

Relatifs et référents inclus dans un *SN* : des paramètres pour présélectionner la saisie

Laurence Kister

ATILF - CNRS UMR 7118

IUT Nancy-Charlemagne - Université Nancy 2 - Département Info-Com

2ter, boulevard Charlemagne - 54000 Nancy

Laurence.Kister@univ-nancy2.fr

Résumé

Notre objectif est de repérer l'existence de régularités pour prévoir l'attachement du relatif et de son référent introduit par un *SN* de la forme *dét. N1 de (dét.) N2* en vue d'un traitement automatique. Pour évaluer les préférences, nous avons entrepris une analyse sur corpus. L'examen des occurrences examinées laisse entrevoir des variations en fonction de paramètres d'ordre fonctionnel, syntaxique et sémantiques : déterminants, traits sémantiques, saillance, relation établie par l'utilisation de *de*, position grammaticale du *SN* qui introduit le référent dans le discours...

Mots clés

Anaphore - *SN* complexes - référence – préposition *de* – relatif *qui*

L'analyse sur corpus que nous proposons tend à identifier les paramètres qui permettent de prévoir l'attachement d'un relatif et d'un référent introduit par un *dét.* *N1 de (dét.) N2* en vue de l'automatisation du processus de reconnaissance des anaphores. Un rapide examen des occurrences repérées laisse entrevoir des variations en fonction des déterminants, des traits sémantiques et de la position grammaticale du *SN* qui introduit le référent. Nous limitons l'analyse aux anaphores en *qui* pour deux raisons : des psycholinguistes ont montré une saisie préférentielle du *N1* quand le référent est introduit par un *SN* (Mitchell, Cuetos, Zagar, 1990 - Zagar, 1995 - Zagar, Pinte, Rativeau, 1997, entre autres), des analyses d'ordre général sur corpus (Baltazart, Kister, 2000) montrent que l'anaphore en *qui* d'un référent introduit par un *SN* complexe est très fréquente (48 %). Notre corpus se compose d'anaphores extraites d'un corpus constitué lors de travaux antérieurs (Baltazart, Kister, 2000) et d'occurrences repérées spécialement pour cette analyse. Il s'agit de textes scientifiques et techniques sous forme de monographies (économie, mathématiques, sciences de l'information, histoire de l'industrie et des techniques, arts, sciences de la terre, sport, chimie, philosophie et droit), de mémoires, d'articles de revues spécialisées (informatique, sciences de l'information et de la documentation), de rapports d'activité, de romans, d'articles (presse quotidienne, magazines de large diffusion, journaux d'information destinés à des clients ou des adhérents, presse interne d'entreprise) et de dépêches relatant des catastrophes naturelles. Il comporte 1613 anaphores en *qui* dont le référent est introduit par un *SN* complexe. Comme l'anaphore sur *SN* simple c'est « une expression dont l'interprétation référentielle dépend d'une autre expression (ou d'autres expressions) mentionnée dans le contexte et généralement appelée son antécédent ». Il s'agit d'un « processus qui indique une référence à un référent déjà connu par l'interlocuteur, c'est-à-dire un référent « présent » ou déjà manifeste dans la mémoire immédiate » (Kleiber, 94). Un examen rapide des résultats permet de constater que même si l'hypothèse du *linguistic tuning*¹ proposée par les psycholinguistes paraît vérifiée avec 57 % d'appariement du relatif à l'ensemble du *SN*, on ne peut passer sous silence les variations dues aux paramètres évoqués ci-dessus. Le tableau 1 reprend la répartition pour les sept combinaisons présentes au moins trente fois dans notre corpus. Si on s'intéresse aux combinaisons retenues, on remarque que la reprise du *SN* est favorisée pour quatre constructions (*Entre le traitement brut de la réalité et la fiction, le documentaire est la forme de télévision qui s'approche le plus de la réalité*) tandis que celle du *N2* l'est pour trois autres (*J'ai lu dans l'un de ces bouquins l'histoire d'une fille qui avait eu la vocation d'aller soigner les lépreux*). Pour ces *SN* le test du Chi 2 indique une corrélation entre le type de déterminant et l'attachement préféré. On note trois formes pour lesquelles le *N2* suit directement la préposition : *un N1 de N2*, *ce N1 de N2* et *le N1 de N2*. Ce biais s'explique en partie par le fait que ces *SN* comportent des expressions lexicalisées (*pomme de terre, lampe de poche...*) où le *N2* est un « attribut » du *N1* et des structures « partitives » (*un morceau de pain, une part de tarte...*). Quand le *N2* est précédé d'un défini (*un N1 du N2* et *le N1 du N2*), l'attachement dépend du déterminant qui introduit le *SN*. Ainsi, pour un *SN* précédé d'un indéfini, la saisie du *SN* est plus fréquente. La rupture discursive introduite par le *SN* complexe indéfini, comme pour les anaphores sur *SN* indéfini simple, traduit un nouveau thème de discours.

1 Il s'agit de l'évaluation des mouvements oculaires et des temps de lecture lors des retours en arrière quand le texte lu présente des ambiguïtés (le *qui*) et risque de conduire à une mauvaise interprétation de l'une des entités linguistiques.

Type de SN	Nb d'occurrences	Reprise du SN	Reprise du N2
<i>le N du N</i>	667	46 %	54 %
<i>le N de N</i>	395	76 %	24 %
<i>le N d'un N</i>	216	19 %	81 %
<i>un N de N</i>	181	91 %	9 %
<i>un N du N</i>	59	78 %	22 %
<i>un N de ce N</i>	36	14 %	86 %
<i>ce N de N</i>	30	87 %	13 %

Tableau 1 : Répartition des anaphores selon le type de *dét. N1 de (dét.) N2*

L'indéfini a une valeur de présentation qui permet la première mention d'une entité qui acquiert le statut d'élément saillant et peut faire l'objet d'une anaphore. Lorsque le SN est introduit par un défini, le N2 fait, le plus souvent, l'objet d'une anaphore. Il fonctionne comme un déterminatif qui complète un N1 insuffisamment déterminé : le N2 permet d'identifier clairement le N1 et devient saillant. Un indéfini devant le N2 introduit un nouveau référent : il opère une rupture de la chaîne discursive et introduit ou change le thème de discours. Quand Flaux (1992) montre que *la fermière* et *la fille du fermier* désignent un même référent qui invite celui qui interprète à chercher dans le contexte un repère pour l'identifier, elle explique en partie les résultats des expérimentations psycholinguistiques sur les relations pour lesquelles les deux N portent le trait *animé*. « L'anaphore prépositionnelle » est une coréférence entre les deux SN qui autorise une reprise de *fermière* ou sur *fille*. Pour le N1 d'un N2 l'argument de Flaux (1992) selon lequel le N1 d'un N2 peut faire l'objet d'une interprétation indéfinie et fonctionner comme *un N* peut être utilisé. Le SN introduit par un indéfini est suffisamment saillant pour faire l'objet d'une anaphore. Ainsi, l'anaphore porte respectivement sur le SN et le N2. Kleiber (1994) rappelle que les « indéfinis doivent être considérés comme les autres SN, ce sont des SN référentiels, dont l'objet de référence peut être requis par une expression anaphorique ». La construction *un N1 de ce N2*, quant à elle, montre que l'emploi du démonstratif devant le N2 attire l'anaphore. Les propriétés sémantiques des N mis en relation par la préposition *de* ont une influence sur les possibilités d'anaphores. Nous avons retenu les trois traits les plus fréquents : *abstrait*, *animé* et *inanimé*. Les SN où le N1 joue un rôle de *quantifieur* (*un groupe de gamins*, *une poignée de sable...*) où le N1 quantifie le N2 et fonctionne comme un déterminant nominal (Buvet, 1993 et 1994, Flaux, 1997 – Benninger, 1993 et 1997) qui entraîne une saisie préférentielle du N2 (Kister, 1995) ont été écartés. Il en va de même pour les SN qui contiennent une date (*la loi de 1957*, *la guerre de 1914/18...*), un ou plusieurs noms propres (y compris sous forme de sigles), ou un numéral et un cardinal (*un groupe de trois enfants*, *le premier de l'an...*). Le tableau 2 présente la répartition en fonction des traits sémantiques pour les combinaisons dont le nombre d'occurrences est d'au moins soixante dix neuf.

Traits sémantiques des <i>N</i>	Nb d'occurrences	Reprise du <i>SN</i>	Reprise de <i>N2</i>
abstrait – abstrait	1049	59 %	41 %
abstrait – animé	234	24 %	76 %
abstrait – inanimé	118	53 %	47 %
animé – abstrait	79	86 %	14 %

Tableau 2 : Répartition selon les traits sémantiques des *N* inclus dans le *SN*

L'influence du trait *animé* paraît évidente. Nos résultats sont à confronter à l'hypothèse de Flaux (1997) qui montre que dans les groupes où le *N1* est perçu comme la tête de la structure prépositionnelle le *N2* est dépourvu de véritable référence : *N2* a une valeur de caractérisation (*un visage de boxeur...*). Ces constructions sont identifiables par la mise au pluriel : le *N2* ne varie pas quand le nombre du *N1* varie (*des bols de soupe, ces crétins d'automobiliste...*). Elle distingue les *N1 prep. N2* qui ont une « tête concrète » de ceux qui ont une tête « abstraite ». Dans les groupes à « tête concrète » le *N2* entretient avec le *N1* une relation qui fait que le groupe réfère « dans son entier » et qu'il est « unique ». Ces constructions apparaissent au pluriel lorsque la relation se répète : *les époux de Marie* (deux époux successifs), *les maires de cette ville...* L'unicité est plus fréquente pour les « têtes abstraites » qui désignent un événement qui n'atteint un individu qu'une seule fois (*la naissance de Pierre, l'élection de Jean* (pour un mandat donné)) ou des faits liés à des phénomènes empiriques et des connaissances acquises (*la découverte de l'Amérique...*).

L'idée d'observer un éventuel rôle de la position grammaticale découle des hypothèses selon lesquelles l'anaphore privilégie les entités qui occupent des positions identiques. Pour les occurrences sélectionnées *qui* est toujours sujet de la subordonnée : seule la position du *SN* est à examiner. Une préférence pour la saisie du *SN* apparaît en positions *sujet* (15 %), *objet direct* (47 %) et *objet indirect* (14 %). La situation est inversée pour la catégorie *autres* pour laquelle les différents compléments doivent être distingués pour affiner l'analyse (Tableau 3).

Position du <i>SN</i>	Nb d'occurrences	Reprise du <i>SN</i>	Reprise du <i>N2</i>
Objet direct	759	55 %	45 %
Sujet	241	74 %	26 %
Objet indirect	232	60 %	40 %
Autres	381	45 %	55 %

Tableau 3 : Répartition en fonction de la position grammaticale du *SN*

Quand on associe les combinaisons de déterminants et les traits sémantiques des *N*, on obtient la répartition présentée par le tableau 4 qui prend en compte les résultats pour les sept combinaisons de déterminants et les quatre combinaisons de traits représentatifs.

	abstrait-abstrait		abstrait-animé		abstrait-inanimé		animé-abstrait	
	<i>SN</i>	<i>N2</i>	<i>SN</i>	<i>N2</i>	<i>SN</i>	<i>N2</i>	<i>SN</i>	<i>N2</i>
<i>le N1 du N2</i>	46 %	54 %	24 %	76 %	45 %	55 %	86 %	14 %
<i>le N1 de N2</i>	80 %	20 %	39 %	61 %	78 %	22 %	11 %	89 %
<i>le N1 d'un N2</i>	22 %	78 %	9 %	91 %	14 %	86 %	67 %	33 %
<i>un N1 de N2</i>	93 %	7 %	44 %	56 %	67 %	33 %	90 %	10 %
<i>un N1 du N2</i>	79 %	21 %	66 %	34 %	75 %	25 %	78 %	22 %
<i>le N1 de ce N2</i>	12 %	88 %	-	100 %	50 %	50 %	-	-
<i>ce N1 de N2</i>	90 %	10 %	-	-	86 %		100 %	-

Tableau 4 : Répartition des reprises selon des traits sémantiques des *N* et les déterminants

Cette répartition nous conduit à formuler plusieurs hypothèses. Comme cela a été proposé aux cours de travaux antérieurs pour les anaphores sur *SN* simple (Ariel, 1990 – Charolles, 1991, entre autres), il existe une tendance à rattacher le relatif à l'élément le plus saillant. Le biais d'attachement qu'on observe est sensible à ce qu'on désigne par « focalisation ». Le processus d'interprétation référentielle est alors un phénomène essentiellement pragmatique : les calculs inférentiels s'effectuent à partir du contexte et du savoir partagé plutôt que par des règles fixes et conventionnelles (Reichler-Beguelin, 1988 et 1989 – Reboul, 1989a et b, 1990 et 1991). Reboul signale que « plus le sens (...) d'une expression référentielle est réduit et plus accessible ou saillant doit être son référent ». Le pronom est dépourvu de sens et renvoie à un élément saillant du contexte antérieur : une analyse sémantique s'impose. Notre approche est confrontée au même type de problèmes que la plupart des travaux qui s'intéressent aux aspects linguistiques ou à l'analyse en vue d'une application en « intelligence » humaine : elle occupe « une position intermédiaire entre des analyses quasiment linguistiques et les approches résolument pragmatiques » (Kleiber, 1994). En effet, il est indispensable de prendre en compte des contraintes linguistiques et des facteurs structurels pour proposer des préférences. Les difficultés à rendre compte des facteurs qui permettent d'identifier un référent inclus dans un *dét. N1 de (dét.) N2* sont dues à la multitude des structures anaphoriques qu'il est possible de rencontrer : la préposition *de* admet différentes valeurs d'où différents fonctionnements référentiels, les déterminants inclus dans la structure prépositionnelle permettent différents modes de donation du référent, la saillance d'un constituant ou de l'ensemble de la structure est possible, les traits sémantiques des *N* contribuent à rendre saillant un constituant ou le *SN*, la position grammaticale de la structure complexe joue un rôle... Afin de déterminer les paramètres d'ordre sémantique et pragmatique il est nécessaire de s'intéresser aux contextes gauche et droit (Apothéloz, 1995) : parfois, c'est l'emploi de certains verbes après le relatif qui permet d'effectuer la saisie du référent d'un anaphorique d'où la nécessité de s'intéresser au prédicat.

Références

- Apothéloz D. (1995), Rôle et fonctionnement de l'anaphorique dans la dynamique textuelle, *Langue et culture*, Genève, Droz.
- Ariel M. (1990), *Accessing noun phrase antecedent*, London and New York, Routledge
- Baltazart D., Kister L. (2000), Is it possible to predetermine a referent included in a French N de N structure ?, *Corpus-based and Computational Approaches to Discourse Anaphora*, S. Botley and A.M. McEnery eds, *Studies in Corpus Linguistics*, p. 65-80.
- Benninger C. (1993), Les substantifs quantificateur en -ée, *Le nombre*, *Faits de langue*, 2, p. 79-84.
- Benninger C. (1997), *De la quantité aux substantifs quantifieurs*, Thèse de Doctorat, Université de Strasbourg.
- Buvet P.A. (1993), *Les déterminants nominaux quantifieurs*, Thèse de Doctorat en Linguistique, Laboratoire de Linguistique informatique, Université Paris XIII.
- Buvet P.A. (1994), Détermination : les noms, *Linguisticae Investigationes*, XVIII, 1, p. 121-150.
- Flaux N. (1992), Les syntagmes nominaux du type **le fils d'un paysan** : référence définie ou indéfinie ?, *Le Français Moderne*, LX, 1 p. 23-45.
- Flaux N. (1997), Les déterminants et le nombre, Entre général et particulier : les déterminants, N. Flaux, D. van de Velde et W. de Mulder eds, *Etudes littéraires et linguistiques*, Arras, Artois Presses Université, p.15-82
- Kister L. (1999), Identification des chaînes de référence dans les systèmes automatiques : le cas des anaphores dans les *N de N*, *VExTAL*, 22-24 novembre 1999, Venise.
- Kister L. (1995), Accessibilité pronominale des *dét. N1 de (dét.) N2* : le rôle de la détermination, *Linguisticae Investigationes*, XIX, I, p. 107-121.
- Kleiber G. (1994), *Anaphores et pronoms*, Champs linguistiques, Louvain-la-Neuve, Editions Duculot
- Mitchell D.C., Cuetos F., Zagar D. (1990), *Reading in Different Languages : is there a Universal Mechanism for Parsing Sentences*, *Comprehension Processes in Reading*, D.A. Balota, G.B. Flores d'Arcais and K. Rayner (eds.), Hillsdale, Lawrence Erlbaum, Associates, Inc.
- Reboul A. (1989a), Résolution de l'anaphore pronominale : sémantique et/ou pragmatique, *Cahiers de linguistique française*, 10, pp. 77-100.
- Reboul A. (1989b), Résolution automatique de l'anaphore pronominale, *Modèles du discours. Recherches actuelles en Suisse romande*, C. Rubattel éd., Berne, Peter Lang, pp. 173-192.
- Reboul A. (1990), Pragmatique de l'anaphore pronominale, *Sigma*, 12/13, pp. 197-231.
- Reboul A. (1991), Le système des pronoms personnels en français contemporains, *Tranel*, 17, pp. 123-146.
- Reichler-Béguelin M.J. (1988), Anaphore, cataphore et mémoire discursive, *Pratiques*, 57, pp. 15-43.
- Reichler-Béguelin M.J. (1989), Anaphore, connecteurs et processus inférentiels, *Modèles du discours. Recherches actuelles en Suisse romande*, C. Rubattel éd., Berne, Peter Lang, pp. 303-336.
- Zagar D. (1995), *La lecture : processus de base*, Habilitation à diriger les recherches, U.F.R. de Sciences Humaines - Université de Bourgogne, LEAD - CNRS Ura 1938.
- Zagar D., Pynte J., Rativeau S. (1997), Evidence for Early-closure Attachment on First-pas Reading Times in French, *The Quarterly Journal of Experimental Psychology*, 2, pp. 421-438.