent, ils remplacent un groupe nominal entier, avec ou sans modifieurs. Les pronoms anaphoriques sont ceux qui reprennent en partie ou entièrement les propriétés lexicales d'un élément déjà évoqué. Ils nous permettent d'assurer une cohérence à notre discours sans répéter à chaque fois les entités discursives que l'on évoque<sup>5</sup>.

Les pronoms ne font pas uniquement référence à des entités sous forme nominale. Les anaphores abstraites notamment reprennent des entités sous forme de phrase simples. Il nous importe, dans le cadre de cette étude, d'examiner les pronoms anaphoriques qui peuvent avoir un statut événementiel. Ce sont les pronoms qui ont comme antécédent un événement exprimé sous forme phrastique.

Les phrases dans leur ensemble sont dépourvues de traits morphologiques de nombre et de genre. On peut alors supposer qu'un pronom qui anaphorise un élément phrastique doit aussi être dépourvu de ces traits morphologiques afin d'assurer l'accord nécessaire pour la reprise de son antécédent. En français, ce sont les pronoms tels que *cela* et sa forme abrégée *ça*, mais aussi *ce* et *le* qui sont

<sup>5</sup> Voir 1.2 sur l'anaphore, ci-dessus.

neutres, ainsi que *y* et *en*. Les pronoms qui nous intéressent sont ceux qui peuvent apparaître en position argumentale de conteneurs événementiels<sup>6</sup>. Cela veut dire qu'ils peuvent être sujet des verbes comme *arriver*, *se produire* et *avoir lieu* ou bien en position objet de verbes comme *manquer*, *rater*, *assister à*, et *être témoin de*. On dira que les pronoms en position argumentale de tels verbes sont *sous la portée* d'un conteneur événementiel. Regardons quelques exemples pour illustrer la situation. On note la relation anaphorique par des indices alphabétiques et on délimite l'entité événementielle (l'antécédent) avec des crochets :

(10) [Le mur de Berlin est tombé]<sub>i</sub>. C<sub>i</sub> s'est arrivé en 1989.

(11) [La falaise s'est écroulé]<sub>i</sub>. Cela<sub>i</sub> s'est produit très vite.

(12) [André a présenté son algorithme]<sub>i</sub>. C'est dommage d'avoir raté ça<sub>i</sub>.

(13) [Léa a embrassé Max]<sub>i</sub>. Paul en<sub>i</sub> a été témoin.

(14) [Oswald a assassiné Kennedy]<sub>i</sub>. Tout le monde y<sub>i</sub> a assisté.

On remarque que le pronom *cela/ça* apparaît généralement en position sujet, mais qu'il peut aussi être objet direct. Le pronom *ce* ne peut occuper que cette place, alors que les pronoms *le*, *y* et *en* ne peuvent apparaître qu'en position d'objet antéposé.

Le statut événementiel des pronoms *en*, *y* et *le* est discutable, mais on peut avancer qu'ils représentent respectivement des formes de *de ça*, *à ça* et *ça* objet postposé. De plus ces pronoms sont la plupart du temps employés pour reprendre des entités nominales, alors que les pronoms *cela/ça* et *ce* sont les plus fréquents pour la reprise d'éléments phrastiques. On se concentrera, pour ces raisons, sur ces derniers.

### 3.1.2. Conteneurs des expressions anaphoriques événementielles

La notion de *conteneur événementiel* de Vendler, que l'on a évoquée ci-dessus, regroupe les verbes qui requièrent un sujet événementiel (*se passer*, *se produire*, *arriver* et *avoir lieu*), ou un objet événementiel (*assister à*, *être témoin de*, et *manquer*). Les arguments de ces verbes doivent être soit sous forme nominale, *la chute du mur de Berlin a eu lieu*, soit pronominale, *ça a eu lieu*. Le conteneur détermine le type d'entité qu'un pronom peut reprendre. Prenons un exemple :

(15) Jan était content que le mur de Berlin tombe. Ça s'est passé en 1989.

Il y a plusieurs types d'entités différentes ici – un état exprimé par le verbe *être content* et un fait, qui consiste en la description d'un événement, représenté par la complétive. Le conteneur *se passer* nécessite un sujet événementiel. Le pronom en position sujet de ce verbe doit anaphoriser un événement pour assurer la cohérence du discours. Le seul choix est alors *le mur de Berlin tombe*.

---

<sup>6</sup> Voir 3.1.2. pour une description de la notion de conteneurs événementiels.



Il existe d'autres types de conteneurs, qui requièrent la présence d'autres entités. Le verbe *croire*, par exemple, est un conteneur propositionnel. Il implique une incertitude, son complément n'est pas forcément considéré vrai. Il ne consiste pas en l'assertion d'un énoncé par rapport au monde, mais juste en la description d'un état psychologique de son sujet. On peut *croire*, *espérer* ou *penser* toute sorte de choses, sans qu'elles ne soient vraies.

Par contraste, le verbe *savoir* implique la véracité de son argument. Savoir une chose implique que cette chose est vraie. Le verbe *savoir* est un conteneur factuel. De même pour les expressions qui impliquent la vérité de leurs arguments, telles que *être vrai que* et *être un fait que*.

(16) Jean sait que Marie lui ment.

(17) Il est vrai que le Président de la République est sourd.

(18) C'est un fait que deux Français ont bombardé le Rainbow Warrior en 1985.

Les verbes tels que *connaître* (et encore *savoir*), *détester*, et *aimer* expriment des états, en l'occurrence des états psychologiques ou émotionnels. D'autres verbes, comme *contenir*, *posséder* et *avoir* expriment aussi des états tout comme le verbe *être*, notamment lorsqu'il est suivi d'un adjectif. Les phrases suivantes représentent toutes des états (mais aussi des faits ou des propositions) :

(19) Chirac connaît Villepin.

(20) La boîte contient des chocolats.

(21) Max possède une voiture.

(22) Le Président de la République est sourd.

Les conteneurs s'avèrent utiles car ils permettent de sélectionner le type d'antécédent d'un pronom. Les cas de reprise anaphorique suivants sont alors impossibles car ils engendrent une incohérence entre un conteneur événementiel et un antécédent de type différent :

(23) #Léa croit que [la lune est verte]<sub>i</sub>. Ça<sub>i</sub> arrive souvent. (proposition)

(24) #Max sait que [la Terre est ronde]<sub>i</sub>. Ça<sub>i</sub> arrive souvent. (fait)

(25) #[Jean connaît Marie]<sub>i</sub>. Ça<sub>i</sub> arrive souvent. (état)

D'un autre côté, les exemples de reprise suivants sont acceptables, car les conteneurs sont compatibles :

(26) Max sait [que la Terre est ronde]<sub>i</sub>. C'<sub>i</sub>est un fait bien connu.

(27) Léa croit [que la lune est verte]<sub>i</sub>, mais cela<sub>i</sub> est faux.

Dans le premier exemple, le conteneur du pronom est la construction *être un fait* qui demande un complément de nature factuel. En remplaçant le pronom par l'antécédent, on obtient :

(28) Que la terre est ronde est un fait bien connu.

(29) \*La terre est ronde est un fait bien connu.

De la substitution de la proposition subordonnée enchâssée résulte une phrase

agrammaticale, car l'expression *être un fait* demande une complétive (ou un sujet nominal). L'antécédent doit être une proposition grammaticale, et dans ce cas correspondre à la complétive elle-même.

Dans le deuxième exemple, le conteneur est l'expression *être faux*, qui attend un argument de nature propositionnelle. Par substitution de l'anaphore on obtient :

(30) Que la lune soit<sup>7</sup> verte est faux.

(31) \*La lune est/soit verte est faux.

Comme ces expressions verbales demandent une complétive, il n'est pas possible de la substituer par une phrase simple. Les pronoms, dans ces exemples, reprennent une complétive de type propositionnel ou factuel<sup>8</sup>.

Pour les états, la situation est moins nette, comme on le voit à travers les deux exemples suivants :

(32) [Jean connaît Marie]<sub>i</sub> et ce<sub>i</sub> depuis longtemps.

(33) [Jean est malade]<sub>i</sub>. Ça<sub>i</sub> lui arrive souvent.

Un conteneur événementiel peut aussi avoir sous sa portée un état. La différence, qu'il est important de noter, réside dans la nature du verbe et non pas dans la nature du conteneur. Il est donc nécessaire de distinguer les verbes statifs des verbes événementiels, car même si le conteneur peut être une bonne indication, il y a des cas d'exception.

La présence d'un conteneur événementiel fournit des contraintes sémantiques sur la nature de l'antécédent potentiel et aide à distinguer un bon candidat d'un mauvais. Un conteneur événementiel exclut les antécédents propositionnels et factuels, mais peut avoir sous sa portée, dans certains cas, un état<sup>9</sup>. Dans ces cas, il faut recourir à d'autres indices, notamment la nature du verbe, afin de pouvoir distinguer un antécédent statif d'un antécédent événementiel.

### **3.2. Forme de l'antécédent**

Une fois que l'on a établi la nature d'un pronom, il importe de trouver un antécédent avec lequel il est compatible. Dans cette section on examinera les formes que peut prendre un antécédent événementiel, sachant déjà que l'on ne considère que les antécédents phrastiques, et non les antécédents nominaux. On prendra en compte les propriétés lexicales des verbes qui dénotent des événements, ainsi que certaines des constructions syntaxiques dans lesquelles ils

7 Pour que la phrase reste grammaticale il faut changer le mode du verbe au subjonctif.

8 Les complétives en position sujet, comme on le verra plus tard, ne représentent que des propositions. Le complémenteur *que* en début de phrase peut être considéré comme un conteneur propositionnel.

9 Cela renforce le principe de regroupement des événements et des états sous le terme d'*éventualité*.

peuvent apparaître. On terminera par une brève étude d'éléments de nature aspectuelle, les temps verbaux et la modalité, en étudiant leur effet sur la relation d'anaphore.

### 3.2.1. Les phrases événementielles

Un événement selon la typologie que l'on vient de discuter est une entité individuelle qui a des propriétés spatio-temporelles. Cela suggérerait que, pour désigner un événement, un groupe verbal doit être tensé, puisqu'à l'infinitif un verbe est dépourvu de temps. Un verbe ou groupe verbal à l'infinitif représente plutôt un type d'action ou d'événement, mais ne peut désigner une instance. En termes de référence, cela veut dire qu'un infinitif a uniquement une référence virtuelle alors qu'un groupe verbal fléchi peut représenter une instance particulière ayant une référence actuelle. *Manger à midi* et même *Max manger à midi*<sup>10</sup> représentent un type d'événement, alors que *Max a mangé à midi* est une instance particulière.

Un candidat antécédent événementiel comportera soit un verbe temporellement fléchi, représentant une instance d'événement, soit un verbe à l'infinitif, qui représentera uniquement une anaphore virtuelle.

Un point crucial à noter est que n'importe quel verbe ne peut pas désigner un événement. Traditionnellement, en linguistique et en philosophie, on divise les phrases simples en deux classes. Vendler, et puis plus tard Terrence Parsons (1989), a classifié les verbes de l'anglais selon qu'ils décrivent des états ou des événements.

Les états sont des entités abstraites qui « tiennent »<sup>11</sup> ou pas à un moment donné, qui n'ont pas de fin explicite et qui n'engendrent pas de changement perceptible. La phrase *la voiture est rouge* en est un exemple, ainsi que *la syntaxe est ennuyeuse*. Ces phrases sont simplement des constats. La première décrit une propriété de la voiture, et la deuxième décrit une propriété de la syntaxe (telle qu'elle est perçue par le locuteur). Ces phrases n'impliquent pas un changement des choses. Il n'y a pas de sens à demander combien de temps un état a pris, mais on peut très bien demander combien de temps il a duré. Ainsi, la question *ça fait combien de temps que la voiture est rouge ?* est naturelle, alors que *combien de temps la voiture a mis pour être rouge ?* ne l'est pas. En plus, on ne peut pas

---

10 On rappelle que les constructions à montée et à contrôle ont un infinitif avec un sujet. C'est aussi le cas des phrases qui décrivent des situations. C'est l'absence de temps verbal qui empêche une référence actuelle et pas l'absence d'une localisation temporelle. On ne traitera pas ces constructions et on se concentrera plutôt sur les verbes fléchis.

11 Un meilleur terme n'a pu être trouvé en français qui capture la propriété des états d'exister. Une adaptation de l'anglais *to hold* est utilisée.

demander si un état est terminé : *quand est-ce que la voiture a fini d'être rouge ?*

Les événements sont opposés aux états car ils engendrent un changement (ils ont des effets causaux). Les verbes événementiels sont encore sous-classifiés en *activités, accomplissements et achèvements*.

Les **activités** n'ont pas de fin ni de but explicites. Elles expriment des périodes de temps non uniques et indéfinies. Les phrases simples comme *Jean peint, l'avion vole* et *Schumacher roule vite* sont des exemples d'activités. Les verbes ou expressions verbales qui décrivent des activités peuvent ne pas avoir de complément, ou avoir un complément indéfini. La phrase *Jean mange* exprime une activité. En revanche, *Jean mange une pomme* est un événement singulier, un accomplissement, car, quand qu'il aura fini de manger la pomme, l'événement sera terminé. *Jean mange des pommes* (on ne sait pas combien) est une activité, car la fin de l'acte n'est pas explicite.

Lorsque l'objet d'une activité est exprimé, elle devient un accomplissement. L'objet désigne le but de l'acte. Les **accomplissements** impliquent des périodes de temps uniques et définies. *Jean peint* dénote une activité alors que *Jean peint la maison* décrit un accomplissement. Les verbes qui ont un but explicite, comme les verbes d'accomplissement, sont *téliques*. Les verbes d'activités sont *atéliques*.

La distinction entre les accomplissements et les achèvements est assez fine. La phrase *l'ambassadeur est arrivé à Moscou* dénote un achèvement. Les **achèvements** ne marquent pas la durée, alors que c'est le cas pour les accomplissements. Un accomplissement peut avoir une durée, de nature télique ou atélique :

(34) Jean a peint la maison pendant une heure. (atélique)

(35) Jean a peint la maison en une heure. (télique)

Un achèvement n'est compatible avec l'expression de la durée :

(36) \*L'ambassadeur est arrivé pendant une heure à Moscou.  
(durée de l'arrivée)

(37) \*L'ambassadeur est arrivé en une heure à Moscou.  
(durée de l'arrivée)

*Peindre* est une action que l'on peut faire pendant une certaine durée alors que *arriver* ne l'est pas.

Les verbes psychologiques ou de connaissance, comme *penser, savoir* et *connaître*, ne peuvent décrire que des états, qui n'impliquent aucun changement et n'ont pas d'effets causaux. A un moment donné, un état tient ou ne tient pas, il n'y a pas de milieu. Alors, la phrase *Jean pense que la terre est ronde* ne peut pas être la description d'un événement car elle ne consiste qu'en la description d'un état psychologique de son sujet. Les verbes intentionnels, tels que *vouloir, désirer, avoir envie, souhaiter*, de façon similaire, ne désignent que des états psychologiques. C'est aussi le cas des verbes comme *aimer, adorer, détester* et

*hair*.

De même, les verbes de possession *avoir*, *posséder*, et *appartenir* ainsi que les verbes *habiter*, *rester*, *demeurer* et *exister* ne peuvent pas désigner d'événements car ils n'impliquent pas un changement dans le monde. Ce sont des verbes statifs. *Philippe a une 206 rouge* ne désigne pas un événement, mais un état. La phrase n'implique aucun changement et l'état n'a pas d'effets causaux. En l'occurrence c'est aussi la description d'un fait. Un autre verbe statif est *être*. *Le plomb est plus dense que le fer* est la description d'un état contenue dans un fait. Par contraste, *la lune est faite de camembert* est la description d'un état contenue dans une proposition.

Type	Télique	Durée	Exemples
État	non	pendant une heure *en une heure	être + ADJ, penser, savoir, croire, aimer, connaître
Activité	non	pendant une heure *en une heure	travailler, peindre, ronfler, chanter, courir, manger
Accomplissement	oui	*pendant une heure en une heure	aller à + GN, manger + GN, raconter + GN, peindre + GN, chanter + GN †
Achèvement	oui/non	*pendant une heure *en une heure	atteindre, gagner, perdre, réussir, mourir, reconnaître, dire

Tableau 1 : La classification des verbes selon Vendler. † GN non partitif.

Les descriptions d'événements que l'on considère comme antécédents consistent donc en un verbe événementiel tensé, qui aura des arguments et d'éventuels modifieurs. Le marqueur principal est celui du verbe événementiel, qui indique à lui seul la présence de la description d'un événement dans le discours.

### 3.2.2. Les contrefactuelles

La proposition subordonnée circonstancielle, comme le nom l'indique, introduit une circonstance pour le verbe d'une proposition principale : un temps, une manière, un but, un lieu ou une condition. Les circonstanciels conditionnelles sont celles qui imposent une condition sur un verbe. On s'intéresse aux cas qui sont introduits par la conjonction de coordination *si*. Les circonstanciels conditionnelles qui présupposent la fausseté de la proposition subordonnée et de

la proposition principale sont appelées contrefactuelles<sup>12</sup>. Elles décrivent alors une situation hypothétique. Intuitivement, on serait donc tenté de considérer que les contrefactuelles ne contiennent que des entités abstraites de nature propositionnelle, même si ces dernières contiennent des événements. Ce que l'on souhaite examiner ici concerne justement le statut événementiel des énoncés contrefactuels et vise à déterminer si ce genre de construction a des conséquences pour la reprise d'événements enchâssés.

Les contrefactuelles peuvent avoir deux formes différentes, chacune avec une signification différente.

*Si + imparfait / conditionnel présent* : *Si Max venait on lui donnerait ses cadeaux*. Le verbe de la subordonnée est à l'imparfait de l'indicatif. Celui de la proposition principale (postposée) est au conditionnel présent. Cette forme exprime une situation hypothétique qui est contraire à l'état actuel des choses.

*Si + plus-que-parfait / conditionnel passé ou présent* : *Si Max était venu on serait déjà parti. Si Léa avait fait ses devoirs elle irait avec les autres*. Ici la subordonnée est au plus-que-parfait et la principale postposée est au conditionnel passé ou au conditionnel présent. La principale exprime une hypothèse sur un « fait » dans le passé (la subordonnée) qui s'est avérée fausse.

Dans les deux cas, la nature hypothétique de la conjonction *si* fait qu'un événement enchâssé dans la proposition subordonnée ne s'est pas produit. De plus, le verbe de la principale est conjugué au conditionnel, et la proposition ne consiste qu'en la description d'un événement envisagé. On doit se demander si un événement qui ne s'est pas produit consiste en une entité qui est introduite au niveau discursif, et si cette entité est disponible à la référence anaphorique ou pas. Y a-t-il des conséquences pour la reprise anaphorique d'événements enchâssés dans des contrefactuelles ?

D'abord on vérifiera si un événement dans la subordonnée peut être disponible comme antécédent.

(38) Si [Max venait]<sub>i</sub>, Léa serait contente. C<sub>i</sub>'est arrivé hier, mais elle était triste.

La subordonnée consiste en la description d'un événement alors que la proposition principale consiste en la description d'un état. Le pronom *ce* de la phrase suivante est sous la portée d'un conteneur événementiel, ce qui fait que son antécédent ne peut être que de type événementiel. La seule possibilité est la proposition *Max venait*. Mais ce n'est pas tout à fait le cas, car c'est un événement hypothétique qui ne s'est pas produit. On ne peut pas dire qu'un événement qui ne s'est pas produit est arrivé hier. Ce genre de reprise ne peut être de type *actuel*, car l'antécédent n'est pas un événement instancié. La subordonnée consiste plutôt en

---

<sup>12</sup> La littérature sur ce point est assez développée, mais on ne rentrera pas dans ce débat ici. Citons l'œuvre de David Lewis (1973) pour une discussion du sujet.

la description d'un type d'événement. Ce cas de reprise relève de l'anaphore virtuelle. Il y a coréférence virtuelle entre le pronom *ce* et l'événement enchâssé dans la subordonnée.

On examine les autres cas de figure :

(39) Si [Max était venu]<sub>i</sub>, Léa aurait été contente. C<sub>i</sub>'est arrivé hier, mais elle était triste.

(40) Si [Max était venu]<sub>i</sub>, Léa serait contente. C<sub>i</sub>'est arrivé hier, mais elle était triste.

Là encore, il s'agit d'une relation d'anaphore virtuelle où le pronom est en coréférence virtuelle avec la proposition exprimée par l'événement hypothétique enchâssé dans la subordonnée.

Maintenant, et de la même manière, on souhaite tester la reprise d'un événement enchâssé dans la principale d'une proposition contrefactuelle.

(41) Si Max savait lire, [on lui offrirait un livre]<sub>i</sub>. C<sub>i</sub>'est déjà arrivé, alors qu'il est analphabète.

Dans cet exemple c'est la subordonnée qui contient la description d'un état. Dans la principale se trouve la description d'un événement. Alors, le seul élément pouvant être antécédent du pronom *ce* dans ce cas est la proposition *on lui offrirait un livre*. On a encore affaire à un événement hypothétique, cette fois marqué par un verbe au conditionnel. La principale exprime un événement qui se produirait si la condition exprimée par la subordonnée était remplie. Comme pour le cas précédent, il ne peut y avoir coréférence actuelle entre le pronom et son antécédent événementiel. La principale exprime encore un type d'événement et non pas une instance particulière. Il y a donc une relation d'anaphore avec coréférence virtuelle entre le pronom et son antécédent.

(42) Si Max avait su lire, [on lui offrirait un livre]<sub>i</sub>. C<sub>i</sub>'est déjà arrivé, alors qu'il est analphabète.

(43) Si Max avait su lire, [on lui aurait offert un livre]<sub>i</sub>. C<sub>i</sub>'est déjà arrivé, alors qu'il est analphabète.

Les deux autres cas de figure ont le même comportement, ce qui montre que les différents temps verbaux des contrefactuelles n'ont pas d'effet sur la relation anaphorique.

Après cet examen des contrefactuelles, on peut conclure que cette construction particulière ne bloque pas la référence anaphorique à des événements se trouvant à l'intérieur de la principale ou de la subordonnée. Étant donnée la nature fortement hypothétique des contrefactuelles, la relation d'anaphore dans ces cas ne peut être que virtuelle.

### 3.2.3. Les propositions subordonnées complétives

Les propositions subordonnées complétives « complètent », dans une phrase principale, un nom, un verbe ou un adjectif. Sous ses différentes formes, une complétive peut assumer divers rôles syntaxiques. Elle peut avoir une fonction de sujet (*Que tu viennes m'agace*), d'attribut (*L'essentiel est que tu viennes*), de complément de nom (*Je ne regrette pas le fait que tu sois venu*), de complément d'adjectif (*Je suis content que tu viennes*), de complément d'objet direct (*Tu m'as dit que tu viens*) ou indirect (*Je m'oppose à ce que tu viennes*) du verbe.

La plupart de ces formes prennent un verbe au subjonctif, mode qui s'emploie le plus souvent dans des propositions subordonnées complétives, relatives ou circonstancielles. Les antécédents que l'on considère sont de nature phrastique. On se concentrera dans un premier temps sur une étude des complétives à l'indicatif, qui ont le statut de phrase indépendante. On passera par la suite aux complétives au subjonctif.

Les complétives introduisent différents types d'entités selon la nature de leur verbe principal :

- *savoir, être vrai + que* : fait
- *penser, croire, prétendre, affirmer, nier, déclarer, dire, craindre, espérer + que*: proposition
- *voir, regarder, entendre, sentir + que* : fait
- *paraître, sembler + que* : proposition
- *faire en sorte que, faire que* : événement ou état

On cherche à déterminer si les complétives peuvent introduire des événements qui sont disponibles pour la référence anaphorique, ou si la sémantique du verbe principal pose des contraintes sur la nature de la complétive en question.

(44) Pierre croit que Paul est tombé.

Dans ce cas, le verbe principal *croire* introduit une complétive contenant un événement. Isolée, la phrase *Paul est tombé* décrit une entité de type événementiel. Mais ce qui détermine qu'il s'agit d'une complétive propositionnelle est la nature du verbe principal. Le verbe *croire* est un conteneur propositionnel.

(45) Pierre croit que [Paul est tombé]<sub>i</sub>. Ce<sub>i</sub> serait arrivé hier.

Lorsque l'anaphore apparaît sous la portée d'un conteneur événementiel, l'événement enchâssé peut être sélectionné comme antécédent, et c'est le seul choix possible dans ce cas.

L'exemple suivant est similaire au précédent :

(46) Marie sait que le mur de Berlin est tombé.

Là encore, la complétive introduit la description d'un événement. La seule différence ici est que le verbe principal implique la véracité de la proposition introduite par la complétive. Si l'on *sait* quelque chose c'est que cette chose est



vraie *a priori*.

(47) Marie sait que [le mur de Berlin est tombé]<sub>i</sub>. C'<sub>i</sub>est arrivé en 1989.

La reprise anaphorique de l'événement enchâssé n'est pas problématique. Il s'agit dans ce cas d'un événement qui s'est produit et dont *Marie* sait qu'il s'est produit.

(48) Paul a vu que Marie a ouvert la porte.

Dans ce cas aussi, le complément du verbe *voir* est une complétive qui introduit un événement. Le verbe principal est un verbe de perception à complétive dont l'interprétation n'est pas la même que pour son emploi en phrase simple. La perception n'est pas directe et on peut la considérer comme l'expression d'une « attitude épistémique par rapport au fait rapporté »<sup>13</sup>, la perception du résultat d'un événement. Les verbes de perception à complétive sont en fait des verbes factuels, et leur interprétation est plutôt celle de *savoir*.

(49) Paul a vu que [Marie a ouvert la porte]<sub>i</sub>. Il a été témoin de ça<sub>i</sub>.

La reprise anaphorique dans ce cas ne pose pas de problème. L'antécédent du pronom *ça* est l'événement *Marie a ouvert la porte*. Comme le verbe *voir* ici a une interprétation factuelle, il n'est pas un candidat antécédent potentiel.

Passons maintenant aux complétives qui contiennent un verbe au subjonctif. Il est intéressant d'examiner ces constructions afin de déterminer si le mode subjonctif du verbe aura des conséquences pour la reprise anaphorique d'événements enchâssés. Les verbes tels que *s'attendre à*, *être* suivi d'un adjectif, et les expressions adjectivales en *être* avec pour sujet le *il* impersonnel, prennent tous un verbe au subjonctif. Lorsqu'une complétive a la fonction de sujet syntaxique, son verbe est aussi au subjonctif.

(50) Je m'attendais à ce que [Léa s'endorme vite]<sub>i</sub>. C'<sub>i</sub>est arrivé cette nuit.

(51) Je suis content que [Léa s'endorme vite]<sub>i</sub>. C'<sub>i</sub>est arrivé cette nuit.

(52) Il était important que [Léa s'endorme vite]<sub>i</sub>. C'<sub>i</sub>est arrivé cette nuit.

(53) Que [Léa s'endorme vite]<sub>i</sub> m'a fait plaisir. C'<sub>i</sub>est arrivé cette nuit.

Dans ces exemples, où les complétives sont introduites par des conteneurs divers, la complétive contient un événement enchâssé. La reprise anaphorique est tout à fait naturelle, car le pronom est sous la portée d'un conteneur qui oblige l'antécédent à être événementiel. Ni le type de conteneur de la complétive ni le mode subjonctif du verbe, ne posent de contraintes.

(54) [Que Léa soit malade m'a fait pleurer]<sub>i</sub>. C'<sub>i</sub>est arrivé hier.

Cependant, si la complétive occupe la position sujet d'un verbe, elle joue le rôle d'un groupe nominal et contribue, en tant qu'actant, à la description d'un événement. Dans le cas ci-dessus, la phrase entière est considérée comme antécédent du pronom. La complétive introduit la description d'un état. Que se passe-t-il si la complétive contient, elle aussi, un événement?

13 Amsili et al. (2005) p.5.

(55) [Que Léa parte a fait pleurer Max]<sub>i</sub>. C<sub>i</sub>'est arrivé hier.

(56) Que [Léa parte]<sub>i</sub> a fait pleurer Max. C<sub>i</sub>'est arrivé hier.

Dans ce cas il y a bien une ambiguïté. Le pronom *ce* peut avoir comme antécédent soit la phrase principale en entier, soit l'événement enchâssé dans la complétive.

Il faut pourtant noter que les verbes pouvant prendre une complétive en position sujet représentent un groupe restreint. En effet, une complétive de cette forme représente un élément de nature propositionnelle ou factuelle, et ce sont des entités qui n'ont pas les propriétés causales qu'ont les entités animées concrètes et les événements. On peut les faire précéder de *le fait*, *l'idée* ou bien *la proposition*. Par exemple :

(57) Le fait que Léa parte a fait pleurer Max.

On peut alors conclure que, lorsqu'une complétive occupe la position de sujet syntaxique, le complémenteur *que* a un statut de conteneur propositionnel ou factuel, mais que cela n'empêche pas la reprise d'un événement enchâssé dans la complétive.

Poursuivons maintenant avec l'expression verbale *faire en sorte que*, qui nécessite aussi un verbe au subjonctif dans sa complétive subordonnée. *Faire en sorte que* est bien un conteneur événementiel. Il dénote précisément un acte de la part de son sujet, qui implique un changement dans l'état des choses.

(58) [Jean a fait en sorte que la voiture soit rouge]<sub>i</sub>. Ça<sub>i</sub> s'est passé hier.

Ici, il n'y a aucun doute : le pronom reprend l'événement *Jean a fait en sorte que la voiture soit rouge*. La complétive consiste en la description d'un état, la combinaison *être* + adjectif étant une construction stative, qui ne peut par conséquent pas constituer l'antécédent d'un pronom sous la portée d'un conteneur événementiel.

En revanche, si la phrase contient un événement enchâssé, déterminer l'antécédent devient problématique.

(59) [Jean a fait en sorte que la voiture explose]<sub>i</sub>. Paul a raté ça<sub>i</sub>.

(60) Jean a fait en sorte que [la voiture explose]<sub>i</sub>. Paul a raté ça<sub>i</sub>.

L'événement enchâssé est disponible pour la reprise anaphorique, comme l'est la phrase entière. L'acte de *faire en sorte que* et *l'explosion* sont cause et effet, mais ils n'est pas nécessaire qu'ils soient temporellement proches. La phrase est ambiguë car dans le premier cas Jean a mis une bombe sous le capot, sans que l'explosion ait lieu, mais elle se produira ultérieurement. L'antécédent du pronom est l'acte effectué par Jean. Dans le deuxième, cas l'explosion s'est produite et c'est cet événement qui est l'antécédent du pronom.

Pour réduire l'effet de cette ambiguïté, on a recours au fait que *faire en sorte que* exprime un acte de la part d'un être animé (ou par extension un groupe d'êtres animés). L'acte de *faire en sorte que* quelque chose se produise relève de

l'intentionnalité, une propriété que l'on associe à des êtres ayant une volonté. La reprise anaphorique serait plus naturelle dans un conteneur qui exprime cette propriété du sujet. On admettrait plus facilement la combinaison de phrases suivante, où le pronom *ça* est en position d'objet direct du verbe *faire* :

(61) [Jean a fait en sorte que la voiture explose]<sub>i</sub>. Il a fait ça<sub>i</sub> hier.

Ou encore avec le pronom *le* en position d'objet direct :

(62) [Jean a fait en sorte que la voiture explose]<sub>i</sub>. Il l<sub>i</sub>'a fait hier.

L'événement enchâssé n'est pas un antécédent possible dans ces cas. L'antécédent préféré dans les constructions avec *faire en sorte que* est l'événement enchâssé quand le sujet du verbe événementiel enchâssé (ici *exploser*) n'est pas un être animé, et que l'anaphore est en position sujet d'un conteneur événementiel comme les verbes *arriver*, *se passer* ou *avoir lieu*.

Il en va de même pour l'expression *faire que*, qui se comporte de la même façon :

(63) [Jean a fait que la voiture explose]<sub>i</sub>. Ça<sub>i</sub> s'est passé hier.

(64) Jean a fait que [la voiture explose]<sub>i</sub>. Ça<sub>i</sub> s'est passé hier.

L'ambiguïté est la même. On peut trancher sur le choix de l'antécédent en examinant la nature des sujets de chacun des verbes événementiels en question.

Pendant, lorsque les sujets de deux verbes événementiels sont de nature animée, il devient plus difficile de prendre une décision.

(65) [Jean a fait en sorte que Max appelle Léa]<sub>i</sub>. C'<sub>i</sub>est arrivé hier.

(66) Jean a fait en sorte que [Max appelle Léa]<sub>i</sub>. C'<sub>i</sub>est arrivé hier.

Ces deux cas de figure sont naturels. Il n'est donc pas possible de choisir un antécédent au lieu de l'autre sur la base de la nature des arguments sujet.

Finalement, sur le plan syntaxique, une propriété des complétives est qu'elles empêchent la relation anaphorique événementielle entre un pronom qu'elles contiennent et la proposition principale dans laquelle elles apparaissent. Un pronom sous la portée d'un conteneur événementiel à l'intérieur d'une complétive ne peut pas faire référence au verbe principal de la phrase. Les exemples de reprise ci-dessous sont alors incohérents :

(67) #[Le ministre affirme]<sub>i</sub> que cela<sub>i</sub> s'est passé hier.

(68) #[Jean s'opposait]<sub>i</sub> à ce que ça<sub>i</sub> se produise.

(69) #[La police avait fait en sorte]<sub>i</sub> que ce<sub>i</sub> soit arrivé en pleine nuit.

Dans tous les cas, une anaphore événementielle contenue dans une complétive ne peut avoir comme antécédent l'événement de la phrase principale qui la contient. Cette observation vaut pour tous les niveaux d'enchâssement :

(70) #[Jean a affirmé que Max a fait en sorte que Marie dise]<sub>i</sub> que c'<sub>i</sub>est arrivé.

(71) #[Jean a affirmé que Max a fait en sorte]<sub>i</sub> que Marie dise que c'<sub>i</sub>est arrivé.

(72) #[Jean a affirmé]<sub>i</sub> que Max a fait en sorte que Marie dise que c'est arrivé.

Intuitivement, cela représente une contrainte sur les antécédents événementiels phrastiques analogue à celle décrite par la théorie du liage (Chomsky (1981)) pour les expressions nominales<sup>14</sup>.

Cependant, cette contrainte ne semble pas valoir pour les anaphores interphrastiques. Les événements enchâssés sous de multiples complétives peuvent être repris par des pronoms en dehors de la proposition qui contient son antécédent.

(73) Jean sait que Max croit que [Marie a affirmé que c'est fini]<sub>i</sub>. C'<sub>i</sub>est arrivé hier.

(74) Jean sait que [Max a affirmé que Marie croit que c'est fini]<sub>i</sub>. C'<sub>i</sub>est arrivé hier.

(75) [Jean a affirmé que Max croit que Marie sait que c'est fini]<sub>i</sub>. C'<sub>i</sub>est arrivé hier.

Il faut avouer que ces énoncés ne sont pas particulièrement naturels. Lorsqu'on construit un discours, on essaie d'éviter les enchâssement trop profonds de complétives ou des relatives. En général en français on accepte naturellement deux niveaux de complétives, l'une enchâssée dans l'autre. Dès que l'on en a une troisième, l'énoncé risque d'être incompréhensible.

Cependant, même avec un seul enchâssement, il peut y avoir une ambiguïté lorsqu'on essaie de déterminer l'antécédent d'une anaphore événementielle, comme l'illustre l'exemple suivant :

(76) [Le ministre a affirmé que le gouvernement a levé les restrictions]<sub>i</sub>. C'<sub>i</sub>est arrivé hier.

(77) Le ministre a affirmé que [le gouvernement a levé les restrictions]<sub>i</sub>. C'<sub>i</sub>est arrivé hier.

Il y a en effet deux antécédents possibles. L'un consiste en la phrase entière, où le verbe principal est *affirmer* et l'autre est la proposition complétive enchâssée, *le gouvernement a levé les restrictions*. On pourrait argumenter qu'on a une tendance à accepter le deuxième plus facilement que le premier comme antécédent du pronom, mais cela n'indique qu'une préférence qui relève de l'interaction entre la forme du conteneur événementiel et la forme de l'événement en question. La combinaison de conteneur et verbe antécédent est certainement pertinente, mais on ne l'examinera pas ici.

L'ensemble de ces données nous mène à tirer plusieurs conclusions sur les complétives par rapport à la relation anaphorique événementielle. On a vu que les complétives de conteneurs propositionnels et factuels peuvent contenir des

---

14 Cette même contrainte s'applique aux interrogatives et aux relatives : #[Max dit]<sub>i</sub> comment c'<sub>i</sub>est arrivé. #[Max a parlé de l'endroit]<sub>i</sub> où c'<sub>i</sub>est arrivé.

descriptions d'événements. Ces événements sont disponibles dans certains cas à la référence anaphorique et peuvent être sélectionnés par un pronom qu'ils précédent, pourvu que ce pronom apparaisse sous la portée d'un conteneur événementiel. On a trouvé que le mode subjonctif du verbe dans une complétive n'empêche pas la relation d'anaphore et qu'une anaphore dans une subordonnée complétive ne peut reprendre l'événement de sa principale. Il reste cependant certaines constructions qui engendrent des ambiguïtés. Dans certains cas il est possible d'établir une préférence pour un antécédent et dans d'autres il n'est pas évident de favoriser un antécédent ou un autre.

#### 3.2.4. Les propositions subordonnées relatives

Les propositions subordonnées relatives sont également intéressantes lorsqu'on examine l'anaphore. Une relative est introduite par un pronom relatif simple (*qui, que, quoi, dont, où*), ou composé (*lequel, laquelle, auquel, duquel*) qui la relie à la proposition principale. Les relatives sont la plupart du temps employées en expansion nominale.

On cherche à établir si les relatives posent des contraintes sur la reprise anaphorique des événements qu'elles contiennent.

Commençons par le premier exemple, où une relative dépend de l'objet direct du verbe principal. Dans tous les cas, on a la description d'un état (*détester, connaître, aimer, être*) avec une relative qui contient un événement enchâssé (*tuer, parler, assassiner*). Si les relatives **objet** permettent la reprise anaphorique, le seul antécédent possible dans la phrase sera bien l'événement enchâssé. On marque par *t* la présence de la trace pour expliciter la présence du sujet, de l'objet direct ou de l'objet indirect à l'intérieur de la relative.

(78) Max déteste l'homme<sub>i</sub> qui [*t*<sub>i</sub> a tué Kennedy]<sub>j</sub>. C<sub>j</sub>'est arrivé en 1963 à Dallas.

(79) Jacqueline connaissait l'homme<sub>i</sub> qu'[Oswald a tué *t*<sub>i</sub>]<sub>j</sub>. C<sub>j</sub>'est arrivé en 1963 à Dallas.

(80) Jean habite près de l'endroit<sub>i</sub> où [Oswald a assassiné Kennedy *t*<sub>i</sub>]<sub>j</sub>. C<sub>j</sub>'est arrivé en 1963 à Dallas.

La reprise de l'événement enchâssé ne semble pas problématique, même si dans certains cas elle n'est pas entièrement naturelle.

Maintenant, regardons les relatives qui dépendent de l'argument **sujet** du verbe principal. On voudrait savoir si l'événement enchâssé est disponible pour la reprise anaphorique. Une nouvelle fois, il n'y a qu'un seul antécédent événementiel potentiel.

(81) #L'homme<sub>i</sub> qui [*t*<sub>i</sub> a tué Kennedy]<sub>j</sub> aime la bière. C<sub>j</sub>'est arrivé en

1963 à Dallas.

(82) #L'homme<sub>i</sub> qu' [Oswald a tué  $t_i$ ]<sub>j</sub> aimait la bière. C<sub>j</sub>'est arrivé en 1963 à Dallas.

(83) #La ville<sub>i</sub> où [Oswald a assassiné Kennedy  $t_i$ ]<sub>j</sub> a un métro. C<sub>j</sub>'est arrivé en 1963 à Dallas.

Les exemples ci-dessus ne sont pas naturels. Même si le seul événement présent est celui qui se trouve dans la relative, le « focus » de l'énoncé est sur le verbe de la principale<sup>15</sup>. Le discours est incohérent lorsqu'on essaie de reprendre un événement enchâssé dans une relative en position sujet.

Cependant, l'incohérence dépend également de la nature prédicative de la principale. L'exemple suivant ne semble engendrer aucune incohérence :

(84) L'homme qu'[Oswald a tué  $t_i$ ]<sub>j</sub> était le Président. C<sub>j</sub>'est arrivé en 1963 à Dallas.

En général, on peut conclure que les relatives en position objet ne posent pas de restriction pour la reprise anaphorique d'événements qu'elles contiennent. En revanche, les relatives sujet bloquent souvent la reprise anaphorique car elles forment souvent une frontière attentionnelle entre l'événement enchâssé et la proposition principale.

### 3.2.5. Événements coordonnés

La catégorie des conjonctions de coordination regroupe les mots invariables qui établissent une relation de coordination entre deux éléments de même nature et de même fonction syntaxiques (deux mots, deux syntagmes, deux propositions, ou même deux phrases).

Par *événements coordonnés*, on entend un groupe de deux ou plusieurs événements simples connectés par une conjonction de coordination (*et, mais, ou* etc). Cela comprend également les listes d'événements simples séparés par des virgules. Dans les deux cas, les événements en question doivent être au même temps<sup>16</sup>. Lorsqu'on parle d'événements, on entend les événements sous forme phrastique, ce qui veut dire que l'on considère des coordinations de deux phrases ou de deux propositions (IP<sub>[+/-TEMPS]</sub>). La présence d'une conjonction de coordination entre deux éléments de type événementiel peut indiquer la synchronicité des événements qu'elle relie, exprimer un ordre temporel dans une suite, ou montrer un contraste. Il faut souligner l'importance des temps verbaux dans la coordination d'événements. Les temps verbaux doivent être compatibles pour que la conjonction d'événements soit possible. Les exemples que l'on

---

15 Selon la théorie du centrage il y a eu un changement important dans l'état « attentionnel ».

16 Y compris deux verbes à l'infinitif.

considère ici seront compatibles sur le plan temporel, mais on réservera une discussion sur les temps pour la suite<sup>17</sup>.

Prenons un premier exemple :

(85) Les manifestants se sont rassemblés et ont appelé à la grève.

Cette phrase est une contraction de *Les manifestants se sont rassemblés et les manifestants ont appelé à la grève*. La conjonction permet l'élision du sujet dans la deuxième proposition. La forme non élidée est aussi un événement composé, mais elle n'est pas très naturelle, car on a tendance à éviter ce genre de répétition lorsqu'on construit un discours.

Il existe aussi des cas où l'auxiliaire n'est pas répété car les deux verbes se conjuguent de la même manière. La phrase suivante équivaut à *Jean a mangé et Jean a rangé la cuisine* :

(86) Jean a mangé et rangé la cuisine.

Dans une relation anaphorique, la conjonction nous permet de faire référence aux deux événements :

(87) [Les manifestants se sont rassemblés et ont appelé à la grève]<sub>i</sub>. Ça<sub>i</sub> s'est passé ce matin.

(88) [Jean a mangé et rangé la cuisine]<sub>i</sub>. C'<sub>i</sub>est arrivé ce matin.

Dans le cas où les sujets de chaque verbe sont différents, on les explicitera :

(89) Les manifestants se sont rassemblés et le syndicat a appelé à la grève.

Les deux événements sont indépendants, mais la conjonction les associe de façon à permettre une reprise anaphorique de l'ensemble des deux :

(90) [Les manifestants se sont rassemblés et le syndicat a appelé à la grève]<sub>i</sub>. Ça<sub>i</sub> s'est passé ce matin.

Non seulement la conjonction *et* permet la reprise de l'ensemble, mais elle l'oblige. La reprise de l'un ou l'autre des événements coordonnés n'est pas possible. Cela explique pourquoi les cas de reprise suivants ne sont pas naturels :

(91) #Jean a mangé et [rangé la cuisine]<sub>i</sub>. C'<sub>i</sub>est arrivé ce matin.

(92) #[Jean a mangé]<sub>i</sub> et rangé la cuisine. C'<sub>i</sub>est arrivé ce matin.

La coordination est souvent présente à la fin d'une liste d'éléments séparés par des virgules où les dernier et avant-dernier éléments sont coordonnés. La virgule joue le même rôle que la coordination *et* dans ce cas :

(93) [Jean a mangé, fait la vaisselle et rangé la cuisine]<sub>i</sub>. C'<sub>i</sub>est arrivé hier.

L'anaphore, dans ces cas, reprend toute la liste.

Considérons maintenant d'autres conjonctions. Prenons d'abord la conjonction *ou* qui exprime la disjonction de deux propositions.

(94) Jean a menti ou Marie s'est trompée. C'est arrivé hier.

---

<sup>17</sup> En effet, on considère uniquement les événements qui ne sont marqués ni par des temps verbaux différents, ni par des modificateurs temporels adverbiaux.

Dans ce cas il y a une ambiguïté due à l'impossibilité de déterminer lequel des événements s'est produit. L'un ou l'autre, ou bien les deux auraient pu logiquement se produire. Mais étant donné qu'à partir du contexte limité on n'a pas le moyen de décider, la disjonction des événements est considérée comme un ensemble. Dans ce cas c'est l'ensemble en entier qui est disponible comme antécédent potentiel et non pas ses constituants.

La conjonction *mais* permet d'exprimer une opposition entre deux éléments de même nature.

(95) Les prisonniers se sont évadés mais les autorités les ont repris.

La conjonction *mais* établit bien une opposition entre deux événements qui se sont produits dans le passé. Maintenant, que se passe-t-il lorsqu'on enchaîne avec une reprise anaphorique événementielle?

(96) Les prisonniers se sont évadés mais les autorités les ont repris. C'est arrivé hier.

Dans cet exemple, on peut facilement considérer que l'antécédent est l'ensemble de la phrase qui précède l'anaphore. Les événements en question ont des restrictions quant à leur localisation temporelle. L'événement d'*être repris* ne peut se passer sans *évasion* préalable. C'est la suite logique de la narration. En plus, le sujet, *les prisonniers*, de la première proposition est anaphorisé par le pronom *les* en position objet de la deuxième. Cela assure une cohérence narrative à la séquence et les deux sont étroitement liés. Cependant ce n'est pas toujours le cas, et on peut très bien avoir deux événements coordonnés par *mais* qui n'ont pas ce genre de restriction.

(97) L'ONU a lutté pour la paix mais les États-Unis ont déclaré la guerre. C'est arrivé hier.

Dans cet exemple l'antécédent le plus naturel est celui qui figure dans la deuxième proposition de la première phrase. Une nouvelle fois, il s'agit plus d'une préférence que d'une règle fixe. On peut tenter d'expliquer ce fait en faisant référence à la nature des verbes *lutter pour la paix* et *déclarer la guerre* en rapport avec le conteneur, dans ce cas *arriver*. La première expression est une activité alors que la deuxième consiste en un événement ponctuel, un achèvement. La phrase *c'est arrivé hier* réfère à un événement ponctuel dans le passé, d'où la préférence. On voit encore la pertinence de la combinaison de conteneur et verbe antécédent. La classification de Vendler explique pourquoi un conteneur événementiel sélectionne un certain type de verbe pour une anaphore se trouvant sous sa portée.



### 3.2.6. Les temps verbaux

L'usage des différents temps verbaux est un facteur qui influence beaucoup la structure et la cohérence du discours. Tout discours se situe dans un certain cadre temporel qui peut évoluer. On souhaite déterminer ici si les temps verbaux des événements dans un discours ont des conséquences pour la reprise anaphorique de ces événements. On ne se livre pas à une étude du discours dynamique, mais plutôt à une inspection empirique superficielle de l'effet des temps verbaux sur la relation anaphorique événementielle. En particulier, on souhaite considérer la relation entre le temps verbal d'un conteneur événementiel et les antécédents qu'il peut potentiellement sélectionner. Le but est de vérifier s'il existe des préférences. On se concentrera sur les conteneurs au passé composé. On regardera également l'effet des temps sur la conjonction d'événements.

(98) #[Max fait la vaisselle]<sub>i</sub>. C<sub>i</sub>'est arrivé hier.

(99) #[Max faisait la vaisselle]<sub>i</sub>. C<sub>i</sub>'est arrivé hier.

(100) #[Max avait fait la vaisselle]<sub>i</sub>. C<sub>i</sub>'est arrivé hier.

(101) #[Max fera la vaisselle]<sub>i</sub>. C<sub>i</sub>'est arrivé hier.

(102) #[Max aura fait la vaisselle]<sub>i</sub>. C<sub>i</sub>'est arrivé hier.

Avec une interprétation simple, toutes les relations d'anaphores ci-dessus sont bloquées par une incompatibilité au niveau des temps.

Le présent peut être interprété de différentes façons. La première est une interprétation d'habitude. Ainsi, dans cet exemple, *faire la vaisselle* est un acte que Max effectue de façon habituelle. Selon cette interprétation, la combinaison est cohérente. On voit mieux cet effet dans les séquences d'événements au présent qui peuvent être reprises au passé par une anaphore virtuelle :

(103) [Max fait la vaisselle, range sa chambre et s'affaisse sur le canapé]<sub>i</sub>.

C<sub>i</sub>'est arrivé hier (comme tous les jours).

Une autre interprétation est simplement que le présent relate ce qui est en train de se passer. C'est cette interprétation que l'on considère dans le premier exemple ci-dessus et qui est incompatible avec une reprise dans un conteneur au passé composé.

L'incohérence des combinaisons du futur et du passé est évidente, car ces temps sont aux deux extrêmes du continuum temporel.

L'incompatibilité de l'imparfait et du passé composé vient du fait que le premier représente un aspect de continuité ou de répétition dans le passé, tandis que le deuxième situe un verbe de façon ponctuelle dans le passé. Les deux sont incompatibles dans le contexte de cet exemple.

Un constat que l'on peut être amené à faire est que l'anaphore actuelle ne semble pas être possible lorsque les temps verbaux sont différents. En revanche, lorsqu'un conteneur et l'antécédent d'un pronom se trouvent sous sa portée ont le

même temps verbal, l'anaphore actuelle est possible. On fait remarquer le contraste entre les exemples précédents et le suivant, qui est tout à fait naturel :

(104) [Max a fait la vaisselle]<sub>i</sub>. C'<sub>i</sub>est arrivé hier.

Ici, la relation anaphorique repose sur la coréférence actuelle ainsi que sur la coréférence virtuelle. L'occurrence de l'événement décrit dans la première phrase est exactement la même dans la deuxième phrase. Examinons rapidement d'autres temps verbaux pour tester cette hypothèse :

(105) [Max fait la vaisselle]<sub>i</sub>. Ça se passe dans la cuisine.

(106) [Max faisait la vaisselle]<sub>i</sub>. Ça se passait dans la cuisine.<sup>18</sup>

(107) [Max fera la vaisselle]<sub>i</sub>. Ça se passera dans la cuisine.

(108) [Max aura fait la vaisselle]<sub>i</sub>. Ça se sera passé dans la cuisine.

Les temps étant les mêmes à chaque fois, ces exemples sont naturels et contiennent une relation d'anaphore actuelle.

Finalement on examine l'effet des temps verbaux sur la conjonction d'événements (ce que l'on a appelé des événements coordonnés). On veut savoir si les combinaisons de différents temps verbaux ont des conséquences pour leur coordination. Regardons quelques exemples, avec des temps du passé :

(109) #Léa regardait la télé et Max a mangé.

(110) #Max a regardé la télé mais Léa mangeait.

(111) #Marie dormait ou Jean a bu une bière.

Lorsqu'on essaie de coordonner deux événements dont les temps verbaux sont différents, les énoncés résultants ne sont pas très naturels. On ne peut pas les regrouper dans un ensemble qui soit disponible comme antécédent pour une relation anaphorique :

(112) #Marie dormait et Jean a bu une bière. C'est arrivé hier.

On a montré quelques cas simples d'incohérence. Certaines combinaisons ne sont pas du tout naturelles, alors que d'autres dépendent de l'interprétation que l'on donne aux temps. En effet, l'interprétation des temps verbaux et le contexte sont des facteurs importants pour la cohérence et le sens d'un discours. Pour un seul énoncé, il est souvent possible de trouver une lecture cohérente et une qui ne l'est pas. Étant donné la complexité de la situation et le nombre de paramètres on ne cherche pas à proposer une solution ici.

### 3.2.7. La modalité

Le phénomène de la modalité dans un énoncé, selon les linguistes, consiste en

---

<sup>18</sup> On peut discuter du statut événementiel des énoncés à l'imparfait, qui est parfois considéré comme un temps statif. Ici, on le traite comme événementiel.

l'expression d'une attitude envers le contenu de cet énoncé. Elle est exprimée de diverses façons : par des flexions verbales (subjonctif, conditionnel), par certains adverbes (*peut-être, sûrement, éventuellement*), par des noms (*possibilité, certitude*), par des verbes modaux (*devoir*) ou encore par des verbes à *modalité épistémique*. La modalité épistémique décrit le niveau de certitude exprimé dans un énoncé. Les verbes comme *savoir, croire, penser, affirmer, douter* et *soupçonner* expriment une modalité épistémique.

On souhaite maintenant examiner l'effet de la modalité dans des situations de reprise anaphorique événementielle. On se concentrera sur une étude des verbes à modalité épistémique car ils constituent des conteneurs propositionnels ou factuels au sens de Vendler.

(113) Léa croit que la prof raconte des bobards.

Le verbe *croire* représente un certain degré d'incertitude. Ce qui est cru n'est pas certain et décrit plutôt un état psychologique de son sujet. Maintenant, examinons ce qui se passe lorsqu'on essaie de reprendre un événement enchâssé avec un pronom qui apparaît sous la portée d'un conteneur événementiel :

(114) #Léa croit que [la prof raconte des bobards]<sub>i</sub>. C<sub>i</sub>'est arrivé hier.

L'événement en question est que *la prof raconte des bobards* et le pronom *ce* (élide en *c'*) est sous la portée du conteneur *arriver*. On ne peut pas admettre ce genre de reprise car il y a une incohérence entre le fait que la première phrase n'affirme pas que l'événement s'est produit et le fait que la deuxième l'affirme. La première phrase exprime une certaine modalité, mais pas la deuxième.

(115) Léa croit que [la prof raconte des bobards]<sub>i</sub>. Ce<sub>i</sub> serait arrivé hier.

Dans cet exemple, on évite cette incohérence car il y a une modalité dans les deux phrases, qui sont compatibles. L'emploi du conditionnel dans le conteneur événementiel exprime une possibilité, mais pas la nécessité de la proposition. La reprise est possible car les deux contextes n'impliquent pas la vérité des propositions qu'ils expriment. Ces énoncés sont sujets à des phénomènes de *subordination modale*<sup>19</sup> qui imposent des contraintes qui assurent la cohérence du discours.

La modalité peut avoir des conséquences sur le type de relation anaphorique événementielle. Lorsqu'il y a une incertitude sur un événement, il n'est pas sûr qu'il se soit produit. Ainsi, une relation anaphorique avec cet événement hypothétique ne peut être que virtuelle, comme on a vu ci-dessus pour les constructions contrefactuelles. Il y a coréférence virtuelle, mais pas actuelle dans les cas d'énoncés incertains. En revanche, le contraire est vrai lorsqu'il s'agit d'un énoncé qui n'a pas d'incertitude. Quand un événement apparaît dans un contexte de certitude, une reprise anaphorique peut mettre en jeu une relation de

---

<sup>19</sup> La notion de *subordination modale* a été introduite par Craige Roberts (1989) pour regrouper un ensemble de phénomènes portant sur l'effet de la modalité.

coréférence actuelle ainsi que virtuelle. Il est possible de reprendre l'instance spécifique de l'événement en question.

On peut conclure que lorsqu'il s'agit d'un verbe d'attitude propositionnelle, la reprise d'un événement enchâssé dans une subordonnée complétive n'est possible que lorsque le pronom apparaît dans un contexte ayant une modalité compatible avec celle de la complétive. La phrase contenant la reprise anaphorique ne doit pas affirmer la véracité de la proposition qu'elle exprime. Sur le plan linguistique, cela impliquerait la présence d'une expression qui dénote l'incertitude (*peut-être, éventuellement*) ou bien d'un temps ou mode verbal (conditionnel ou subjonctif).

### **3.3. Formes non anaphoriques**

Un point important à noter est que toutes les occurrences de pronoms n'impliquent pas forcément une relation anaphorique. Dans certains cas, les pronoms que l'on utilise dans le discours obtiennent une référence de façon déictique, c'est-à-dire par rapport au contexte d'énonciation ; parfois, ils n'ont pas de référence, ce sont des pronoms impersonnels. Les pronoms impersonnels ne sont pas référentiels. Ils occupent simplement une position argumentale (le plus souvent de sujet syntaxique) d'un verbe, sans avoir de contenu sémantique. Ils ne sont donc pas anaphoriques. Il existe également des constructions où un pronom, sans être qualifié d'impersonnel, n'est pas anaphorique.

Dans une opération de résolution d'anaphores, il est donc nécessaire de distinguer les pronoms qui ont un antécédent potentiel de ceux qui n'en ont pas. On mentionne plusieurs cas de figure, sans prétendre fournir une liste exhaustive.

#### 3.3.1. Le *ça* impersonnel

Il y a plusieurs cas où le pronom *ça* (ou *cela*) est considéré impersonnel. Les exemples qui suivent sont tirés en grande partie de Corblin (1994).

Les verbes météorologiques : Les occurrences du pronom *cela/ça* dont il ne faudra pas chercher l'antécédent sont celles que l'on peut considérer comme un emploi *impersonnel*. En français, les verbes météorologiques tombent dans cette catégorie. Les verbes tels que *pleuvoir, neiger, cailler* et *faire beau* prennent tous un sujet impersonnel, en général le pronom *il*. Cependant, dans un langage familier certains verbes météorologiques peuvent prendre le pronom *ça* impersonnel en position sujet.

(116) *Ça* caille ce matin !

(117) *Ça* vase.

(118) Ça neige.

Expressions figées : On trouve certaines expressions familières qui contiennent un verbe avec un sujet impersonnel sous la forme du pronom *ça*. Ce sont des expressions que l'on emploie le plus souvent à l'oral comme *ça barde, ça va, ça roule, ça boume* etc. Ces verbes, lorsqu'ils sont employés dans des expressions de ce genre, ne peuvent en effet prendre que le pronom *ça* comme sujet. On pourrait discuter du statut impersonnel de *ça* dans ce cas, en disant qu'il a une référence vague qu'il obtient de façon déictique. Mais en tous cas, il n'est pas anaphorique.

Pronoms « vagues » : On trouve aussi en français des emplois du pronom *ça* que Francis Corblin (1994) appelle « à sujet indistinct ». Ce sont des instances du pronom qui ne sont pas impersonnelles car elles ont une référence, mais cette référence provient du contexte d'énonciation. Une nouvelle fois, il s'agit d'un usage familier que l'on entend dans la langue parlée. Voici un exemple :

(119) Ça dort là-dedans.

Ici le locuteur souhaite décrire la situation de façon générale. Le *ça*, qui est déictique, ne reprend pas un agent spécifique et l'énoncé décrit plutôt ce qui se passe dans le contexte d'énonciation. Le référent du pronom *ça* n'est pas explicité dans le discours. Même si on ne peut pas qualifier ces emplois d'impersonnels, ils ne font pas intervenir une relation anaphorique, et c'est cela qui est important.

Ça anaphorique non événementiel (une exception) :

Dans un langage familier on peut trouver des occurrences du pronom *ça* qui remplacent un pronom de troisième personne en position sujet. Si *ça* ne peut pas être anaphorique en position sujet de certains verbes à complétive (voir ci-dessous), il peut être anaphorique en position sujet de verbes simples.

(120) Les gens poussent pour accoster. Ça tape dans tous les sens, ça hurle, ça pousse.

Dans cet exemple, de style très marqué, le pronom *ça* remplace le *ils* qui reprendrait le groupe nominal *les gens*. Cela nous montrent que *ça* peut être anaphorique en position sujet et qu'il peut reprendre, dans un contexte très particulier, une entité nominale. Une des grandes difficultés est de distinguer les cas où son antécédent est non événementiel des cas où il l'est.

(121) Les gens poussent pour accoster. Ça m'énerve.

Cela revient à la nature du conteneur sous la portée duquel il apparaît. Le verbe *énervé* ici est un conteneur de type propositionnel ou factuel, alors le pronom ne peut pas reprendre un événement.

Pronoms sujet de verbe à complétive : Un autre cas où on peut considérer que

*ça* est non anaphorique est quand il est sujet d'un verbe à complétive :

(122) *Ça* m'étonnerait qu'il vienne.

(123) *Ça* fait plaisir de te voir.

(124) *Ça* lui arrive de se tromper.

On peut considérer le pronom *ça* comme un emploi impersonnel ici. Il figure en position de sujet syntaxique du verbe principal, dont le sujet « sémantique » est la complétive déplacée. Dans ce cas, il est non seulement non anaphorique, mais non référentiel. Les troisième et quatrième de ces exemples sont particulièrement pertinents car le pronom *ça* est en position sujet d'un conteneur événementiel, mais il ne peut pas être anaphorique à cause de la construction.

On peut aussi décrire le *ça* comme reprenant ici la complétive, et comme étant bien référentiel. Mais, ce serait un exemple de cataphore, et non pas d'anaphore, car le pronom est coréférentiel avec un élément qui apparaît après lui dans le discours. De plus, il n'est pas évident que le pronom, dans ces exemples, ait un statut événementiel, car il reprend une complétive qui est plutôt de type propositionnel. Dans les deux premiers exemples, les verbes *étonner* et *faire plaisir* ne sont pas des conteneurs événementiels. Ce cas de figure est alors à ignorer dans la recherche de l'antécédent.

### 3.3.2. Le *ce* non anaphorique

Il y a certains contextes où le pronom *ce* n'est pas anaphorique. Les relatives périphrastiques (Riegel et al. (1994) pp 487-88), les constructions clivées et interrogatives peuvent contenir des occurrences du pronom *ce* qui n'ont pas d'antécédents.

Les relatives périphrastiques : Ce genre de construction (aussi appelées relatives indéfinies) n'ont qu'un sens très général. Elles sont introduites par le pronom *ce*, invariable, qui est suivi du pronom relatif *que*, *qui*, *dont* ou *quoi* précédé d'une préposition. Le contenu sémantique du pronom n'est pas explicité : Les seuls contraintes sont pour une entité inanimée dont les propriétés sont déterminées par la relative. Par exemple :

(125) Ce qui est rare est cher.

(126) Ce que je dis déplaît au gens.

(127) Ce dont je parle est très sérieux.

(128) Ce à quoi je pense est très drôle.

(129) Léa aime tout ce qui est cher.

(130) Jean ne croit pas ce que je dis.

(131) Max réfléchit à ce dont je parle.

(132) Pierre ne sait pas ce à quoi je pense.

La relative périphrastique a un statut de groupe nominal. Elle peut apparaître en position sujet, en position de complément d'objet direct d'un verbe, ou bien en position d'objet indirect.

Il ne faut pas confondre ce genre de construction avec les relatives définies, où les occurrences de *ce* sont anaphoriques :

(133) Max a marché sur une banane, ce qui l'a fait glisser.

(134) Max a marché sur une banane, ce que Marie a trouvé très marrant.

(135) Max a mangé un champignon. C'est ce qui l'a rendu malade.

Dans ces exemples, le pronom obtient sa référence grâce à un élément qui lui est antérieur dans le discours. Dans le cas des relatives périphrastiques, on remarque que chaque phrase ne consiste qu'en une seule proposition grammaticale, alors que dans le cas des relatives définies, il y en a (au moins) deux. La relative définie est en fonction appositive et a une proposition entière ou une phrase comme antécédent.

Formes interrogatives : Les tournures interrogatives en *est-ce que* contiennent un pronom *ce* non anaphorique. La construction en elle-même sert de marqueur interrogatif et le pronom n'a pas de référent.

(136) Est-ce que tu viens demain?

(137) Qu'est-ce que tu fais demain?

(138) À quelle heure est-ce que tu viendras?

(139) Combien de temps est-ce que tu vas rester?

Constructions clivées : Les constructions clivées permettent l'extraction d'un élément d'une proposition interrogative ou déclarative<sup>20</sup>. L'élément est extrait et encadré par *c'est* (ou éventuellement *est-ce* pour une interrogative) et par un pronom relatif *que* ou *qui*.

Prenons les phrases déclaratives suivantes :

(140) Max a marché sur une peau de banane.

(141) Léa veut voir le nouveau film d'Almodóvar.

Elles peuvent se transformer en phrases clivées :

(142) C'est Max qui a marché sur une peau de banane.  
(déclarative)

(143) Est-ce sur une peau de banane que Max a marché?  
(interrogative)

(144) C'est le nouveau film d'Almodóvar que Léa veut voir.  
(déclarative)

---

<sup>20</sup> Les constructions clivées ne sont pas possibles pour les propositions impératives à cause de l'absence de sujet.

(145) Est-ce Léa qui veut voir le nouveau film d'Almodóvar?  
(interrogative).

L'élément extrait peut être le sujet, l'objet direct, mais aussi un complément circonstanciel (groupe prépositionnel (GP), adverbe, subordonnée ou gérondif). Ce qui nous importe ici est le fait que dans tous les cas, le pronom *ce* du *c'est* est non anaphorique. Il n'a pas de contenu sémantique et la construction sert comme support syntaxique à l'extraction.

Autres cas : Il existe d'autres expressions où figure un pronom *ce* non anaphorique, notamment celles avec *c'est...que* avec un groupe adjectival intercalé. *C'est dommage que, c'est vrai que, c'est important que, c'est un fait que* par exemple. Certaines d'entre elles prennent normalement la forme pronominale *il*, qui peut être remplacée par *ce*, surtout lorsque le niveau de langage est informel. Une nouvelle fois, comme pour les pronoms sujet de complétives, le statut référentiel du pronom est discutable. On peut très bien considérer qu'il reprend la complétive en entier, cas où il est de type factuel ou propositionnel. En tous cas on peut être sûr de pouvoir affirmer qu'il s'agit d'un emploi non anaphorique, car la complétive que le pronom reprend figure obligatoirement après lui, ce qui relève d'une relation cataphorique.

Les complétives subordonnées qui ont la fonction d'objet indirect prennent aussi un pronom *ce* non anaphorique qui remplit plutôt une fonction de support syntaxique à la construction : *Je m'attends à ce que tu viennes, je m'oppose à ce que tu partes*.

On a repéré plusieurs contextes où le pronom *ce* n'a pas de contenu référentiel, et par conséquent, ne peut pas avoir un antécédent dans le discours. Il y a également des contextes où le pronom peut être non anaphorique, même s'il est référentiel.



## **4. Algorithmes pour la résolution d'anaphores**

De multiples approches au problème de la résolution automatique d'anaphores ont été conçues. La plupart d'entre elles se focalisent sur les anaphores individuelles (à antécédent nominal). La résolution d'anaphores abstraites (à antécédent phrastique) a été moins étudiée, mais il existe néanmoins quelques algorithmes permettant de les résoudre.

On examinera ici trois algorithmes pour la résolution d'anaphores individuelles et un pour la résolution d'anaphores abstraites (qui intègre aussi la résolution d'anaphores individuelles).

### **4.1. Résolution d'anaphores nominales**

Parmi les algorithmes pour la résolution d'anaphores individuelles, celui de Jerry Hobbs, qui fonctionne uniquement sur base de critères syntaxiques est un des précurseurs.

Plus récemment, Shalom Lappin et Herbert J. Leass ont formulé un algorithme qui exploite des informations morpho-syntaxiques et sémantiques. Cette approche hybride a eu un certain degré de succès et reste une des références dans le domaine.

Le troisième algorithme que l'on présentera est celui de Ruslan Mitkov, qui repose sur des données d'entrée assez pauvres.

#### **4.1.1. Hobbs**

Jerry Hobbs (1978) a élaboré un algorithme sur la base de contraintes imposées par la syntaxe de l'anglais. L'algorithme prend en entrée un arbre syntaxique complet et correct qu'il parcourt à la recherche d'antécédents en leur appliquant diverses contraintes syntaxiques et morphologiques. Au niveau intraphrastique, l'algorithme consiste en un parcours en largeur de gauche à droite avec une préférence pour l'antécédent le plus proche de l'anaphore. Un parcours en largeur est aussi effectué au niveau interphrastique, avec une préférence pour les sujets comme antécédents. En effectuant le parcours, l'algorithme fait l'inventaire des antécédents possibles, qu'il vérifie ensuite en appliquant des contraintes d'accord morphologique – traits de genre et nombre. Il applique également des contraintes syntaxiques, qui sont basées sur la condition B de la théorie du liage. Ce sont notamment :

- I. Un pronom non réfléchi et son antécédent ne peuvent pas apparaître dans la même phrase simple.

## II. L'antécédent d'un pronom doit précéder ou c-commander le pronom.

Les pronoms traités par cet algorithme sont les pronoms personnels *he, she, it* et *they*. Le taux de réussite global était assez élevé, il est de 88.3%. Cependant, il échoue sur certains cas, tels que la reprise d'éléments phrastiques :

- (146) [Ford was in trouble]<sub>i</sub> and he knew it<sub>i</sub>. (Hobbs 1978)  
[Ford avait des ennuis]<sub>i</sub> et il le<sub>i</sub> savait.

Il ne traite pas non plus les classes de noms comme *picture* :

- (147) John<sub>i</sub> saw a picture of himself<sub>i</sub>. (Hobbs 1978)  
John<sub>i</sub> a vu une image de lui-même<sub>i</sub>.

L'algorithme effectue aussi une recherche pour des antécédents de pronoms cataphoriques, qui apparaissent après le pronom. Cependant il ne cherche pas en dessous des niveaux GN ou P, ce qui fait que les phrases comme la suivante ne sont pas traitées :

- (148) Mary<sub>i</sub> sacked out in his<sub>j</sub> apartment before Sam<sub>j</sub> kicked her<sub>i</sub> out. (Hobbs 1978)

Mary<sub>i</sub> a dormi dans l'appartement de Sam<sub>j</sub> avant qu'il<sub>j</sub> ne la<sub>i</sub> vire.

L'approche de Hobbs, qui date maintenant de presque trente ans, a donné des résultats assez bons. Cependant, il est évident qu'il ne peut pas résoudre les pronoms dans certaines constructions. Les approches plus récentes, telles que celle de Lappin et Leass, prennent en compte d'autres paramètres que la syntaxe, ce qui peut aider à améliorer la couverture du système de résolution.

### 4.1.2. Lappin & Leass

Shalom Lappin et Herbert J. Leass (1994) proposent un algorithme pour l'identification des antécédents nominaux de pronoms de troisième personne (*he, she, they, it*) et d'anaphores réflexives et réciproques (*himself, herself, themselves, itself*) en anglais.

L'algorithme se base sur l'utilisation d'informations de nature syntaxique et morphologique. Il utilise un modèle qui calcule dynamiquement la saillance d'un antécédent potentiel sur la base de différents facteurs. A chaque facteur est attribué un indice différent selon son utilité dans la procédure de résolution. Cette mesure de saillance pondérée est utilisée afin de classer les candidats potentiels pour déterminer une préférence. Avant l'application des mesures de saillance, les pronoms non-anaphoriques sont éliminés. Les pronoms pléonastiques, comme *it* dans des constructions avec un adjectif modal, sont mis à l'écart : *it is necessary/essential/sufficient to/that*. Les occurrences de *it* avec un verbe cognitif ou d'attitude propositionnelle sont aussi exclues, car le pronom est impersonnel : *it is recommended/believed/known/expected that*.

Ensuite, un filtre de contraintes morphologiques et syntaxiques élimine les candidats qui ne satisfont pas les contraintes de la Théorie du liage ou les contraintes d'accord de genre et nombre. Ce filtre est mis en place afin d'exclure la coréférence dans les exemples suivants :

- (149) The woman<sub>i</sub> said that he<sub>i</sub> is funny.
- (150) She<sub>i</sub> likes her<sub>i</sub>.
- (151) She<sub>i</sub> sat near her<sub>i</sub>.
- (152) This is the man he<sub>i</sub> said John<sub>i</sub> wrote about.
- (153) His<sub>i</sub> portrait of John<sub>i</sub> is interesting.

Tous les référents qui forment une chaîne anaphorique sont regroupés dans des classes d'équivalence. À chaque classe d'équivalence est attribué un poids, qui correspond à la somme de tous les facteurs de saillance qui sont associés à au moins un membre de la classe d'équivalence.

Au fur et à mesure de la résolution, les valeurs de saillance des antécédents potentiels sont diminuées selon certaines règles. Dans une classe d'équivalence, c'est l'antécédent le plus saillant qui est choisi comme le référent d'un pronom. Dans les situations où deux ou plusieurs antécédents ont la même mesure de saillance, c'est celui qui est le plus proche du pronom qui l'emporte. Les antécédents intraphrastiques ont la priorité sur ceux qui sont interphrastiques.

Les facteurs de saillance utilisés dans l'algorithme sont principalement des propriétés structurales ou syntaxiques. Chaque facteur permet, selon sa pertinence, d'augmenter le score des antécédents potentiels:

- Sentence recency : exprime le fait que plus un antécédent est proche de l'anaphore, plus il est saillant. La valeur de ce facteur est diminuée de moitié pour chaque nouvelle phrase. (100)
- Subject emphasis : un antécédent en position sujet est plus saillant que ceux ayant d'autres rôles grammaticaux. (80).
- Existential emphasis : un élément nominal dans une construction existentielle (« There is... ») est saillant. (70).
- Accusative emphasis : un élément en position d'objet direct est saillant, mais pas autant qu'un sujet. (50).
- Indirect object, oblique complement emphasis : un objet indirect est moins saillant qu'un élément objet direct. (40).
- Head noun emphasis : une tête nominale dans un groupe nominal complexe est plus saillante qu'un nom non tête, qui est pénalisé.
- Non adverbial emphasis : ce facteur pénalise les GN dans des constructions adverbiales.

Une implémentation de l'algorithme a été effectuée en Prolog et testée sur un corpus de manuels informatiques. Les tests ont donné un taux de réussite de 86%, 4% de plus que l'algorithme de Hobbs sur le même corpus.

Cette approche donne d'assez bons résultats. Le système de Lappin et Leass a pour inconvénient de nécessiter diverses données d'entrée et de nombreux niveaux d'analyse. Leur algorithme fonctionne avec une analyse syntaxique et morphologique, et nécessite aussi une analyse des rôles sémantiques des GN dans le texte. Cela rend une implémentation assez coûteuse en temps et en code. Un des algorithmes de Mitkov, que l'on examinera par la suite, a été conçu pour contourner ce problème.

#### 4.1.3. Mitkov

Le principe de l'algorithme de Ruslan Mitkov (1998) est de minimiser l'utilisation de données syntaxiques et sémantiques, qui sont assez coûteuses en termes de développement, pour la résolution d'anaphores. Le but est de faciliter l'implémentation tout en assurant un bon taux de réussite sur le traitement de manuels techniques, et de permettre l'adaptation d'une langue à l'autre. Cette approche ne nécessite ni parsing syntaxique ni analyse sémantique, mais prend simplement en entrée la sortie d'un étiqueteur morpho-syntaxique. L'algorithme consiste en l'application de simples heuristiques préférentielles (« antecedent indicators ») basées sur des données empiriques.

L'algorithme se déroule de la façon suivante :

Les groupes nominaux apparaissant au plus à une distance de deux phrases de l'anaphore à résoudre sont d'abord identifiés. Ensuite, l'accord des traits de genre et nombre entre chacun des GN et l'anaphore est vérifié. C'est ici que les heuristiques sont appliquées en séquence pour calculer un score pour les GN antécédent potentiels. À chacun des candidats est attribué un score (-1, 0, 1, ou 2) pour chaque heuristique, qui contribue à une somme totale. Le candidat avec le meilleur score total est sélectionné comme antécédent.

Les heuristiques utilisées dans le processus sont en rapport avec la saillance, la répétition d'expressions, la distance référentielle, et la topologie lexicale du texte. En voici l'inventaire :

- Definiteness : la définitude est un facteur qui favorise un antécédent. Les groupes nominaux définis sont des antécédents plus probables que les indéfinis. Les GN définis ont un score de 0 et les indéfinis sont pénalisés avec -1.
- Givenness : les GN dans les phrases précédentes qui représentent le thème de l'énoncé (« the given information ») sont considérés comme de bons candidats. L'heuristique utilisée pour déterminer le thème est que c'est le premier GN dans une phrase non-impérative. Cela est dû à l'observation que, dans un texte, le thème apparaît en premier, et crée un lien avec le texte

- précédent. Les GN thème ont un score de 1, les autres 0.
- Indicating verbs : si un verbe appartient à l'ensemble des verbes « indicateurs », le premier GN qui le suit est considéré comme le candidat préféré. L'ensemble de ces verbes est  $V = \{discuss, present, illustrate, identify, summarise, examine, describe, define, show, check, develop, review, report, outline, consider, investigate, explore, assess, analyse, synthesise, study, survey, deal, cover\}$ . Les observations montrent que les noms qui suivent ces verbes sont particulièrement saillants, et c'est cela qui fait que ces verbes sont de bons indicateurs.
  - Lexical reiteration : la répétition lexicale attribue 2 points aux groupes nominaux qui sont répétés deux fois ou plus dans la même phrase, 1 point si répété une fois, 0 points sinon. Cela comprend les synonymes qui peuvent être précédés de pronoms démonstratifs (*The cartridge...this cartridge*), et les GN ayant la même tête lexicale (*The toner bottle, the bottle of toner, the bottle*).
  - Section heading preference : si un GN apparaît dans le titre d'une section du document, alors il est favorisé d'un point, 0 sinon.
  - « Non-prepositional » noun phrases : un groupe nominal qui ne fait pas partie d'un groupe prépositionnel (0 points) est préférable à un GN qui fait partie d'un GP (-1 point).
  - Collocation pattern preference : ce critère favorise les GN qui apparaissent dans une construction identique au pronom. La préférence est limitée aux contextes de la forme « GN (pronom), verbe » et « verbe, GN (pronom) ». Ex. *Press the key<sub>i</sub> down...press it<sub>i</sub> again.*
  - Immediate reference : cette préférence rend compte d'une construction fréquente dans les manuels techniques. Dans la suite « ...(You) V<sub>1</sub> GN ... CONJ (you) V<sub>2</sub> it (CONJ (you) V<sub>3</sub> it) » où CONJ représente une conjonction (*and, or, before, after...*), le GN qui suit V<sub>1</sub> est un antécédent très probable pour le pronom *it* qui apparaît immédiatement après V<sub>2</sub>. On lui accorde 2 points, 0 aux autres GN.
  - Referential distance : moins un GN est éloigné d'une anaphore, plus il est probable comme antécédent. Dans les phrases complexes, les GN dans la proposition précédente sont les meilleurs candidats (2 points), suivi des GN dans la phrase précédente (1), ensuite deux phrases avant (0) et finalement 3 phrases avant (-1). Pour les anaphores dans les phrases simples, les GN dans la phrase précédente ont le meilleur score (1 point), ensuite ceux qui se trouvent 2 phrases avant (0), et finalement 3 phrases avant (-1).
  - Term preference : les GN qui appartiennent au domaine couvert par le texte sont favorisés (1 point) par rapport à ceux qui n'y appartiennent pas (0).

Les pronoms non anaphoriques, comme le *it* pléonastique, sont éliminés par un

« filtre référentiel ». Les cataphores ne sont pas traitées par l'algorithme, qui se décrit de la façon suivante :

- Examiner les trois phrases précédant l'anaphore (si disponibles) et rechercher les GN à gauche de l'anaphore.
- Éliminer de la liste de candidats potentiels : ceux dont les traits de nombre et genre sont incohérents avec l'anaphore.
- Appliquer les heuristiques préférentielles et assigner des points. Le candidat avec le score total le plus élevé est l'antécédent. Si deux candidats ont le même score, celui avec le meilleur score pour « immediate reference » est préféré. Si cette préférence n'aide pas à trancher, alors c'est « collocation pattern » qui décide; sinon « indicating verbs », sinon choisir le candidat le plus récent.

Les évaluations effectuées par l'auteur ont montré un taux de réussite assez favorable (89,7%) sur un corpus de manuels techniques. De plus, il a été adapté au polonais et à l'arabe avec des taux de réussite encore meilleurs (93,3% et 95,2% respectivement).

#### **4.2. Résolution d'anaphores abstraites**

Il existe relativement peu d'algorithmes pour la résolution d'anaphores abstraites. Un de ces algorithmes, celui de Eckert et Strube, a pour but la résolution non seulement des anaphores individuelles, mais également des anaphores événementielles, statives, propositionnelles et factuelles. Ces dernières sont des anaphores qui ont des antécédents phrastiques.

##### 4.2.1. Eckert & Strube

L'algorithme de Miriam Eckert et Michael Strube (1999) est une extension de l'algorithme formulé par Strube (1998). La différence est qu'il vise non seulement à repérer les antécédents d'anaphores individuelles, mais aussi ceux d'anaphores abstraites (« discourse deictic anaphora »), dans les dialogues oraux. Les auteurs essaient de rendre compte des particularités du discours oral – comme l'attention de locuteurs multiples, la discontinuité, les énoncés sans entités discursives.

Ils partent de la supposition que le type de l'antécédent d'une anaphore abstraite est déterminé par le contexte dans lequel elle apparaît (la notion de *conteneur*). Le cadre de sous-catégorisation du contexte prédicatif de l'anaphore est utilisé pour déterminer le type du référent (propositionnel, événementiel, factuel, statif etc). De même, les arguments de certains verbes sont de préférence des entités nominales concrètes. La notion de conteneur est très importante dans cet

algorithme. L'algorithme classe une anaphore selon qu'elle représente *préférentiellement* une entité individuelle ou abstraite.

Comme il s'agit de dialogues oraux, le discours est divisé en segments selon des critères pragmatiques, plutôt que d'être divisé selon la structure syntaxique des énoncés. De plus, le domaine de recherche des antécédents est défini selon le modèle de discours, au lieu d'être délimité syntaxiquement. Des entités deviennent disponibles pour une référence anaphorique seulement lorsqu'elles établissent un lien entre les interlocuteurs.

Les pronoms impersonnels (sujets de verbes météorologiques, de verbes à montée) sont exclus du traitement.

Une liste de toutes les entités nominales disponibles pour la référence anaphorique est établie (la « S-list » de Strube (1998)). L'algorithme utilise également une liste pour stocker tous les référents des anaphores abstraites. Celle-ci, la « A-list », est remplie seulement dans le cas où un pronom abstrait ou démonstratif est résolu, et son contenu est mis à jour au fur et à mesure du dialogue.

Un ordre précis pour la recherche d'antécédents est défini. La recherche se poursuit dans cet ordre :

1. Dans la A-list, qui contient des entités abstraites déjà anaphorisées
2. Dans le même segment de discours : dans la phrase à gauche de celle qui contient l'anaphore
3. Dans le segment de discours précédent : la proposition la plus à droite (et subordonnées à sa droite)
4. Dans les segments de discours précédents : dans la première phrase complète à partir de la droite

L'algorithme comprend deux branches : une pour la résolution de pronoms personnels et l'autre pour la résolution de démonstratifs. Les deux font appel à des fonctions pour la résolution d'anaphores individuelles et abstraites. À chaque occurrence d'un pronom dans le discours, le type du pronom est déterminé (abstrait ou individuel). Ensuite, l'antécédent est recherché selon le contenu du pronom et selon la séquence de recherche. Le pronom est coindexé avec un élément qui satisfait les contraintes imposées par le contenu. L'antécédent est ajouté à la A-list.

La performance de l'algorithme a été évaluée sur un corpus de dialogues de standards téléphoniques avec un taux de rappel de 70% et une précision 63,6%. La valeur relativement basse de la précision laisse conclure aux auteurs que la classification des pronoms n'était pas très efficace.

Ce qui pourrait éventuellement améliorer les résultats d'une telle approche serait de traiter les différents types d'entités abstraites séparément. Cela éviterait d'éventuels conflits lors de la classification des antécédents, surtout dans la

mesure où les propositions et les faits peuvent contenir des entités d'autres types.

### **4.3. Un algorithme pour la résolution d'anaphores événementielles**

L'anaphore individuelle étant la plus traitée, il existe des algorithmes qui permettent de la résoudre, et ce avec de bons taux de réussite. Pour ces approches, le type d'entité sémantique de l'antécédent n'est pas dans les préoccupations, et c'est principalement la nature syntaxique et configurationnelle qui est centrale. Dans le cadre de cette étude, on s'est penché uniquement sur les antécédents qui sont des entités de type événementiel sous forme phrastique.

La notion sémantique centrale de cette étude, due à Vendler, est celle de conteneur événementiel, qui contraint une anaphore à avoir un événement comme antécédent. Les phénomènes linguistiques que l'on a examinés ont largement porté sur la syntaxe. On s'est concentré sur une étude de certaines constructions et sur leurs effets sur la relation anaphorique événementielle. On a brièvement touché aux notions aspectuelles de temps et de modalité afin de déterminer leur influence sur la reprise anaphorique d'événements.

Le but de cette étude linguistique a été de dégager des indications sur d'éventuelles contraintes ou préférences qui pourraient faire partie d'une procédure de résolution. Ce qui suit est la description d'un algorithme pour la résolution d'anaphores événementielles qui se base sur les données que l'on a présentées.

#### 4.3.1. Algorithme

On supposera comme entrée un texte lemmatisé où chaque unité lexicale a été étiquetée avec sa catégorie grammaticale, ainsi que ses traits morphologiques de genre, nombre, personne, temps ou mode, selon le cas. Le texte sera découpé en propositions numérotés. Chaque unité lexicale et signe de ponctuation aura un indice correspondant à sa position dans une proposition. Le texte sera également traité pour déterminer les rôles grammaticaux des éléments. Les sujets, objets directs, objets indirects et obliques seront mis en évidence.

La première étape sera d'éliminer les pronoms non anaphoriques. Ensuite les verbes événementiels dans le texte seront repérés, et puis les événements coordonnés. Les pronoms à résoudre seront ensuite détectés en fonction des conteneurs. C'est à ce stade que la procédure de résolution intervient. On propose d'utiliser un système de décision comme celui de Lappin et Leass ou Mitkov. Cette approche consiste en l'attribution d'un score à chaque antécédent candidat, en pondérant des facteurs différents, ce qui représente bien la notion de



préférence. Les facteurs que l'on a établis ont chacun un système de pondération différent.

1. Élimination de pronoms non anaphoriques : Comme tous les pronoms ne sont pas anaphoriques, il est nécessaire de distinguer les pronoms qui n'ont pas d'antécédent de ceux qui en ont un. Dans un processus de résolution d'anaphores cette distinction est importante car elle permet d'éviter une tentative de résolution lorsqu'il n'y a pas d'antécédent.

*Ça* non anaphorique :

Verbes météorologiques – *ça caille, ça vase, ça flotte, ça neige, etc.*

Expressions figées – *ça barde, ça va, ça boume, ça roule, etc.*

Sujet de complétive – *ça arrive que Léa se trompe, ça m'étonnerait qu'il vienne, ça arrive de se tromper.*

*Ce* non anaphorique :

Relatives périphrastiques - *Ce qui est rare est cher* (sujet), *Léa aime tout ce qui est cher* (objet).

Formes interrogatives en *est-ce que* - *Est-ce que tu viens ? Qu'est-ce que tu manges ?*

Constructions clivées – *C'est Max qui vient, est-ce Max que Léa aime?*

Construction en *ce + être + que* – *C'est dommage que, c'est vrai que, c'est un fait que.*

Complétives objet indirect – *Je m'attends à ce qu'il vienne,*

**Règle** : Tout pronom non anaphorique doit être marqué comme tel avant la procédure de résolution.

**Implémentation** : On disposera d'une liste des verbes qui prennent un sujet *ça* non anaphorique. La liste contiendra les lemmes de ces verbes, ainsi que leur cadre de sous catégorisation : LEMME <SUJ,COD,COI>. Par exemple :

ALLER <"ça",\_,\_>

ARRIVER <"ça",\_,CP>

BARDER <"ça",\_,\_>

ETONNER <"ça",\_,CP>

Si le sujet d'un verbe dans la liste est *ça* ou *cela*, alors il sera étiqueté non anaphorique, #NANA. Comme on suppose disposer des informations sur les rôles grammaticaux des GN, à chaque fois que l'on trouve dans le texte un verbe de la

liste, on vérifiera si cette occurrence à un *ça* sujet<sup>21</sup>.

Pour les *ce* non anaphoriques le filtrage se fera sur la base de la structure. Toute occurrence de *ce* dans les constructions *c'est que*, *ce qui/que* + VP, *ce* + *être* + GN + *qui/que*, et *ce* + *être* + GA + *que* sera étiquetée #NANA. Les *ce* des interrogatives *est-ce* et des constructions GP + *ce* + *que* seront traités de la même manière.

2. Repérer les verbes événementiels : Les événements sous forme phrastique sont repérables grâce à leur verbe. Les verbes événementiels se distinguent principalement des verbes statifs. Les verbes événementiels décrivent des activités, des accomplissements ou des achèvements selon la classification de Vendler. Cette distinction est assez importante car elle permet de distinguer un antécédent événementiel d'un antécédent statif.

Des exemples de verbes statifs sont *savoir*, *connaître*, *se souvenir*, *posséder*, *avoir*, *survivre*, *exister*, *aimer*, *vouloir* et le verbe *être* suivi d'un adjectif. En contraste, les verbes événementiels dénotent des activités comme *manger*, *peindre*, *lire* et *boire*, des accomplissements comme *manger une pomme*, *peindre la maison*, *lire le journal* et *boire une bière*, ou bien des achèvements comme *atteindre le sommet* ou *gagner (une course)*.

Compiler une liste des verbes événementiels représente un travail extrêmement long et fastidieux. Une telle liste regrouperait les formes ou expressions verbales qui représentent des événements, par opposition aux états. Pour l'instant aucune liste telle ne semble disponible. Les linguistes se sont occupés de formuler la distinction mais aucun travail lexical exhaustif n'a été effectué à cette fin.

**Règle** : Tout verbe événementiel signale la présence de la description d'un événement et doit être disponible comme antécédent pour une relation anaphorique.

**Implémentation** : Pour cette étude on supposera qu'une liste de verbes événementiels est disponible. La liste contiendra les lemmes des verbes. La liste ne distinguera pas les différents types de verbes événementiels (*activité*, *accomplissement*, *achèvement*), car ce niveau de détail n'est pas nécessaire.

Lors du passage dans le texte, toute occurrence d'un verbe dont le lemme figure dans la liste sera étiquetée. On pourrait envisager d'utiliser l'étiquette #VEVT.

---

21 Si on ne disposait pas des rôles grammaticaux on pourrait procéder ainsi : à chaque occurrence d'un verbe répertorié dans la liste, si le premier élément (pro)nominal à sa gauche et dans la même phrase a la forme *cela/ça*, ce pronom sera étiqueté non anaphorique.

3. Repérer les événements coordonnés : Une liste ou ensemble de deux ou plusieurs événements au même temps verbal et reliés par une conjonction de coordination peuvent être repris par un seul pronom. Les conjonctions comme *et*, *mais* et *ou* servent à relier deux événements au même temps verbal et qui ont soit le même sujet, soit un sujet différent :

(154) [Jean arrive et repart]<sub>i</sub>. Ça<sub>i</sub> se passe souvent.

(155) [Jean est arrivé mais Marie est partie]<sub>i</sub>. C'<sub>i</sub>est arrivé ce matin.

(156) [Jean est arrivé, Marie est partie et Paul a appelé]<sub>i</sub>. C'<sub>i</sub>est arrivé ce matin.

(157) [Jean est arrivé et Marie est partie mais Paul a appelé]<sub>i</sub>. C'<sub>i</sub>est arrivé ce matin.

Dans tous ces cas le pronom *ce* (élide en *c'*) fait référence de façon non ambiguë à l'ensemble des événements. On appellera les antécédents de cette forme des *groupes événementiels*.

Il faut noter que les temps verbaux des différents événements sont toujours les mêmes. Ce point est important car si les temps des verbes étaient distincts, ils ne pourraient pas former un groupe événementiel.

**Règle** : Les événements ayant le même temps verbal et qui sont reliés par une conjonction de coordination forment un seul *groupe événementiel*. Cela comprend les listes d'événements séparés par une virgule.

**Implémentation** : On envisagera une liste qui stockera tous les groupes événementiels trouvés dans un texte. Chaque entrée dans la liste consistera en la graphie de chacun des verbes qui appartiennent au groupe accompagné du numéro de la proposition dans lequel il apparaît et son indice. Une fois que l'on a étiqueté les verbes événementiels, on procédera à la délimitation des éventuels groupes événementiels qui seront des potentiels antécédents. De gauche à droite dans une proposition, si deux verbes événementiels sont séparés par une virgule ou par une conjonction de coordination, on crée une entrée dans la liste. Chaque verbe événementiel coordonné sera ajouté à la même entrée.

Pour le texte *Jean est arrivé mais Marie est partie. C'est arrivé ce matin*, l'entrée de liste aurait le contenu suivant :

arrivé(0,2)	// le premier élément (proposition, indice)
partie(1,2)	// le deuxième élément (proposition, indice)

Lors de la recherche de l'antécédent d'un pronom, l'unité disponible sera désormais le groupe événementiel au lieu de ses constituants seuls.

4. Repérer les pronoms à résoudre : La notion de conteneur événementiel réfère aux contextes linguistiques qui requièrent l'intervention d'un événement. Les conteneurs événementiels sont des verbes qui prennent un événement en position argumentale. Les verbes *se passer*, *avoir lieu* et *se produire* prennent en position sujet un événement sous forme nominale ou pronominale :

(158) La chute du mur/ça s'est passé en 1989.

(159) La Révolution/C'est arrivé en 1789.

(160) La finale/Ça aura lieu demain.

Les verbes *assister à*, *être témoin de*, *participer à* et *manquer* prennent un événement en position objet direct, encore sous forme nominale ou pronominale :

(161) J'ai assisté à la chute du mur/ça.

(162) Nous avons participé à la Révolution/cela.

(163) Jean a manqué la finale/ça.

Les pronoms dans ces exemples n'ont pas d'interprétation sans plus de contexte. Le conteneur impose des restrictions sur le type d'antécédent qu'ils peuvent reprendre. Dans ce cas un événement, qui peut être sous forme phrastique :

(164) [Le mur de Berlin est tombé]<sub>i</sub>. C<sub>i</sub>'est arrivé en 1989.

(165) [Le peuple s'est révolté contre la monarchie]<sub>i</sub>. Ça<sub>i</sub> s'est produit en 1789.

(166) [Barça a battu Arsenal]<sub>i</sub>. Jean est déçu d'avoir manqué ça<sub>i</sub>.

**Règle** : Tout pronom en position argumentale d'un conteneur événementiel doit avoir comme antécédent un événement, qu'il soit sous forme nominale ou sous forme phrastique. Cependant, on s'intéresse exclusivement aux antécédents phrastiques.

**Implémentation** : à chaque occurrence d'un conteneur événementiel, on vérifiera s'il a un pronom (*cela/ça, ce*) en position argumentale qui ne porte pas l'étiquette #NANA. Si un de ces pronoms figure comme argument, il sera étiqueté #AEVT pour *anaphore événementielle*. Lors de la résolution, cette étiquette indiquera qu'il faut rechercher un antécédent.

5. Évaluation des antécédent : pour chaque occurrence d'un pronom, le système procède à l'application de scores pour les événements qui apparaissent jusqu'à dix propositions avant une anaphore. À chaque événement est attribué un score pour chacun d'un nombre de facteurs de préférence. Un score total est calculé : il correspond à la somme de tous les scores attribués selon ces facteurs.

L'événement qui a le score le plus élevé est sélectionné comme antécédent. Dans le cas où deux antécédents ont le même score, celui qui est le plus proche de l'anaphore est préféré.

Les facteurs qui contribuent à la sélection sont détaillés ci-dessous.

Dans toutes les approches de la résolution d'anaphores, individuelles et abstraites, la notion de *distance* constitue un facteur central. Elle est calculée en nombre de phrases ou de propositions à gauche d'une anaphore à résoudre. La distance, en ces termes, représente le fait que les entités qui ont été évoquées le plus récemment dans un discours sont les plus saillantes. Il est souhaitable d'intégrer le paramètre de distance dans cet algorithme. Pour ce faire, on adoptera une approche similaire à celle de Mitkov. On attribuera un score initial de 10 aux antécédents, qui diminuera d'un point selon leur distance de l'anaphore à résoudre. On comptera la distance en terme de nombre de propositions grammaticales à gauche d'une anaphore. Voici un exemple :

(167) Paul [a acheté] un billet pour l'Argentine. Il [a payé] avec sa carte bleue. On [a volé] son portefeuille avec le billet dedans. C'est arrivé en ville ce matin.

Le premier antécédent candidat, *a volé*, recevra un score de 9. Le deuxième, *a payé*, aura 8. Le troisième, *a volé*, aura un score de 7 points pour le facteur de distance.

**Règle** : plus un antécédent est proche d'une anaphore, plus il est préférable comme antécédent (10, décrémente de 1 pour chaque nouvelle proposition).

Si la notion de conteneur est très importante, une relation anaphorique événementielle peut, dans un cas particulier, exister sans la présence d'un conteneur événementiel. C'est notamment le cas lorsqu'on a la description d'un événement qualifié par un modifieur dans une construction avec *et ce* ou *et cela/ça* suivi d'un groupe prépositionnel :

(168) [Geneviève a emmené Max à l'hôpital]<sub>i</sub>, et ce<sub>i</sub> dans la voiture de son père.

(169) [L'application compresse même les films très longs en très bonne qualité]<sub>i</sub>, et ce<sub>i</sub> à une grande vitesse.

(170) [L'équipe a lancé la procédure]<sub>i</sub>, et cela<sub>i</sub> avec l'approbation du patron.

Chaque exemple contient une proposition principale qui consiste en la description d'un événement (un verbe événementiel accompagné de ses compléments et d'éventuels modifieurs). Dans la proposition coordonnée introduite par *et*, figure un pronom anaphorique *ce*, *cela/ça* suivi d'un groupe

prépositionnel. Ce groupe prépositionnel est syntaxiquement rattaché au pronom, mais il qualifie sémantiquement l'événement. Le pronom ne peut avoir comme antécédent que l'événement dans la proposition principale.

Ce genre de construction peut aussi qualifier des entités d'autres types, des états par exemple :

(171) Notre objectif est de gagner, et ce depuis le début.

Ce qui différencie ce cas des précédents est l'absence d'un verbe événementiel. La nature de l'antécédent joue un rôle important et les antécédents non événementiels ne sont pas considérés.

**Règle** : les pronoms qui figurent après la description d'un événement dans une construction *et + cela/ça/ce + GP* doivent être résolus. Lorsqu'on a dans le texte une construction de ce type, l'antécédent candidat de préférence est celui qui apparaît directement avant l'anaphore et qui est coordonné avec le GP (5,0).

Certaines contraintes existent sur la reprise anaphorique d'éléments phrastiques.

Les propositions relatives en position sujet représentent parfois une barrière pour la relation anaphorique. Un événement enchâssé dans une relative sujet est moins saillant qu'un événement dans la principale.

(172) #L'homme<sub>i</sub> qui [<sub>t<sub>i</sub></sub> a tué Kennedy]<sub>j</sub> aime la bière. C<sub>j</sub>'est arrivé en 1963 à Dallas.

C'est le verbe dont la relative est sujet qui est plus saillant dans ce cas.

**Règle** : l'événement enchâssé dans une relative sujet est pénalisé (-2) par rapport au verbe de la proposition principale (0).

L'antécédent phrastique d'un pronom dans une proposition relative ou complétive ne peut pas être dans la principale de cette relative. Cela revient à une contrainte similaire à celle exprimée par la théorie du liage pour les entités nominales.

(173) #[Jean a trouvé]<sub>i</sub> que cela<sub>i</sub> s'est passé en Hongrie.

(174) #[Jean a vu]<sub>i</sub> où cela<sub>i</sub> s'est passé.

(175) #[Jean a dit]<sub>i</sub> comment c'<sub>i</sub> est arrivé.

**Règle** : L'antécédent d'un pronom dans une relative, complétive ou interrogative, quel que soit le nombre d'enchâssements, doit être en dehors de la proposition principale. Un événement dans une relative ou complétive est pénalisé pour un pronom qui figure à l'intérieur de la construction. Il reçoit pour ce facteur un score de -10. Les autres candidats ne sont pas affectés (0).

Les complétives (*Que Léa parte*) jouent le rôle d'un groupe nominal. Lorsqu'une complétive apparaît dans une phrase, elle peut être le sujet d'un verbe.

(176) *Que Léa parte a fait pleurer Max. C'est arrivé hier.*

Il y a une ambiguïté due au fait que la phrase contient deux événements. La nature factuelle de la complétive et son statut de sujet mènent à une préférence pour l'événement enchâssé.

**Règle** : dans les complétives sujet qui contiennent un événement enchâssé, il y a une préférence pour cet événement (1) par rapport à celui dont la complétive est sujet (0).

Les complétives de la forme *faire (en sorte) que* sont un cas particulier. Dans une reprise anaphorique événementielle d'un événement enchâssé dans une telle construction, ce dernier est en général préféré à celui que représente la complétive elle-même.

(177) *Jean a fait en sorte que la voiture explose. C'est arrivé hier.*

On explique cela par le fait que les conteneurs événementiels ne sont pas tout à fait cohérents avec le sens de l'expression *faire en sorte*.

**Règle** : dans ces constructions, l'événement enchâssé doit être favorisé (1,0)

Il existe également des contraintes de nature temporelle pour la sélection d'un antécédent. Les temps verbaux peuvent donner des indications utiles. Un pronom dans un conteneur événementiel sélectionne de préférence un événement au même temps verbal.

(178) [*Max avait mis le feu à la maison*]<sub>i</sub>. Il venait de Paris et parlait d'un ton dédaigneux. Paul avait été témoin de cela<sub>i</sub>.

**Règle** : un pronom sélectionne de préférence un antécédent qui est au même temps que le conteneur événementiel sous la portée duquel il apparaît (1,0).

L'effet de la négation n'a pas fait partie de l'étude linguistique ci-dessus. On établira néanmoins une règle qui se base sur la supposition qu'un verbe événementiel nié ne constitue pas un événement.

(179) ?[*Max n'est pas rentré de la soirée chez Léa*]<sub>i</sub>. Ça<sub>i</sub> arrive souvent.

**Règle** : un verbe événementiel dans une construction négative est défavorisé comme antécédent (-5,0)

Chaînes anaphoriques : La notion de chaîne anaphorique désigne la reprise

d'une même entité dans un discours par plusieurs occurrences de pronoms. Lorsqu'on construit un discours, on a tendance à éviter la répétition. On peut répéter un pronom au lieu de mentionner cette même entité à chaque fois qu'on veut en parler. Par exemple :

(180) Il a vendu sa voiture. Il n'*en* avait plus besoin. *Elle* était dans un mauvais état. C'était une 206 rouge.

(181) Il a entendu que son frère a été empoisonné, que *cela* s'est passé en Chine et que *ça* a eu lieu hier.

Dans les deux exemples, les pronoms en italiques reprennent la même entité. Dans le premier cas c'est une entité concrète sous forme nominale, *sa voiture*. Dans le deuxième exemple il s'agit d'une entité abstraite, plus précisément l'événement *son frère a été empoisonné*, qui est reprise. Ce sont des exemples de chaîne anaphorique. Une chaîne anaphorique garantit qu'un antécédent reste saillant dans le discours, même s'il est éloigné d'un pronom. On prendra en compte toute chaîne anaphorique qui consiste en la reprise d'un événement par des anaphores successives qui sont sous la portée d'un conteneur événementiel.

**Règle** : une entité qui a déjà été anaphorisée est plus saillante qu'une autre.

**Implémentation** : On envisagera une adaptation de la « A-list » d'Eckert et Strube. On stockera dans une liste (E-liste) tout verbe événementiel qui a été repris par un pronom. Ces éléments resteront disponibles pour la reprise par un pronom sous la portée d'un conteneur événementiel. La liste stockera la graphie du verbe, son numéro de proposition, son indice dans la proposition, le numéro de proposition de l'anaphore qui l'a repris et son indice :

*verbe* (no. proposition, indice verbe)(no. proposition, indice anaphore)

Cela servira à « avancer » la position de cet événement dans le discours pour représenter le maintien de sa saillance. Par exemple :

(182) Léa [est tombée]<sub>i</sub> au bureau. Paul n'était pas là mais Max [a dit] que c'est arrivé devant tout le monde et que *ça*<sub>i</sub> s'est passé hier.

La E-liste se met à jour dynamiquement, voici son contenu après la résolution du premier pronom, *ce* :

*tombée* (0,2)(1,10)

Lors de l'évaluation des candidats pour le deuxième pronom, *ça*, cet événement, qui est maintenant dans la liste, recevra un score qui correspond à la distance entre le premier pronom et le deuxième.



À la fin de l'analyse du texte, toutes les relations anaphoriques repérées seront répertoriées dans la E-liste.

#### 4.3.2. Description sommaire

##### Éliminer les pronoms non anaphoriques :

- verbes météorologiques
- *ça/cela* barde, va, boume, roule
- *ça/cela* sujet de verbe à complétive
- constructions clivées
- *ce qui/que/dont/à quoi* - relative périphrastique sujet
- *ce qui/que/dont/à quoi* - relative périphrastique objet direct/indirect
- *est-ce que* – interrogatives
- *c'est..qui/que* - clivées déclaratives
- *est-ce..que/qui* - clivées interrogatives
- *à/de ce que* - objet prépositionnel indirect
- *c'est + GA/GP + que*

##### Repérer les verbes événementiels :

- on suppose disposer d'une liste de verbes avec leur cadre de sous-catégorisation

##### Repérer les événements coordonnés :

- regrouper les événements qui ont le même temps verbal et qui sont reliés par une conjonction de coordination, ou par des virgules

##### Baliser les pronoms

- *ça/cela, ce, c'* figurant sous la portée d'un conteneur événementiel
  - arriver, se passer, se produire, avoir lieu* - position sujet
  - rater, louper, manquer* - position objet direct
  - être témoin de, assister à* - position objet prépositionnel indirect
- *ça/cela, ce* figurant dans une construction *et + pronom + SP*
  - ..., *et cela* dans les plus brefs délais
  - ..., *et ça* à minuit
  - ..., *et ce* pour une raison quelconque

##### Évaluation des antécédents

- pour une occurrence de pronom, pour les 10 propositions à sa gauche, commençant par celle qui contient le pronom, attribuer des scores aux verbes

et groupes événementiels selon les critères suivants :

- 1. distance : décrémente de 1 le score pour chaque nouvelle proposition à gauche du pronom. (10,0)
- 2. temps : préférence pour un antécédent au même temps verbal (1,0)
- 3. construction *et + pronom + GP* : l'événement directement précédent cette construction est préférable (5,0).
- 4. complétives, relatives et interrogatives : un événement dans la proposition principale d'une complétive, relative ou interrogative où figure l'anaphore est pénalisé. (-10,0).
- 5. complétive sujet : préférence pour un événement enchâssé dans une complétive sujet, au lieu de la phrase entière. (1,0).
- 6. complétive « *faire (en sorte) que* » : préférence pour un événement enchâssé dans l'expression, au lieu de la principale. (1,0).
- 7. relative sujet : un événement enchâssé dans une relative sujet est pénalisé. (-2,0).
- 8. événement nié : un verbe événementiel dans une construction négative est pénalisé (-5,0).
- 9. chaîne anaphorique : préférence pour un événement déjà anaphorisé. Mettre à jour le facteur de distance pour un verbe dans la E-liste. Sa distance est celle de la dernière anaphore qui le reprend.

Le verbe ou groupe événementiel avec le meilleur score est l'antécédent. Si deux ou plusieurs candidats ont le même score, celui qui a le score le plus élevé pour le facteur de distance est choisi.

## 5. Résultats préliminaires

On a testé l'algorithme à la main sur quelques extraits de textes littéraires obtenus sur le site de la base textuelle Frantext ([www.frantext.fr](http://www.frantext.fr)).

Pour chaque extrait cité, on entoure les verbe événementiels candidats par des crochets. Si le même verbe apparaît plusieurs fois, chaque occurrence est attribuée un indice selon l'ordre d'apparition.

On donne aux pronoms dans le texte un indice (1,2,3...) selon l'ordre de la résolution.

Après chaque texte on liste les candidats potentiels et leurs scores pour chacun des facteurs qui intervient (les autres facteurs n'auront aucun effet sur les scores). Le total pour chaque candidat figure dans la dernière colonne et l'antécédent choisi est indiqué par texte blanc sur un fond noir.

Extrait 1 :

Un anglais qui y [pêchait] à la ligne s'y laissa [choir]. Par un rare bonheur, il [fut sauvé] ; l'eau en [bondissant] le [jeta] comme une paille sur des rochers qui forment un bassin au torrent, d'où le pêcheur [fut pêché]. C'est arrivé l'an dernier.

Pronom C <sub>1</sub>	Facteurs		
	Distance	Temps	Total
Pêchait	5	0	5
Choir	5	0	5
Fut sauvé	7	0	7
Bondissant	8	0	8
Jeta	8	0	8
Fut pêché	9	0	9

Selon l'interprétation la plus naturelle de cet extrait, le pronom fait référence à la suite complexe d'événements. L'algorithme n'applique que deux facteurs, mais le seul qui a un effet est celui de distance. C'est le plus proche des candidats, *fut pêché*, qui est choisi, et son interprétation est naturelle, même si ce n'est pas la meilleure possibilité.

Extrait 2 :

En [parcourant] chaque figure, il [trouve] que son frère n'est plus au monde; qu'il [a été empoisonné], et qu'il [est mort] subitement; que cela<sub>1</sub> est arrivé à la Chine, et que

$c_2$ 'est arrivé dans une capitale de la Chine.

Dans cet extrait il y a deux pronoms à résoudre. Voici le tableau pour le premier, *cela* :

Pronom $c_{1a}$	Facteurs			
Candidat	Distance	Temps	Complétive	Total
Parcourant	6	0	0	6
Trouve	7	0	-10	-3
A été empoisonné	8	1	0	9
Est mort	9	1	0	10

Les facteurs de distance et de temps interviennent encore. Le facteur qui pénalise un antécédent dans la proposition principale d'une complétive figure également. L'antécédent sélectionné est *est mort*, le candidat le plus proche. Un bon antécédent serait aussi la conjonction des candidats *a été empoisonné* et *est mort*, mais l'algorithme ne gère pas deux événements coordonnés qui sont séparés par un enchâssement de complétive. Il choisit alors l'un des deux.

Pronom $c'_2$	Facteurs				
Candidat	Distance	Temps	Complétive	Anaphorisé	Total
Parcourant	5	0	0	0	5
Trouve	6	0	-10	0	-4
A été empoisonné	7	1	0	0	8
Est mort	8	1	0	9	10

Pour le deuxième pronom, les mêmes facteurs sont utilisés. Mais cette fois, un des candidats a déjà été anaphorisé. Pour le facteur de distance, le candidat *est mort* reçoit le score qui correspond à la distance du pronom qui l'a repris (dans la colonne « anaphorisé »). Il reste alors l'antécédent de préférence.

Extrait 3 :

Bientôt nous en avons eu cinq ou six d'accrochés au bordage et qui nous [demandaient] l'aumône. Nos marins [tapaient] dessus à coups de cordes et Joseph avec ses pincettes, et tout cela<sub>1</sub> pour les [décourager] un peu et qu'il n'en [vienne] pas davantage. On leur [donne] quelques piastres, ils les [mettent] dans leur bouche et s'en [retournent] par le même chemin. Mais ce#NANA qu'il y a eu de beau tantôt, c#NANA 'est que  $c_2$ 'est arrivé à un moment où l'on changeait de manoeuvre.

Ici, deux occurrences de pronom *ce* non anaphorique sont repérées et étiquetées #NANA. Le premier pronom à résoudre est *cela*, qui apparaît dans une construction *et + pronom + SP*. Le premier candidat est largement favorisé, ce qui fait qu'il est sélectionné comme antécédent.

Pronom <i>cela</i> <sub>1</sub>	Facteurs		
Candidat	Distance	ET + SP	Total
Demandaient	8	0	8
Tapaient	9	5	14

Dans cet extrait on a un groupe événementiel qui consiste en une liste de trois événements, dont les verbes sont *donne*, *mettent* et *retournent*. L'algorithme les considère comme un seul candidat.

Le deuxième pronom à résoudre est *c'*.

Pronom <i>cela</i> <sub>1</sub>	Facteurs				
Candidat	Distance	Temps	Négation	Anaphorisé	Total
Demandaient	4	0	0	0	4
Tapaient	5	0	0	6	6
Décourager	6	0	0	0	6
Vienne	7	0	-5	0	2
Donne + mettent + retournent	8	0	0	0	8

Cette fois, c'est le groupe événementiel qui est sélectionné comme antécédent. C'est, encore une fois, le candidat le plus proche qui est choisi. Ce n'est pourtant pas le meilleur choix. L'antécédent le plus naturel est encore une suite complexe d'événements qui sont évoqués dès le début du paragraphe et jusqu'à la reprise. Ceci dit, le choix n'est pas tout à fait mauvais, mais ce n'est pas ce qu'un locuteur choisirait comme antécédent.

À partir de ces quelques exemples on constate l'importance du facteur de distance. Il joue un rôle prépondérant dans le choix d'un candidat, ce qui rend les autres facteurs souvent insignifiants. Cela fait que le choix peut être mauvais. Il faudrait trouver le moyen de tempérer l'effet de ce facteur afin de permettre une décision plus équilibrée qui repose sur tous les critères possibles.

Cet algorithme ne permet de repérer que certaines suites simples d'événements. Lorsqu'il existe des liens entre deux événements, notamment de nature temporelle ou causale, il n'en rend pas compte. Cependant, afin de bien évaluer la performance, il serait intéressant de procéder à une implémentation.

## 6. Conclusion et travail futur éventuel

Pour cette étude, on a voulu se placer dans l'optique d'une implémentation d'un programme de TAL, en exploitant des données de nature syntaxique et lexicale. On a également considéré, plus brièvement, les propriétés aspectuelles des modalités et des temps. Ces deux phénomènes sont très complexes et ne se prêtent pas facilement à une implémentation. L'algorithme prend en compte les temps de façon très superficielle mais la modalité n'a pas été intégrée.

La notion la plus importante intégrée à l'algorithme est celle de conteneur. Le contexte prédicatif d'une anaphore impose des contraintes sur le type de l'antécédent. La littérature scientifique et philosophique, depuis Vendler, en témoigne, et des algorithmes qui reposent sur cette notion ont déjà été conçus.

Une première conclusion que l'on peut tirer de cette étude est que la syntaxe ne semble pas imposer un grand nombre de contraintes rigides sur la sélection d'un antécédent phrastique, du moins lorsqu'il est de type événementiel. À part quelques exceptions, les constructions linguistiques que l'on a examinées n'ont pas constitué de barrière à la relation anaphorique entre un pronom démonstratif neutre sous la portée d'un conteneur événementiel et un antécédent événementiel phrastique. La principale contrainte que l'on a observée, sur les subordinées complétives, relatives et interrogatives, ressemble aux contraintes décrites par la théorie du liage sur les entités nominales<sup>22</sup>. Cette analogie nécessite d'être explorée plus en détail.

On a constaté que certaines propriétés morpho-syntaxiques ou sémantiques peuvent déterminer la nature de la relation anaphorique pronominale, c'est-à-dire si elle engendre une coréférence actuelle ou uniquement virtuelle. Le mode conditionnel des verbes et la conjonction *si* en sont des exemples.

En ce qui concerne le matériau de reprise, les pronoms *cela/ça* et *ce* peuvent tous faire l'objet d'une relation anaphorique virtuelle ou actuelle pour les entités abstraites. En comparaison, pour les anaphores individuelles, il y a une division entre les pronoms personnels (*il, elle, le, la, lui*), qui représentent une reprise actuelle, et le pronom *en* qui peut être virtuel.

Les temps verbaux et la modalité fournissent des contraintes sur le type de relation, mais aussi sur la sélection de l'antécédent. Ce sont des propriétés propres aux antécédents phrastiques. Il serait intéressant d'effectuer une étude plus approfondie de l'effet qu'ont les temps verbaux et la modalité sur la relation

---

<sup>22</sup> Voir Chomsky (1981).

d'anaphore abstraite. Cela permettrait de dégager des caractéristiques propres à ce genre de relation. De plus, une analyse d'autres facteurs sémantiques pourrait être utile. On sait, par exemple, que la quantification et la négation posent des contraintes pour l'anaphore individuelle, mais cela n'est *a priori* pas le cas pour l'anaphore abstraite<sup>23</sup>.

L'algorithme que l'on a formulé ici est assez simple : après avoir déterminé les antécédents candidats, il évalue la relation qui existe entre une anaphore donnée et chacun de ces candidats. Par contre, il ne prend que très peu en compte la relation qui peut exister entre deux candidats. On peut considérer qu'il relève d'une approche impérative assez statique et qu'il ne vise pas à représenter la façon dont les humains interprètent un discours. De plus, il consiste à effectuer des calculs pour chaque occurrence d'un pronom, ce qui n'est pas très optimisé.

Une des idées principales de la théorie du centrage est de capturer le modèle « attentionnel » d'un locuteur par rapport aux relations qui existent entre les entités discursives. Ce genre de modèle se rapproche d'une représentation du système cognitif humain, et rend compte de la façon dont les locuteurs construisent et interprètent un discours. Dans un modèle de centrage, les entités discursives sont introduites au fur et à mesure du discours et deviennent plus ou moins saillantes selon les relations qu'elles entretiennent. Celle qui est la plus saillante (appelé le centre) au moment de l'apparition d'une anaphore est sélectionnée comme antécédent<sup>24</sup>. Ce type d'approche est alors plus dynamique et ne demande pas d'effectuer les mêmes calculs de façon répétitive.

Afin de bien représenter le phénomène d'anaphore abstraite, il est pertinent de se placer dans un cadre théorique qui permette de représenter les contraintes qui existent entre les éléments phrastiques. La DRT a été conçue pour rendre compte des anaphores non liées (anaphore interphrastique). Ce formalisme de sémantique dynamique permet d'analyser un discours en faisant des mises-à-jour du contexte avec chaque nouvelle phrase prise en entrée. Lorsqu'une entité discursive est introduite dans le discours, elle reste disponible à une reprise anaphorique selon des critères structurels. En DRT, une entité discursive événementielle, représentée par une variable existentiellement quantifiée, est disponible pour la reprise anaphorique de la même façon que les entités individuelles. Le processus de résolution des anaphores est intégré dans le formalisme. Il pourrait maintenant être bénéfique de se tourner vers un formalisme comme la DRT, qui a, de plus,

---

23 Voir annexe sur la quantification temporelle.

24 Cependant, il faut noter que le centrage, tel qu'il est formulé, ne traite que les entités discursives individuelles. Il serait nécessaire d'adapter le modèle afin de rendre compte des entités abstraites.

déjà fait l'objet de diverses implémentations (notamment dans Blackburn & Bos (2005)).

Une approche hybride de sémantique dynamique et centrage prendrait en compte à la fois les relations qui existent entre les entités discursives et celle qui existe entre une anaphore et un antécédent potentiel. Elle pourrait intégrer les critères sémantiques évoqués plus haut pour disposer d'une plus large couverture au niveau des heuristiques. La DRT permettrait de rendre compte des entités abstraites événementiels en tant qu'antécédents interphrastiques potentiels. Cette approche viserait une représentation dynamique de la façon dont on construit et interprète un discours.



## 7. Références

Amsili, Pascal; Pascal Denis & Laurent Roussarie (2005). *Anaphores abstraites en français : représentation formelle*. *Traitement Automatique des Langues*, 46:1. pp 15-39.

Asher, Nicolas (1993). *Reference to Abstract Objects in Discourse*. Kluwer, Dordrecht.

Bach, Emmon (1986). *The Algebra of Events*. *Linguistics and Philosophy* 9. pp 5-16.

Benz, Katia (2001). *Résolution anaphorique : Ébauche d'un cadre formel hybride Centrage / DRT*. Mémoire de DEA. Université Paris 7, France.

Blackburn, Patrick & Johan Bos (2005). *Representation and Inference for Natural Language. A First Course in Computational Semantics*. CSLI, Stanford.

Chomsky, Noam (1981). *Lectures on Government and Binding*. Foris, Dordrecht.

Corblin, Francis (1994). *Existe-t-il un « ça » impersonnel en français?* *L'information grammaticale* 62, juin 1994. pp 45-47.

Corblin, Francis (1995). *Les formes de reprise dans le discours*. Presses Universitaires de Rennes, France.

Cornish, Francis (1986). *Anaphoric Relations in English and French*. Croom Helm, London.

Davidson, Donald (1967). *La forme logique des phrases d'action, traduction française de P. Engel* (1993), in *Actions et Événements*. Presses Universitaires de France, Paris.

Eckert, Miriam & Michael Strube (1999). *Resolving Discourse Deictic Anaphora in Dialogues*. *Proceedings of the ninth conference of EACL*. pp 38-44.

Eckert, Miriam & Michael Strube (2001). *Dialogue Acts, Synchronising Units and Anaphora Resolution*. *Journal of Semantics*, 17:1. pp 51-89.

- Godart-Wendling, Béatrice (2000). *Comment ça réfère ?* Revue de Sémantique et Pragmatique, 7. Presses Universitaires d'Orléans. pp 105-121.
- Grosz, Barbara J., Aravind K. Joshi & Scott Weinstein (1995). *Centering : A Framework for Modeling the Local Coherence of Discourse*. Computational Linguistics, 21:2. pp 203-225.
- Hobbs, Jerry R. (1978). *Resolving Pronoun References*. Lingua, Volume 44. pp 311-338.
- Kleiber, George (1983). *Les démonstratifs (dé)montrent-ils? Sur le sens référentiel des adjectifs et pronoms démonstratifs*. Le français moderne 2. pp 99-117.
- Lappin, Shalom & Herbert J. Leass (1994). *An Algorithm for Pronominal Anaphora Resolution*. Computational Linguistics, 20:4. pp 335-361.
- Lewis, David (1973). *Counterfactuals*. Harvard University Press, Cambridge, MA.
- Milner, Jean-Claude (1982). *Ordres et raisons de langue*. Seuil, Paris.
- Mitkov, Ruslan (1998). *Robust Pronoun resolution With Limited Knowledge*. Annual Meeting of the ACL, Proceedings of the 36th annual meeting on Association for Computational Linguistics, 2. pp 869-875.
- Parsons, Terence (1989). *The Progressive in English : Events, States and Processes*. Linguistics and Philosophy, 12:2. pp 213-241.
- Parsons, Terence (1994). *Events in the Semantics of English : A Study in Subatomic Semantics*. MIT Press, Cambridge, MA.
- Riegel, Martin; Jean-Christophe Pellat & René Rioul (1994). *Grammaire méthodique du français*. Presses universitaires de France, Paris.
- Roberts, Craige (1989). *Modal Subordination and Pronominal Anaphora in Discourse*. Linguistics and Philosophy, 12:6. 683-721.

Schuster, Ethel (1988). *Anaphoric Reference to Events and Actions: A Representation and Its Advantages*. Proceedings of the 12th conference on Computational Linguistics, 2. pp 602-607.

Schuster, Ethel (1992). *Pronominal Reference to Events & Actions : Computational Foundations*. Thèse de doctorat. University of Pennsylvania, School of Engineering & Applied Science, Computer & Information Science Department.

Vendler, Zeno (1957). *Verbs and Times*. The Philosophical Review, 66. pp 143-160.

Vendler, Zeno (1967). *Linguistics and Philosophy*. Cornell University Press, Ithaca, N.Y.

## 8. Annexe

Cette annexe contient les résultats de quelques exercices effectués par des francophones. Le but des exercices, qui n'ont pas toujours été bien conçus, était de vérifier si certains phénomènes mettaient en valeur des préférences pour un antécédent événementiel ou pas. Ces résultats, très sommaires, ne servent qu'à donner une indication et n'ont pas forcément été intégrés dans l'algorithme.

### 8.1. Les temps verbaux

Pour chaque exemple, le lecteur devait répondre à la question *qu'est-ce qui est arrivé hier?* Les antécédents possibles sont *on dit que Paul gagne (D)*, *Paul gagne (G)* ou *les deux (DG)*. La colonne MAX% indique le pourcentage de réponses favorables pour l'antécédent préféré.

Exemple	D	G	DG	MAX %
1. On dit que Paul gagne. C'est arrivé hier.	0	3	0	100
2. On dit que Paul a gagné. C'est arrivé hier.	0	3	0	100
3. On a dit que Paul gagne. C'est arrivé hier.	1	2	0	66,67
4. On a dit que Paul a gagné. C'est arrivé hier.	0	2	1	66,67
5. On a dit que Paul gagnerait. C'est arrivé hier.	0	2	1	66,67
6. On disait que Paul allait gagner. C'est arrivé hier.	0	3	0	100
7. On disait que Paul gagnerait. C'est arrivé hier.	0	3	0	100

On remarque une nette préférence pour l'antécédent le plus proche, quel que soit le temps verbal. Cependant certaines suites de phrases ne sont pas naturelles et les participants ne l'ont pas signalé, se concentrant plutôt sur la réponse à la question. Les résultats ne sont donc pas représentatifs de la naturalité des énoncés.

### 8.2. Quantification temporelle

Les instructions étaient les mêmes que pour l'exercice précédent. La colonne intitulée « ? » indique que le participant n'a pas pu prendre une décision quant à l'antécédent préféré.

Exemple	D	G	DG	?	MAX %
1a. On dit que Paul gagne souvent. C'est arrivé hier.	0	9	0	0	100
1b. On dit souvent que Paul gagne. C'est arrivé hier.	5	2	2	0	55,56
1c. On dit souvent que Paul gagne parfois. C'est arrivé hier.	5	4	0	0	55,56
2a. On dit que Paul a souvent gagné. C'est arrivé hier.	0	9	0	0	100
2b. On dit souvent que Paul a gagné. C'est arrivé hier.	7	1	1	0	77,78
2c. On dit souvent que Paul a parfois gagné. C'est arrivé hier.	5	2	0	2	55,56
3a. On a dit que Paul gagne souvent. C'est arrivé hier.	2	6	0	1	66,67
3b. On a dit souvent que Paul gagne. C'est arrivé hier.	5	0	2	2	55,56
3c. On a dit souvent que Paul gagne parfois. C'est arrivé hier.	4	2	0	3	44,44
4a. On a dit que Paul a souvent gagné. C'est arrivé hier.	0	5	4	0	55,56
4b. On a dit souvent que Paul a gagné. C'est arrivé hier.	3	2	4	0	44,44
4c. On a dit souvent que Paul a parfois gagné. C'est arrivé hier.	4	3	1	1	44,44
5a. On a dit que Paul gagnerait souvent. C'est arrivé hier.	1	7	1	0	77,78
5b. On a souvent dit que Paul gagnerait. C'est arrivé hier.	2	7	0	0	77,78
5c. On a souvent dit que Paul gagnerait parfois. C'est arrivé hier.	3	5	1	0	55,56
6a. On disait souvent que Paul allait gagner. C'est arrivé hier.	1	7	1	0	77,78
6b. On disait que Paul allait souvent gagner. C'est arrivé hier.	1	7	0	1	77,78
6c. On disait souvent que Paul allait parfois gagner. C'est arrivé hier.	3	5	1	0	55,56
7a. On disait que Paul gagnerait souvent. C'est arrivé hier.	0	8	1	0	88,89
7b. On disait souvent que Paul gagnerait. C'est arrivé hier.	1	7	1	0	77,78
7c. On disait souvent que Paul gagnerait parfois. C'est arrivé hier.	3	6	0	0	66,67

Les exemples 1 à 4 peuvent mener à la conclusion que l'antécédent préféré est celui qui est quantifié par l'adverbe temporel. Cependant les exemples 5 à 7 montrent que dans certains cas c'est systématiquement le dernier candidat qui est l'antécédent le plus naturel. Cependant, cela peut être dû aux combinaisons des temps verbaux et non pas à un effet de la quantification temporelle. Il faudrait mieux isoler les paramètres afin d'avoir des résultats plus utiles.