



HAL
open science

En quoi le pronom “ on ” a-t-il une valeur anaphorique ? Le cas des successions d’occurrences de “ on ”

Marine Delaborde, Frédéric Landragin

► To cite this version:

Marine Delaborde, Frédéric Landragin. En quoi le pronom “ on ” a-t-il une valeur anaphorique ? Le cas des successions d’occurrences de “ on ”. Les cahiers de praxématique, 2019, La gestion de l’anaphore en discours : complexités et enjeux, 72, pp.1-18. hal-02161902

HAL Id: hal-02161902

<https://hal.science/hal-02161902>

Submitted on 21 Jun 2019

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.

En quoi le pronom « on » a-t-il une valeur anaphorique ?

Le cas des successions d'occurrences de « on »

To what extent is French pronoun "on" anaphoric? A study of occurrences of "on" in succession

Marine Delaborde and Frédéric Landragin



Electronic version

URL: <http://journals.openedition.org/praxematique/5464>

ISSN: 2111-5044

Publisher

Presses universitaires de la Méditerranée

Brought to you by Ecole Normale Supérieure Paris



Electronic reference

Marine Delaborde and Frédéric Landragin, « En quoi le pronom « on » a-t-il une valeur anaphorique ? », *Cahiers de praxématique* [Online], 72 | 2019, Online since 20 June 2019, connection on 21 June 2019.

URL : <http://journals.openedition.org/praxematique/5464>

This text was automatically generated on 21 June 2019.

Tous droits réservés

En quoi le pronom « on » a-t-il une valeur anaphorique ?

Le cas des successions d'occurrences de « on »

To what extent is French pronoun “on” anaphoric? A study of occurrences of “on” in succession

Marine Delaborde and Frédéric Landragin

Introduction

- La procédure d'annotation manuelle du corpus Democrat (Landragin, 2016), et la publication du manuel d'annotation associé¹, donnent l'occasion de discuter des problèmes posés par l'annotation référentielle d'occurrences du pronom « on ». Les contraintes de la mise en œuvre de ce corpus écrit, regroupant divers textes narratifs, entraînent en effet l'attribution d'un référent (individu ou groupe d'individus) à chaque pronom « on », y compris dans les cas où l'annotateur hésite entre plusieurs possibilités, par exemple entre plusieurs groupes possibles. Lorsque plusieurs occurrences de « on » se succèdent et que ces occurrences semblent à première vue coréférentes, chaque « on » s'appuyant sur la référence du précédent, la procédure d'annotation révèle encore plus la nature des contraintes qui l'accompagnent : il est tentant de leur attribuer le même référent et de construire ainsi une seule chaîne de référence. Mais une seconde lecture du texte, plus attentive, peut soulever quantité de questions : dans quelle mesure la référence de « on » peut-elle s'appuyer sur un antécédent préalable ? « On » a-t-il une valeur anaphorique ? Comment rendre compte de l'intuition de l'annotateur quand elle ne rentre pas dans les « cases » d'un manuel d'annotation, par exemple quand plusieurs occurrences consécutives de « on » s'accompagnent d'un glissement progressif vers le générique ? Au-delà des ambiguïtés possibles et présentes également pour diverses expressions référentielles, la résolution de la référence de « on » peut-elle se caractériser par un certain flou ? Pour y voir plus clair parmi tous ces phénomènes et tenter de répondre à ces questions, nous proposons un ensemble d'analyses et de pistes de

réflexion portant sur les successions d'occurrences de « on » afin de montrer l'intérêt de traiter ces cas de manière adaptée au niveau de l'annotation en corpus.

- 2 Dans la mesure où les notions convoquées sont nombreuses et où il est parfois difficile de faire la part des choses entre flou et ambiguïté, entre anaphore et coréférence, nous commençons cet article par une partie portant sur la terminologie. Nous pouvons alors décrire de manière précise les exemples rencontrés dans le corpus *Democrat* (Partie 2), et en proposer des analyses linguistiques détaillées qui nous amènent à une schématisation de la (co)référence du pronom « on », mettant en évidence sa complexité référentielle (Partie 3). Prendre en compte cette complexité référentielle dans un schéma d'annotation à la manière de Charolles (2014) ou dans un système de traitement automatique des langues (TAL) permettrait de ne pas perdre d'information, mais aurait des conséquences sur la complexité des algorithmes et leurs résultats. Ce sont ces aspects que nous discutons rapidement dans une section portant sur le décalage entre la modélisation et la gestion de corpus (Partie 4) puis dans la discussion (Partie 5). Nous présentons alors notre principale conclusion et quelques perspectives de recherche.

1. Les notions impliquées et leurs définitions

1.1. Anaphore, coréférence et chaîne de référence

- 3 La référence est l'acte de langage consistant à désigner un objet extralinguistique qui existe dans le monde (Frege, 1892) ou que l'on peut se représenter (Charolles, 2002) par le biais d'une expression référentielle. Ces expressions peuvent être reliées syntaxiquement, sémantiquement et surtout pragmatiquement par différentes relations comme l'anaphore ou la coréférence par exemple.
- 4 Le phénomène linguistique de l'anaphore implique de se référer à un segment du discours pour pouvoir en interpréter un autre. Dans l'exemple [1], le pronom « elle » est l'anaphorique du nom propre « Marthe ». Pour désigner le segment auquel on se réfère, il est courant d'utiliser le terme d'*antécédent* car il précède généralement l'anaphorique. Toutefois, Tesnière (1959) observe que cet *antécédent* peut parfois se trouver après l'anaphorique (cataphore), c'est pourquoi il préfère utiliser l'expression de *source sémantique* qui met l'accent sur le rôle de ce segment plutôt que sur sa localisation. Ducrot & Todorov (1972 : 363) parlent quant à eux d'*interprétant* de l'anaphorique mais il est aussi parfois appelé *contrôleur de l'anaphorique*, (co)référent ou encore *référé* (Reichler-Béguelin, 1988). L'anaphore est donc une relation asymétrique et non autonome (Milner, 1976 : 65) dans la mesure où il est impossible d'interpréter l'anaphorique sans se reporter à l'antécédent. Il est aussi important de noter que si un anaphorique a besoin de son antécédent pour être interprété, ils ne désignent pas obligatoirement le même référent (Kleiber, 1988). C'est le cas notamment des anaphores associatives dans lesquelles l'anaphorique est interprétable grâce à l'intégration de certaines relations de contiguïté dans le discours comme les relations méronymiques, locatives, fonctionnelles et actancielles (Kleiber, 2001). C'est ce que l'on voit dans l'exemple [2], où « les gros barreaux rouillés » sont interprétables en se référant à « la grille du potager » même s'il ne s'agit pas exactement du même référent (relation partie-tout).

[1] Marthe se figura que je m'ennuyais de plus en plus. Elle se sentait prête à tout pour me distraire.

RADIGUET Raymond, *Le Diable au corps*, 1923

[2] Les nouveaux amis l'accompagnèrent jusqu'à la grille du potager par où il était venu. À pleines mains, il empoigna les gros barreaux rouillés, se hissa au faite, enjamba les lances pointues avec adresse, se laissa tomber de haut, et tout courant, s'en fut parmi les grands sapins.

AUDOUX Marguerite, *Douce Lumière*, 1937

- 5 Lorsque deux segments du discours désignent la même entité, elles sont reliées par une relation symétrique de coréférence (Corblin, 1985). Dans l'exemple [3], « une lanterne de fer » et « ce luminaire » sont coréférents parce qu'ils ont le même référent. Il s'agit aussi d'une anaphore en raison de la présence du démonstratif « ce » qui renvoie au groupe nominal précédent. Les notions de coréférence et d'anaphore sont proches et peuvent se recouper. Dans l'exemple [1], « Marthe » et « elle » sont ainsi des expressions référentielles coréférentes. Nous avons vu que les anaphores peuvent être coréférentielles ou non, et parallèlement une coréférence peut être anaphorique ou non : selon Kleiber (1988 : 3), « toute expression coréférentielle d'une expression antérieure n'est pas nécessairement une expression anaphorique ». Dans l'exemple [4], il n'y a pas besoin de la première occurrence « du soleil » pour interpréter la seconde : « le soleil ». Cette relation de coréférence est par conséquent non anaphorique.

[3] La cuisine de l'auberge n'était éclairée que par une lanterne de fer suspendue au plafond. Le squelette de ce luminaire dessinait une large étoile d'ombre tremblotante sur tout l'intérieur de la pièce, et rejetait sa pâle clarté vers les solives enfumées du plafond.

SAND Georges, Pauline, 1881

[4] La terre tourne autour du soleil. Le soleil est en effet le centre de l'univers. (Kleiber, 1988 : 3)

- 6 La coréférence est donc la relation entre plusieurs expressions référentielles qui désignent le même référent. Elles forment ainsi une *chaîne de référence* (Chastain, 1975), aussi parfois appelée *chaîne de coréférence(s)*, dont les « maillons » sont les mentions coréférentes. Ces mentions peuvent être anaphoriques ou non. Le nombre minimum de maillons est un sujet qui divise. Pour certains auteurs (Corblin, 1995 ; Schnedecker 1992), une chaîne de référence comporte au moins 3 maillons, les notions de coréférence et d'anaphore étant utilisées pour décrire les relations entre seulement deux expressions référentielles. De manière générale en traitement automatique des langues, et comme pour Desoyer *et al.* (2014) par exemple, une chaîne de référence à deux maillons est tout à fait acceptable, pour ne pas dire essentielle : dans le corpus ANCOR (Muzerelle *et al.*, 2013), les chaînes comportant deux maillons sont plus nombreuses que les chaînes à trois maillons, elles-mêmes plus nombreuses que les chaînes à quatre maillons et ainsi de suite. Dans l'exemple [5], « Noël », « il », « il » et « Noël » forment une chaîne de référence qui a pour référent le personnage de Noël.

[5] Elle pensait à Noël. S'il était là, il saurait bien la défendre. Mais la ferme des Barray était peu éloignée du village, et Noël n'avait rien à faire sur la route qui conduisait au Verger, distant de plus d'un kilomètre.

AUDOUX Marguerite, *Douce Lumière*, 1937

[6] Mes frères plaisaient ma petite sœur. Les roues de sa bicyclette ont à peine quarante centimètres de diamètre : « On te laissera seule sur la route. ».

RADIGUET Raymond, *Le Diable au corps*, 1923

- 7 Dans l'exemple [6], « Mes frères » et « on » forment aussi une chaîne de référence bien que l'utilisation de ce pronom soit particulièrement propice à une interprétation incertaine du référent.

1.2. L'incertitude interprétative

- 8 La notion de référence implique l'identification d'un référent (Charolles, 2002). Cependant, il arrive parfois qu'un référent soit difficile à identifier pour différentes raisons. Ainsi, une expression linguistique est ambiguë s'il y a un choix à opérer entre plusieurs significations distinctes (Fuchs, 1996). L'ambiguïté peut se trouver au niveau lexical comme avec les cas d'homonymie ou de polysémie, mais elle peut aussi être syntaxique lorsque plusieurs analyses distinctes sont possibles pour une même unité. Selon Fuchs (1996 : 10), il s'agit d'un phénomène « prédictible en langue », c'est-à-dire qu'il est possible de décrire une ambiguïté en fonction des entrées d'un dictionnaire ou d'une grammaire. Il y a donc ambiguïté lorsque plusieurs sens qui s'excluent mutuellement sont associés à une seule forme linguistique. Dans le langage courant, la notion d'ambiguïté est souvent associée à des énoncés flous. Or, le flou relève plutôt de la sous-détermination : lorsque le sens d'une expression linguistique est univoque mais que son interprétation reste ouverte à cause d'un « manque de précision référentielle » (Fuchs, 1996). Mais on peut aussi se demander si le flou ne peut pas parfois relever de la sur-détermination : lorsqu'une expression linguistique est associée à plusieurs significations juxtaposées les unes aux autres, comme dans certains jeux de mots pour lesquels le sens du mot est à interpréter avec toute sa polysémie. Qu'il s'agisse de sous-détermination ou de sur-détermination, le flou n'est pas de l'ambiguïté car si plusieurs sens peuvent être associés à une unique forme, aucun choix n'est imposé.
- 9 Dans le cas d'une anaphore, l'ambiguïté réside dans le choix de l'antécédent de l'élément anaphorique. Dans l'exemple [7], le choix du référent du pronom « il » est à faire entre les noms propres « Jean » et « Paul ». Il existe différentes stratégies de résolution d'une anaphore (Kail & Léveillé, 1977) : choisir le syntagme nominal (SN) qui possède les mêmes marques lexicales que le pronom (genre et nombre), choisir le SN qui a les mêmes relations fonctionnelles ou encore choisir le SN le plus proche du pronom. Dans le cas de l'exemple [7], la première stratégie n'est pas pertinente car les marques lexicales sont les mêmes pour « Pierre » et « Paul » et la seconde stratégie est en contradiction avec la troisième, il s'agit donc d'une anaphore ambiguë. Dans l'exemple [8], l'anaphore relève plutôt de la sur-détermination car le pronom « elle » peut référer au pronom « lui » (s'il est féminin) mais aussi à « son idée » et finalement aux deux en même temps parce que l'idée et la personne qui la produit sont des notions fortement reliées.

[7] *Jean est en colère contre Paul. Il ne l'accompagne pas à la fête.*

[8] *Je lui ai dit sur un ton de plaisanterie que son idée était intéressante, qu'elle montrait les choses sous un angle auquel en effet on n'est pas habitué.*

ROMAINS Jules, *Les Hommes de bonne volonté*, 1939; exemple discuté dans (Landragin, 2007).

- 10 Avant de montrer comment des analyses de « on » font intervenir ces notions, définissons ce pronom en nous appuyant sur les grammaires du français.

1.3. « On » : un pronom impersonnel humain ? Objet d'étude et problématisation

- 11 Le pronom « on » est difficile à catégoriser (Landragin & Tanguy, 2014 : 99) : il peut être considéré comme un *pronom personnel* (Charaudeau, 1992), un *pronom indéfini* (Sandfeld, 1965 ; Grevisse & Goosse, 2002), un *pronom personnel indéfini* (TLFi) ou encore un *pronom impersonnel humain* (Cabredo Hofherr, 2008 : 35). Selon les grammaires du français (Riegel *et al.*, 1994), « on » est toujours nominal, en position sujet, et désigne des référents humains animés – avec des exceptions qui sont répertoriées – singuliers ou pluriels, masculins et/ou féminins, il est donc épïcène. Mais souvent, les référents de ce pronom sont particulièrement difficiles à identifier de manière précise. On ne sait pas toujours s'il est inclusif ou exclusif : l'énonciateur est-il compris dans le référent ? Il est parfois aussi compliqué de savoir si un « on » est générique ou spécifique : le référent représente-t-il l'ensemble d'une catégorie ou un humain bien identifié ? Ce pronom peut même être dans une position intermédiaire comme dans l'exemple [9] qui suit.

[9] On a besoin de travailler. (interlocuteurs dans une queue devant Pôle Emploi).

- 12 Comme le souligne Gjesdal (2008 :15), les analyses classiques du pronom « on » lui accordent deux valeurs discursives principales qui le rendent sémantiquement « complexe » : le personnel et l'indéfini. L'emploi personnel de ce pronom correspond particulièrement au langage familier. Dans ce cas, « on » concorde avec tous les autres pronoms avec lesquels il peut être substitué, particulièrement les pronoms « je » et « nous ». L'emploi indéfini du pronom « on » correspond sémantiquement à une ou des « personne(s) sans référence spécifique », substituable à des pronoms indéfinis (ex. « quelqu'un »). Il correspond syntaxiquement à des « constructions impersonnelles et passives ». Muller (1970 : 52-54) parlait aussi d'une troisième « valeur stylistique » pour laquelle l'utilisation de « on » a « toujours une intention affective » comme l'ironie, l'affection ou la tendresse, entre autres. D'un point de vue discursif, « on » peut avoir différentes fonctions comme la distanciation du référent ou encore le maintien du flou par l'occultation de certaines informations.
- 13 Le flou référentiel a des conséquences sur les successions de « on » et leur interprétation coréférentielle (ou non). Cela contribue à faire potentiellement dépendre l'interprétation du pronom « on » de celle d'un antécédent antérieur. Mais les avis divergent sur le caractère anaphorique de « on ». Certains pensent que ce dernier n'existe pas : selon Atlani (1984), « on » n'a aucune valeur anaphorique, c'est-à-dire qu'il va réactualiser sa référence à chaque occurrence. Pour Bouguerra (1999), il y a une « vacance référentielle originelle » pour « on » : c'est une « forme vide » que « l'allocutaire compétent » doit remplir. Dans les grammaires du français, il est question de référence anaphorique pour tous les pronoms. Et plusieurs auteurs (Fløttum *et al.*, 2007 ; Gjesdal, 2008) parlent de complexité référentielle pour « on », avec des possibilités de sur-détermination plutôt que de sous-détermination. Dans la mesure où elle met en avant la complexité référentielle de « on », nous nous plaçons dans la lignée théorique de ces derniers auteurs. Nous suivons l'hypothèse que le pronom « on » s'appuie partiellement sur la référence d'un antécédent, et qu'il peut donc avoir une valeur anaphorique et non-autonome. Dans l'exemple [10], il y a un facteur anaphorique qui intervient dans

l'interprétation du pronom « on » car il renvoie à l'antécédent « mon ami et moi ». Selon Milner (1982 : 33), un pronom ne peut pas « fonctionner comme le premier terme d'une relation d'anaphore » et un terme ne peut pas être l'anaphorique de lui-même. Deux « on » consécutifs ne constituent donc pas une anaphore mais une coréférence.

[10] Mon ami et moi étions arrivés à 10 heures. On est repartis à 11 heures.

- 14 Le pronom « on » peut être anaphorique, et peut par conséquent faire partie d'un ensemble d'expressions référentielles reliées par des anaphores coréférentielles. En linguistique de corpus outillée et en traitement automatique des langues, il y a coréférence lorsque les référents sont strictement identiques, et il y a construction d'une chaîne uniquement lorsque toutes les relations sont strictes : c'est le cas par exemple dans le projet ANR Democrat. Mais avec le pronom « on », il est souvent difficile d'établir un lien de coréférence stricte, et on doit faire face à un dilemme : ignorer ces occurrences ou les rendre strictes.
- 15 Il est possible qu'un lecteur peu attentif considère comme coréférentielles deux occurrences successives de « on » sémantiquement proches (cf. les deux dernières occurrences de l'exemple [11]). Cependant, une lecture plus fine peut amener à introduire de l'incertitude dans la référence – de l'une ou l'autre occurrence, voire toutes – et à casser ainsi la relation coréférentielle. En effet, les deux premiers « on » de l'exemple [11] font référence à un groupe bien identifié qui comporte deux autres référents : le narrateur et sa sœur. La coréférence est stricte, notamment en raison du phénomène de répétition de la séquence « on l'arrête ». Le troisième « on » relève du discours rapporté et ne coréfère pas de manière stricte au groupe précédent comprenant le narrateur et sa sœur. En effet, il peut aussi avoir une interprétation générique bien que l'adverbe déictique « ici » renforce la référence au groupe qui inclut le narrateur en ancrant la référence dans la narration. Le dernier « on » ne coréfère pas non plus de manière stricte au premier groupe car il peut aussi avoir une interprétation générique.

[11] Elle parle aussi avec une sentimentalité criante. Ma sœur et moi on l'arrête. On l'arrête à temps. Alors elle dit on ne me laisse pas parler ici. Mais ce ne sont pas des paroles qu'on a envie d'entendre, je ne sais pas pourquoi.
AKERMAN Chantal, *Ma mère rit*, 2013

2. Analyse en corpus des successions d'occurrences de « on »

Dans notre corpus, constitué de quatre textes issus du corpus Democrat, nous nous sommes intéressés à la place du pronom « on » par rapport aux autres pronoms avec lesquels il peut être substitué. Les occurrences des pronoms personnels (4034 au total) sont répertoriées dans le Tableau 1. Les pronoms indéfinis sont quant à eux moins nombreux (365 au total) : « tout » (198), « quelque » (66), « aucun » (28), « chaque » (26), « chacun » (16), « plusieurs » (15), « quelqu'un » (9) et « certain » (7).

Tableau 1 : Les pronoms personnels dans le sous-corpus Democrat.

Pronoms personnels (lemmes)	Nombre d'occurrences	Pronoms personnels (lemmes)	Nombre d'occurrences

elle	781	vous	82
je	640	moi	82
se	565	tu	70
il	507	te	59
me	322	toi	21
lui	250	eux	18
la le	178	elle-même	16
on	133	lui-même	7
Y	127	soi	5
nous	125	moi-même	4
en	115	nous-mêmes	1
la	108	toi-même	1
le	87	SOMME	4304

Le pronom « on » apparaît à la 8^{ème} place sur les 25 lemmes de pronoms personnels identifiés dans notre corpus par treetagger (Schmid, 1994). Il apparaît moins que des pronoms de première et troisième personne du singulier mais plus que le pronom « nous ».

- 16 Lorsque le pronom « on » apparaît plusieurs fois de manière successive dans un texte, fait-il toujours référence à la même entité extralinguistique ? En suivant une approche descriptive sur corpus, fondée sur un ensemble d'occurrences de « on », nous tenterons de décrire le fonctionnement des exemples et d'identifier les paramètres intervenant lors de la résolution de la référence. Nous verrons que des expressions référentielles peuvent désigner le même référent de manière stricte mais aussi de manière plus vague : l'un des référents peut avoir en plus une valeur générique ou en inclure un autre. Cependant, une succession d'occurrences de « on » n'implique pas obligatoirement de relation de coréférence.

2.1. Même référent ?

- 17 Il arrive que le référent désigné par le pronom « on » ne soit pas explicité précédemment par un groupe nominal par exemple. Dans ce cas, une succession de « on » coréférents sera identifiée comme une chaîne de référence et non comme une chaîne anaphorique. C'est le cas dans les deux exemples [12] et [13]. Dans l'exemple [12], les paragraphes précédents permettent d'identifier le maire, la mère D... (l'aveugle), Pauline et Laurence (l'actrice) comme étant les protagonistes du dîner qui sont aussi les référents des pronoms « on » de la dernière phrase. Les deux « on » de cette dernière phrase coréférent de manière stricte. Le deuxième « on » fait partie d'une proposition coordonnée à celle qui contient le premier « on » et ils sont tous les deux les agents de verbes qui décrivent

l'enchaînement d'actions de fin de repas. Dans l'exemple [13], le contexte permet aussi d'identifier un référent aux pronoms « on » de manière non explicite. Avec ce pronom, Thierry Jeanroy s'inclut de manière claire dans le groupe auquel les « on » font référence, mais il désigne plus largement un groupe flou qui comprendrait à la fois les employés du bar mais aussi les commerçants les plus volontaires du Ballon d'Alsace sans en connaître précisément la proportion. Si ce groupe peut paraître un peu vague, les deux « on » coréférent de manière stricte entre eux car le référent désigné est le même : l'action représentée par le verbe de la première phrase (bricoler) est la conséquence du fait d'avoir fait une croix sur la neige.

[12] Elle fut si aimable et si jolie dans ce badinage, que le bon maire en tomba amoureux comme un fou, voulut lui baiser la main, et ne se retira que lorsque **madame D... et Pauline lui eurent promis de le faire dîner chez elles ce même jour avec la belle actrice de la capitale**. Le dîner fut fort gai. Laurence essaya de se débarrasser des impressions tristes qu'elle avait reçues, et voulut récompenser l'aveugle du sacrifice qu'elle lui faisait de ses préjugés en lui donnant quelques heures d'enjouement. Elle raconta mille historiettes plaisantes sur ses voyages en province, et même, au dessert, elle consentit à réciter à M. le maire des tirades de vers classiques qui le jetèrent dans un délire d'enthousiasme dont madame la mairesse eût été sans doute fort effrayée. Jamais l'aveugle ne s'était autant amusée ; Pauline était singulièrement agitée ; elle s'étonnait de se sentir triste au milieu de sa joie. Laurence, tout en voulant divertir les autres, avait fini par se divertir elle-même. Elle se croyait rajeunie de dix ans en se retrouvant dans ce monde de ses souvenirs, où elle croyait parfois être encore en rêve.

On était passé de la salle à manger au salon, et on achevait de prendre le café, lorsqu'un bruit de socques dans l'escalier annonça l'approche d'une visite.

SAND Georges, *Pauline*, 1881

[13] Pourtant, mardi, il y avait bien quelques randonneurs, courageux ou inconscients, à passer devant la ferme-auberge du Ballon, les visages figés par la pluie cinglante et les jeans gorgés d'eau.

Il manque juste le beau temps. Le soleil, c'est l'or du Ballon » résume Thierry Jeanroy, dont le bar des Démineurs est installé à cheval entre Vosges et Territoire de Belfort.

« On bricole. La saison est vraiment médiocre. La neige, on fait une croix dessus mais il faudrait tout de même que le temps devienne sec.

L'Est Républicain, 02/01/2013

2.2. Générique, spécifique, un peu entre les deux ?

- 18 Le pronom « on » peut permettre au locuteur de prendre de la distance avec un énoncé et de faire des descriptions ou des généralités. Dans ce cas, lorsqu'il y a une succession de « on », ils peuvent très bien coréférent de manière stricte comme dans l'exemple [14] et garder une valeur seulement générique, mais il peut aussi y avoir un référent particulier désigné en plus.

[14] Je ne sais quelle timidité, si ce n'est celle que l'on éprouve en face de ce qu'on n'a jamais fait, me retenait.

RADIGUET Raymond, *Le Diable au corps*, 1923

- 19 Le premier « on » de l'exemple [15] est générique. Le second est aussi générique mais il ne coréférent pas avec le premier « on ». En revanche, il inclut le narrateur : la phrase est du discours direct où le personnage se parle à lui-même. Dans l'exemple [16], il y a une

chaîne de référence composée de quatre « on ». Ils ont une valeur générique, ce qui permet une distanciation par rapport au référent, mais le fait d'utiliser le féminin pour l'adjectif « seule » inclut le personnage dans la référence en plus de la valeur générique. On retrouve ce phénomène dans l'exemple [17], où les deux premiers « on » coréfèrent de manière stricte et désignent « l'aveugle » tout en gardant une valeur générique, comme si cela pouvait toucher tout le monde au réveil. Le troisième « on » ne coréfère pas avec les autres bien qu'il ait une valeur générique car il permet d'éviter une tournure passive. Mais il inclut peut-être aussi le personnage de l'aveugle car c'est elle qui se demande comment l'appeler.

[15] Lorsque quelque chose, venu de l'extérieur, m'obligeait à penser moins paresseusement à Marthe, j'y pensais sans amour, avec la mélancolie que l'on éprouve pour ce qui aurait pu être. « Bah ! me disais-je, c'eût été trop beau. On ne peut à la fois choisir le lit et coucher dedans.

RADIGUET Raymond, *Le Diable au corps*, 1923

[16] J'en étais victime ; mais je rougissais de les raconter. Quand on est si loin de toute amitié, si seule, si triste, toute démarche difficile devient impossible. On s'observe, on se craint soi-même, et l'on se suicide dans la peur de se laisser mourir.

SAND Georges, *Pauline*, 1881

[17] Allons, reprit l'aveugle, qui craignait instinctivement de déplaire à sa fille, en raison du besoin qu'elle avait de son dévouement, laissez-moi le temps de me remettre un peu ; je suis si surprise ! et comme cela, au réveil, on ne sait trop ce qu'on dit... Je ne voudrais pas vous faire de chagrin, mademoiselle... ou madame... Comment vous appelle-t-on maintenant !

SAND Georges, *Pauline*, 1881

- 20 Le pronom « on » est parfois utilisé pour décrire des situations ou des paysages, il est alors souvent suivi par le verbe « voir ». C'est le cas dans les exemples [18] et [19]. Il y a une répétition de « On la voyait » dans l'exemple [18] qui appuie la coréférence des deux premiers « on ». Ils ont une valeur générique en raison de la description mais ils sont fortement reliés aux deux personnages présents dans la pièce. Le troisième « on » est du discours direct et inclut le personnage de Noël ainsi que le personnage féminin, mais il a aussi une valeur descriptive. Il y a une longue chaîne de référence dans l'exemple [19] qui désigne Roxane Herbrecht et Nicolas Burcey et qui est principalement composée de plusieurs « nous ». Le pronom « on » permet de prendre un peu de distance en revêtant un rôle descriptif plus générique tout en référant aux deux personnages.

[18] Elle souleva le rideau et fut tout étonnée de découvrir la sapinière si proche. On la voyait bien, la sapinière. On la voyait sur une grande étendue, et, tout là-bas, les arbres, par places, se mettaient en rang comme les petites filles à l'école. La voix joyeuse de Noël éclata soudain :

— Ah ! enfin on voit clair !

AUDOUX Marguerite, *Douce Lumière*, 1937

2.3. Inclusion d'un référent dans un autre

- 21 L'exemple [19] contient une chaîne de référence qui désigne « les méchants », mais s'agit-il des « filles et garçons regagnant leur demeure » au complet ou seulement de certains ? On a ici aussi une chaîne anaphorique floue. Dans l'exemple [20], « Jean-Louis Fretti » parle des veillées trappeurs organisées pour les touristes par les accompagnateurs en

montagne du Ballon d'Alsace. Le premier « on » fait partie d'un discours narratif produit par le journaliste et a une valeur générique bien qu'il désigne aussi les accompagnateurs et les touristes du Ballon. Le premier « nous » réfère aux organisateurs mais les deux « nous » qui suivent incluent en plus les touristes. Le « on » qui suit permet d'indiquer qu'il s'agit d'une cabane où il faut s'éclairer à la bougie mais réfère aussi aux personnes qui s'y trouvent (les accompagnateurs et les touristes). Avec « les gens », Jean-Louis Fretti s'exclut de la référence mais le « on » qui suit pourrait bien l'inclure. Cette chaîne de référence floue désigne un groupe large, composé des accompagnateurs et des touristes et chaque maillon de la chaîne met l'accent sur une partie de ce groupe ou y ajoute une information. Dans l'exemple [21], les deux premiers « on » sont les anaphoriques de l'antécédent « toutes ces femmes », les deux derniers réfèrent à ce groupe de femmes moins les plus déçues.

[19] La classe finie, d'autres tourments l'attendaient sur la route qu'elle suivait en compagnie de filles et garçons regagnant leur demeure. Toutes les malices étaient bonnes à faire à cette gnangnan qui ne se défendait pas et ne se méfiait jamais. On la poussait brusquement dans un fossé vaseux, ou dans un buisson plein d'épines d'où elle sortait salie et déchirée. Quand vint la neige, elle fut toute désignée pour recevoir les boules, qu'on lui jetait de préférence au visage. Elle pensait à Noël. S'il était là, il saurait bien la défendre. Mais la ferme des Barray était peu éloignée du village, et Noël n'avait rien à faire sur la route qui conduisait au Verger, distant de plus d'un kilomètre. Il y avait bien Marguerite Dupré, une grande qui prenait parfois sa défense, mais alors c'était elle qu'on attaquait, Marguerite Dupré, dont la maison n'était pas très éloignée de celle de la petite, prenait, en même temps qu'elle, le même sentier. Mais, arrivée là, Douce ne craignait plus rien, elle courait plus vite qu'une oie et avançait facilement les méchants.

AUDOUX Marguerite, *Douce Lumière*, 1937

[20] A la Gentiane, fidèle au poste dans la cabane en bois **des accompagnateurs en montagne du Ballon**, Jean-Louis Fretti lit et note sur son journal les températures et le temps de chaque jour. Le vocabulaire est malheureusement peu varié : pluie, brouillard, vent. Il confirme la présence **des touristes** : « **Les gens** ont réservé souvent plusieurs mois à l'avance pour Nouvel An. Ils sont là. Il y avait 150 personnes à Saint-Maurice, dimanche, pour le pot d'accueil des nouveaux arrivants. Le Langenberg est bien rempli. Le potentiel est là... ». Reste qu'avec un temps pareil, on peut faire un trait sur les sorties en raquette, les parcours VTT et même les sorties pédestres. « Seules les veillées trappeurs fonctionnent. **Nous** en organisons une chaque soir. **Nous** partons par petits groupes et marchons environ 1,2 km pour **nous** retrouver dans un chalet en bois où on s'éclaire à la bougie autour d'un feu de bois. **Les gens** font des grillades et on discute du massif, des légendes locales, des animaux... ».

L'Est Républicain, 02/01/2013

[21] Toutes ces femmes se tenaient d'un côté du salon comme un régiment en déroute, et de l'autre côté, entourée de Pauline, de sa mère et de quelques hommes de bon sens qui ne craignaient pas de causer respectueusement avec elle, Laurence siégeait comme une reine affable qui sourit à son peuple et le tient à distance. Les rôles étaient bien changés, et le malaise croissait d'un côté, tandis que la véritable dignité triomphait de l'autre. On n'osait plus chuchoter, on n'osait même plus regarder, si ce n'est à la dérobée. Enfin, quand le départ des plus déçues eut éclairci les rangs, on osa s'approcher, mendier une parole, un regard, toucher, demander l'adresse de la lingère, le prix des bijoux, le nom des pièces de théâtre le plus à la mode à Paris, et des billets de spectacle pour le premier voyage qu'on ferait à la

capitale.
SAND Georges, *Pauline*, 1881

2.4. Référents différents

- 22 Lorsque le même pronom apparaît successivement de manière rapprochée, il ne désigne pas nécessairement le même référent. Dans l'exemple [22], le premier « on » désigne les visiteurs et le deuxième réfère à la personne qui aurait déplacé le calendrier, ils ne sont pas coréférents. Dans l'exemple [23], le premier « on » désigne le porteur du message et le second réfère aux serviteurs. Ces deux « on » permettent l'occultation du procès de l'énonciation en omettant de nommer certains protagonistes pour mettre l'accent sur l'action.

[22] Il y a deux ans que l'on n'est venu me tenir compagnie le soir, et c'est un merveilleux hasard qui m'amène toute la ville à la fois. Est-ce qu'on aurait dérangé le calendrier, et ma fête, que je croyais passée il y a six mois, tomberait-elle aujourd'hui ?

SAND Georges, *Pauline*, 1881

[23] Une prompte libéralité de la jeune dame la décida à aller réveiller le chef, et, après un quart d'heure d'attente, qui parut mortellement long à notre voyageuse, on vint enfin lui apprendre que mademoiselle Pauline D... n'était point mariée, et qu'elle habitait toujours la ville. Aussitôt l'étrangère ordonna qu'on mît sa voiture sous la remise et qu'on lui préparât une chambre.

SAND Georges, *Pauline*, 1881

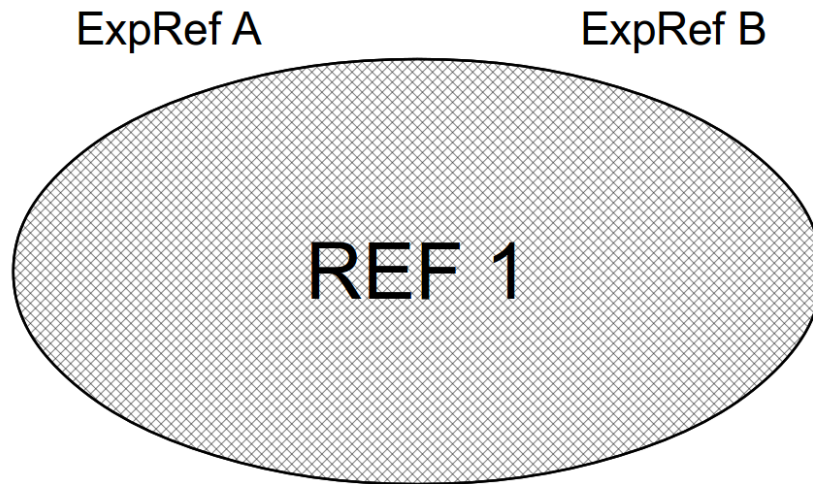
3. Modélisation de l'interprétation de « on »

- 23 Ces différentes catégories de référence des successions d'occurrences de « on » font appel à des interprétations référentielles différentes qui seront schématisées dans cette partie. Cette schématisation à la fois graphique et formelle ne correspond pas à une formalisation complète comme on pourrait le faire à l'aide d'un formalisme tel que la DRT (*Discourse Representation Theory* ou Théorie des représentations du discours, Kamp, 1981), mais plutôt à une modélisation qui permettra d'aboutir à terme à une formalisation.

3.1. Même référent

- 24 Lorsque deux expressions réfèrent à la même entité extralinguistique, elles coréfèrent de manière stricte. On peut utiliser la figure 1 pour schématiser la référence des deux « on » de l'exemple [12] : le premier « on » correspond à l'expression référentielle A (ExpRef A) et le deuxième « on » à l'expression référentielle B (ExpRef B), les deux ayant pour référent (Ref 1) les invités du dîner.

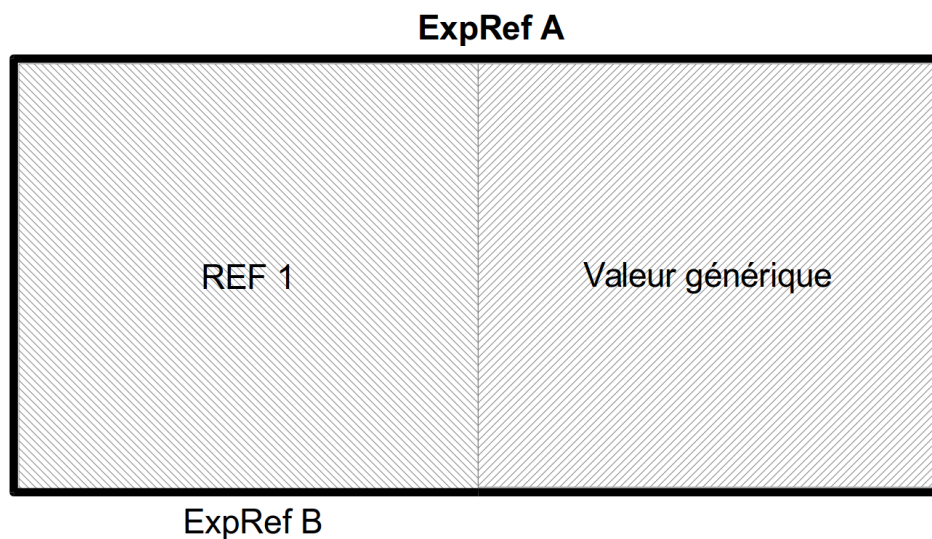
Figure 1 : Même référent



3.2. Générique, spécifique, un peu entre les deux

- 25 Pour schématiser la référence d'une expression qui a pour référent une entité précise associée à une valeur générique, on peut se reporter à la figure 2. Avec l'exemple [16], tous les « on » correspondent à l'expression référentielle A (ExpRef A) car ils réfèrent à la fois au narrateur (ExpRef B) tout en ayant une valeur générique.

Figure 2 : Entre générique et spécifique

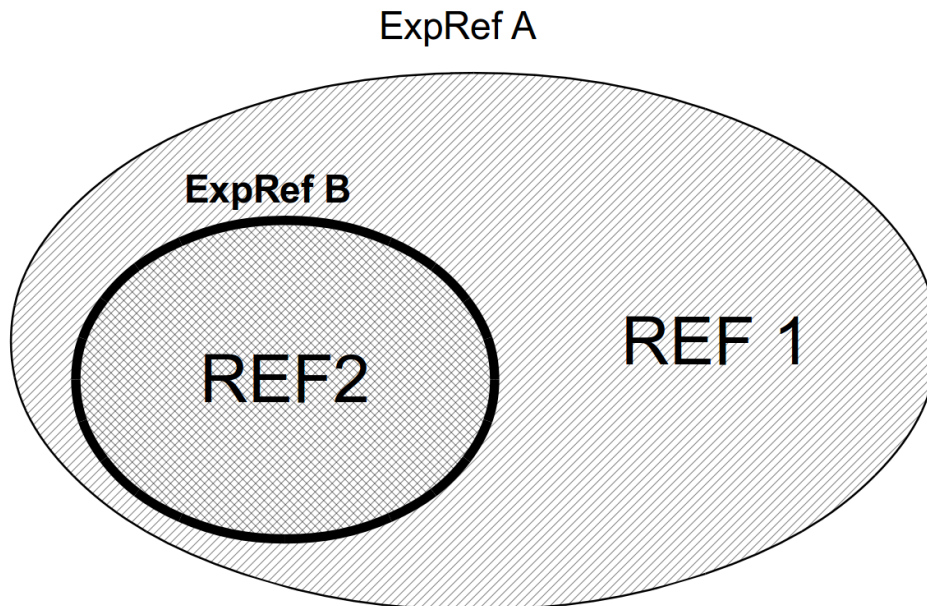


3.3. Inclusion d'un référent dans un autre

- 26 Dans l'exemple [19], les deux « on » de la chaîne de référence correspondent à l'expression référentielle B (ExpRef B) car le référent (Ref 2) est une sous-partie de

l'expression référentielle A « toutes les dames des maisons voisines » ayant pour référent le groupe au complet (Ref 1).

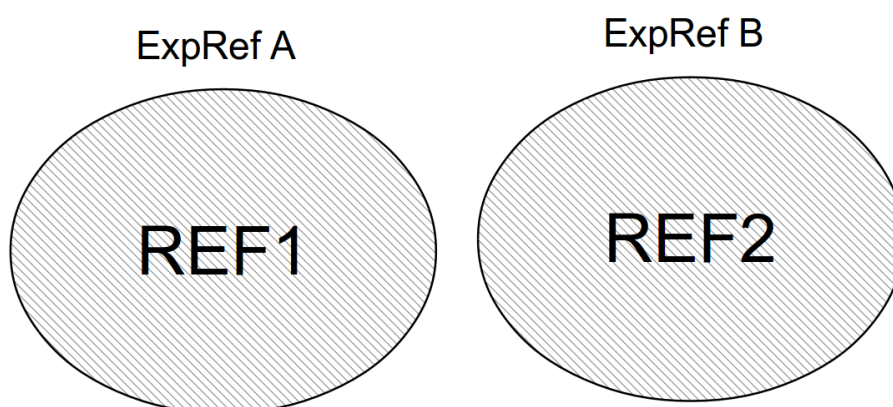
Figure 3 : Inclusion d'un référent dans un autre



3.4. Référents différents

- 27 Lorsque deux expressions ont chacune un référent différent, comme dans l'exemple [23], la schématisation est simple : le premier « on » (ExpRef A) renvoie à un référent (REF 1) différent de celui (REF 2) du deuxième « on » (ExpRef B).

Figure 4 : Référents différents



4. Décalage entre la modélisation et la gestion de corpus : le cas du projet Democrat

28 Les exemples collectés pour ce travail sont issus de quatre extraits de romans. Ces textes narratifs de 10 000 mots chacun ont aussi été annotés en chaînes de référence pour le projet ANR Democrat. Ce projet a pour but de décrire et modéliser les chaînes de référence à l'écrit en français. L'un des objectifs de ce projet est de fournir un corpus annoté en chaînes de référence qui puisse servir entre autres de corpus de référence pour les campagnes d'évaluation portant sur la coréférence. Ce corpus, constitué de textes libres de droits de genres variés et d'époques différentes, sera disponible en mars 2019. Selon le manuel d'annotation du projet Democrat, une chaîne de référence n'existe que si la coréférence est stricte entre les maillons. Dans l'exemple [22] par exemple, selon le schéma d'annotation Democrat, il y a deux chaînes de référence distinctes alors que ces deux référents sont sémantiquement très proches. Il y a donc une perte d'information due à un schéma trop strict.

5. Discussion

- 29 Lorsque la coréférence n'est pas stricte mais que les référents sont proches, il serait intéressant de pouvoir garder cette information au niveau de l'annotation. En effet, il est dommage de perdre des informations en séparant des chaînes qui peuvent être reliées, mais il ne faut pas tomber dans l'écueil inverse qui serait de faire une seule chaîne contenant des expressions référentielles qui ne sont pas coréférentes. Alors comment tenir compte de l'incertitude interprétative du pronom « on » ?
- 30 Pour garder l'information de la proximité sémantique de deux référents lorsque la coréférence n'est pas stricte, il est possible d'enrichir le schéma d'annotation déjà existant pour le projet Democrat. L'enrichissement ne se situe pas au niveau du référent mais au niveau des relations entre les maillons, afin de préciser si la coréférence est stricte ou non. Cette précision peut être implémentée par le biais d'un indice du degré de coréférence. La relation entre deux maillons désignant le même référent aurait une propriété « coréférence » dont la valeur serait « stricte ». La relation entre deux maillons désignant le même référent, mais dont l'un des deux aurait une valeur générique associée en plus à ce référent, aurait une propriété « coréférence » dont la valeur serait « spécifique + générique ». On pourrait en outre préciser lequel des maillons possède en plus cette valeur générique avec une propriété « généricité » dont la valeur serait égale à l'identifiant du maillon dont l'interprétation est plus générique. La relation entre deux maillons dont l'un inclut l'autre aurait une propriété « coréférence » dont la valeur serait « inclusive ». On pourrait aussi préciser quel maillon inclut l'autre avec une propriété « référent complet » dont la valeur serait égale à l'identifiant du maillon qui inclut l'autre.
- 31 Annoter des chaînes de références avec de la coréférence qui n'est pas stricte entraîne des conséquences lors de l'utilisation du corpus annoté de la sorte. Le premier problème à envisager concerne l'accord inter-annotateurs. En effet, lorsqu'il y a de l'incertitude interprétative pour la référence d'une expression, il y a plus de chances pour que les annotateurs ne tombent pas d'accord sur le référent à choisir. Si la tâche paraît

compliquée pour un humain, elle risque de l'être encore plus pour une machine. Pour les systèmes de traitement automatique des langues par exemple, la tâche de détection automatique des chaînes de références incluant de la coréférence floue ajoute un niveau de complexité.

Conclusion et perspectives

- 32 Nous considérons donc que « on » a une valeur référentielle complexe dont les approches orientées corpus et traitement automatique ne rendent pas pleinement compte. Nous montrons tout l'intérêt d'introduire la notion d'incertitude interprétative dans les analyses de la référence de « on », notamment quand plusieurs occurrences se suivent et se répondent l'une l'autre. En français, il serait intéressant de comparer « on » aux autres pronoms avec lesquels il peut être substitué afin de pouvoir mieux mesurer son degré d'incertitude référentielle. Un même travail pourrait être effectué avec une analyse contrastive des équivalents de « on » dans d'autres langues, éventuellement avec un corpus aligné.
- 33 Nous proposons un premier modèle de la référence de « on », qu'il serait bon de concrétiser sous la forme d'un schéma d'annotation adapté. Il est intéressant de proposer enfin des perspectives de recherche et de développement, concernant notamment les conséquences de l'introduction d'un tel schéma d'annotation. D'une part, la mise au point de critères discriminant clairement nos quatre catégories et d'autre part, la révision des méthodes de calcul de l'accord inter-annotateurs et panorama des difficultés prévisibles quant à la prise en compte des différentes valeurs référentielles de « on » en traitement automatique des langues.

BIBLIOGRAPHY

- ATLANI F., 1984, « On l'illusionniste », in *La langue au ras du texte*, A. Gresillon et J.-L. Lebrave (éds), Lille, Presses Universitaire de Lille, 13-29.
- BOUGUERRA T., 1999, « L'autre je(u) du on », in J. Bres, R. Delamotte-Legrand, F. Madray-Lesigne, P. Siblot (éds.), *L'autre en discours*, Montpellier, Publications de l'Université de Paul Valéry, Montpellier III.
- CABREDO HOFFHER P., 2008, « Les pronoms impersonnels humains : syntaxe et interprétation », *Modèles linguistiques*, 57.
- CHARAUDEAU P., 1992, *Grammaire du sens et de l'expression*, Paris, Hachette.
- CHAROLLES M., 2014, « Annotation des expressions référentielles et profondeur de traitement », in M. Fossard & M.-J. Beguelin (eds.), *Nouvelles perspectives sur l'anaphore : points de vue linguistique, psycholinguistique et acquisitionnel*, Bern, New York, Peter Lang, 55-98.
- CHAROLLES M., 2002, *La référence et les expressions référentielles en français*, Ophrys.

- CHASTAIN C., 1975, « Reference and context », in K. Gunderson (ed.), *Language, Mind and Knowledge*, Minneapolis, 194-269.
- CORBLIN F., 1995, *Les formes de reprise dans le discours. Anaphores et chaînes de référence*, Rennes, Presses Universitaires de Rennes.
- CORBLIN F., 1985, « Remarques sur la notion d'anaphore », *Revue québécoise de linguistique* 15(1), 173-195.
- DESOYER A., LANDRAGIN F., TELLIER I., LEFEUVRE A., & ANTOINE J.-Y., 2014, « Les coréférences à l'oral : une expérience d'apprentissage automatique sur le corpus ANCOR », *Traitement Automatique des Langues* 55(2), 97-121.
- DUCROT O., & TODOROV T., 1972, *Dictionnaire encyclopédique des sciences du langage*, Paris, Seuil.
- FLØTTUM K., JONASSON K., & NOREN C., 2007, *On : pronom à facettes*, Paris, De Boeck/Duculot.
- FREGE G., 1892, *Über Sinn und Bedeutung*, Leipzig, Pfeffer.
- FUCHS C., 1996, *Les ambiguïtés du français*, Paris-Gap, Ophrys.
- GJESDAL A. M., 2008, *Étude sémantique du pronom ON dans une perspective textuelle et contextuelle*, Thèse de Doctorat, Université de Bergen.
- GREVISSE M. & GOOSSE A., 2002, *Le Bon usage*, Paris, De Boeck/Duculot.
- KAIL M. & LÉVEILLÉ M., 1977, "Compréhension de la coréférence des pronoms personnels chez l'enfant et l'adulte", *L'année psychologique*, 77(1), 79-94.
- KAMP H., 1981 (rééd. 2002), "A theory of truth and semantic representation" in P. Portner & B. H. Partee (eds), *Formal semantics The essential readings*, Hoboken N.J., Wiley-Blackwell, 189-222.
- KLEIBER G., 2001, *L'anaphore associative*, Paris, Puf.
- KLEIBER G., 1988, « Peut-on définir une catégorie générale de l'anaphore ? », *Vox Romanica* 47, 1.
- LANDRAGIN F., 2016, « Description, modélisation et détection automatique des chaînes de référence (DEMOCRAT) », *Bulletin de l'AFIA* 92, 11-15.
- LANDRAGIN F., 2007, « L'anaphore à antécédent flou : une caractérisation et ses conséquences sur l'annotation des relations anaphoriques », *Journée d'étude de l'Association pour le Traitement Automatique des Langues (ATALA) sur la résolution des anaphores*, Paris.
- LANDRAGIN F. & TANGUY N., 2014, « Référence et coréférence du pronom indéfini on », *Langages*, 195(3), 99-115.
- MILNER J. C., 1982, *Ordres et raisons de langue*, Paris, Seuil.
- MILNER J. C., 1976, « Réflexions sur la référence », *Langue Française* 30, 63-73.
- MULLER C., 1970, « Sur les emplois personnels de l'indéfini on », *Revue de linguistique romane* 34, 48-55.
- MUZERELLE J., LEFEUVRE A., ANTOINE J.-Y., SCHANG E., MAUREL D., VILLANEAU J., & ESHKOL I., 2013 « ANCOR, premier corpus de français parlé d'envergure annoté en coréférence et distribué librement. », in ATALA (éd.), *TALN'2013, 20e conférence sur le Traitement Automatique des Langues Naturelles*, Les Sables d'Olonne, 555-563.
- REICHLER-BÉGUELIN M.-J., 1988, « Anaphore, cataphore et mémoire discursive », *Pratiques* 57(1), 15-43.

- RIEGEL M., PELLAT J.-C. & RIOUL R., 1994, *Grammaire méthodique du français*, Paris, Presses Universitaires de France.
- SANDBELD K., 1965, *Syntaxe du français contemporain : Les pronoms (Tome 1)*, Paris, Champion.
- SCHMID H., 1994, « Probabilistic Part-of-Speech Tagging Using Decision Trees », *Proceedings of International Conference on New Methods in Language Processing*, Manchester, UK.
- SCHNEDECKER C., 1992, *Référence et discours : chaînes de référence et redénomination (essai sur l'emploi en seconde mention du nom propre)*, Thèse de Doctorat, Strasbourg II.
- TESNIERE L., 1959, *Eléments de syntaxe structurale*, Paris, Klincksieck.
- TLFi (*Trésor de la langue française informatisé*), ATILF - CNRS & Université de Lorraine, [En ligne] consulté le 12 novembre 2018. URL : <http://www.atilf.fr/tlfi>

NOTES

1. <http://www.lattice.cnrs.fr/democrat/>
-

ABSTRACTS

When several occurrences of the French pronoun “on” succeed one another and seem to be coreferential at first sight, it is tempting to attribute them the same referent and, therefore, to construct a single coreference chain. A second and more careful reading may however shed light on the fact that coreference is not always strict. Following other annotation projects, the annotation manual of the ANR Democrat project recommends creating distinct chains. With the help of corpus examples, we show the benefits of an intermediate solution.

Lorsque plusieurs occurrences de « on » se succèdent et qu'elles semblent à première vue coréférentes, il est tentant de leur attribuer le même référent et de construire ainsi une seule chaîne de référence. Mais une seconde lecture plus attentive peut mettre en évidence le fait que la coréférence n'est pas toujours stricte. A l'instar d'autres projets d'annotation manuelle, le manuel d'annotation du projet ANR Democrat préconise dans ce cas de créer plusieurs chaînes de référence. En nous fondant sur des exemples de corpus, nous montrons tout l'intérêt d'une solution intermédiaire.

INDEX

Mots-clés: référence, anaphore, coréférence, flou référentiel, ambiguïté, pronom impersonnel humain

Keywords: reference, anaphora, coreference, fuzzy reference, ambiguity, human impersonal pronoun

AUTHORS

MARINE DELABORDE

Lattice, CNRS, ENS, Université de Paris 3, PSL Research University, USPC

FRÉDÉRIC LANDRAGIN

Lattice, CNRS, ENS, Université de Paris 3, PSL Research University, USPC