Le repérage de l'hyperonymie par un faisceau d'indices : mise en question de la notion de " marqueur "

Anne Condamines, Marie-Paule Jacques

▶ **To cite this version:**

Anne Condamines, Marie-Paule Jacques. Le repérage de l'hyperonymie par un faisceau d'indices : mise en question de la notion de " marqueur ". Semaine de la connaissance, Jun 2006, 30062006, France. halshs-01321031

HAL Id: halshs-01321031
https://shs.hal.science/halshs-01321031

Submitted on 24 May 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Le repérage de l'hyperonymie par un faisceau d'indices : mise en question de la notion de « marqueur »

Anne Condamines[1] et Marie-Paule Jacques[1]

[1] ERSS (UMR 5610, CNRS et Université de Toulouse Le Mirail), 5 allées Antonio Machado 31058 Toulouse CEDEX, anne.condamines@univ-tlse2.fr, marie-paule.jacques@univ-tlse2.fr

Résumé

Cet article s'intéresse à la construction de la relation d'hyperonymie en montrant qu'elle se fait souvent en faisant intervenir différents indices discursifs plutôt qu'un seul marqueur de relation. Ce faisant, cet article s'interroge sur la notion de marqueur de relation au regard de la complexité des fonctionnements discursifs.

Mots clés : anaphore infidèle, discours, hyperonymie, marqueur de relation.

1 Introduction

Depuis une quinzaine d'années, les textes sont utilisés comme vecteurs de connaissances pour construire des terminologies ou des ontologies. L'utilisation de textes est venue se substituer à l'interrogation d'experts ce qui a permis de mettre en œuvre des méthodes d'extraction automatique de connaissances. La thématique de la construction de ressources terminologiques, soutenue par une demande sociétale de la gestion de la documentation très prégnante, s'est ainsi fortement développée. Concrètement, construire une terminologie ou une ontologie à partir d'un corpus consiste à élaborer des réseaux terminologiques en mettant en œuvre une analyse des contextes contenant des « termes candidats » et en cherchant en particulier à repérer les « marqueurs » de telle ou telle relation. Il est évident que ce mode de représentation ne constitue pas un modèle unique d'explication du fonctionnement de la connaissance mais il est très répandu et parfaitement adapté aux traitements automatiques. Toutefois, de nombreux travaux en TAL (Traitement Automatique des Langues) ont eu tendance à considérer comme connaissances les seules portions de discours (les « marqueurs ») directement associables à une représentation relationnelle. Nous souhaitons ré-examiner cette position, qui tient compte surtout des possibilités des traitements automatiques, en la confrontant à la réalité du fonctionnement discursif. Nous nous intéressons particulièrement à la construction de la relation d'hyperonymie en montrant qu'elle se fait souvent en faisant intervenir des indices qui, par leur convergence, permettent cette construction.

2 Présentation du problème

2.1 La notion de marqueur de relation conceptuelle

La notion de marqueur de relation est apparue sous divers noms d'abord en sémantique lexicale plutôt à orientation cognitive puis en traitement automatique des langues. Ainsi, Cruse parle de « diagnostic frames » [9]et Ahmad de « knowledge probes » [1]. Dans les deux cas, le choix dénominatif met l'accent sur la notion d'accès à une connaissance favorisée par certaines structures plutôt que sur le fonctionnement discursif de ces éléments.

La définition et la mise en oeuvre des marqueurs de telle ou telle relation ont fait l'objet de nombreux travaux, en particulier en TAL et en IC (Ingénierie des Connaissances) [3], [13], [15], mais elle s'est accompagnée d'une vision souvent simplificatrice des fonctionnements. On a eu ainsi tendance à considérer le couple marqueur/relation comme une association très stable entre un élément langagier et une interprétation en terme de relation (cette relation étant une « connaissance »). Ainsi, un marqueur serait un élément langagier dont la recherche systématique dans un corpus permettrait de trouver une relation en générant le moins de bruit possible.

Cette vision figée du fonctionnement des marqueurs a parfois conduit à penser qu'était riche en connaissances un texte riche en marqueurs (souvent seulement compris comme des formes repérables automatiquement). Or, l'analyse fine des textes permet de mettre en évidence deux caractéristiques de fonctionnement qui mettent en question ce point de vue sur les marqueurs.

La première caractéristique concerne la nécessaire prise en compte du contexte linguistique et extra-linguistique.

Tout d'abord, et c'est un élément souvent mentionné, un marqueur est la plupart du temps constitué d'un élément pivot qui ne peut fonctionner comme marqueur que dans des contextes syntaxiques donnés.

Condamines et Jacques

Prenons l'exemple de la structure *être-un*, généralement associée à une hyperonymie. L'hyperonymie ne sera avérée qu'avec certains déterminants auxquels on pourra associer une valeur « générique » (par exemple, pas *une partie de* ou *quelques*) et à condition que le nom en position d'attribut soit modifié. Ainsi dans :

Une tocante est une montre

on n'a pas de relation d'hyperonymie mais une relation d'équivalence.

Ensuite, au-delà de la notion de contexte linguistique, le contexte extra-linguistique peut être déterminant dans l'attribution d'un rôle de marqueur à tel ou tel élément. Le genre textuel peut ainsi avoir un rôle majeur (parfois supérieur au rôle du contexte syntaxique) dans la possibilité d'interpréter telle ou telle préposition avec une certaine valeur relationnelle. C'est le cas de *chez* et de *avec* [5], [6].

En effet, pour ces deux prépositions, une interprétation méronymique est possible mais elle n'est massivement présente que dans des textes qui relèvent d'un genre bien précis. Pour *chez,* les textes doivent relever du domaine des sciences naturelles (zoologie et biologie). Pour *avec,* les textes doivent avoir un objectif de mise en saillance commerciale (exemple : catalogue de jouets/petites annonces) et/ou perceptive (exemple : description d'itinéraires, catalogues touristiques).

Exemples :

Chez les colobinés, le nez fait saillie sur la lèvre supérieure (encyclopaedia Universalis) (*nez* partie-de *colobinés*)

Porteur évolutif avec canne, repose-pieds et ceinture de sécurité amovible (catalogue) (*canne, repose-pieds, ceinture de sécurité* partie-de *porteur évolutif*).

La deuxième caractéristique concerne le fait que la décision de construire une relation se fait très souvent non pas en utilisant un passage unique du texte (trop ambigü ou trop imprécis) mais un ensemble de contextes, souvent disséminés en différents endroits du corpus et qui servent d'autant d'indices pour permettre la construction d'une relation.

C'est précisément sur les cas de ce second type que porte le travail présenté ici et plus précisément sur la construction de la relation d'hyperonymie.

2.2 Le cas de la construction de l'hyperonymie

Que ce soit pour l'élaboration de terminologies ou d'ontologies, la relation hyperonymique est considérée comme essentielle pour représenter la connaissance, en particulier parce qu'elle est la seule à permettre l'héritage de propriétés, ce qui constitue une économie représentationnelle majeure. A cause de son importance, les « marqueurs » qui peuvent lui être associés ont faits l'objet de descriptions nombreuses. On peut ainsi faire remonter la première description à Aristote qui a mis en évidence la présence d'un hyperonyme (genus) dans les définitions.

"Il faut en définissant poser l'objet dans son genre et alors seulement y rattacher les différences" (Aristote, *Topiques* VI, 1)

Mais cette structure, qui permet de repérer de manière assez fiable une hyperonymie n'est pas toujours présente dans les textes, en particulier dans ceux qui sont rédigés par des spécialistes à destination d'autres spécialistes. Pour autant, il n'est pas impossible de construire des relations hyperonymiques à partir de ces textes, en mettant en oeuvre des ressources discursives, sans doute moins transparentes mais qui constituent autant d'indices d'une possible interprétation hyperonymique.

Le point de départ de notre étude est constitué par une structure discursive qui, dans certains cas, peut mettre en oeuvre une relation d'hyperonymie et de ce fait peut être considérée comme un marqueur d'hyperonymie : il s'agit de l'anaphore infidèle.

L'anaphore nominale infidèle est un procédé discursif qui permet de reprendre une partie d'un contexte par un nom ou un groupe nominal qui n'a aucun élément de forme commun avec ce contexte [2].

Il est connu que la reprise anaphorique infidèle peut faire intervenir une relation d'hyperonymie entre l'élément anaphorique et un élément nominal qui apparaît dans le contexte gauche [8]. Par exemple :

A bomb exploded yesterday in a car *belonging to a well-known magistrate. No-one was hurt, but* the vehicle *was completely destroyed.*

Mais on sait aussi que cette anaphore ne se met pas toujours en place entre deux noms mais souvent entre un nom (l'anaphorique) et un antécédent qui peut prendre d'autres formes ou même ne pas correspondre à une forme grammaticale identifiable mais à une portion de discours comme dans :

La pression atmosphérique avait été évaluée il y a une trentaine d'années à 1.12 de celle de l'atmosphère terrestre. On a réduit ***cette appréciation*** *car Mariner 4 a trouvé qu'elle équivalait à 6 milibars.* (GEO)

Une étude menée par Anne Condamines [7] a montré que le fonctionnement de l'anaphore infidèle n'est pas très différent dans les corpus spécialisés et dans les corpus non-spécialisés en tout cas en ce qui concerne la nature de l'élément anaphorisé. L'étude a été menée sur 5 corpus différents :

- un manuel de géomorphologie (206 000 mots) (GEO)
- un manuel technique fourni par EDF (148 100 mots) (GDP)
- un manuel de spécification en génie logiciel, fourni par EDF (451 000 mots) (Moug)
- un roman (Bel Ami), 170 200 mots (Bel A.)
- des extraits du Monde Diplomatique (110 700 mots) (LMD).

Concrètement, seuls les SN démonstratifs dont le N ne figure pas dans le paragraphe précédent ont été retenus,

ceci afin d'éviter les reprises fidèles. Les contextes ont été recherchés de façon automatique et, s'il est incontestable que l'anaphore infidèle par hyperonyme est tout autant possible avec des SN définis, il demeure que le SN défini est un bien moins bon indicateur d'anaphore que le SN démonstratif, dans la mesure où il peut jouer d'autres rôles discursifs que l'anaphore. Rechercher de façon automatique non seulement les SN démonstratifs mais aussi les SN définis dont le N ne figure pas dans le paragraphe précédent aurait sans doute augmenté le nombre de contextes pertinents mais aurait aussi augmenté excessivement le nombre de réponses, dans la mesure où l'on n'a pas de moyen de filtrer automatiquement seulement les SN anaphoriques parmi tous les SN définis.

Dans le même esprit, tous les SN démonstratifs n'ont pas été retenus, une « stop-list » a été appliquée afin d'écarter les « ce N » présentant une haute probabilité d'avoir une valeur déictique en raison de la valeur sémantique du N : des noms qui renvoient au lieu et au temps de l'énonciation, tels que *heure, instant, jour, année, date, matin, soir, mois,* etc. ; des noms qui renvoient au support écrit lui-même : *article, manuel, ouvrage, chapitre,* etc. ; enfin des noms apparaissant dans des structures figées généralement employées pour la progression du discours : *base, but, cas, condition,* etc.

L'étude des résultats obtenus a mis en évidence que l'élément anaphorisé pouvait être de différentes natures :

- lien avec un mot identifiable
  - l'anaphorique est un hyperonyme : hy
  - l'anaphorique est un synonyme : syn
  - l'anaphorique est une nominalisation déverbale (*gérer/gestion, étudier/étude*) : Dév
  - l'anaphorique est un déadjectival (*mat/matité*) : déadj
  - l'anaphorique est un dénominal (très rare) : dén
  - l'anaphorique est une figure : fig.
- lien avec une portion de discours, cas appelés supplétifs : Supp.

Le tableau suivant montre comment se répartissent ces modes d'anaphorisation selon les 5 corpus.

TAB. 1 – *répartition des occurrences anaphoriques (en pourcentages)*

Corpus	**GEO**	**GDP**	**Moug**	**LMD**	**Bel A.**
Hy	26%	32%	60%	19%	15,5%
Supp.	50%	55%	31,5%	64,5%	47%
Syn	10%	5,5%	4,5%	9%	14,5%
Dév	9%	5,5%	4%	1%	1%
Déadj	2%	2%	0%	1%	0%
Dén	0%	0%	0%	0,5%	0%
Fig	3%	0%	0%	5%	22%
Total	100% (266)	100% (246)	100% (107)	100% (415)	100% (305)

Le repérage de l'hyperonymie

On voit que l'anaphore basée sur une relation hyperonymique n'est pas beaucoup plus présente dans les corpus spécialisés que dans les corpus non-spécialisés, hormis dans le manuel de génie logiciel (Moug). Ce résultat est assez étonnant. En effet, on aurait pu s'attendre à ce que, les relations conceptuelles étant réputées plus stables dans les domaines spécialisés (techniques en particulier) que dans les domaines généraux, la relation d'hyperonymie soit plus souvent convoquée pour les processus d'anaphorisation dans les corpus spécialisés.

Ce résultat oblige à un constat majeur concernant le rôle du genre textuel. En effet, autant le fonctionnement de certains marqueurs est fortement déterminé par la nature du genre textuel dans lequel il apparaît (voir ci-dessus *avec* et *chez*), autant ce fonctionnement discursif (« reprise anaphorique ») ne semble pas tellement influencé par le genre.

Même si elle est finalement assez peu présente, il est intéressant d'étudier comment fonctionne cette hyperonymie dans cette structure discursive de reprise anaphorique. Il faut bien comprendre que l'hyperonymie n'est pas réellement « marquée » par la reprise anaphorique. Elle est plutôt présupposée (supposée connue par le locuteur et l'interlocuteur) et implicite dans le discours. Cela signifie que *a priori*, elle pourrait ne pas être identifiable par un non expert du domaine qui ne la connaîtrait pas déjà – notons que la préposition *chez* fonctionne aussi de cette manière (faire appel à une connaissance supposée connue) : elle n'est pas à proprement parler un marqueur de méronymie mais on sait qu'elle est souvent utilisée dans des contextes méronymiques.

Toutefois, ce qu'a montré l'étude, c'est que la reprise anaphorique déclenche un phénomène plus ou moins conscient de recherche de référent (l'élément anaphorisé). Le locuteur peut ne pas aller jusqu'au bout de cette recherche mais le linguiste ou terminologue peut mettre en œuvre un processus de recherche d'un référent potentiel, qui lui paraîtra plus ou moins évident selon les contextes linguistiques et selon sa compétence dans le domaine.

Ainsi, l'anaphore infidèle joue un rôle de déclencheur d'une recherche d'un éventuel hyponyme et peut constituer un premier indice précieux pour la construction d'une relation d'hyperonymie.

D'autres indices peuvent jouer un rôle dans cette « enquête », jusqu'à ce qu'une relation claire et cohérente soit identifiée ou au contraire invalidée.

Comme un enquêteur, le linguiste peut ainsi utiliser un faisceau d'indices qui vont lui permettre de tirer du texte bien plus de connaissances qu'une lecture très superficielle pouvait le laisser penser.

3 Expérimentation sur un corpus spécialisé

Nous montrons ici comment, à partir de ces contextes potentiellement porteurs de l'expression d'une relation d'hyperonymie, il est possible d'opérer une recherche ciblée dans un corpus afin d'obtenir les éléments de décision sur la validité de cette relation.

3.1 Mise en place de l'étude

Nous avons mis en œuvre cette démarche de recherche de connaissances à partir de déclencheurs et d'indices sur un corpus spécialisé dans le domaine du vol libre, plus particulièrement du parapente. Le corpus est constitué de 50 articles (environ 80 000 mots) provenant de deux rubriques de magazines spécialisés dans le vol libre : la rubrique « essais de voiles » et la rubrique « météo ». Ces articles ne s'adressent *a priori* pas à des néophytes mais à des pratiquants expérimentés, ce qui leur confère certaines particularités discursives : ils ne comportent que très peu de contextes définitoires – et on sait combien les énoncés définitoires sont précieux pour l'élaboration de taxinomies et/ou d'ontologies – ; ils sont riches en réductions de termes, c'est-à-dire en occurrences de termes complexes dont un constituant est effacé (par exemple, *le thermique* pour *l'ascendance thermique* ou *la fermeture* pour *la fermeture asymétrique* [12]). Nous verrons plus loin comment ces caractéristiques interviennent pour la recherche de relations d'hyperonymie.

Comme pour l'étude citée précédemment [7], les contextes servant de déclencheurs sont constitués d'occurrences de SN démonstratifs dont le N ne figure pas dans le même paragraphe, ceci afin d'éviter les reprises fidèles (nous avons repris le même protocole d'extraction automatique décrit dans la section 2.2).

Nous avons ainsi recueilli 228 contextes pour lesquels un antécédent a systématiquement été recherché. C'est ensuite en fonction de la nature de cet antécédent que nous avons classé l'élément anaphorique, c'est-à-dire le SN démonstratif. Nous nous intéresserons plus loin aux difficultés qui peuvent se rencontrer dans cette situation d'interprétation de la relation anaphorique en corpus spécialisé, nous commentons d'abord les résultats de ce classement, qui apporte un nouvel éclairage sur le fonctionnement de l'anaphore infidèle.

3.2 Résultats

Au terme de l'analyse, les 228 occurrences recueillies se répartissent en 17% d'occurrences non anaphoriques et 83% d'occurrences anaphoriques.

Par occurrences non anaphoriques, nous entendons des SN démonstratifs qui ne peuvent être reliés à une partie identifiable du contexte précédent, en ce qu'il jouent sur un autre niveau que celui de la co-référence. Voici par exemple les premières phrases de l'un des textes du corpus :

Cet hiver *j'ai croisé Bruce et Robbie sur les sites de Roquebrune, Gourdon et Saint-André. Ils travaillaient à leur gamme, à grands coups de wings et de fermetures. Depuis, c'est prêt. Voici l'Electron, modèle Intermédiaire.*

Cet hiver se range dans la catégorie « Déictique », dans la mesure où il sert à faire référence à un temps commun à l'auteur et au lecteur.

Le tableau suivant indique la répartition des occurrences anaphoriques selon les catégories indiquées dans la section 2.2.

TAB. 2 – *répartition des occurrences anaphoriques dans le corpus de vol libre*

Classement	Nombre d'occurrences	Pourcentage
Hyperonyme	112	58,9%
Supplétif	71	37,4%
Synonyme	6	3,2%
Déverbal	1	0,5%
	190	100%

On notera que la relation d'hyperonymie est bien présente, davantage que dans les corpus présentés dans le Tab. 1, hormis le corpus Moug, qui faisait figure d'exception avec 60%.

Intéressons-nous plus précisément à ces hyperonymes. Les 112 occurrences se répartissent sur 38 hyperonymes différents (nous n'avons pas considéré comme différentes les formes singulier et pluriel d'un même mot). Ceux-ci sont indiqués dans le tableau suivant :

TAB. 3 – *hyperonymes relevés*

aile	18	constructeur allemand	1
machine	16	débattement	1
catégorie d'aile	12	face	1
axe	10	importateur	1
configuration	5	marché germanique	1
catégorie	4	marque	1
constructeur	4	modèle Small	1
voile	4	nuages	1
manoeuvre	3	parapente intermédiaire	1
style d'aile	3	phase	1
accessoire	2	phénomène	1
allure	2	position	1
modèle	2	principe	1
produit	2	qualités	1
vitesse	2	style de parapente	1
altitudes	1	style de voile	1
casquette	1	tissus	1
charge	1	valeur	1
choses	1	voile neuve	1

L'une des questions que l'on peut se poser est d'évaluer si cette approche des textes est réellement complémentaire ou se révèle redondante par rapport à d'autres méthodes de repérage de la relation d'hyperonymie. Pour cela, nous avons projeté sur ce corpus quelques marqueurs « classiques » de définition [14] et d'hyperonymie [4].

Les premiers n'ont moissonné que peu de définitions, moins d'une dizaine, et dans lesquelles on ne trouve qu'une seule fois *aile* comme *genus* de la définition et une seule fois *marque* :

Serial Class, ce terme vous évoque-t-il quelque chose ? Il désigne ***ces ailes*** *homologuées Afnor Performance ou DHV 2-3, qui jouent en Coupe du monde, au milieu des grands protos teigneux...*

Gradient est ***la marque*** *tchèque d'Ondrej Dupal et de Vaclav Sykora.*

Le marqueur *est-un* s'est avéré un peu plus productif en fournissant 30 contextes exploitables. Sur ces 30 contextes, 19 relient la voile essayée dans l'article et l'hyperonyme *aile* (11 contextes) ou l'hyperonyme *machine* (8 contextes). Deux autres relient le sac de la voile essayée à l'hyperonyme *produit*, ce qui est strictement similaire à l'une des occurrences de *ce produit* en reprise anaphorique infidèle.

Ce qui signifie que les 34 autres hyperonymes identifiés par l'analyse des anaphores infidèles ne peuvent être atteints par la projection d'autres marqueurs.

Il apparaît toutefois que certains sont en même temps la tête de termes complexes, ce qui peut éventuellement permettre de les positionner comme hyperonymes (à condition que le sens du terme complexe soit lui-même compositionnel). C'est le cas de *aile*, qui entre dans des composés tels que *aile standard*, *aile intermédiaire*, *aile compétition*, etc. ; de *axe* qui permet de former *axe de roulis*, *axe de tangage*, *axe de lacet* ; de *manœuvre* pour *manœuvre de descente rapide*, *manœuvre symétrique, manœuvre asymétrique* ; de *phase*, tête de *phase parachutale*, *phase de décollage* mais l'inventaire s'arrête là.

En première analyse, il semble donc que, sur des corpus tels que celui qui a été analysé dans cette étude, qui s'avère assez pauvre en énoncés définitoires et en marqueurs « classiques » de la relation, l'examen des anaphores démonstratives infidèles constitue une approche complémentaire pour la construction de relations d'hyperonymie.

Toutefois, la relation ne s'impose pas toujours à la simple lecture du SN démonstratif, même en incluant dans cette lecture une portion un peu plus large de co-texte. Pour le spécialiste, l'interprétation ne pose aucun problème (puisque, rappelons-le, s'il est possible de procéder à une anaphore avec un hyperonyme, c'est parce que la relation d'hyperonymie entre les deux termes est présupposée, elle préexiste à son expression dans le texte), mais pour le non-spécialiste, certaines difficultés surgissent, nous les évoquons maintenant.

4 Difficultés se posant au non-spécialiste

4.1 Interprétation de la relation anaphorique

Il nous faut préciser que, pour la validité de la démarche, les 228 occurrences de SN démonstratifs ont été examinées par une non-spécialiste de vol libre. En effet, ce que nous cherchons à évaluer, c'est la pertinence d'amorcer la recherche de relations conceptuelles (plus précisément ici, la relation hyperonymique) par certaines configurations discursives *a priori* favorables à l'expression de cette relation dans différents genres textuels et dans différents domaines. L'idée sous-jacente est que ces configurations discursives présentent des propriétés suffisamment stables pour constituer un point d'appui pour un lecteur non spécialiste du domaine de la connaissance concerné (documentaliste, terminologue, ingénieur de la connaissance, etc.).

Et effectivement, pour de nombreuses occurrences, il est relativement aisé de déterminer l'antécédent, en voici quelques exemples (pour la lisibilité des extraits, nous mettons l'anaphorique en caractère gras, soulignons l'antécédent et indiquons la relation entre parenthèses) :

En fait, le bord d'attaque en intrados frétille un peu et retransmet ***ce mouvement*** *aux élévateurs.* (dev/hy)

A bout d'accélérateur, ça déménage à presque 50 km/h, en butée de poulies. La voile reste solide, et j'avoue avoir abusé de ***cet accessoire****, histoire de s'enlever plus vite du décor.* (hy)

Plus de 7 heures de vols sous la Saphir 2000 et l'exploration de zones turbulentes (histoire de tester la masse d'air pour mes élèves) m'incite à dire que ***cette voile*** *est simple à gérer.* (hy)

Ensuite la Sigma 4 accélère très franchement pour arriver à 48 km/h au premier barreau et à la vitesse maxi en butée de poulies au second barreau. Le constructeur limite sur ses sellettes la vitesse max à 47 km/h. A ***ces allures****, très vigilant en traversant des zones thermiques fortes, nous n'avons pas eu la moindre alerte et pas un coin de voile ne frissonne.* (hy)

Mais tous les contextes ne sont pas aussi clairs. On peut identifier plusieurs types de difficultés, qui montrent les limites de la conception d'un marqueur comme un passage unique du texte et qui montrent la nécessité d'une recherche complémentaire dans le texte. Les deux premiers types concernent l'interprétation au niveau du texte lui-même, le dernier (section suivante) fait entrer en jeu des choix de représentation.

1. Interprétation de la relation

Dans certains contextes, l'anaphore infidèle marque non une relation d'hyperonymie mais une équivalence : l'élément anaphorique est équivalent sur le plan sémantique à l'élément anaphorisé, en voici un exemple dans le corpus :

Cela prend un "certain temps", typiquement entre quelques minutes et un quart d'heure... C'est ***cet intervalle de temps*** *séparant chaque bouffée thermique que l'on appelle le "cycle thermique", et qui dépend d'une foule de paramètres !*

Ici, l'équivalence est aisée à reconnaître dans la mesure où elle ne fait pas intervenir de notions spécifiques au domaine. Mais dans le cas inverse, lorsque les notions mobilisées relèvent du savoir spécialisé, la distinction entre l'hyperonymie et l'équivalence devient délicate :

Un écopage dissymétrique se rattrape facilement en recentrage/commande et n'interrompt pas le décollage. L'aile reste très légère aux commandes et aux élévateurs durant toute ***cette phase du début de vol****.*

Un non expert du domaine peut difficilement décider au vu de ce contexte seul si le décollage est une phase du début de vol (en faisant l'hypothèse qu'il y aurait plusieurs phases du début de vol et que le décollage serait l'une d'elles), auquel cas on est en présence d'une hyperonymie, ou si *décollage = phase du début de vol*.

Deux parades sont envisageables : 1. s'adresser à un expert du domaine ; 2. chercher dans le corpus d'autres contextes levant l'ambiguïté et permettant de prendre une décision. Une recherche dans le corpus pour identifier les différentes phases montre l'existence des termes *phase de décollage*, *phase du décollage*, *phase de vol*. On constate ainsi que le *décollage* est une *phase*. Le rapprochement de l'extrait ci-dessus avec les contextes de *phase de/du décollage* montre l'équivalence des expressions.

Malgré une certaine légèreté aux élévateurs et aux freins dans toute cette phase du décollage, la Bellus fait partie des ailes faciles à mettre en l'air.

La Syrius est bien présente aux élévateurs et aux commandes pendant toute la phase de décollage.

En effet, le reste de la phrase fait à chaque fois appel aux mêmes notions : légèreté *vs* présence aux élévateurs et aux commandes pendant/durant/dans la phase de décollage/début de vol, ce qui indique que la signification de ces expressions est la même.

2. Identification de l'antécédent

Une autre difficulté, plus fréquemment rencontrée, tient à l'identification de l'antécédent de l'anaphorique. Les extraits suivants sont exemplaires de cette difficulté :

Les B, pour quoi faire ? C'est vrai, j'en parle à tous les essais, alors que depuis quelques temps, cette pratique se marginalise, présentant vraiment peu d'intérêt en situation turbulente ou ventée. Et en situation calme, en avons-nous vraiment besoin ? Alors, dans rubrique essais à venir, je pense que ***cette manoeuvre*** *n'apparaîtra plus.*

Très difficile de rater un décollage avec la Taïga. Un soupçon de frein, et me voilà en l'air. Première constatation : ça dépote. J'entends par là que la vitesse mains hautes de 38 km/h stabilisés me surprend presque pour ***cette catégorie d'aile****.*

Le premier plané m'a séduit par l'impression de glisse efficace. En tangage, la Saphir se montre bien amortie, de sorte qu'elle n'oscille pas sur ***cet axe****. En roulis, attention à votre réglage de ventrale !*

Là encore, la difficulté vient de la non expertise du domaine qui engendre l'incompréhension des anaphoriques : que représente le terme *manœuvre*, à quoi fait référence l'expression *catégorie d'aile*, de quel *axe* s'agit-il ? De la même façon que précédemment, on peut s'adresser à un expert du domaine ou effectuer une recherche dans le texte pour mieux cerner la signification des différents termes. Nous avons déjà souligné la pauvreté de ce corpus en énoncés définitoires, c'est donc plutôt par le rapprochement d'occurrences des termes concernés et par la projection sur le corpus d'autres marqueurs d'hyperonymie qu'émergera leur signification.

Pour *manœuvre*, les autres occurrences sont très rapidement explicites :

Une ***manoeuvre*** *peu classique sur cette aile : les B.*

Les ***manoeuvres*** *de descente rapide aux "B" ne posent aucun problème à cette machine qui en sort instantanément, pratiquement sans abattre, au relâché rapide.*

Toutes les ***manoeuvres*** *symétriques style B, frontales, sont gentilles, sans surprise ni excès dans les réactions.*

Les B, physiques, conduisent à un fort rétrécissement dans la corde, sans conséquence dans l'envergure, avec une Vz de 8 m/s et un peu d'instabilité. Le relâcher rapide ne donne pas d'abattée significative. Au relâché lent la voile a repris seule son vol avant la fin de la ***manoeuvre*** *!*

En route pour quelques vracs durant lesquels je laisse faire la voile. La frontale ? insignifiant ! La fermeture asymétrique à plus de 50 % ? pas même un quart de tour que tout est arrêté, ouvert et à plat. Trop gentil ! Les B ? Ah non, on en parle plus, c'est fini ! Notez seulement la douceur de l'abattée de sortie. Décidément, L'Epsilon se montre vraiment très sage... Et à défaut de thermiques, je répète ***ces manoeuvres****, confirmant ses qualités.*

D'ailleurs, l'Arcus fut la préférée de toute l'équipe d'Aérogliss durant l'été, particulièrement lorsqu'il s'agissait de démontrer aux élèves certaines évolutions, style B, wing-overs radicaux, fermetures asymétriques et autres brusqueries. Elle reste homogène et compréhensive, sans aucun flou dans ***ces manoeuvres****.*

Ces différents contextes permettent de déterminer ce qui est appelé *manœuvre* dans ce domaine : *les B*, *la frontale* (en fait, *fermeture frontale*, c'est un des cas de réduction de terme complexe que nous avons évoqués plus haut), *la fermeture asymétrique*, *les wing-overs radicaux*. L'incertitude concernant l'antécédent de l'extrait initial est alors levée : l'antécédent est *Les B* et on est bien devant une relation d'hyperonymie entre *manœuvre(s)* et une série d'hyponymes.

Le deuxième extrait conduit à analyser les différents emplois de *catégorie*. Il faut souligner qu'avec *cette aile* et *cette machine*, *cette catégorie* est le SN démonstratif le plus fréquent de notre sélection : 16 occurrences (dont 12 pour *cette catégorie d'aile*). La résolution de sa signification apporte donc une réponse non à une interrogation isolée, mais à la question posée par toute une série de contextes. Certains co-occurrents de *catégorie* permettent de cerner sa valeur :

La Syrius s'arrête net à l'enfoncement des commandes. Les performances sont bonnes dans ***la catégorie des ailes standard*** *[…]*

L'Allegra permet au pilote de progresser en transition par de bonnes performances réellement utilisables. Elle peut se placer en excellente deuxième voile, pour des pilotes peu assidus, soucieux de rester en ***catégorie standard****.*

*A ajouter à la série "Response, Bonanza, Sigma 4", dans les ailes à essayer "****catégorie sport****".*

*Plus maniable que la Futura mais moins vive que la Bagheera, l'Allégra m'a permis de jouer en l'air. Tantôt sur la tranche, tantôt tout en finesse, elle permet l'abus, sans sanction. Une très bonne aile pour qui ne veut pas d'aile "****catégorie compétition****", ni même "Performance".*

La réponse amortie et diluée en virage la place tout à fait dans la ***catégorie intermédiaire performante*** *et tempère des actions brutales à la commande.*

L'Avax Rx marche très bien en compétition, la Saphir 2000 continue son chemin sans soucis et l'Onyx se propose d'ouvrir les cieux aux jeunes apprentis pilotes, en jouant la carte de la ***catégorie intermédiaire****.*

Elle fait partie des meilleures dans la ***catégorie DHV2*** *et vient presque chercher des noises à sa grande soeur Tension...*

Ces extraits évoquent différentes catégories qui représentent, comme le laissait supposer l'expression *cette catégorie d'aile*, une catégorisation des ailes. On trouve d'ailleurs dans les textes, à l'instar des deux premiers extraits, une alternance *catégorie standard / aile standard*, *catégorie intermédiaire / aile intermédiaire*, *catégorie compétition / aile compétition*, *catégorie DHV2 / aile DHV2*, etc., ou encore, *ailes labellisées standard* (la plupart des ailes reçoivent, par un organisme officiel, une labellisation qui indique le niveau requis pour les utiliser). L'anaphore *cette catégorie d'aile* s'ancre alors sur la mention de l'aile essayée elle-même, de la même manière que si on avait, dans un article consacré au dalmatien, une anaphore de la forme *cette espèce de chien*.

Le troisième extrait pose une difficulté d'un autre ordre. Là encore, des indices peuvent être trouvés par l'analyse des occurrences de *axe* :

L'amortissement sur ***l'axe de tangage*** *est important sans être ennuyeux surtout au delà de 20°.*

Sur l'axe de tangage*, à la charge alaire de l'essai, l'aile atteint facilement 36 km/h bras hauts et, plus progressivement, 45 km/h accéléré, des valeurs parfaitement dans l'ordre de grandeur de ce qu'annonce le constructeur.*

L'aile communique par des mouvements secs vite amortis, ***sur l'axe de roulis****.*

Le mot *axe* apparaît comme tête de deux termes complexes : *axe de tangage* et *axe de roulis*. Le premier figure dans le co-texte de *amortissement*, le second dans le co-texte du participe passé de *amortir*. En poursuivant l'analyse par la recherche de contextes rassemblant *amortissement* et *tangage* ou *roulis*, on obtient de nouveaux composés, *amortissement du/en tangage* et *amortissement du/en roulis* qui, ajoutés au fait que l'on observe à nouveau une occurrence de *cet axe* dans la phrase suivant l'une des occurrences, conduisent à la conclusion d'une équivalence entre *axe de tangage* et *tangage*, et entre *axe de roulis* et *roulis*.

Normalement tenue aux commandes, l'Azimut a un amortissement du tangage bon sans être massif. Elle reste donc assez vivante sur ***cet axe****, acceptant volontiers de rentrer dans l'ascendance.*

L'amortissement du roulis se manifeste dès le début de l'inclinaison, sans zone de liberté à la sellette.

L'amortissement en tangage facilite l'apprentissage…

L'amortissement en roulis est très présent après les 15 premiers degrés d'inclinaison où la Blues est laissée "libre" et donc assez communicative. Il est important en tangage.

Cette conclusion est confirmée par le contexte suivant dans lequel *ces axes* reprend *en roulis / tangage*.

En vol, la stabilité sur trajectoire en roulis / tangage est frappante. Les déstabilisations volontaires sur ***ces axes*** *sont rapidement amorties en une oscillation sans intervention du pilote.*

En fait, il s'agit du phénomène de réduction de termes complexes, fréquent dans ces textes, qui touche ici la tête de ces termes. D'autres exemples de cette réduction dans ce corpus sont les couples *frontale/fermeture frontale*, *asymétrique/fermeture asymétrique*, *thermique/ascendance thermique*, *dynamique/ascendance dynamique*, *ventrale/sangle ventrale*, etc. (nous avons dénombré plus d'une douzaine de termes susceptibles de voir leur tête effacée dans ce corpus [12]). Lors de l'analyse de cette réduction, il avait été souligné que le constituant effacé peut ressurgir dans la suite du discours sous forme d'une anaphore, par exemple :

Dans ***ce thermique****, mélange d'instabilité et de confluence, je ne fais que régler doucement l'inclinaison de l'aile. […] Pleinement concentré sur la forme bizarroïde de* ***cette ascendance****, je laisse beaucoup faire !*

C'est exactement la même configuration que l'on observe avec les occurrences de *en tangage / sur cet axe*, le terme *axe* n'est autre que le constituant effacé du terme *axe de tangage*.

Ce même phénomène de réduction est aussi à l'œuvre dans l'extrait suivant, qui correspond formellement aux contextes que nous avons examinés (pour mémoire, un SN démonstratif dont le N ne figure pas dans le paragraphe précédent) :

Je m'éloigne du relief et en profite pour détailler la voile. Bord d'attaque légèrement convexe, ***cloisons diagonales*** *partielles qui relient l'intrados du point de suspente, à l'extrados, de chaque côté de la cloison porteuse, […].* ***Ces diagonales*** *s'arrêtent à une cinquantaine de centimètres du bord de fuite.*

Mais contrairement aux premiers exemples, le terme réduit n'est pas l'antécédent mais l'anaphorique, fournissant ainsi un bel exemple des configurations par lesquelles s'établit l'équivalence entre un terme complexe et un terme réduit.

Ces deux types de contextes dans lesquels intervient le phénomène de réduction de termes complexes sont typiquement des écueils pour le lecteur non familier du domaine en ce qu'ils brouillent la relation de co-référence.

Cependant, tout le développement qui précède montre que, face à des difficultés locales d'identification d'un antécédent, l'analyse des autres occurrences du nom employé pour l'anaphore livre généralement la clé de l'énigme. En ce sens, pour l'analyste qui ne connaît pas le domaine dont il a à établir la taxinomie, la configuration étudiée ici constitue un fil à tirer, une entrée possible dans un corpus et dans le réseau de relations conceptuelles du domaine.

A côté des difficultés d'interprétation qui viennent d'être décrites, se posent d'autres questions, moins liées à la méconnaissance du domaine qu'à la pertinence de la relation pour la construction d'un réseau terminologique.

4.2 Pertinence de la relation pour un réseau terminologique

Différents cas de figure surgissent à l'analyse de la configuration étudiée, qui conduisent à engager la réflexion sur ce qu'il convient de retenir pour la construction du réseau terminologique. En effet, pour chacun des contextes qui suivent, il ne va pas de soi d'inclure soit l'antécédent, soit l'anaphorique, dans une taxinomie, et ce bien que dans chaque cas, la relation identifiée puisse effectivement être considérée comme une forme d'hyperonymie.

1. Cas des termes très longs

Il milite pour le développement d'ailes aux comportements hors domaine de vol labellisées standard et cherche à appliquer ***ce principe*** *avec plus ou moins de bonheur à l'ensemble des machines de ses deux marques : Trekking et Custom Sail.*

Le problème des termes très longs s'inscrit dans la problématique globale à laquelle nous sommes confrontés lorsque nous voulons passer d'un discours à une représentation relationnelle. Il s'agit en effet de s'interroger sur les possibilités de discrétiser le discours pour n'en retenir que des portions auxquelles on donnera un statut de terme ou de relation. Dans le cas des termes longs, le problème est de savoir si cette longueur est compatible avec le fait de dénommer un concept ; il y a donc une dimension cognitive non-négligeable. En pratique, on utilise le corpus pour vérifier si ce même terme est utilisé ailleurs ou, au contraire, si, dans d'autres parties du corpus, c'est seulement une partie du syntagme qui apparaît ce qui est un argument pour scinder ce syntagme et en faire plusieurs termes.

2. Cas des figures

Après quelques années de travail comme team leader de l'équipe d'Allemagne delta, aéronef qu'il pratique depuis 20 ans, et un titre de champion du monde […] sous ***cette casquette****, il revenait au parapente, un milieu professionnel qu'il connaît parfaitement.*

Le problème des figures ne peut pas être réglé une fois pour toutes. En effet, on sait qu'un des principes de création terminologique consiste en des emprunts métaphoriques d'un domaine à l'autre. Si bien que, dans certains cas, ce qui apparaît comme une figure aux yeux d'un néophyte est en réalité le signe de la présence d'un terme, parfaitement identifié par les spécialistes. L'exemple présenté ici relève d'une autre problématique. La métonymie utilisée (la casquette comme signe d'appartenance à une équipe), assez courante dans le langage commun, ne peut pas relever d'un fonctionnement terminologique. Mais dans certain cas, le départ entre figure pertinente d'un point de vue terminologique et figure pertinente du seul point de vue discursif peut être difficile à faire.

3. Cas des termes contenant des variables

Au réglage d'origine le débattement, plutôt court dans l'ensemble, est de 15 cm pour 30 km/h et elle demande 1.5 kg en vol droit à ***cette vitesse****.*

Ce cas de figure se retrouve avec nombre de termes impliquant une mesure, par exemple, dans le corpus : *vitesse*, *allure*, *altitude*, *valeur*… A priori, il semble difficile d'inclure tel quel le terme sous-ordonné : la relation d'hyperonymie ne s'établit pas seulement entre *vitesse* et *30 km/h*, mais plus généralement entre *vitesse* et *X km/h*. Selon nous, ce seront surtout les objectifs de modélisation qui dicteront la conduite à tenir et la façon de rendre compte de cette relation.

4. Cas des termes qui ne sont pas de la même catégorie grammaticale

Le Nom du SN démonstratif est en fait susceptible de reprendre des antécédents de diverses catégories grammaticales, ce qui peut poser un sérieux problème de modélisation, car il est généralement admis que les différents éléments d'une même taxinomie sont de même catégorie.

- Nom et adverbial

Le choix de l'emplacement du secours (et donc du harnais ou de la sellette) devrait être lucidement déterminé par chacun selon les avantages et les inconvénients qu'il y voit. […] En haut du dos, derrière la tête : la double poignée sur l'épaule est accessible des deux mains. […] On n'a pas encore vu d'accident aux cervicales mais ***cette disposition*** *du secours est très récente. […] Latéralement : la solution massivement choisie pour les secours équipés d'extracteurs divers […] Mais* ***cette disposition*** *a cassé la colonne de Jeff Fauchier…*

Sur le plan conceptuel, il est indéniable que *en haut du dos, derrière la tête* ou *latéralement* dénotent une position ou une *disposition*. Il semblerait donc préférable d'intégrer ces différents éléments que le contexte montre reliés.

- Nom et adjectif

C'est donc une machine particulièrement maniable et très manoeuvrante, vivante sans être vivace. Par ***ces quali-***

tés *la Syrius est très facile à faire monter en thermique, à recentrer, à piloter dans l'ascendance.*

Là encore, la relation conceptuelle est indéniable et il semblerait contre-productif de ne pas l'intégrer dans une taxinomie. La situation est toutefois dans ce dernier cas plus favorable que précédemment, car des adjectifs peuvent aisément être transformés en nom, les auteurs des textes analysés ne se privent d'ailleurs pas de cette possibilité puisque l'on trouve aussi :

La maniabilité et une manoeuvrabilité hors pair sont donc au rendez-vous. ***Ces qualités de vol*** *ajoutées à une inversion de virage en 2 s en font une aile très efficace en thermique.*

De manière générale, tous ces contextes engagent à élargir la réflexion quant aux éléments à inclure dans un réseau conceptuel. Même si une tradition ancienne met les noms et groupes nominaux à l'honneur, les textes témoignent de la multiplicité des formes sous lesquelles des notions et des relations pertinentes pour un domaine trouvent à s'exprimer. Une même notion sous-jacente peut facilement être mobilisée par un nom, un adjectif, une préposition [10], et être de ce fait reliée dans le discours à un élément de catégorie grammaticale différente, même lorsque aucune opération de dérivation n'a lieu entre les deux termes, ce qui serait le cas avec, par exemple, un antécédent verbal et une anaphore déverbale (cas où l'anaphorique est une nominalisation déverbale comme dans *j'ai volé pendant une heure / ce vol*).

5 Conclusion

Les résultats que nous avons présentés dans cet article confirment deux éléments que nous avons proposés en introduction. D'une part, les marqueurs « classiques » d'hyperonymie ne sont pas toujours présents dans un corpus spécialisé ; d'autre part, d'autres fonctionnements discursifs, qui relèvent plutôt de l'implicite, peuvent prendre le relais dans ce cas de déficit et, en se combinant, permettre de construire quand même une relation d'hyperonymie.

Nous avons ainsi montré que l'anaphore infidèle peut jouer un rôle majeur de déclencheur de recherche d'hyperonymie, ce fonctionnement discursif étant complété par d'autres portions du corpus étudié.

La notion de marqueur de relation apparaît ainsi comme devant être nuancée. Elle ne peut concerner les seules portions de textes qui donnent accès directement à une relation (une « connaissance », pour bon nombre de travaux en TAL ou IC) mais on doit considérer aussi les structures discursives qui mettent en œuvre une relation sans la déclarer. Evidemment ces cas sont beaucoup plus difficiles à prendre en compte d'un point de vue automatique. En effet, beaucoup de ces structures n'ont pas été décrites ou en tout cas pas décrites en tant que marqueurs potentiels. De ce fait, bon nombre d'entre elles sont encore imprédictibles. Une autre difficulté pour une mise en œuvre automatique est liée au fait que plusieurs indices doivent souvent être pris en considération pour construire une hyperonymie. Mais il faudrait examiner dans quel ordre il est le plus pertinent de les rechercher. Lorsque la relation peut être considérée comme sûre à l'aide d'un seul indice (par exemple, celui de l'anaphore infidèle), il est inutile, car coûteux, de rechercher d'autres indices.

Il apparaît finalement que, plus que jamais, des analyses linguistiques fines sont nécessaires pour établir des ponts entre des phénomènes discursifs relativement bien décrits et la possibilité d'utiliser ces phénomènes dans une perspective d'extraction automatique d'information.

Références

[1] K. Ahmad et H. Fulford. Knowledge Processing : Semantic Relations and their Use in Elaborating Terminology. *Computing Sciences Report CS-92-Guildford: University of Surrey,* 1992.

[2] D. Apothéloz et M.-J. Reichler-Béguelin. Interpretations and functions of demonstrative NPs in indirect anaphora. *Journal of Pragmatics*, vol 31, pages 363-397, 1999.

[3] N. Aussenac-Gilles et P. Séguéla. Les relations sémantiques : du linguistique au formel. *Cahiers de Grammaire*, vol 25, pages 175-198, 2001.

[4] A. Borillo. Exploration automatisée de textes de spécialité : repérage et identification de la relation lexicale d'hyperonymie. *LINX,* vol 34-35, pages 113-124, 1996.

[5] A. Condamines. *Chez* dans un corpus de sciences naturelles : un marqueur de méronymie ? *Cahiers de Lexicologie*, vol°77, pages 165-187, 2000.

[6] A. Condamines. *Avec* et l'expression de la méronymie : l'importance du genre textuel. In G. Kleiber et C. Schnedecker (eds), *La partition en langue et en discours*, à paraître.

[7] A. Condamines. Anaphore nominale infidèle et hyperonymie : le rôle du genre textuel. *Revue de Sémantique et Pragmatique*, vol°18, à paraître.

[8] F. Cornish. *Anaphoric relations in English and French, a Discourse Perspective*. London: Croom Helm, 1986.

[9] D.A. Cruse. *Lexical Semantics*. Cambridge: University Press, 1986.

[10] G. Gross. Sémantique lexicale et connecteurs. *Langages* vol 136, pages 76-84, 1999.

[11] M.A. Hearst. Automatic Acquisition of Hyponyms From Large Text Corpora. *Proccedings of the 14th International Conference on Computational Linguistics*, Nantes, pages 539-545, 1992.

[12] M.-P. Jacques. *Approche en discours de la réduction des termes complexes dans les textes spécialisés*. Doctorat Nouveau Régime, Université Toulouse II Le Mirail, 2003.

[13] E. Morin. Acquisition de patrons lexico-syntaxiques caractéristiques d'une relation sémantique. *Traitement Automatique des Langues*, vol 40(1), pages 143-166, 1999.

[14] J. Rebeyrolle et L. Tanguy. Repérage automatique de structures linguistiques en corpus : le cas des énoncés définitoires. *Cahiers de Grammaire,* vol 25, pages 153-174, 2001.

[15] P. Séguéla. Adaptation semi-automatique d'une base de marqueurs de relations sémantiques sur des corpus spécialisés. *Terminologies nouvelles,* vol 19, pages 52-60, 1999.