



Revue de linguistique, psycholinguistique et informatique

<http://journals.openedition.org/discours/>

Types de chaînes de référence dans les articles de recherche de format IMRAD

Bruno Oberle

LILPA (Linguistique, langues, parole), EA1339

.....
Bruno Oberle, « Types de chaînes de référence dans les articles de recherche de format IMRAD », *Discours* [En ligne], 25 | 2019, mis en ligne le 30 décembre 2019.

.....
URL : <http://journals.openedition.org/discours/10487>

.....
Titre du numéro : *Varia*

Coordination : Laure Sarda & Denis Vigier

Date de réception de l'article : 16/06/2019

Date d'acceptation de l'article : 07/11/2019



Presses
universitaires
de Caen 

Types de chaînes de référence dans les articles de recherche de format IMRAD

Bruno Oberle

LILPA (Linguistique, langues, parole), EA1339

.....

La plupart des articles de recherche en sciences expérimentales présentent un format standardisé, dit « IMRAD » (Introduction, méthodes, résultats et discussion). Chacune de ces parties remplit une fonction spécifique (présenter le cadre de la recherche, décrire les méthodes, rapporter les observations, et les discuter) qui est marquée par une variation de certains phénomènes linguistiques. Notre recherche a pour objectif d'étudier le fonctionnement des chaînes de référence (l'ensemble des expressions linguistiques qui renvoient à la même entité extralinguistique) dans un corpus d'articles IMRAD écrits en français (5 textes, 32 539 mots). Notre analyse fait ressortir l'hétérogénéité des chaînes, tant par le type de leur référent que par la répartition de leurs occurrences dans le texte. Nous montrons que des chaînes de différents types de référents (notamment l'auteur de l'article, les référents abstraits et les groupes d'individus) n'ont pas les mêmes propriétés et ne peuvent pas être décrites de la même façon. De même, nous distinguons différents types de chaînes en fonction de caractéristiques telles que leur saillance, leur étendue dans le texte, la distance entre leurs éléments, etc. Cette double typologie des chaînes de référence nous permet à la fois de modéliser le comportement des principaux types de chaînes mais aussi de proposer une analyse fine de la progression textuelle dans les articles de format IMRAD, et d'expliquer comment certains référents sont présents tout au long du texte, alors que d'autres n'apparaissent que dans certaines sections.

Mots clés: chaîne de référence, coréférence, IMRAD, corpus, article scientifique

.....

Many research articles in experimental sciences present a standardized form known as the IMRAD format, an acronym for "Introduction, Methods, Results and Discussion". Each of these sections has a specific purpose (presenting the framework, describing the methodology, reporting the results and then discussing them), but also specific linguistic features. Our goal is to study coreference chains (all the expressions of a text that refer to the same extra-linguistic entity) in a corpus of French IMRAD articles (5 texts, 32,539 words). We show that different types of chains have specific features and that they cannot be studied without distinction. We first categorize chains by the type of their referent and analyze in particular three of these types: author(s), abstract concepts and groups. We then define types of chains by features such as their prominence, their length in the text, the distance between their elements, etc. This dual typology of coreference chains allows us to describe the main types of chains, but also to give a fine-grained analysis of the textual progression in IMRAD articles, especially how and why some referents are present in several sections while others are limited to one section only.

Keywords: coreference chain, coreference, IMRAD, corpus, scientific article

.....

Ce travail a été réalisé avec le soutien du projet ANR DEMOCRAT (« Description et modélisation des chaînes de référence : outils pour l'annotation de corpus et le traitement automatique », ANR-15-CE38-0008).

1. Introduction

1 La plupart des articles de recherche en sciences expérimentales présentent une structure standardisée appelée «IMRAD», un acronyme anglais pour «*Introduction, Methods, Results and Discussion*» (Swales, 1990 et 2004). Ces termes reprennent les intertitres qui séparent les grandes sections de ces articles.

2 Chacune de ces sections a une «fonction rhétorique» propre (Swales, 1990 et 2004; Gjesdal, 2013) :

- l'introduction présente la question de recherche et l'inscrit dans les travaux antérieurs;
- la méthodologie décrit les données (le corpus), la façon dont elles ont été recueillies, les procédures utilisées pour les traiter;
- les résultats présentent les observations faites à l'issue de l'expérimentation, en notant aussi les anomalies observées;
- la discussion tente de mettre les résultats en perspective en répondant à la question formulée dans l'introduction.

3 Ce format se caractérise par un haut degré de codification de chaque section (Swales, 1990 et 2004; Milard, 2007; Rinck, 2010). Bazerman (1988 : 260) indique que le manuel de l'American Psychological Association est «plein de prescriptions», un peu comme si le fait d'écrire ces articles revenait finalement à remplir un «questionnaire». Pour lui, il n'y a même plus besoin de transitions entre les parties; c'est ce que montre aussi la structure des articles de la revue *Plos Biology*, par exemple, qui place la section «méthodologie» à la fin de l'article (Oberle, 2018a).

4 Cette codification permet de comparer les sections entre elles. Puisque la section «méthodes», par exemple, est très codifiée, on peut comparer entre elles les sections «méthodes» des différents articles. De même, on peut opposer les sections IMRAD les unes avec les autres, par exemple l'introduction à la section «méthodes» ou à la discussion (Swales, 1990 : 177).

5 Cette spécialisation a intéressé les linguistes qui ont montré que chaque section était caractérisée par des phénomènes linguistiques propres, comme :

- la répartition du pronom *on*. Gjesdal (2013) étudie ainsi les différentes valeurs de *on* que l'on peut trouver dans des articles de recherche médicale. Par exemple, l'introduction est surtout caractérisée par la présence du *on* qui désigne l'auteur et du *on* qui inclut d'autres chercheurs, alors que la discussion voit apparaître le *on* qui signifie «tout le monde»;
- le domaine lexical des verbes. Reimerink (2006) trouve ainsi que l'introduction, les résultats et la discussion font plus appel aux verbes «d'existence» alors que la méthodologie utilise plutôt des verbes d'action;

- la distribution des citations et références. Ainsi, Bertin et Atanassova (2014) étudient le type de verbe associé à chaque citation, en fonction de la section d'occurrence;
- les marques d'énonciations. Heslot (1983), par exemple, compare des articles IMRAD américains et français du point de vue de la présence du pronom de première personne, de la voix, des formes temporelles...
- on trouvera dans Swales (1990) une liste plus complète.

6 À notre connaissance, les chaînes de référence n'ont pas été étudiées spécifiquement du point de vue des articles IMRAD.

7 Une «chaîne de référence» est définie comme l'ensemble des *expressions référentielles* qui renvoient à la même entité extralinguistique (Charolles, 1988; Corblin, 1985 et 1995; Schnedecker, 1997; Schnedecker et Landragin, 2014). Les expressions qui la composent sont appelées «maillons» par métaphore. Suivant la définition de Schnedecker (1997), nous ne considérons dans ce travail que les chaînes de plus de trois maillons.

8 Par exemple, le référent «la Grotte I des Treilles» est repris par *la Grotte I des Treilles*, *y* et *ce site* dans l'exemple suivant :

[1] [La Grotte I des Treilles] a été découverte par L. Balsan, M. R. Galzin et J. Maillé en 1933. Louis Balsan [y] a effectué des fouilles [...]. [Ce site] a été attribué au Chalcolithique [...].
(T1¹)

9 Les chaînes de référence sont caractérisées par des indicateurs tels que (Schnedecker et Landragin, 2014; Schnedecker, 2019; Oberle *et al.*, 2018) :

- le nombre de maillons;
- la moyenne de la distance entre deux maillons consécutifs (distance inter-maillonnaire);
- la portée locale ou globale;
- la catégorie et la fonction grammaticales des maillons;
- les patrons (séquences les plus fréquentes de catégories et de fonctions grammaticales des maillons);
- le coefficient de stabilité, qui évalue la stabilité lexicale à l'intérieur d'une chaîne.

10 Plusieurs études (Schnedecker, 2005; Schnedecker et Longo, 2012; Schnedecker et Landragin, 2014; Schnedecker, 2017; Baumer *et al.*, 2019) ont montré que ces paramètres variaient en fonction du genre discursif d'occurrence (par exemple, le

1. Les textes sont numérotés de T0 à T4, la référence exacte est donnée en fin d'article.

portrait journalistique, les faits divers, les recettes de cuisine, les introductions d'articles d'encyclopédies, les incipits de romans, etc.) : le genre textuel a donc une influence sur la composition des chaînes de référence.

11 Puisque les sections des articles de format IMRAD ont chacune des fonctions rhétoriques propres avec des caractéristiques linguistiques différentes, nous faisons l'hypothèse que les chaînes de référence varient en fonction de la section dans laquelle elles apparaissent, comme elles varient en fonction des genres. Nous chercherons donc à distinguer le comportement des chaînes selon la section. Ce faisant, nous dégagerons une première typologie des chaînes de référence.

12 Notre étude se fonde sur un corpus de cinq articles IMRAD en français. Nous commencerons par décrire le format IMRAD plus en détail et la façon dont nous avons collecté notre corpus, avant de proposer une typologie des chaînes qui s'appuie sur les oppositions entre les sections des articles de format IMRAD.

2. Le format IMRAD et la constitution du corpus

13 Le format IMRAD concerne surtout les articles de recherche en sciences « expérimentales » (Swales, 2004), par exemple physique, chimie, biologie, médecine, etc. Ils relatent une expérimentation et en tirent des conclusions.

14 Il s'est développé à partir du XVII^e siècle dans le but de garantir la reproductibilité d'une expérience (Pontille, 2007 ; Swales, 1990) : la standardisation du format garantit la standardisation des procédures expérimentales.

15 À l'époque moderne, en 1979, le format a été codifié par l'American National Standards Institute sous le nom d'*American National Standard for the preparation of scientific papers for written or oral presentation*. C'est donc dans les publications de langue anglaise et dans les sciences de la nature que ce format se développe.

16 Il se répand ensuite dans les sciences humaines, notamment dans la psychologie expérimentale, qui cherchent alors à se distinguer de la tradition philosophique et psychanalytique (Bazerman, 1988) : c'est pourquoi le manuel de style de référence pour le format IMRAD reste le *Publication Manual* de l'American Psychological Association.

17 En France, les publications IMRAD sont très fréquentes dans les sciences de la nature (dites « dures ») mais ces articles ont tendance à être écrits en anglais, même quand ils émanent de chercheurs francophones. En sciences humaines, au contraire, les articles sont plus souvent écrits en français mais ils respectent assez rarement le format IMRAD (Pontille, 2003).

18 Afin d'établir un panorama du paysage francophone (donc excluant toutes les publications anglaises) en sciences humaines et sociales (SHS), nous avons utilisé la plateforme OpenEdition Journals, qui propose plusieurs centaines de revues SHS en libre accès. Nous avons analysé la table des matières des articles des 4 derniers

numéros de 341 revues, soit un total de 21 689 articles. Il n'y a que 255 articles IMRAD en français, soit 2 % à peine du total.

19 Nous avons ensuite considéré les 20 revues qui avaient le plus d'articles IMRAD, et notamment les articles des 10 derniers numéros, soit 1 580 articles : 314 ont une structure IMRAD en français, soit 20 %.

20 Il apparaît donc que le format IMRAD est rare en français en SHS (2 %), mais certaines revues ont un taux conséquent d'articles de ce type (20 %). Ces revues se répartissent dans certaines disciplines :

- les sciences de l'éducation : 45 % ;
- les sciences de la nature ; il s'agit en fait surtout d'articles en archéologie qui utilisent des méthodes des sciences de la nature (physique, chimie, biologie...), ainsi que des articles en géographie physique : 35 % ;
- les autres disciplines (psychologie, linguistique, management, sociologie) : 20 %.

21 Les disciplines littéraires ne comptabilisent aucun article IMRAD, et les « arts et humanités », selon le regroupement d'OpenEdition Journals, ont un taux extrêmement faible de 0,37 %.

22 Parmi ces articles IMRAD en français, nous avons sélectionné cinq textes, en respectant les proportions ci-dessus :

- deux dans les sciences de l'éducation (T2 et T4) ;
- deux en archéologie (T1 et T3) ;
- et un en psychologie (T0).

3. Choix des référents à annoter

23 Jusqu'à présent, les études qui portent sur les chaînes de référence se sont surtout intéressées à une sélection de référents homogènes :

- des référents humains dans les faits divers (Schnecker et Longo, 2012), les portraits journalistiques (Schnecker, 2005), les textes narratifs médiévaux (Glikman *et al.*, 2014), les nouvelles brèves (Baumer, 2017), les récits brefs en français (Obry *et al.*, 2017) ;
- des entités géographiques dans les introductions d'articles d'encyclopédies (Schnecker, 2014) ;
- des objets physiques comme des ingrédients ou des ustensiles de cuisine (Schnecker, 2014).

24 De plus, ces textes sont généralement brefs (de quelques centaines de mots à 2000 mots).

- 25 Or les articles scientifiques sont, du point de vue des référents, plus hétérogènes : on y trouve des individus (si ce n'est le ou les auteurs), groupes, objets, matériaux, concepts abstraits, sans qu'il soit possible de dire lequel est le plus important ; un article sur le burnout au travail, par exemple, aura des référents principaux aussi bien humains et concrets (les travailleurs) que non humains et abstraits (le burnout). Or, on peut s'attendre à des différences importantes dans la composition des chaînes, par exemple la chaîne de l'auteur serait plutôt exprimée à l'aide de pronoms (« nous »), les noms massifs (« l'azote »), plutôt par des noms définis, les entités nommées (« la Grotte I des Treilles »), peut-être avec des noms définis entrecoupés de pronoms et d'anaphores infidèles.
- 26 De plus, les articles font généralement entre 6 000 et 8 000 mots. Ils sont donc beaucoup plus longs, et il faut donc s'attendre à des chaînes très disparates, certaines couvrant l'ensemble du texte, d'autres ne dépassant pas les limites d'un paragraphe, voire d'une phrase.
- 27 On arrive donc là aux limites des analyses « traditionnelles », qui envisagent les chaînes de référence sans considérer les différents types de chaînes. De telles études ne fonctionnent que dans des textes relativement brefs et homogènes, mais semblent difficiles dans des textes aussi longs et hétérogènes que les articles scientifiques.
- 28 C'est pourquoi le reste de la présentation sera organisé comme un parcours entre différents types de chaînes, différentes à la fois par leur référent et par leurs caractéristiques, dont nous montrerons les particularités dans les textes de format IMRAD et chacune des sections.
- 29 À notre connaissance, la seule étude qui esquisse une typologie de référents est celle de Longo et Todirascu (2014), qui concerne des textes juridiques longs (jusqu'à 8 000 mots), avec une distinction entre référents non humains et humains (spécifique, générique, collectif). Il faut également citer le corpus ANCOR-Centre (« Anaphore et coréférence dans les corpus oraux » – Muzerelle *et al.*, 2013), qui indique un type d'entités nommées (selon la classification ESTER2 [Évaluation de systèmes de transcription enrichie d'émissions radiophoniques – Galliano *et al.*, 2005]) pour certains maillons, par exemple les personnes, lieux, fonctions, organisations, produits, etc. On peut alors déduire, pour un certain nombre de chaînes, le type du référent. Mais nous n'avons trouvé aucune étude qui exploite cette information pour faire une analyse des chaînes selon leur type.

4. Caractéristiques des chaînes saillantes

- 30 Nous n'avons cependant pas annoté *tous* les référents des textes², nous avons repéré les plus saillants à partir des mots clés, du titre et du résumé, ainsi que d'une rapide

2. L'annotation de tous les référents, y compris ceux qui ne donnent pas lieu à des relations de coréférence (singletons), aurait permis une étude contrastive entre chaînes et singletons (Recasens *et al.*, 2013). Cela dépasse cependant la portée de cet article.

analyse lexicométrique (mots les plus fréquents). Dans les cinq textes, nous avons annoté³ 89 chaînes, soit une petite vingtaine par texte, avec près de 2000 maillons (tableau 1).

	Texte	Tokens	CR	Maillons
Éducation	T2	7 123	22	622
	T4	8 627	17	422
Archéologie	T1	6 008	20	292
	T3	4 318	15	186
Psychologie	T0	6 463	15	434
TOTAL		32 539	89	1 956

Tableau 1 – Annotation des référents saillants (CR = chaînes de référence)

31 Ces chaînes se caractérisent d’abord par leur portée globale : elles couvrent l’ensemble, ou du moins une grande partie, du texte. Ce n’est pas étonnant, puisque les référents sont saillants, et que c’est justement pour cette raison que nous avons choisi de les annoter.

32 Ce sont cependant des chaînes avec peu de maillons mais très distendues. Avec 22 maillons en moyenne sur l’ensemble du texte (8000 mots en moyenne) et une distance inter-maillonnaire de 333 mots qui excède la longueur moyenne du paragraphe (119 mots), elles n’apparaissent que dans un paragraphe sur trois.

33 Par exemple, le burnout (T0), une chaîne de 32 maillons, couvre 17 paragraphes, du 1^{er} au 53^e (alors que le texte fait 55 paragraphes) : 11 paragraphes n’ont qu’une seule expression s’y rapportant, 5 paragraphes ont 2 à 4 expressions, et seul un paragraphe en a plus (7 expressions). Surtout, la chaîne est tout à fait absente de 38 paragraphes sur 55.

34 Cela explique un coefficient de stabilité élevé⁴, à 0,8 : afin que le texte reste compréhensible, il ne peut pas y avoir de reprise par un terme différent (anaphore infidèle) si le terme précédent se trouve trois paragraphes en arrière. Reprenons l’exemple du burnout : la chaîne comporte 31 désignations nominales, mais seulement deux termes différents – « burnout » (dans 30 cas) et « phénomène » (dans un seul cas).

3. L’annotation a été faite à l’aide de SACR (Script d’annotation des chaînes de référence – Oberle, 2018b).

4. Le coefficient de stabilité (adapté de Perret, 2000) est calculé en ne prenant en compte que la tête syntaxique des groupes nominaux (GN) (les pronoms et déterminants ne sont pas inclus dans le calcul). Le coefficient est compris entre 0 et 1 : plus il est élevé, moins il y a de variations (s’il est à 1, toutes les reprises nominales sont les mêmes, s’il est de 0, elles sont toutes différentes). Nous avons utilisé la formule suivante : $1 - (x-1)/(n-1)$, où x est « le nombre de désignations différentes » et n « le nombre total d’anaphores nominales », pour reprendre les termes de Perret.

35 Il en va de même pour les pronoms : ils comptent pour 13 % (252 sur 1956) des maillons, alors que d'autres études font état de 30 % de pronoms dans des textes plus narratifs comme les faits divers (Schnecker et Longo, 2012), ou même plus de 40 % dans les portraits journalistiques (Schnecker, 2005). Notre chaîne du burnout, elle, ne comporte qu'un seul et unique pronom.

5. Typologie des référents

36 Comme nous l'avons évoqué plus haut, il semble difficile de traiter toutes les chaînes en un bloc, sans faire de distinction quant aux types de référents. Aussi proposons-nous une typologie de référents qui prend en compte à la fois :

- des problèmes référentiels (voir aussi Oberle, 2017), qui se rapprochent plus des problématiques de philosophie du langage, puisqu'on peut difficilement traiter un référent concret comme un référent abstrait, notamment pour ce qui est de déterminer dans quelle mesure deux référents abstraits sont coréférentiels ou non (le réchauffement climatique est-il coréférentiel au changement climatique?) ;
- des problèmes d'annotation spécifiques que nous avons rencontrés, comme l'annotation des groupes qui évoluent au fil du texte (ce sont des référents évolutifs concrets, voir Charolles et Schnecker, 1993) ;
- des spécificités des articles de recherche : nous avons ainsi créé des types pour le ou les auteurs et l'article lui-même, puisque ce sont des référents particuliers dans ce type de textes.

37 Nous proposons les types suivants :

- auteur (ex. : *nous*) ;
- recherche et article (ex. : *notre article*) ;
- entités nommées et référents définis (ex. : *la Grotte des Treilles*) ;
- ensembles (ex. : *les étudiants*) ;
- massifs (ex. : *l'azote*) ;
- référents génériques (ex. : *le vétérinaire rural*) ;
- noms abstraits (ex. : *l'interdisciplinarité*) ;
- noms prédicatifs (ex. : *la modération*, comme dans «La modération par les compétences politiques des facteurs de stress au travail sur la santé psychologique») ;
- ensembles «flous» (sans limites clairement définies, comme dans «les vestiges archéologiques» sans autre précision) ;
- «variables liées» (quand la référence n'est pas spécifique, comme dans «un capitaine d'équipe»).

38 Nous nous concentrerons dans cet article sur trois de ces types : l'auteur, les ensembles, les noms abstraits. Ce sont des référents importants pour les articles scientifiques et ce sont des types qui n'ont pas été étudiés dans les études antérieures. De plus, ils offrent une bonne illustration des différences de caractéristiques des chaînes d'un type à l'autre.

5.1. Les auteurs

39 Contrairement aux textes littéraires, qui se concentrent généralement sur un ou deux personnages, dans les articles scientifiques de type IMRAD, il n'y a guère que le ou les auteurs qui soient des individus uniques, bien définis.

40 La chaîne des auteurs est également intéressante parce qu'elle s'inscrit dans de nombreuses recherches sur l'expression de l'auteur scientifique (dans une perspective sociologique : voir Pontille, 2007 ; dans une perspective linguistique : Grossmann, 2012 ; dans la perspective des textes scientifiques : Loffler-Laurian, 1980 ; Fløttum, 2006).

41 Cette chaîne n'est pas présente dans tous les textes : seuls deux de nos textes en ont une de taille assez importante (tous les textes ont une présence de l'auteur, mais parfois seulement dans les remerciements).

42 Elle est composée presque exclusivement de pronoms (« nous ») et de déterminants (« notre », « nos »), avec, parfois, un nom : « auteur » (« Deux des trois auteurs du présent article [...] »). Les mentions sont presque exclusivement en fonction sujet et, pour les déterminants possessifs, occupent celle de complément du nom (nous avons annoté les déterminants possessifs comme compléments du nom).

43 Cette chaîne n'est pas présente dans toutes les parties (figure 1) : 36 % des maillons sont dans l'introduction, notamment dans l'exposé de la recherche :

- [2] nous proposons ici de contribuer
(T0)
- [3] le cadre dans lequel nous nous situons
(T0)
- [4] nous faisons l'hypothèse
(T0)
- [5] nous présenterons ce modèle en détail
(T0)
- [6] à notre connaissance
(T0, T2)

44 Au contraire, il y en a très peu dans la méthodologie : seulement 6 maillons. C'est un peu surprenant parce qu'on pourrait s'attendre à une partie plus narrative

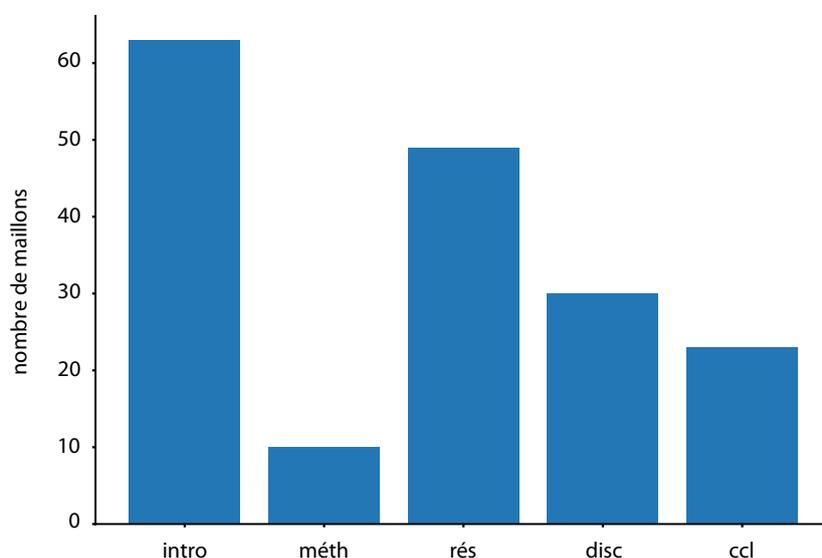


Figure 1 – Répartition des maillons de la chaîne «auteur»

qui décrit l'action de l'auteur pour préparer son expérimentation (« nous avons interrogé tant et tant de personnes », « nous avons utilisé tel ou tel instrument »). Cette partie est en fait beaucoup plus impersonnelle.

45 Nos résultats semblent en désaccord avec ce que d'autres chercheurs ont trouvé. Par exemple, Heslot (1983), examinant des articles de biologie en anglais et en français, trouve pour ces derniers une grande variabilité, mais pas de règle particulière. Dans les articles en anglais, en revanche, l'auteur est présent dans l'introduction et la discussion. Gjesdal (2013), au cours d'une étude qualitative d'articles médicaux, trouve une plus grande fréquence de « nous » et de « on » (dont les valeurs sont souvent associées aux auteurs) dans les méthodes et les résultats (50 % des occurrences sont dans ces sections, contre 16 % dans l'introduction, 30 % dans la discussion et 4 % dans la conclusion).

5.2. Les ensembles

46 Les ensembles sont très présents dans les articles IMRAD. En effet, ce sont surtout des expérimentations statistiques, et les statistiques cherchent à dire quelque chose à propos d'une population (un ensemble) à partir de l'observation d'un ou plusieurs échantillons (d'autres ensembles). L'article IMRAD est donc construit autour d'ensembles, que ce soit des ensembles de personnes, d'objets, de matériaux, etc.

47 En observant la répartition de ces chaînes (figure 2), on constate une concentration de leurs maillons dans les résultats et la discussion (28 % pour chacune de ces sections). Autrement dit, les maillons de ces chaînes sont essentiellement situés dans ces deux sections.

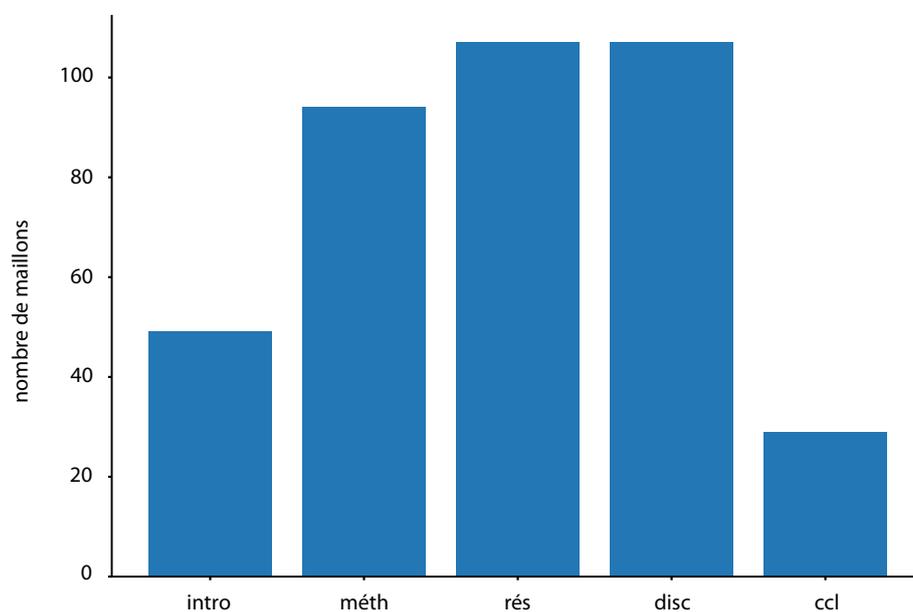


Figure 2 – Répartition des maillons des chaînes d'ensembles

	Nombre de maillons	Pourcentage
Auteur(s)	10	3,13 %
Article	7	2,19 %
Référents définis	39	12,23 %
Ensembles	94	29,47 %
Massifs	16	5,02 %
Référents génériques	2	0,63 %
Noms abstraits	55	17,24 %
Noms prédictifs	10	3,13 %
Ensembles flous	26	8,15 %
Variables liées	60	18,81 %

Tableau 2 – Répartition des maillons de la section « Méthodologie » par types de référents

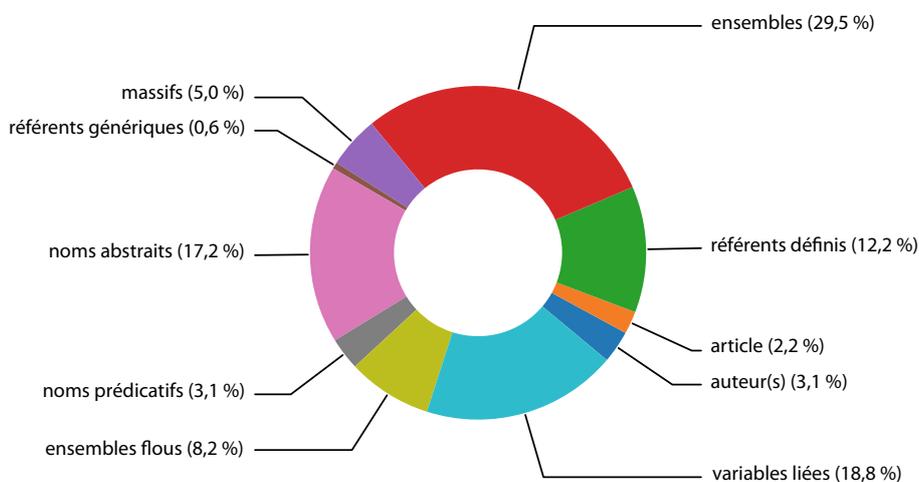


Figure 3 – Répartition des maillons de la section «Méthodologie» par types de référents

48 Si, en revanche, on prend tous les maillons de la méthodologie (figure 3, tableau 2), on constate que la majorité des maillons (près d'un tiers) appartiennent à des chaînes dont le référent est un ensemble. Autrement dit, dans la méthodologie, il est surtout question d'ensembles. En effet, c'est la section qui décrit les groupes de participants : leur composition, leurs tâches, etc.

[7] **Les étudiants** appartiennent à 1 049 établissements différents.
(T2)

[8] une communauté d'**étudiants** de 99 pays différents
(T2)

[9] **Ils** devaient d'abord sélectionner les cinq items [...].
(T4)

[10] **Ils** devaient ensuite procéder de la même manière [...].
(T4)

49 Il est aussi intéressant d'observer la répartition des fonctions de ces maillons selon les sections : par exemple, entre la méthodologie et la conclusion (tableau 3).

	Méthodologie	Conclusion
Compléments du nom	43 % (41 sur 94)	69 % (20 sur 29)
Compléments d'objet	30 % (28 sur 94)	10 % (3 sur 29)

Tableau 3 – Fonctions complément du nom et d'objet dans la méthodologie et la conclusion

50 Cela s'explique par la fonction de chaque section. Dans la méthodologie, on trouve surtout des compléments d'objet et des compléments du nom :

– le complément d'objet est utilisé pour donner des tâches aux participants :

[11] il était demandé **aux participants**
(T4)

[12] il **leur** a été proposé de hiérarchiser les évocations produites
(T4)

– alors que le complément du nom permet de situer les participants :

[13] le niveau **des étudiants**
(T4)

[14] les caractéristiques sociodémographiques **des étudiants**
(T4)

51 Au contraire, dans la conclusion, il n'y a plus que des compléments du nom, qui servent à décrire la population :

[15] les spécificités de **la population d'élèves étudiée**
(T2)

[16] le comportement **des étudiants**
(T2)

52 On ne donne plus de tâches aux participants, ce qui explique la disparition des compléments d'objet.

5.3. Les noms abstraits

53 Nous rangeons sous le terme «noms abstraits» les concepts abstraits comme «l'interdisciplinarité», «l'entrepreneuriat», «l'innovation» ou encore «le burnout». Là encore, ce sont des référents incontournables des articles scientifiques, parce que ces textes sont concernés avant tout par ce type de référents. Les titres des articles sont eux-mêmes des catalogues de noms abstraits :

[17] «**Le rôle modérateur des compétences politiques sur le burnout**»
(3 noms, 3 référents abstraits)

[18] «**La propension à l'interdisciplinarité des étudiants en situation d'innovation**»
(4 noms, 3 référents abstraits)

[19] «**Représentations socio-professionnelles et choix de la spécialisation : le cas de la filière vétérinaire rurale**»
(5 noms, 5 référents abstraits)

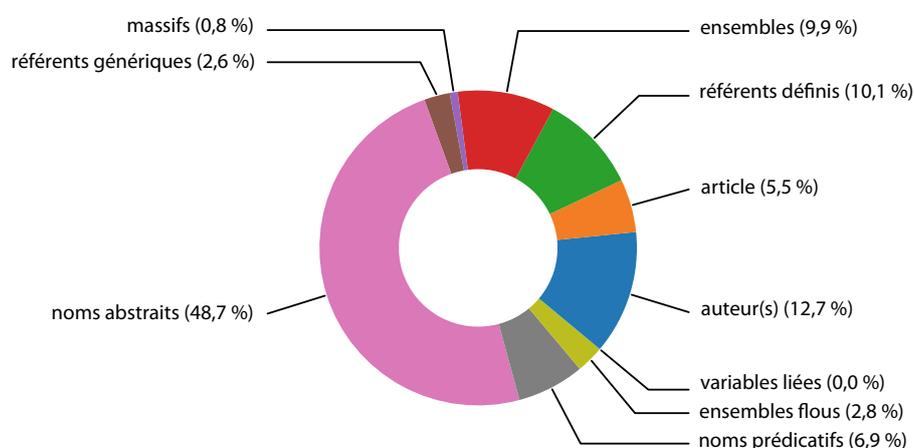


Figure 4 – Répartition des maillons de la section « Introduction » par types de référents

	Nombre de maillons	Pourcentage
Auteur(s)	63	12,73 %
Article	27	5,45 %
Référents définis	50	10,10 %
Ensembles	49	9,90 %
Massifs	4	0,81 %
Référents génériques	13	2,63 %
Noms abstraits	241	48,69 %
Noms prédicatifs	34	6,87 %
Ensembles flous	14	2,83 %
Variables liées	0	0,00 %

Tableau 4 – Répartition des maillons de la section « Introduction » par types de référents

54

Les référents abstraits sont ainsi le sujet, le thème des articles, et c'est la raison pour laquelle on les trouve notamment dans l'introduction. La moitié de tous les maillons de l'introduction (figure 4, tableau 4) renvoient à des concepts abstraits. Là encore, c'est dû à la fonction de l'introduction, qui est de présenter les concepts qui vont être étudiés dans l'article : les définir, les situer par rapport à d'autres recherches (la revue de la littérature), présenter les hypothèses que l'on fait au sujet de ces référents abstraits.

	Nombre de maillons	Pourcentage
Nom indéfini	22	3,32 %
Nom sans déterminant	129	19,49 %
Nom possessif	8	1,21 %
Nom défini	452	68,28 %
Déterminant possessif	3	0,45 %
Nom démonstratif	19	2,87 %
Pronom personnel	17	2,57 %
Pronom zéro	4	0,60 %
Pronom réfléchi	1	0,15 %
Pronom relatif	6	0,91 %
Pronom démonstratif	1	0,15 %

Tableau 5 – Répartition des maillons des chaînes de noms abstraits par catégories grammaticales

55 Les chaînes de ces référents sont surtout composées (tableau 5) de GN définis (68 %) et de GN sans déterminants (20 %) («facteurs de stress», «niveau de burnout») : c'est une particularité de certains des référents abstraits de ne quasi jamais apparaître sous la forme d'indéfinis, notamment dans les articles de recherche : on voit mal «un burnout», «une interdisciplinarité».

56 Au niveau des fonctions, un maillon sur deux est un complément du nom («le modèle du burnout» [T0], «les dimensions du burnout» [T0], «le goût des étudiants pour l'interdisciplinarité» [T2]). Il y a peu de sujets (17 % [110 sur 662]) : ce sont des concepts qu'on analyse mais qui ne sont que rarement directement «acteurs» («Le burnout découle [...]»).

57 En guise de bilan, rappelons que les types de référents (auteurs, ensembles, noms abstraits) initient des chaînes très différentes, au moins du point de vue de la composition des chaînes, de la fonction des maillons, et de leur répartition dans les différentes sections IMRAD.

6. Les chaînes de paragraphes

6.1. Introduction

58 L'étude des chaînes saillantes est toutefois limitée par le fait que ce sont des chaînes avec peu de maillons mais très distendues, puisqu'elles n'apparaissent en

moyenne qu'une fois tous les trois paragraphes. D'autres chaînes, au contraire, sont moins saillantes au niveau du texte, mais jouent un rôle important au niveau du paragraphe.

59 Or, le paragraphe forme une unité de sens (Bessonnat, 1988) qui a une influence importante sur les relations anaphoriques. Par exemple, Schnedecker (1997) montre que le nom propre coïncide généralement avec l'initiale des paragraphes, et Ariel (1990) montre que 90 % des pronoms ont leur antécédent dans le même paragraphe.

60 C'est pourquoi nous avons réannoté les textes⁵ avec ce que nous avons appelé des « chaînes de paragraphes », comprises dans les limites d'un paragraphe : dès lors que, dans les limites d'un paragraphe, trois expressions référentielles ou plus renvoient au même référent, nous avons annoté la chaîne, quelle qu'elle fût, sans considération de type ou de saillance.

61 On peut supposer que les processus cognitifs sont différents entre les chaînes de paragraphe où le référent apparaît plusieurs fois à très peu de distance d'intervalle (avec une distance entre les maillons de 40 mots en moyenne) et les chaînes de texte où le référent n'apparaît qu'une fois tous les trois paragraphes (avec une distance entre les maillons de 333 mots en moyenne).

62 Manifestement, ce ne sont pas les mêmes types de chaînes : ni sur le plan cognitif, ni sur le plan de la composition. C'est pourquoi il nous a paru important de les annoter et de les étudier de façon distincte.

6.2. Caractéristiques des chaînes de paragraphes

63 Nous avons trouvé 204 chaînes de ce type, soit une cinquantaine de chaînes par texte, avec 914 maillons (tableau 6). Ces chaînes sont plus brèves : 4,5 maillons en moyenne.

	Texte	Tokens	CR	Maillons
Éducation	T2	7123	82	396
	T4	8627	78	343
Archéologie	T1	6008	35	147
	T3	4318	9	28
Psychologie	T0	6463	--	--
TOTAL		32539	204	914

Tableau 6 – Annotation des chaînes de paragraphes

5. À l'exception du texte T0.

64 Le nombre de chaînes est inférieur au nombre de paragraphes, ce qui signifie que tous les paragraphes n'ont pas de chaînes qui leur sont propres. Ces paragraphes sans chaînes sont même majoritaires, puisqu'ils représentent 62 % des paragraphes. Ce qui signifie qu'il y a une concentration importante des chaînes dans quelques paragraphes.

65 Puisque l'espace plus réduit du paragraphe favorise les reprises pronominales et anaphores infidèles, le coefficient de stabilité baisse un peu (0,74 contre 0,8 auparavant) et il y a un peu plus de pronoms (17 % [159 sur 926] contre 13 % [252 sur 1956] auparavant).

[20] Dans un premier temps, un recueil du corpus sémantique de la représentation a été réalisé auprès de [38 étudiants] [...]. Il a été demandé [aux participants] de noter spontanément au plus dix mots ou expressions qu'[ils] associaient à l'expression stimulus: «activité vétérinaire rurale».

(T4)

[21] [La Grotte I des Treilles] a été découverte par L. Balsan, M. R. Galzin et J. Maillé en 1933. Louis Balsan [y] a effectué des fouilles dans la salle B [...]. [Ce site] a été attribué au Chalcolithique par Costantini [...].

(T1)

66 Bien évidemment, la distance inter-maillonnaire est bien plus faible et inférieure à la longueur des paragraphes (40 mots).

67 Il faut cependant noter qu'il y a un contraste important entre les chaînes et qu'on peut en isoler certaines qui ont une distance inter-maillonnaire de moins de 20 mots.

6.3. Chaînes éphémères

68 C'est ce que l'on peut appeler des «chaînes éphémères», au sens où leur cycle de vie est extrêmement bref: elles sont dans une phrase ou même dans une proposition, et elles se composent d'un nom suivi de pronoms ou de déterminants possessifs, souvent coordonnés.

[22] En effet, il n'a pas été possible, pour des raisons de disponibilité, d'inclure des étudiants préparant [le concours C], alors que [sa] spécificité et [son] mode de recrutement pourraient avoir un impact sur les représentations des étudiants le préparant.

(T4)

[23] [...] [tous les arguments] [qui] viennent d'être mentionnés et [qui] auraient été susceptibles d'expliquer l'aversion des étudiants pour l'interdisciplinarité.

(T2)

69 Ces chaînes doivent être catégorisées à part, parce qu'elles n'ont sans doute pas la même importance que des chaînes pour lesquelles il y a plusieurs reprises

nominales. Ce sont en quelque sorte des chaînes «par accident», dues à la présence d'une coordination.

6.4. Chaînes uniques et chaînes liées

70 Les chaînes de paragraphe sont définies comme des chaînes qui ne dépassent pas les limites d'un paragraphe. Un même référent peut dès lors initier :

- une seule chaîne dans un seul paragraphe du texte, ce que nous appelons une «chaîne unique» ;
- des chaînes dans plusieurs paragraphes du texte, ce que nous appelons des «chaînes liées».

71 Il y a 105 référents qui donnent lieu à des chaînes de paragraphes. Les trois quarts de ces référents ne sont présents que dans un seul paragraphe : ils représentent 79 chaînes (soit 39 % de toutes les chaînes de paragraphes). Les référents du quart restant sont présents chacun dans plusieurs paragraphes : cela donne 132 chaînes au total (soit 61 % des chaînes). Pour résumer : un quart des référents initient plus de 61 % des chaînes de paragraphes.

72 Ces deux types de chaînes s'inscrivent dans deux perspectives différentes et c'est la raison pour laquelle nous pensons qu'il est important de les séparer. Les chaînes liées sont ouvertes sur le reste du texte puisqu'elles sont liées aux paragraphes précédents et suivants, en ce sens elles sont le support transitoire de l'information, qui apparaît avant et qui continue après, et elles participent donc à la cohérence et à la cohésion du texte. Au contraire, les chaînes uniques ont un cycle de vie entièrement compris dans un paragraphe : elles sont donc repliées sur elles-mêmes, en lien avec le thème du paragraphe plutôt que celui du texte.

73 On peut donc s'attendre à des comportements différents. Par exemple, les chaînes uniques ont plus de pronoms (22 % [79 sur 322] contre 15 % [89 sur 604] pour les chaînes liées), notamment des relatifs, ce qui indique un fort taux d'expansions. Un patron fréquent est «GN défini > relatif > GN défini/démonstratif», avec une relative déterminative, puisque la chaîne commence par un défini. Le référent est ainsi introduit et décrit, puis disparaît :

[24] C'est le cas notamment des appareils monospire comme [le MS2D de Bartington] [qui] a été utilisé pour la plupart des études sur le projet Canal Seine-Nord Europe. [Cet appareil] permet une mesure point par point de la susceptibilité magnétique volumique [...].
(T3)

74 Mais il nous semble plus intéressant de nous attarder sur la distribution de ces différents types de chaînes. Ainsi, le graphique (figure 5) montre une répartition différente des chaînes liées selon les sections IMRAD : fortement présentes dans l'introduction, elles sont beaucoup plus rares dans la méthodologie, puis leur proportion remonte progressivement.

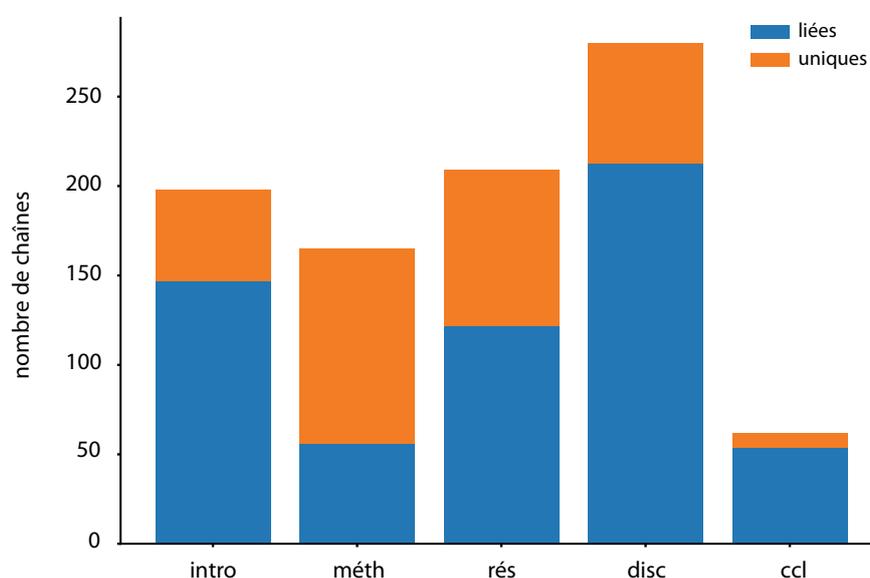


Figure 5 – Répartition des chaînes liées et des chaînes uniques dans les sections

75 Ceci est dû à la fonction de chaque section. Dans la méthodologie, près des deux tiers (62 % [23 sur 37]) des chaînes sont des chaînes uniques, parce que cette section présente outils et procédures de façon successive.

76 Par exemple, il y a des paragraphes pour la description de chacun des groupes de participants (chaîne en gras) et de leurs tâches (chaîne en italiques) :

[25] Dans un premier temps, un recueil du corpus sémantique de la représentation a été réalisé auprès de **[38 étudiants]** [...] à l'aide de la méthode du réseau d'association [...]. Il a été demandé **[aux participants]** de noter spontanément au plus *[dix mots]* ou courtes expressions *[qu']* **[ils]** associaient à l'expression stimulus : « activité vétérinaire rurale ». **[Ils]** devaient aussi numéroter en chiffres arabes, au fur et à mesure de *[leur]* apparition, *[ces mots ou expressions]*.

(T4)

77 Les 38 participants ne sont mentionnés qu'une seule fois dans la suite du texte, dans un autre paragraphe, mais sans donner une chaîne (c'est-à-dire sans apparaître trois fois).

78 Le motif de la figure 5 s'explique lorsqu'on s'intéresse aux référents qui apparaissent dans les différentes sections IMRAD. La figure 6 montre la répartition des référents dans le texte T4, qui est la relation d'une étude sur les choix de la filière de spécialisation des étudiants vétérinaires en fonction de leurs représentations socio-professionnelles.

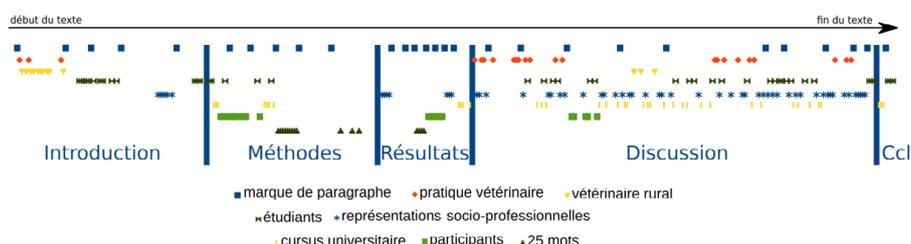


Figure 6 – Cartographie des chaînes liées dans le texte T4

79 L'introduction présente les principaux concepts : « la pratique vétérinaire », « le
vétérinaire rural » (au sens générique, comme un modèle : « l'exercice des vétérinaires
ruraux »), « les représentations socio-professionnelles » ; ainsi que la population
étudiée : « les étudiants vétérinaires ».

80 Dans la méthodologie, on introduit les participants et leurs tâches, sans reprendre
les référents de l'introduction : on se concentre sur des procédures, c'est pourquoi
il y a très peu de référents partagés avec d'autres sections.

81 Les résultats décrivent l'aboutissement de l'expérimentation : on reprend donc
les référents qui ont été introduits dans la méthodologie (les participants), sans
reprendre ceux de l'introduction.

82 La discussion est le moment de l'extrapolation : on se sert des résultats de
l'expérimentation pour dire quelque chose de la population générale et répondre aux
questions formulées dans l'introduction. Il y a donc reprise des principaux référents
(mais pas tous) des résultats et de l'introduction.

83 Enfin, la conclusion a un statut à part, parce qu'elle ne fait pas partie officiellement
du format IMRAD, et pourtant elle est présente dans presque tous les articles IMRAD.
Elle oscille entre la synthèse de l'article et les perspectives. On pourrait s'attendre
à ce que la plupart des chaînes soient reprises dans la conclusion. Ici, la conclusion
est très courte, si bien que si beaucoup de référents sont présents, peu initient des
chaînes (c'est-à-dire que peu sont présents au moins trois fois par paragraphe). C'est
pourquoi il n'y a que deux référents qui apparaissent dans la conclusion.

7. Conclusion

84 À travers l'étude des sections des articles de recherche de format IMRAD, dont
nous avons montré que chacune avait des chaînes de référence aux caractéristiques
propres, nous avons établi une première typologie de chaînes.

85 D'abord en ce qui concerne les types de référents. Nous avons présenté une liste
d'une dizaine de types, et nous avons analysé trois d'entre eux (le ou les auteurs,
les ensembles, et les concepts abstraits) en montrant que les chaînes qu'ils initient
ont des comportements bien différents.

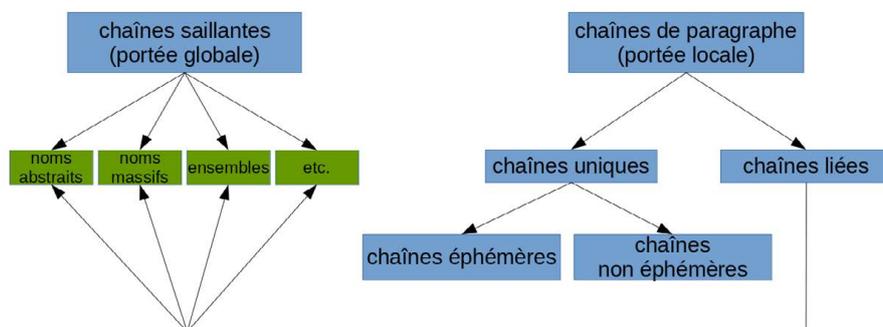


Figure 7 – Types de chaînes

86 Ensuite en ce qui concerne la composition des chaînes (figure 7). Nous avons opposé les chaînes saillantes aux chaînes de paragraphe, elles-mêmes divisées entre chaînes uniques (éphémères ou non) et chaînes liées, support d'une information qui traverse plusieurs paragraphes, sinon l'ensemble du texte. Nous avons illustré ce dernier type de chaîne en analysant la façon dont les chaînes liées étaient reprises, ou non, d'une section IMRAD à une autre.

87 Ce travail montre que les chaînes de référence peuvent difficilement être étudiées sans faire de différences entre leurs types. Il s'agit d'une étude exploratoire, cependant, et il conviendrait de vérifier les types et leurs critères sur un corpus plus grand, et étendu à d'autres genres de textes.

Références bibliographiques

Corpus

- Texte T0** : DAGOT, L., BORTEYROU, X., GRÉGOIRE, C. et VALLÉE, B. 2014. Le rôle modérateur des compétences politiques sur le burnout. *Revue internationale de psychologie sociale* 27 (2) : 5-34.
- Texte T1** : HERRSCHER, E., LHEUREUX, J., GOUDE, G., DABERNAT, H. et DURANTHON, F. 2013. Les pratiques de subsistance de la population Néolithique final de la grotte I des Treilles (commune de Saint-Jean-et-Saint-Paul, Aveyron). *Préhistoires méditerranéennes* 4. En ligne à l'adresse suivante : <https://journals.openedition.org/pm/783>.
- Texte T2** : HOUY, T., ATTAL, Y. et MELAMED, Y. 2014. La propension à l'interdisciplinarité des étudiants en situation d'innovation. *Revue internationale de pédagogie de l'enseignement supérieur* 30 (2). En ligne à l'adresse suivante : <https://journals.openedition.org/ripes/825>.
- Texte T3** : HULIN, G., BROES, F. et FECHNER, K. 2012. Caractérisation de phénomènes anthropiques par la mesure de paramètres magnétiques sur surface décapée : Premiers résultats sur le projet Canal Seine-Nord Europe. *ArcheoSciences* 36 : 61-70. En ligne à l'adresse suivante : <https://journals.openedition.org/archeosciences/3744>.

Texte T4: DERNAT, S. et SIMÉONE, A. 2014. Représentations socio-professionnelles et choix de la spécialisation : le cas de la filière vétérinaire rurale. *Revue internationale de pédagogie de l'enseignement supérieur* 30 (2). En ligne à l'adresse suivante : <https://journals.openedition.org/ripes/832>.

Études

- ARIEL, M. 1990. *Accessing Noun-Phrase Antecedents*. Londres – New York : Routledge.
- BAUMER, E. 2017. Chaînes de référence et point de vue dans la fiction littéraire : le cas des nouvelles courtes. *Langue française* 195 : 73-90.
- BAUMER, E., DIAS, D. et SCHNEDECKER, C. 2019. Les chaînes de référence dans un corpus contrastif (allemand-anglais-français) de romans réalistes du 19^e siècle : analyse quantitative et qualitative. Contribution présentée au *Colloque international PhraseoRom – Phraséologie et stylistique de la langue littéraire (13-15 mars 2019, Erlangen, Allemagne)*.
- BAZERMAN, C. 1988. *Shaping Written Knowledge: The Genre and Activity of the Experimental Article in Science*. Madison : University of Wisconsin Press.
- BERTIN, M. et ATANASSOVA, I. 2014. A Study of Lexical Distribution in Citation Contexts through the IMRAD Standard. In P. MAYR, P. SCHAEER, A. SCHARNHORST, B. LARSEN et P. MUTSCHKE (éd.), *Proceedings of the First Workshop on Bibliometric-Enhanced Information Retrieval (BIR 2014) Co-located with 36th European Conference on Information Retrieval (ECIR 2014)*. 5-12. En ligne à l'adresse suivante : <http://ceur-ws.org/Vol-1143/paper1.pdf>.
- BESSONNAT, D. 1988. Le découpage en paragraphes et ses fonctions. *Pratiques : linguistique, littérature, didactique* 57 : 81-105.
- CHAROLLES, M. 1988. Les plans d'organisation textuelle : périodes, chaînes, portées et séquences. *Pratiques : linguistique, littérature, didactique* 57 : 3-13.
- CHAROLLES, M. et SCHNEDECKER, C. 1993. Coréférence et identité : le problème des référents évolutifs. *Langages* 27 (112) : 106-126.
- CORBLIN, F. 1985. Les chaînes de référence : analyse linguistique et traitement automatique. *Intellectica* 1 : 123-143.
- CORBLIN, F. 1995. *Les formes de reprise dans le discours : anaphores et chaînes de référence*. Rennes : Presses universitaires de Rennes.
- FLØTTUM, K. 2006. Les « personnes » dans le discours scientifique : le cas du pronom *on*. In M. OLSEN et E. H. SWIATEK (éd.), *XVII^e congrès des Romanistes scandinaves*. Roskilde : Roskilde Universitet.
- GALLIANO, S., GEOFFROIS, E., MOSTEFA, D., CHOUKRI, K., BONASTRE, J.-F. et GRAVIER, G. 2005. The ESTER Phase II Evaluation Campaign for the Rich Transcription of French Broadcast News. In *Proceedings of the 9th European Conference on Speech Communication and Technology (Interspeech 2005 – Eurospeech)*. ISCA (International Speech Communication Association) Online Archive : 1149-1152. En ligne à l'adresse suivante : https://www.isca-speech.org/archive/archive_papers/interspeech_2005/i05_1149.pdf.
- GJESDAL, A. M. 2013. The Influence of Genre Constraints on Author Representation in Medical Research Articles. The French Indefinite Pronoun *On* in IMRAD Research Articles. *Discours* 12 : 1-23. En ligne à l'adresse suivante : <https://journals.openedition.org/discours/8770>.

- GLIKMAN, J., GUILLOT-BARBANCE, C. et OBRY, V. 2014. Les chaînes de référence dans un corpus de textes narratifs médiévaux: traits généraux et facteurs de variation. *Langages* 195 : 43-60.
- GROSSMANN, F. 2012. Pourquoi et comment cela change? Standardisation et variation dans le champ des discours scientifiques. *Pratiques. Linguistique, littérature, didactique* 153-154: 141-160. En ligne à l'adresse suivante : <https://journals.openedition.org/pratiques/1976>.
- HESLOT, J. 1983. Récit et commentaire dans un article scientifique. *DRLAV – Documentation et recherche en linguistique allemande contemporaine* 29 : 133-154.
- LOFFLER-LAURIAN, A.-M. 1980. L'expression du locuteur dans les discours scientifiques. «Je», «nous», et «on» dans quelques textes de chimie et de physique. *Revue de linguistique romane* 44 (173-174) : 135-157.
- LONGO, L. et TODIRASCU, A. 2014. Vers une typologie des chaînes de référence dans des textes administratifs et juridiques. *Langages* 195 : 79-98.
- OBERLE, B. 2017. *Étude des chaînes de référence dans les articles de recherche de format IMRAD: problèmes d'annotation, analyse quantitative et qualitative*. Mémoire de master. Université de Strasbourg.
- OBERLE, B. 2018a. Research Articles from Plos Biology: A Textual Data Analysis. In X. WANG (éd.), *Analyse textuelle: études des chaînes référentielles en chinois et en français*. Pékin: Beijing Language and Culture University Press: 60-73.
- OBERLE, B. 2018b. SACR: A Drag-and-Drop Based Tool for Coreference Annotation. In N. CALZOLARI, K. CHOUKRI, C. CIERI, T. DECLERCK, K. HASIDA, H. ISAHARA, B. MAEGAARD, J. MARIANI, A. MORENO, J. ODJIK, S. PIPERIDIS et T. TOKUNAGA (éd.), *Proceedings of the 11th International Conference on Language Resources and Evaluation – LREC 2018*. Paris: European Language Resources Association (ELRA): 389-394. En ligne à l'adresse suivante : <https://www.aclweb.org/anthology/L18-1059>.
- OBERLE, B., SCHNEDECKER, C., BAUMER, E., CAPIN, D., GLIKMAN, J., GUO, C., REVOL, T., TODIRASCU, A. et TUSHKOVA, J. 2018. Les chaînes de référence dans les textes encyclopédiques du 12^e au 21^e siècle: étude longitudinale. *Travaux de linguistique* 77 : 67-141.
- OBRY, V., GLIKMAN, J., GUILLOT-BARBANCE, C. et PINCEMIN, B. 2017. Les chaînes de référence dans les récits brefs en français: étude diachronique (XIII^e-XVI^e s.). *Langue française* 195 : 91-110.
- MILARD, B. 2007. La mise en forme des publications scientifiques: entre routines, contraintes et organisation de l'expérience collective. In F. GAUDEZ et G. PEUCHLESTRADE (éd.), *Sociologie des arts, sociologie des sciences: actes du colloque international de Toulouse, 2004*. Paris: L'Harmattan : 203-212.
- MUZERELLE, J., LEFEUVRE, A., ANTOINE, J.-Y., SCHANG, E., MAUREL, D., VILLANEAU, J. et ESHKOL, I. 2013. ANCOR: premier corpus de français parlé d'envergure annoté en coréférence et distribué librement. In E. MORIN et Y. ESTÈVE (éd.), *Proceedings of TALN 2013 (Volume 2: Short Papers)*. Paris: Association pour le traitement automatique des langues (ATALA) : 555-563. En ligne à l'adresse suivante : <https://www.aclweb.org/anthology/F13-2007>.
- PERRET, M. 2000. Quelques remarques sur l'anaphore nominale aux XIV^e et XV^e siècles. *L'information grammaticale* 87 : 17-23.

- PONTILLE, D. 2003. Formats d'écriture et mondes scientifiques. Le cas de la sociologie. *Questions de communication* 3 : 55-67.
- PONTILLE, D. 2007. Matérialité des écrits scientifiques et travail de frontières: le cas du format IMRAD. In P. HERT et M. PAUL-CAVALLIER (éd.), *Sciences et frontières: délimitations du savoir, objets et passages*. Louvain-la-Neuve: Éditions modulaires européennes: 229-253.
- RECASENS, M., MARNEFFE, M.-C. DE et POTTS, C. 2013. The Life and Death of Discourse Entities: Identifying Singleton Mentions. In L. VANDERWENDE, H. DAUMÉ III et K. KIRCHHOFF (éd.), *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies – NAACL HLT 2013*. Stroudsburg: Association for Computational Linguistics (ACL): 627-633. En ligne à l'adresse suivante: <https://www.aclweb.org/anthology/N13-1000>.
- RÉGENT, O. 1980. Approche comparative des discours de spécialité pour l'entraînement à l'anglais écrit. *Mélanges CRAPPEL – Revue en didactique des langues et sociolinguistique* 1980: 117-135. En ligne à l'adresse suivante: <http://www.atilf.fr/IMG/pdf/melanges/6regent.pdf>.
- REIMERINK, A. 2006. The Use of Verbs in Research Articles: Corpus Analysis for Scientific Writing and Translation. *New Voices in Translation Studies* 2 : 9-27. En ligne à l'adresse suivante: <https://www.iatis.org/images/stories/publications/new-voices/Issue2-2006/reimerink-paper-2006.pdf>.
- RINCK, F. 2010. L'analyse linguistique des enjeux de connaissance dans le discours scientifique: un état des lieux. *Revue d'anthropologie des connaissances* 4 (3) : 427-450.
- SCHNEDECKER, C. 1997. *Nom propre et chaînes de référence*. Metz – Paris: Université de Metz – Klincksieck.
- SCHNEDECKER, C. 2005. Les chaînes de référence dans les portraits journalistiques: éléments de description. *Travaux de linguistique* 51 : 85-133.
- SCHNEDECKER, C. 2014. Chaînes de référence et variations selon le genre. *Langages* 195 : 23-42.
- SCHNEDECKER, C. 2017. Les chaînes de référence: une configuration d'indices pour distinguer et identifier les genres textuels. *Langue française* 195 : 53-72.
- SCHNEDECKER, C. 2019. De l'intérêt de la notion de chaîne de référence par rapport à celle d'anaphore et de coréférence. *Cahiers de praxématique* 72. En ligne à l'adresse suivante: <https://journals.openedition.org/praxematique/5339>.
- SCHNEDECKER, C. et LANDRAGIN, F. 2014. Les chaînes de référence: présentation. *Langages* 195 : 3-22.
- SCHNEDECKER, C. et LONGO, L. 2012. Impact des genres sur la composition des chaînes de référence: le cas des faits divers. In F. NEVEU, V. MUNI TOKE, P. BLUMENTHAL, T. KLINGLER, P. LIGAS, S. PRÉVOST et S. TESTON-BONNARD (éd.), *SHS Web of Conferences. Actes du 3^e congrès mondial de Linguistique française – CMLF 2012 (Lyon, France, 4-7 juillet 2012)*. Les Ulis: EDP Sciences. Vol. 1 : 1957-1972. En ligne à l'adresse suivante: https://www.shs-conferences.org/articles/shsconf/pdf/2012/01/shsconf_cmlf12_000061.pdf.
- SWALES, J. M. 1990. *Genre Analysis: English in Academic and Research Settings*. Cambridge – New York: Cambridge University Press.
- SWALES, J. M. 2004. *Research Genres: Explorations and Applications*. Cambridge – New York: Cambridge University Press.