Les chaînes de référence : présentation

Catherine Schnedecker, Frédéric Landragin

▶ **To cite this version:**

Catherine Schnedecker, Frédéric Landragin. Les chaînes de référence : présentation. Langages, 2014, Les chaînes de référence, 3 (195), pp.3-22. halshs-01069451

HAL Id: halshs-01069451
https://shs.hal.science/halshs-01069451

Submitted on 29 Sep 2014

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.

Les chaînes de référence : présentation

– DRAFT –

Catherine Schnedecker[1], Frédéric Landragin[2]
[1]LILPA, Fonctionnements Discursifs & Traduction, Université de Strasbourg
[2]Lattice, CNRS/ENS/Université de Paris 3 Sorbonne Nouvelle

1. Introduction

Depuis Chastain (1975) on appelle *chaîne de référence* :

- la suite des expressions d'un texte entre lesquelles l'interprétation construit une relation d'identité référentielle (F. Corblin, 1995, 123)
- <les> suites d'expressions coréférentielles [...]. Seules peuvent appartenir (donner lieu à) une chaîne des expressions employées référentiellement, c'est-à-dire toutes et rien que les expressions nominales (ou pronominales) permettant d'identifier un individu (un objet de discours) quelle que soit sa forme d'existence (personne humaine, événement, entité abstraite) (M. Charolles, 1988a, 8)

On dira ainsi que les expressions nominales et pronominales en gras de (1) constituent les « maillons » de la chaîne de référence renvoyant à la personne nommée Muyassar :

1. Originaire du Tadjikistan, **Muyassar** a été tabassé en décembre 2011 par des hooligans du Spartak Moscou. **Il** subit de plein fouet le racisme dont sont victimes les Tadjiks en Russie.

   Grand et athlétique, le buste droit, la voix posée, **Muyassar** dégage une certaine fierté. **Le jeune homme** est né au Tadjikistan il y a 25 ans, dans la région montagneuse du Pamir. En 2006, **son** diplôme de commercial en poche, **il** met le cap sur Moscou. Pour un jeune tadjik, débuter sa carrière professionnelle au pays est presque illusoire. "A la fin de la guerre civile, l'économie du Tadjikistan s'est effondrée, explique Tohir Kalandarov, anthropologue d'origine tadjique à l'Académie des sciences de Moscou. Aujourd'hui encore, le marché de l'emploi est bouché". **Muyassar** fait partie des plus de 500.000 immigrés économiques travaillant à Moscou.

   Au total, près d'un million de Tadjiks ont quitté leur pays pour gagner leur vie en Russie et subvenir aux besoins de leurs familles restées au Tadjikistan. Après avoir enchaîné les petits boulots, **le jeune homme** travaille depuis deux ans comme installateur de climatiseurs, un emploi qui **lui** convient bien, comme **sa** vie dans la capitale russe : "Maintenant, j'aime bien Moscou. Mais au début, c'était dur. Je ne parlais pas bien la langue. Heureusement, la solidarité entre Tadjiks m'a beaucoup aidé." (*Nouvel Observateur*, 29/02/12)

Comme le montre également cet exemple, les liens qui unissent les différents maillons et sous-tendent la chaîne de référence sont par nature différents : certains reposent sur une relation d'anaphore coréférentielle : c'est le cas des formes pronominales, du déterminant possessif ou du SN *le jeune homme*, qui ne s'interprètent pas sans le recours au cotexte ; d'autres, comme le nom propre, sont autonomes du point de vue de la référence, s'interprètent « directement » : elles sont donc coréférentielles non anaphoriques.
Dans cette optique, seules sont intéressés par les chaînes de référence, les types d'anaphores susceptibles d'établir la coréférence comme :

- les anaphores[1] dites fidèles, c'est-à-dire celles qui réinstancient la tête lexicale d'un SN antérieur : *un homme….l'/cet homme* ;
- les anaphores dites infidèles, qui se constituent d'anaphores nominales dont la tête lexicale varie par rapport à celle d'un SN préalable, incluant : les anaphores dites hyperonymiques (*un bœuf ...l'animal, Ma Porsche…. Cette voiture*), les anaphores recatégorisantes (ou reclassifiantes) (*Paul ...cette andouille/cet agrégé de philo*), et, les anaphores pronominales (*Paul…Jacques ...Il/Celui-ci/ ce dernier/le second*)[2].

2. Chaînes de référence : acquis et problèmes

2.1. La nature et la longueur d'une chaîne de référence

Si les chaînes de référence semblent, au vu de l'exemple (1), se présenter assez simplement, la notion elle-même pose un certain nombre de problèmes. A commencer par celui de sa nécessité et de sa légitimité mêmes. En effet, on peut se demander en quoi elle est opportune ou ce qu'elle apporte. De fait, « chaîne permet de dépasser les contextes de simple succession de deux termes auxquels se limite le plus souvent le linguiste qui sort du domaine phrastique » (Corblin 1987, 7). Autrement dit, la notion s'impose dès l'instant où l'on travaille sur du « long terme référentiel » et non plus sur des paires d'enchaînements référentiels[3], ce qui suppose que le nombre-plancher de maillons soit au moins égal à 3 sans quoi et les notions d'a*naphore* et de *coréférence* suffisent amplement à décrire les phénomènes.

2.2. La nature des maillons d'une chaîne de référence

Un second problème tient à la nature des maillons : d'après la définition de Charolles rappelée *supra*, seuls les maillons dotés d'une forme linguistique seraient aptes à composer une chaîne. Cette conception pourrait paraître trop stricte au regard de tous les éléments textuels susceptibles de « rappeler » un référent dans un texte.

Ainsi, outre les maillons en gras dans (2), d'autres éléments contribuent à informer le récepteur sur le référent : les appositions qui drainent tout un ensemble d'informations non négligeables mais aussi les anaphores dites zéro, voire les phénomènes d'accord ou de coréférence implicite entre un segment référentiel et des participes présents (en ponctuant en (2)) ou passé (*née dans les paillettes*…), voire des verbes conjugués :

2. Drew Barrymore « Les amours à distance c'est l'histoire de ma vie »

   Née dans les paillettes du 7e art

   Elle boit **son** thé glacé à la paille et 0 répond du tac au tac, en ponctuant régulièrement **ses** phrases du fameux « F word ». **Drew Barrymore**, née dans les paillettes du 7e art, a le vocabulaire et l'énergie d'une New-Yorkaise. Plus menue que prévu – en slim gris, T-shirt Mickey Mouse et derbys noirs –, le teint naturel et le sourcil sérieux, **elle** est aux antipodes de l'image que l'on se faisait d'**elle**. Une hippie-chic frivole et bouillonnante ? Pas seulement. **Drew**, l'aventurière extravertie capable de parcourir le globe avec **sa** meilleure amie (Cameron

[1] Nous reprenons ici la définition de Milner (1982) : « Il y a relation d'anaphore entre deux unités A et B quand l'interprétation de B dépend crucialement de l'existence de A, au point que l'on peut dire que l'unité B n'est interprétable que dans la mesure où elle reprend entièrement ou partiellement A ».

[2] Les anaphores dites indirectes (anaphores associatives, génériques, etc.) sont concernées à un autre niveau.

[3] Comme cela a été le cas de certains travaux sur l'anaphore et la coréférence dans les années 1990, pour des raisons méthodologiques tout à fait compréhensibles d'ailleurs mais beaucoup discutées.

Diaz) pour une émission écolo, ou de montrer **ses** seins à David Letterman en plein talkshow télé, **s**'est visiblement assagie. […]. *(Elle*, 20/08/10)

Rappelons, à ce point de vue, que, dans son échelle d'accessibilité référentielle, Ariel inscrit les « ø, réfléchis, traces QU- et accords » (cf. Figure 1)[4]. Entre ces deux positions, l'une très stricte l'autre sans doute trop accueillante, Landragin (2011) adopte une position intermédiaire consistant à ignorer les phénomènes d'accord verbaux et participiaux[5] et à marquer la différence entre maillons avec *vs* sans forme linguistique par l'étiquette de *maillon faible* ou *indice*. Concernant les appositions, formes de prédications secondes, on peut considérer que, même dans les cas où la relation instituée *via* la prédication est de nature identificationnelle (Van Peteghem, 1991) comme dans *Jean est mon frère/Jean, mon frère*, …, il y a une différence de force référentielle entre le SN référentiel à proprement parler et l' « attribut » (cf. Kleiber, 1981).

Accessibilité basse

↑
Nom propre complet avec modifieur
Nom propre complet
Description définie longue
Description définie brève
Nom de famille
Prénom
Démonstratif d'éloignement avec modifieur
Démonstratif de proximité avec modifieur
Démonstratif d'éloignement
Démonstratif de proximité
Pronom tonique
Pronom atone
Pronom clitique
ø, réfléchis, traces QU- et accords
↓

Accessibilité haute

Figure 1. Echelle d'accessibilité (Ariel, 1990).

2.3. Les limites des chaînes de référence

Un troisième problème concerne les limites à assigner aux chaînes. Doit-on considérer que les limites d'une chaîne de référence coïncident avec celles de son texte d'occurrence ? C'est la position implicitement adoptée dans les études menées sur des textes courts. Mais elle semble plus difficile à tenir dès lors qu'on s'intéresse à des chaînes plus longues comme celles de roman: celle renvoyant à Etienne Lantier dans *Germinal* court-elle effectivement sur les quelque 300-400 pages du roman ? Une telle position, du fait qu'elle ne se préoccupe pas du coût de traitement cognitif d'une très (très) longue suite d'expressions référentielles, paraît peu réaliste. Elle l'est d'autant moins que les textes font, comme on sait, l'objet de nombreuses formes de découpages : typographiques (titraille, chapitres, paragraphes[6], etc.) ou sémantiques (comme les « domaines discursifs » de Charolles, 1997, 2009 ; Vigier, 2005, entre autres ; Charolles & Vigier, 2005 notamment), mais aussi ceux qui dissocient

[4] Voir aussi Cornish (1986).

[5] « Compte tenu de l'aspect grammaticalisé et nécessaire de ces phénomènes, nous choisissons de les ignorer ». (Landragin 2011, 8).

[6] Dont les instructions et coûts de traitement ont été bien établis par les linguistes et psycholinguistes (cf. Stark, 1988 ; Hinds, 1979 ; Passerault et Chesnet, 1991 : 160).

épisodes/événements, etc. ou encore l'épineuse question des discours rapportés directement (cf. Guillot, 2009)[7] qui coïncident souvent avec des marquages référentiels particuliers.
De ce point de vue, deux grandes tendances se dégagent de la littérature : d'un côté, les approches « collocatives » constatant simplement et prudemment « une correspondance régulière entre les contextes informationnels et les types de stratégie référentielle » (Marslen-Wilson *et al.,* 1982, 349) et, d'un autre côté, des approches plus déterministes (appelées «approches structurales » par Toole (1996, 267) ou « modèles hiérarchiques » par Huang (2002, 309) qui reconnaissent que les unités textuelles conditionnent le choix des expressions référentielles[8] – par exemple, le découpage en paragraphe imposerait la présence de formes dites de faible accessibilité référentielle (cf. également Ariel, 1990) – et, partant, à considérer que ces unités textuelles servent potentiellement à délimiter les chaînes de référence. Une conception alternative consiste à considérer que certaines expressions référentielles (p.e. le nom propre ou les SN démonstratifs, Charolles, 1995b, 1997b ; De Mulder, 2001 ; Kleiber, 1994 ; Schnedecker, 2005) instituent des ruptures dans la chaîne marquant ainsi un changement de saisie référentielle et, partant, une « sorte de paragraphe pragmatique » propice à un redémarrage de la chaîne. Bref, le débat, toujours d'actualité et que ne peuvent trancher, à notre avis, que des expérimentations psycholinguistiques, révèle un clivage résultant du rôle accordé aux chaînes de référence dans la structuration textuelle : soit leur rôle est secondaire, car conditionné par d'autres facteurs ; soit il est primordial si on leur accorde la capacité à induire d'autres formes de découpage.
Ce point montre à quel point les chaînes concourent – quelle que soit l'importance qu'on leur accorde – à la structuration textuelle : le montre la coïncidence systématique de certaines expressions référentielles (comme le toponyme) avec d'autres plans de l'organisation textuelle (Charolles, 1988, 1995) comme le découpage typographique et/ou sémantique de l'extrait :

3. **L'Argentine** est un pays industrialisé souvent considéré comme émergent même si certains organismes ne reconnaissent pas cette définition, le pays ayant été un des plus riches de la planète jusqu'au début du XXe siècle mais étant souvent frappé par des crises économiques comme en 1989 ou en 2001. **L'Argentine** fait partie du G20. Souffrant d'inflation et d'ingérence financière, le pays doit souvent faire appel aux organisations économiques internationales telles que le FMI.

   L'Argentine est la seconde puissance économique d'Amérique du Sud, derrière le Brésil. Le pays possède une importante richesse agricole, de nombreuses capacités industrielles et un certain potentiel minier, pourtant l'Argentine connaît d'importants problèmes économiques. Le chômage et le bas niveau de vie continuent de marquer le pays, pourtant largement plus développé que les autres nations du tiers monde.

   L'Argentine est le pays le plus développé du continent latino-américain en 2005 selon les données des Nations-Unis fournies en 2007 et se rapproche des standards européens de niveau de vie. […]. (http://fr.wikipedia.org/wiki/Argentine#Environnement).

On voit là l'importance des chaînes de référence non seulement dans la structuration des textes mais aussi dans l'établissement de la cohésion référentielle. En effet, leur absence en (4) (un texte d'enfant) en rend difficile le suivi : les thèmes se suivent et se juxtaposent sans liens référentiels manifestes, ce qui génère un sentiment de « décousu » :

[7] Par exemple dans *Paul dit : « Je suis malade* », l'on comprend que *je* coréfère à *Paul* mais, du fait de ses instructions particulières, il paraît difficile d'inscrire le pronom personnel de première personne dans la même chaîne que celle de *Paul* en raison de sa « token réfléxivité » *i.e*. de sa capacité à ne renvoyer qu'à celui qui dit « je ».

[8] Voir également Fox (1987).

4. Sortie à la dune

Nous sommes partis de l'école à bicyclette pour aller à la côte découvrir la dune. L'euphorbe est une plante toxique, elle fait gonfler la langue et étouffe. Les plantes les plus répandues sont l'oyat et le chiendent. Les plus vieilles dunes ont plus de cinq mille ans. La criste marine est une plante comestible. Madame l'inspectrice accompagnait M. Berger. Nous avons continué la visite avec eux. Puis nous sommes rentrés à l'école. (Exemple de Turco, cité par Reichler-Béguelin *et al*., 1988, 132)

2.4. Le matériau lexical des chaînes de référence : variations en tous genres…

Une quatrième difficulté des chaînes de référence tient à la diversité du matériau lexical des SN qui les composent, le cas échéant. C'est ce qu'illustre (5) où la tête lexicale de chacun des SN référant à L. Saha change constamment[9] :

5. Libre de s'engager avec le club de **son** choix depuis le résiliation de **son** contrat avec Sunderland, **Louis Saha** (34 ans) rejoint la Lazio Rome, assure le club italien sur son site. **L'attaquant français**, qui a passé ce mercredi sa visite médicale au sein de la capitale italienne, va être présenté à 19h00 au centre d'entraînement de Formello. Plus tôt dans la journée, **l'ancien joueur de Fulham, Manchester United et d'Everton** avait annoncé **son** départ pour Rome sur **son** compte *Twitter*.

Louis Saha, **qui s'**est engagé jusqu'au 30 juin prochain, renforce le secteur offensif du club entraîné par Vladimir Petkovic, qui vient de perdre son meilleur buteur, Miroslav Klose (dix réalisations), pour sept à huit semaines. **Le Français** va donc découvrir un troisième championnat après la Ligue 1 et la Premier League. (*L'équipe*, 06/02/2013)

Or, cette variation dans le matériel lexical n'a rien de trivial car elle peut menacer l'interprétation des chaînes de référence. Dans (6) ci-dessus, un lecteur totalement étranger au football, ignorant les aléas de la carrière de Saha, tablera sur la cohérence textuelle et le fait qu'il n'y a dans l'extrait qu'un référent saillant pour pallier ses déficiences culturelles footballistiques. Dans le cas de lecteurs non experts, comme de jeunes apprenants, la variation lexicale a pour effet que ceux-ci considèrent qu'il y a autant de référents que de SN (cf. Masseron & Schnedecker, 1988 ; Laparra, 1989). Cela étant, la variété de matériau lexicale dépend de trois facteurs : le genre discursif d'occurrence des chaînes de référence, la date et la langue de rédaction du texte.

2.4.1. Variation selon le genre discursif

Pour ce qui concerne le premier paramètre, le genre discursif, de nombreuses études ont démontré qu'il conditionne la composition des chaînes de référence. Ces études ont porté soit sur un genre particulier (p.ex. le portrait journalistique (Jenkins, 2002 ; Schnedecker, 2005), le fait divers (Schnedecker & Longo, 2012), les documents techniques (Dupont & Bestgen, 2006), les discours instructionnels (Maes, Arts & Noordman, 2004) ; soit sur la comparaison de genres en nombre variable (presse, romans, textes administratifs (Longo & Todirascu, 2010 ; Longo, 2013) ; nouvelles *vs* portraits journalistiques (Baumer, 2012), textes romanesques/informatifs/instructionnels/journalistiques – quitte à limiter l'analyse à certaines catégories d'anaphores (hyperonymiques (Condamines, 2005), pronominales (Tutin, 2002)) -. Ainsi Condamines (2005) démontre-elle, à partir d'une étude sur corpus, que les catégories d'anaphores varient en fonction du genre textuel. Par exemple, l'emploi des anaphores hyperonymiques (cf. tableau (1)) arrive, dans la majorité des genres, en seconde position, après les anaphores dites par supplétisme[10]. Seul, le genre de discours instructionnel

[9] Cf. Corblin, (1995, 167, 175).

[10] Du type de Aurélien, qui *a juré* à sa maîtresse de ne la point toucher, découvre que *ce serment*…

technique, *Méthode et outils de génie logiciel pour l'informatique scientifique*, fait exception puisqu'elles y sont utilisées à une hauteur de 60%. Par contraste, leur part dans le roman est 4 fois moins élevée que dans les textes « techniques »

Corpus	Hy	Supp.	Syn	Dév	Déadj	Dén	Fig	total
Géo[11]	26%	50%	10%	9%	2%	0	3%	100% (266)
GDP[12]	32%	55%	5.5%	5.5%	2%	0	0	100% (246)
Moug[13]	**60%**	31.5%	4.5%	4%	0	0	0	100% (107)
LMD[14]	19%	64.5%	9%	1%	1%	0.5%	5%	100% (415)
Bel A[15].	15.5%	47%	14.5%	1%	0	0	22%	100% (305)

Tableau 1. Incidence des genres sur la répartition des types d'anaphore (Condamines, 2005, 45). Explicitations des abbréviations : Hy = hyperonymes ; Supp. = supplétifs ; Syn = synonymes ; Dév = déverbaux ; Déadj = dérivés d'un adjectif ; Dén = dérivés d'un nom ; Fig = figures.

Longo (2013) montre de nettes différence dans la composition des chaînes de référence selon les genres (en l'occurrence des extraits de presse, des textes à caractère informatif pur, un roman), en prenant en considération plusieurs critères comme la longueur des chaînes, la distance inter-maillonnaire, la catégorie grammaticale du 1er maillon et celle des autres (cf. tableau 2) :

corpus / critères	**Le Monde**	**Le Monde Diplomatique**	**Acquis Communautaire**	**Les trois mousquetaires**	**La Documentation Française**
Longueur moyenne	4	3,7	**3**	**9**	3,4
Distance moyenne entre antécédents (nb de phrases)	0,8	0,9	0,6	0,4	1,1
Catégorie 1er maillon	Noms propres	SN définis	SN indéfinis	SN indéfinis	SN définis
Fréquence des maillons	30 % Np	50 % SN définis	40 % SN indéfinis	- 36 % pronoms - 28 % possessifs	- 33 % pronoms - 33 % SN définis
Correspondance thème -1er maillon	80 %	100 %	60 %	60 %	40 %

Tableau 2. Variation dans la composition des chaînes de référence selon leurs genres d'occurrence, d'après Longo (2013)

[11] Précis de géomorphologie de 206700 mots.

[12] Guide de planification de 148000 mots.

[13] *Méthode et outils de génie logiciel pour l'informatique scientifique* (45100 mots).

[14] *Le Monde Diplomatique* (1989). 110700 mots.

[15] *Bel Ami* de Maupassant, 170200 mots.

2.4.2. Variation selon l'époque

La date de composition des textes influe également sur la composition des chaînes de référence. Même si leur étude n'a jamais été menée en tant que telle sur les périodes anciennes, plusieurs phénomènes ont déjà été observés comme la relative stabilité désignationnelle des expressions référentielles dans les textes médiévaux. Perret (2000, 17) signale, en effet, que « l'ancien français privilégie les reprises nominales répétitives et semble préférer la stabilité voire la rigidité désignationnelle ». En calculant ce qu'elle nomme le *coefficient de stabilité*[16], elle montre en effet que celui-ci est plus élevé dans les textes anciens que dans les textes du début du 20ième siècle (elle prend l'exemple de Proust) et que, au sein des textes anciens, la variation du genre joue également un rôle puisque les romans en prose et les romans en vers (*Jehan de Saintré, Mélusine, Jehan de Paris*) ont un coefficient de stabilité supérieur à celui des nouvelles (*Cent nouvelles nouvelles*). Par ailleurs, certaines unités grammaticales comme *ledit* ou *lequel* désormais moins usitées, avaient la première une fonction désambiguïsante et de mise en saillance dans la prose de la fin du Moyen Âge (Mortelmans, 2008), la seconde, fréquente en tête de phrase (fin Moyen Âge-16ème siècle) jouait le rôle de « marque d'intégration textuelle » et de charnière narrative (Kuyumcuyan, 2010). Enfin, le procédé anticipatoire de la cataphore dans le français écrit jusqu'au 19ème siècle était sinon absent du moins rare (Combettes, 2006).

2.4.3. Variation selon la langue

Enfin, la variété du matériau lexical dépend – troisième et dernier paramètre – des langues. En comparant l'usage anaphorique dans des textes de langue romane *vs* germanique, certains auteurs ont observé une tendance des langues romanes à exploiter les anaphores dites infidèles de manière plus importante que les langues germaniques. Cela accréditerait donc l'idée que l'emploi des expressions coréférentielles dépend d'une contrainte typologique. L'opposition inter-langues se trouve dans trois études :

- l'une menée par Korzen (cité par Lundquist, 2005, 75 *et seq*.) où la comparaison entre l'italien et le danois montre une propension de l'italien à utiliser l'anaphore infidèle dans les proportions précisées dans le tableau (3) qui sont de l'ordre du simple au quadruple ;
- l'autre, réalisée par Lundquist (2005), qui accuse un écart beaucoup plus important, allant de 1 à 6.4 :

	Langue romane	Langue germanique
Skytte & Korzen (2000)[17]	40% AI (italien)	10% AI (danois)
Lundquist (2005)[18]	16% AI (français)	2.5% AI (danois)

Tableau 3. Comparaison de la proportion d'anaphores infidèles (AI) dans les langues romanes *vs* germaniques.

La troisième étude fait l'objet d'une thèse récente sur les chaînes de référence étudiées dans un corpus (fiction *vs* presse) et dans des langues contrastées (français *vs* anglais). Baumer (2012) montre également que la part des anaphores nominales diffère selon la langue et

[16] Obtenu en divisant pour un référent donné le nombre total d'anaphores nominales par le nombre de désignations différentes (art.cit., 17).

[17] Dans les textes écrits. En note 1, Lundquist signale que cette proportion diminue de moitié : 18% en italien et 6% en danois.

[18] Il s'agit en l'occurrence de textes de presse centrés sur des personnalités politiques.

qu'elles sont deux fois moins employées en anglais qu'en français (9.4% en français *vs* 4.8% en anglais).

3. Des typologies de chaînes et enjeux

Compte tenu des paramètres qui viennent d'être évoqués, ajoutés à la multiplicité des genres textuels, il n'est pas étonnant que l'on ne dispose pas (encore) à l'heure actuelle de typologies des chaînes de référence, qui permettraient de les classer selon leur mode de composition et d'en proposer une sorte de modélisation. Dans cette section nous allons explorer trois points de vue qui pourraient amorcer la réflexion dans ce sens : premièrement, celui de Givón (1983), article fondateur sur la notion de saillance et de continuité référentielle qui lui est liée, deuxièmement le point de vue de Schnedecker (2006) avec la proposition d'une typologie des transitions d'un référent à un autre, et troisièmement le point de vue général du domaine du traitement automatique des langues, qui se focalise sur la notion d'entité nommée et appréhende la coréférence dans ce cadre, avec une visée applicative dont les caractéristiques peuvent nourrir la réflexion linguistique.

3.1. Les caractérisations de Givón (1983) et leurs limites

Givón (1983) a proposé pour les topiques dits « continus » (ou référents dit saillants) *vs* « discontinus » (référents non saillants) des caractérisations schématisées ci-dessous, prenant en compte les paramètres de la distance, persistance et interférence (Givón, *op. cit*. 1983, 12 ; 1989, 214) :

- **Schéma 1. Configuration d'un référent saillant** : p1... p20} **SN1**... il_1... il_1... il_1... il_1... $il_{(10)}$ **SN2**...il_1...il_1... {p1... p15} **SN1'...**
- **Schéma 2. Configuration d'un référent non saillant** : **SN_1**... il_1.... **SN_2**… il_2.... il_2.... { } **SN_3**... il_3... { }

En d'autres termes, la chaîne de référence d'un topique « continu » devrait être majoritairement composée de pronoms personnels, sur une distance « longue » (*i.e*. de 10 phrases consécutives, au moins selon Givón) alors que celles d'un topique « discontinu » du fait qu'elles entrent en compétition avec d'autres chaînes, seraient plus courtes et instancieraient davantage des SN que les chaînes de topiques continus.

Si ce genre de caractérisation correspond à l'intuition que l'on peut avoir de ce que sont des chaînes de référence, force est de constater qu'il ne coïncide qu'avec les chaînes de certains genres textuels : les genres à « personnages dominants » identifiés (6 et 7) :

6. Comme le leader des étudiants, **le doyen des études** éprouvait le besoin, moral et politique, d'observer ce qui se passait dans l'anti-institution. Pour des raisons fort différentes, en majorité dues à **sa** loyauté et à **son** attachement pour Gerard Wijnnobel, **il** savait qu'**il** devait surveiller ce que faisait et prêchait Eva Wijnnobel. **Il** savait qu'**il** n'était pas taillé pour s'acquitter de tâches d'infiltration ou de confrontation. **Il** était par instinct profondément libéral et partisan du laisser-faire. **Il** avait accepté le poste de doyen parce que, d'une part, tout le monde était obligé d'assumer à tour de rôle une fonction d'autorité et, d'autre part, **il** voulait rendre les choses plus faciles et plus libres pour les étudiants. **Il** savait mieux **s'**y prendre avec des réactions instinctives de nervosité dans **son** propre camp — restrictions, répressions, exclusions — qu'avec une opposition doctrinaire qui s'opposait pour le plaisir de s'opposer, ce qu'**il** percevait mais **ø** n'arrivait absolument pas à comprendre. (A.S. Byatt, *Une femme qui siffle*)
7. **Il** rompt dix ans de silence

   Bowie, pape de la pop

Depuis *Reality*, paru en 2003 on **le** croyait à la retraite. A 66 ans, **David Bowie** revient avec *The Next Day*. Pour raconter l'odyssée du **Protée du rock**, Fabrice Pliskin a choisi six de **ses** plus mémorables chansons.

Habemus popam. **Son** silence aura duré dix ans, moins longtemps que le silence éternel de Garbo ou de Rimbaud, mais deux fois plus que le silence de Lennon (1975-1980). Comme **son** compatriote, **il** aura consacré **sa** longue retraite à expérimenter les rites de la paternité à New York. **Bowie** avait publié *Reality*, **son** dernier album studio, en 2003, avant de subir en urgence une opération du cœur l'année suivante. A 66 ans, **il** revient avec *The Next Day* : « Me voici. Pas tout à fait mourant », rugit-**il** en haine des macabres rumeurs. […] (*Nouvel Observateur*, 07/03/2013)

En revanche, dans des textes expositifs-informatifs comme (8), le référent que l'on peut identifier comme saillant – le changement climatique – en raison du nombre d'occurrences d'expressions référentielles notamment, se présente sous forme de SN dont la tête lexicale reste identique, pratiquement sans reprise pronominale, fortement distants et la chaîne subit la concurrence de nombreux autres référents :

8. Les effets d'**un climat en mutation**.

   Les effets **du changement climatique** sont plus ou moins graves selon les régions. Les régions les plus vulnérables d'Europe sont l'Europe du Sud, le bassin méditerranéen, les régions ultrapériphériques et l'Arctique. Les zones de montagne, et en particulier les Alpes, les zones côtières et urbaines et les plaines inondables densément peuplées sont confrontées à des problèmes spécifiques. À l'extérieur de l'Europe, les pays en développement (dont les petits États insulaires) demeureront particulièrement vulnérables.

   Le changement climatique aura des répercussions sur un certain nombre de secteurs. Dans le secteur agricole, **les changements climatiques prévus** auront des retombées sur les rendements agricoles, la conduite de l'élevage et la localisation de la production. La probabilité et la gravité croissantes des phénomènes météorologiques extrêmes augmenteront considérablement le risque de mauvaises récoltes.

   Le changement climatique aura également une incidence sur les sols, avec la destruction de la matière organique, un facteur essentiel du processus de fertilisation des sols. **Il** pourrait aussi modifier l'état sanitaire et la productivité des forêts, ainsi que la distribution géographique de certaines essences (Commission des Communautés Européennes, *Livre blanc – adaptation au changement climatique : vers un cadre d'action européen*, 2009)

- **Schéma 3. Configuration d'un référent saillant non humain dans un texte non narratif :** $\S_{1\ SN1} \ldots \S_{2\ SN2} \cdots\cdots \S_{N\ SNn} \cdots$

Autrement dit, il reste à éprouver la validité de la caractérisation de Givón sur des genres de textes différents pour déterminer auxquels elle correspond le mieux. Les enjeux sont d'autant plus importants que la modélisation des chaînes de référence pourrait servir en retour d'indice linguistique fort pour déterminer les genres textuels.

3.2. Les typologies des transitions référentielles ou des relations entre chaînes

En revanche, dans le domaine de la typologie des chaînes de référence, les travaux sont plus avancés en ce qui concerne la cohabitation de chaînes de référence renvoyant à des référents distincts dans les textes (là encore principalement narratifs). En effet, l'une des complications provient du fait que le déroulement des chaînes de référence est difficilement prévisible : il est délicat, à partir d'une occurrence référentielle donnée, de préjuger du fait qu'elle sera ou non réinstanciée dans la suite du discours ou des interruptions qui, le cas échéant, en perturbent le cours.

Tout dépend de la nature du « premier maillon ». Certaines expressions référentielles, comme le nom propre, sont de bons indicateurs sur l'importance du référent et partant, du fait qu'il fera vraisemblablement l'objet d'une chaîne de référence. Inversement les pronoms indéfinis, comme *quelqu'un*, signalant que leur référent est non spécifique ou spécifique mais indéterminé n'augurent pas la mise en chaîne[19].

La question se complique dès l'instant où interfèrent les unes avec les autres dans un même texte plusieurs chaînes. On peut néanmoins proposer un classement (encore provisoire) de leur mode de cohabitation :

A. la succession (9) rend compte de ce qu'une chaîne disparaît dès l'instant où celle qui lui succède apparaît dans le texte :

9. Johannesbourg, années 70. **Un professeur d'histoire** vit une vie de famille sans histoire, entouré de l'affection des **siens** et aveugle aux problèmes politiques et sociaux engendrés par l'"apartheid", régime ségrégationniste qui sévit en Union sud-africaine. **Cet homme paisible** a *un jardinier noir*, paisible *lui aussi*, et soumis à la fatalité. *Ce jardinier* a un fils et ce fils participe à une manifestation d'écoliers violemment réprimée par la police. On apprendra que l'enfant est arrêté, puis qu'il est mort. (Résumé d'*Une saison blanche et sèche*, *Télérama*, 15.01.92).

B. l'entrecroisement correspond aux cas les plus fréquents, sans doute où deux référents, voire plus, sont instanciés, consécutivement et/ou simultanément, avec des modifications de leur statut syntaxique (par exemple dans (10) : Paulette d'abord sujet devient objet quand intervient le second référent Michel Dollé, puis redevient sujet :

10. En juin 1940, les parents de **la petite Paulette** (5 ans) sont tués sur une route de l'Exode**. La petite fille se** retrouve seule dans un champ, portant le cadavre de **son** chien. *Un jeune paysan*, Michel Dollé (11 ans) **l'**amène à la ferme de **ses** parents, qui décident de **la** garder. **Paulette** veut faire comme les adultes et donner une sépulture à son chien. Pour **elle**, *Michel* organise un cimetière d'animaux. (Résumé de *Jeux interdits, Télérama*, 26.08.92)

C. la dérivation où l'introduction d'un nouveau référent dans un texte n'opère pas par le biais d'un SN indéfini ou d'un nom propre mais peut dériver d'une autre chaîne par des moyens anaphoriques, comme en (11) une anaphore associative : *l'homme au volant* dépend en effet d'*un accident* ou une anaphore possessive dans (12) :

11. Jack est ingénieur du son. Un soir, alors qu'il se promène près d'une rivière pour enregistrer les bruits de la nature, il assiste à **un accident**. Il sauve *la passagère*, mais l'homme au volant est tué. Il était candidat à l'investiture de son parti pour la magistrature suprême. Il faut absolument taire le fait qu'il voyageait en compagnie *d'une femme, aussi publique que lu*i. De plus, une écoute attentive de la bande sonore que Jack a enregistrée accidentellement prouve que l'accident de la voiture a été précédé d'un coup de feu. (Résumé de *Blow Out*, *Télérama*)

12. *Rose* a 18 ans. Elle vit, dans un village de Camargue avec *son* **père Fernand, qui** ne **s'**est jamais consolé du départ de **sa** femme, *son* frère Vincent, le jeune apprenti du garage, Tintin, et les copains de *son* frère, Sauveur, Baptiste et Antoine. Tous considèrent qu'*elle* se doit à eux, qu'*elle* leur appartient. (Résumé de *Mauvaise fille, Télérama*, 15.01.92)

D. la partition et la fusion constituent une sous-catégorie de dérivation procédant par deux mouvements symétriques, l'un consistant à extraire d'un ensemble préalablement posé (*deux hommes* dans (15)) (indexé par E) des référents individués (R_1 et R_2) (*l'un/l'autre, le plus petit/le plus grand*) ; l'autre à constituer en ensemble des référents présentés isolément (*ils* aux 3ième et 4ième paragraphes) :

13. **Deux hommes**$_{(E)}$ parurent.

[19] Cela n'exclut pas les cas de « dévoilement progressif des personnages » cf. Schnedecker (2006, 410-11).

L'un$_{(R1)}$ venait de la Bastille, *l'autre* $_{(R2)}$ du Jardin des Plantes. *Le plus grand*, vêtu de toile, marchait le chapeau en arrière, le gilet déboutonné et sa cravate à la main. *Le plus petit*, dont le corps disparaissait dans une redingote marron baissait la tête sous une casquette à visière jaune.

Quand **ils**$_{(E)}$ furent arrivés au milieu du boulevard, **ils**$_{(E)}$ **s**'assirent à la même minute, sur le même banc.

Pour **s'**essuyer le front, **ils**$_{(E)}$ retirèrent **leurs**$_{(E)}$ coiffures, que chacun posa près de soi ; et *le petit homme* $_{(R1)}$ aperçut écrit dans le chapeau de *son voisin*$_{(R2)}$: Bouvard ; pendant que *celui-ci* distinguait aisément dans la casquette *du particulier en redingote* le mot : Pécuchet (Flaubert, *incipit* de *Bouvard et Pécuchet*, 1881)

E. le déroulement en parallèle correspond aux cas de topiques multiples (Schnedecker, 2006) où deux référents aussi saillants l'un que l'autre sont repris en parallèle :

14. Pour comprendre, il faut d'abord comparer. **DSK** est un prof d'université. *Fabius* est un énarque normalien. **L'un** est rond et séducteur. *L'autre* est sec et élégant. **Strauss-Kahn** est brouillon. *Fabius* est précis. **Le premier** a de l'humour. *Le second*, de l'ironie. **DSK** ne se méfie de personne. *Fabius se* méfie de tout le monde. **DSK** pense — à tort — n'avoir que des amis. *Fabius* croit — à tort — n'avoir que des ennemis. **L'un** finit toujours par se réconcilier. *L'autre* pardonne parfois, mais n'oublie jamais. **DSK** commet beaucoup d'erreurs qu'il sait toujours réparer. *Fabius* fait peu de fautes, mais elles *lui* coûtent très cher. **Le premier** est un homme de contact. *Le second* est un homme de réseau. **DSK** pense d'abord au monde. *Fabius* part toujours de la France. **Strauss-Kahn** croit au mouvement spontané de la société. *Fabius* est un homme de l'Etat. **L'un** est un chef de bande. *L'autre* est le leader d'un courant. **DSK** est un jospiniste atypique. *Fabius* est fabiusien. ***Ces deux hommes*** ne vivent pas sur la même planète ! (*Nouvel Obs*, 24/01/02)

Il resterait à savoir si ces types d'enchaînement ou transition référentiels sont caractéristiques de genres ou de référents particuliers, d'époques ou de systèmes linguistiques particuliers. Par exemple, Schnedecker (2006) montre que l'usage des pronoms ordinaux et de leur mise en chaîne varie selon le genre textuel : les textes narratifs exploitant les constructions parallèles en alternance visant le contraste (*R1p1*/*R2p2*/*R1p3*/*R2P4, R1pn*/*R2pn+1*…) alors que, dans les textes informatifs-argumentatifs, les éléments sériés, préalablement introduits par un SN pluriel (généralement déterminé par un cardinal), font successivement l'objet d'une chaîne de référence « en bloc » (*RE p*/*R1 p1*/*p2*/*pn*….// *R2 p1*/*p2*/*pn*…).

3.3. Les chaînes de référence en traitement automatique des langues

Nous le voyons, les problèmes linguistiques autour des chaînes de référence sont nombreux et fédèrent un vaste ensemble de questions relatives aux marques linguistiques, aux phénomènes d'accès référentiels et de saillance ou encore à l'organisation textuelle. Dans le domaine du traitement automatique des langues, les recherches et les réalisations informatiques explorent des problématiques quelque peu différentes. Les conférences MUC (*Message Understanding Conference*) et d'autres initiatives relevant de l'extraction d'information – cf. Poibeau (2003) à la fois pour les aspects historiques et techniques – ont mis en avant la notion d'*entité nommée*. Plus que les phénomènes de référence, ce sont les types d'expressions désignant des personnes, des organisations, des lieux, des dates, des quantités, etc., qui ont fait l'objet de classifications. Pour détecter automatiquement les entités nommées dans un texte, les efforts ont porté sur la mise au point de listes de termes-types, de règles, et maintenant d'ensembles de traits pertinents pour entraîner des systèmes d'apprentissage automatique. L'enjeu est de repérer des marqueurs et non de s'intéresser aux types d'accès aux référents ou à leur accessibilité. Plus un programme arrive à détecter d'entités nommées et à les catégoriser avec justesse, plus il obtient des mesures satisfaisantes. Conférences et campagnes d'évaluation se succèdent ainsi, avec parfois plusieurs tâches complémentaires pour diversifier la

compétition. La tâche de détection de chaînes de référence est apparue ainsi. Il s'agit non seulement de repérer des marqueurs, de catégoriser les référents correspondants mais aussi de détecter des relations de coréférence entre marqueurs. On parle alors de « résolution des coréférences », plutôt que d'identification des chaînes de référence, et on reste bien dans des applications d'extraction d'information.

Or la résolution des coréférences implique non seulement la détection des entités nommées mais aussi la résolution des anaphores, devenue une problématique à part entière en traitement automatique des langues (Mitkov, 2002). Les premiers efforts ont porté sur les anaphores pronominales avec, tout d'abord, le problème consistant à distinguer les *il* impersonnels – que le programme doit ignorer – des *il* personnels, que le programme doit repérer et pour lesquels il doit trouver un antécédent « plein » de manière à construire une relation de coréférence. Comme pour la détection des entités nommées, des règles sont programmées pour ce faire. Les programmes informatiques se fondent ainsi sur quelques critères morphosyntaxiques, éventuellement syntaxiques mais surtout sur la matière même du texte (termes utilisés, nombre de mots entre une entité nommée et l'anaphore, etc.). Parfois, ce sont même les approches avec les critères les plus simples qui donnent de bons résultats : dans beaucoup de cas, la dernière entité nommée avec le même genre et le même nombre que le pronom s'avère être l'antécédent, une telle règle constituant un point de départ important pour le programmeur, malgré les contre-exemples célèbres que sont « la sentinelle […] il » ou « le maire […] elle ». Les techniques évoluent – pour une revue complète, se reporter à Poesio *et al*. (2010) –, et font désormais appel à des algorithmes d'apprentissage automatique plutôt que des paramétrages à la main, avec là aussi l'enjeu consistant à identifier un ensemble de traits pertinents pour entraîner le système. Les résultats restent cependant loin de la finesse des analyses linguistiques comme celles de Corblin (1987), Kleiber (1994, 2001), Corblin (1995), Schnedecker (1997), Charolles (2002), etc. Cet écart entre théorie et application est remarqué, et amène certains auteurs à critiquer les contraintes des campagnes d'évaluation dans la mesure où elles obligent à ne s'intéresser qu'à un ensemble restreint de phénomènes (van Deemter & Kibble, 2000).

A l'heure actuelle, une dizaine de systèmes disponibles arrivent à identifier plus ou moins bien les relations de coréférence dans un texte tout venant (cf. par exemple la comparaison visible sous le titre « Coreference Resolution Tools: A First Look » sur le site Web suivant : http://www.minvolai.com/blog/2010/09/coreference-resolution-tools-a-first-look/ – consulté en juin 2014). Les performances sont remarquables compte tenu de la difficulté de la tâche mais les erreurs faites par les systèmes peuvent sembler grossières pour un linguiste : des pronoms sont affectés à des référents non pertinents, des expressions référentielles ne sont même pas repérées, alors qu'une lecture (humaine) donne immédiatement les solutions. Pour améliorer les performances de tels systèmes, les efforts actuels portent sur l'exploitation d'une liste de critères pertinents afin de mettre en œuvre plusieurs passes efficaces de traitement (Ng, 2007 ; Bengtson & Roth, 2008 ; Raghunathan *et al*., 2010 ; Recasens *et al*., 2013), sur le développement de plateformes permettant à chacun de paramétrer les critères de son propre système de résolution de coréférences (Stoyanov *et al*., 2010), ou encore sur l'exploitation d'informations spécifiques au domaine traité (Gilbert & Riloff, 2013). Des efforts sont faits également sur des tâches plus spécifiques, comme la résolution des coréférences événementielles (Cybulska & Vossen, 2013) ou la résolution des coréférences dans des textes multilingues (Zhekova & Kübler, 2013).

4. Contexte de travail : le projet MC4

Ce volume est l'émanation des travaux réalisés dans le cadre d'un projet intitulé MC4 : *Modélisation Contrastive et Computationnelle des Chaînes de Coréférence*, projet de type PEPS – « Projets Exploratoires Premier Soutien » – accordé par deux instituts du CNRS, l'Institut des Sciences Humaines et Sociales, et l'Institut des Sciences de l'Information et de leurs Interactions. Le projet s'est intéressé à la référence et à la coréférence dans des textes écrits, en français médiéval et en français contemporain, avec des objectifs à la fois théoriques et pratiques qui regroupent divers chercheurs en linguistique et en informatique, notamment des laboratoires Lattice, LILPA et ICAR. Démarré début juillet 2011, terminé début 2013, ce numéro spécial en constitue une synthèse et un bilan des résultats.

D'un point de vue chronologique, les travaux effectués dans ce cadre ont consisté à mettre en œuvre un travail d'annotation de corpus, en passant par la construction d'un schéma d'annotation focalisé avant tout sur les phénomènes de référence et incluant de ce fait des aspects morphosyntaxiques, syntaxiques et sémantiques. Ces aspects ont été d'abord explorés et testés sur des textes en français contemporain avant de l'être sur des textes en ancien et en moyen français. De fait, une étape du travail a alors consisté à prendre en compte un ensemble de modifications permettant d'obtenir un schéma d'annotation compatible avec les différents états de langue. Parmi les aspects discutés lors des nombreuses réunions plénières, se trouvent le cas marqué, l'aspect pro-drop (pronom non exprimé), la distinction entre *les chevaliers du roi* et *les chevaliers le roi*, etc. Par ailleurs, certains de ces aspects ainsi que d'autres spécificités du français médiéval, notamment l'impossibilité de la cataphore avant une certaine date, ont amené à formuler de nouvelles hypothèses linguistiques qui ont été ajoutées à une liste correspondant aux objectifs du projet.

C'est l'utilisation d'un même schéma d'annotation qui a permis d'aboutir aux résultats présentés dans les différents articles de ce volume : avec un même cadre de réflexion, les chercheurs, qu'ils soient diachroniciens ou non, spécialistes de traitement automatique des langues ou non, ont pu mettre en parallèle différentes études des chaînes de référence. Bien qu'elles couvrent plusieurs genres textuels (textes non narratifs pour C. Schnedecker et narratifs pour J. Glikman, C. Guillot-Barbance et V. Obry ; juridiques pour D. Capin et L. Longo et A. Todirascu), bien qu'elles couvrent plusieurs époques et même plusieurs types d'ambiguïtés référentielles (comme celles du pronom *on* dans l'étude de F. Landragin et N. Tanguy), ces études reposent toutes sur une définition commune des chaînes de référence, des types et des propriétés des marqueurs de référence. Elles ont de plus bénéficié d'un même cadre logiciel pour l'annotation, la visualisation et l'interrogation des données annotées, à savoir le logiciel ANALEC décrit dans le dernier article du volume (F. Mélanie-Becquet et F. Landragin).

Du point de vue de la constitution de ressources ce projet PEPS prépare en explorant les aspects techniques et leurs conséquences la constitution de ressources en français contemporain et en français médiéval, cette constitution exploitant au mieux les efforts déjà fournis par la communauté. Concernant le français médiéval, les efforts déjà fournis (BFM : Base du Français Médiéval ; SCRMF : *Syntactic Reference Corpus of Medieval French*) portent plus spécifiquement sur les couches morphosyntaxique et syntaxique des annotations, alors que le PEPS prépare avec les phénomènes de référence une nouvelle couche, plus sémantique et pragmatique.

Références bibliographiques

ARIEL M. (1990) *Accessing Noun-Phrase Antecedents*, Theoretical Linguistics Series. Routledge.

BAUMER E. (2012) *Noms propres et anaphores nominales en anglais et en français : étude comparée des chaînes de référence*. Thèse de doctorat, Paris Diderot.

BENGTSON E. & ROTH D. (2008) « Understanding the Value of Features for Coreference Resolution », in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Waikiki, Honolulu, Hawai.

BOUDREAU S. & KITTREDGE R. (2005) « Résolution des anaphores et détermination des chaînes de coréférences. Différences entre variétés de textes », *Traitement Automatique des Langues* 46(1), 41-70.

CHAROLLES M. (1988) « Les plans d'organisation textuelle : périodes, chaînes, portées et séquences », *Pratiques* 57, 3-13.

CHAROLLES M. (1995a) « Cohésion, cohérence et pertinence du discours », *Travaux de Linguistique*, 29, 125-151.

CHAROLLES M. (1995b) « Comment repêcher les derniers ? Analyse des expressions anaphoriques en *ce dernier* », *Pratiques*, 85, 89-112.

CHAROLLES M. (1997) « L'encadrement du discours : univers, champs, domaines et espaces », *Cahier de Recherche Linguistique*, LANDISCO, URA-CNRS 1035 Université Nancy 2, n° 6, 1-73.

CHAROLLES M. (1997b) « Les reprises démonstratives en "ce dernier" dans les textes contemporains de grande diffusion », in O.Välikangas & J. Härmä (eds) *Travaux et Recherches en Linguistique Appliquée*, 1, 7-17.

CHAROLLES M. (2002) *La référence et les expressions référentielles en français*, Paris : Ophrys.

CHAROLLES M. & PERY-WOODLEY M.-P. (2005) « Les adverbiaux cadratifs : introduction », *Langue Française*, 148, 3-8.

CHAROLLES M. (2009), « Les cadres de discours comme marques d'organisation des discours », in F. Venier (ed.) *Tra Pragmatica e Linguistica Testuale*. Edizioni dell'Orso, Alessandria,401-420.

CHAROLLES M. (à paraître) « L'anaphore et l'ancrage des référents dans le discours », in Godard D., Abeillé A. & Delaveau A. (éds), *Grande Grammaire du Français* (chapitre 18), Paris : Actes Sud.

CHAROLLES M. & VIGIER D. (2005) « Les adverbiaux en position préverbale : portée cadrative et organisation des discours », *Langue Française*, 148, 9-30.

CHASTAIN C. (1975), « Reference and Context », in : GUNDERSON K. (ed.), *Language Mind and Knowledge*, Minneapolis : University of Minnesota Press, 194-269.

COMBETTES B. (2006), « Cataphore et texte littéraire : aspects diachroniques », in F. Berlan (éds) *Langue littéraire et changements linguistiques*, Presses de l'Université Paris-Sorbonne, 385-496.

CONDAMINES A. (2005) « Anaphore nominale infidèle et hyperonymie : le rôle du genre textuel ». *Revue de sémantique et pragmatique*, 18, 33-52.

CORBLIN F. (1987) *Indéfini, défini et démonstratif*, Genève : Droz.

CORBLIN F. (1995) *Les formes de reprise dans le discours. Anaphores et chaînes de référence*, Rennes : P. U. Rennes.

CORNISH F. (1986) *Anaphoric Relations in English and French. A Discourse Perspective*, London: Croom Helm.

CORNISH F. (1999) *Anaphora, Discourse, and Understanding: Evidence from English and French*, Oxford: Oxford University Press.

CYBULSKA A. & VOSSEN P. (2013) « Semantic Relations between Events and their Time, Locations and Participants for Event Coreference Resolution », in *Proceedings of Recent Advances in Natural Language Processing (RANLP-2013)*, Hissar, Bulgaria, 156-163.

DE MULDER W. (2001) « Peut-on définir les SN démonstratifs par leurs contextes », in H. Kronning *et al*. (éds), *Langage et référence. Mélanges offerts à K. Jonasson à l'occasion de ses soixante ans*, Uppsala, Acta Universitatis Upsaliensis, 115-123.

DUPONT V. & BESTGEN Y. (2006) « Learning From Technical Documents: The Role of Intermodal Referring Expressions », *Human Factors : The Journal of The Human Factors and Erogonomics Society Summer*, 48/2, 257-64.

FAUCONNIER G. (1974) *La coréférence : syntaxe ou sémantique ?*, Paris : Seuil.

FOX B. (1987), *Discourse structure and anaphora*, Cambridge, Cambridge U. P.

GILBERT N. & RILOFF E. (2013) « Domain-Specific Coreference Resolution with Lexicalized Features », in *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL 2013)*, Sofia, Bulgaria.

GIVÓN T. (1983) Topic continuity in discourse: An introduction, in T. Givón (ed.), *Topic Continuity in Discourse. A Quantitative Cross Language Study*, John Benjamins Publishing.

GIVÓN T. (1989) *Mind, Code, and Context: Essays in Pragmatics*, L. Erlbaum Associates.

GROSZ B., JOSHI A. & WEINSTEIN S. (1995) « Centering: a Framework for Modeling the Local Coherence of Discourse », *Computational Linguistics* 21(2), 203-225.

GUILLOT C. (2009) « Ecrit médiéval et traces d'oralité : l'exemple de l'adverbe *or(e)* » in E. Havu *et al*. (éds), *La langue en contexte. Actes du colloque Représentation du sens linguistique IV (Helsinki, 28-30 mai 2008)*, Helsinki, Société Néophilologique, 267-281.

HABERT B. (2005) « Portrait de linguiste(s) à l'instrument », *Texto!* 104, www.revue-texto.net/Corpus/Publications/Habert/Habert_Portrait.html.

HINDS J. (1979) « Organizational Patterns in Discourse », *Syntax and Semantic*, vol 12.

HOBBS J.-R. (1979) « Coherence and Coreference », *Cognitive Science* 3, 67-90.

HUANG Y. (2002) *Anaphora : A cross linguistic approach*, Oxford, Oxford U. P.

JENKINS C. (2002) Les procédés référentiels dans les portraits journalistiques, *XV Skandinaviske romanistkongress*, Oslo.

KARTTUNEN L. (1976) « Discourse Referents » in MCCAWLEY J.-D. (ed.), *Syntax and Semantics* 7, New York, Academic Press, 363-385.

KLEIBER G. (1981) *Problèmes de référence : descriptions définies et noms propres*, Paris, Klincksieck.

KLEIBER G. (1994) *Anaphores et pronoms*, Louvain la Neuve, Duculot.

KLEIBER G. (2001) *L'anaphore associative*, Paris, PUF.

KLEIBER G. (2002) Marqueurs référentiels et théorie du centrage, *LINX*, 47, 107-119.

KLEIBER G., SCHNEDECKER C. & TYVAERT J.-E. (éds) (1997) *La continuité référentielle*, Paris, Klincksieck.

KUYUMCUYAN A. (2010) Lequel « outil de reprise » : d'un quasi-démonstratif au relatif ?, in *Diachro 5, Le français en diachronie*, Lyon.

LANDRAGIN F. (2011) « Une procédure d'analyse et d'annotation des chaînes de coréférence dans des textes écrits », *Corpus* 10, http://corpus.revues.org.

LAPARRA M. (1989) « Le repérage initial des personnages. Difficultés éprouvées par des élèves réputés mauvais lecteurs », *Pratiques* 60, 59-73.

LEGALLOIS D. (éd.) (2006) *Cognition, Représentation, Langage* 2006 : *Organisation des textes et cohérence des discours*, http://corela.edel.univ-poitiers.fr/.

LONGO L. (2013) *Vers des moteurs de recherche « intelligents » : un outil de détection automatique de thèmes. Méthode basée sur l'identification automatique des chaînes de référence*, Thèse de doctorat, Université de Strasbourg.
LONGO L. & TODIRASCU A. (2010) A. Genre-based Reference Chains Identification for French, *Investigationes Linguisticae*, Volume XXI, 57-75.
LUNDQUIST L. (2005) « Noms, verbes et anaphores (in)fidèles. Pourquoi les Danois sont plus fidèles que les Français ? », *Langue française*, 145, 73-91.
MAES A., ARTS A. & NOORDMAN L. (2004) Reference Management in Instructive Discourse, *Discourse Processes*, 37/2, 117-144.
MARANDIN J.-M. (1988) « A propos de la notion de thème de discours. Eléments d'analyse dans le récit », *Langue Française* 78, 67-87.
MARSLEN-WILSON W., LEVY E. & KOMISARJEVSKI-TYLER L. (1982) « Producing Interpretable Discourse : The Establishment and Maintenance of Reference », in R.J. Jarvella (ed.), *Speech, Place and Action*, New York, J. Wiley and Sons, 339-378.
MASSERON C. & SCHNEDECKER C. (1988) Le mode de désignation des personnages, *Pratiques* 60 : 98-123.
MATHET Y. & WIDLÖCHER A. (2012) « Glozz Annotation Platform », http://www.glozz.org/.
MILNER J.-C. (1982) *Ordres et raisons de langue*, Paris : Seuil.
MITKOV R. (2002), *Anaphora Resolution*, Longmann.
MOESCHLER J. & REBOUL A. (1994) *Dictionnaire encyclopédique de pragmatique*, Paris, Seuil.
MORTELMANS J. (2008) *Ledit in Middle French: Textual function and grammaticalization,* PhD Thesis, Antwerpen.
NG V. (2007) « Shallow Semantics for Coreference Resolution », *International Joint Conference on Artificial Intelligence (IJCAI)*, Hyderabad, India, 1689-1694.
PASSERAULT J.-M. & CHESNET D. (1991) « Le marquage des paragraphes : son rôle dans la gestion des traitements pendant la lecture », *Psychologie Française*, 36-2, 159-165.
PERRET M. (2000), « Quelques remarques sur l'anaphore nominale aux 14° et 15° siècles », *L'information grammaticale,* 87,17-23.
POESIO M., PONZETTO S.P. & VERSLEY Y. (2010) Computational Models of Anaphora Resolution: A Survey. Manuscrit non publié disponible sur la page Web de l'auteur.
POIBEAU T. (2003) *Extraction automatique d'information : Du texte brut au web sémantique*, Hermès-Lavoisier.
RAGHUNATHAN K., LEE H., RANGARAJAN S., CHAMBERS N., SURDEANU M., JURAFSKY D. & MANNING C. (2010) « A Multi-Pass Sieve for Coreference Resolution » in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, MIT, Massachusetts.
RECASENS M., CAN W. & JURAFSKY D. (2013) « Same Referent, Different Words: Unsupervised Mining of Opaque Coreferent Mentions », in *Proceedings of Human Language Technologies: The 11th Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT)*, Atlanta, Georgia, 897-906.
REICHLER-BEGUELIN M.-J., DENERVAUD M. & JESPERSEN J. (1988) *Ecrire en français. Cohésion textuelle et apprentissage de l'expression écrite*, Lausanne, Delachaux et Niestlé.
SCHNEDECKER C. (1997) *Nom propre et chaînes de référence*, Paris, Klincksieck.
SCHNEDECKER C. (2005), Les chaînes de référence dans les portraits journalistiques : éléments de description, *Travaux de linguistique* 51, 2005/2, 85-133.
SCHNEDECKER C. (2006), *De l'un à l'autre et réciproquement… Aspects sémantiques, discursifs et cognitifs des pronoms anaphoriques corrélés* l'un/l'autre *et* le premier/le second, Bruxelles, Duculot.

SCHNEDECKER C. & LONGO L. (2012) « Impact des genres sur la composition des chaînes de référence : le cas des faits divers », in *3ième Congrès Mondial de Linguistique Française*, Lyon.

SKYTTE G. & KORZEN I. (2000, *Italiensk–dansk sprogbrug i komparativt perspektiv. Reference, konnexion og diskursmarkering*. Copenhagen: Samfundslitteratur.

STARK H.A., 1988 « What do Paragraph Markings do ? », *Discourse Processes*, 11, 275-303.

STOYANOV V., CARDIE C., GILBERT N., RILOFF E., BUTLER D. & HYSOM D. (2010) « Coreference Resolution with Reconcile », in *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL)*, Uppsala, Sweden.

TOOLE J. (1996) « The Effect of Genre on Referential Choice », in T. Fretheim & J.K. Gundel (eds.), *Reference and Referent Accessibility,* John Benjamins, Amsterdam, 262-290.

TUTIN A. (2002) « A corpus-based study of pronominal anaphoric expressions in French », in *Proceedings of DAARC 2002 (Discourse Anaphora and Anaphora Resolution),* Lisbon.

VAN DEEMTER K. & KIBBLE R. (2000) « On Coreferring: Coreference Annotation in MUC and Related Schemes », *Computational Linguistics* 26(4), 615-623.

VAN PETEGHEM M. (1991) *Les phrases copulatives dans les langues romanes*, Wilhelmsfeld : Gottfried Egert Verlag.

VICTORRI B. (2012) « ANALEC : logiciel d'annotation et d'analyse de corpus écrits », logiciel téléchargeable sur : http://www.lattice.cnrs.fr/ANALEC.

VIGIER D. (2005) « Les adverbiaux praxéologiques détachés en position initiale et leur portée », *Verbum*, XXVII, 3, Nancy, P. U..

ZHEKOVA D. & KÜBLER S. (2013) « Machine Learning for Mention Head Detection in Multilingual Coreference Resolution », in *Proceedings of Recent Advances in Natural Language Processing (RANLP-2013)*, Hissar, Bulgaria 747-754.