



HAL
open science

Interdistance et instabilité au sein des chaînes de référence : indices textuels ?

Lucie Rousier-Vercruyssen, Frédéric Landragin

► To cite this version:

Lucie Rousier-Vercruyssen, Frédéric Landragin. Interdistance et instabilité au sein des chaînes de référence : indices textuels?. *Discours - Revue de linguistique, psycholinguistique et informatique*, 2019, 25, pp.3-32. 10.4000/discours.10522 . halshs-02533399

HAL Id: halshs-02533399

<https://shs.hal.science/halshs-02533399v1>

Submitted on 6 Apr 2020

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.



Discours

Revue de linguistique, psycholinguistique et informatique. A journal of linguistics, psycholinguistics and computational linguistics

25 | 2019
Varia

Interdistance et instabilité au sein des chaînes de référence : indices textuels ?

Lucie Rousier-Vercruyssen et Frédéric Landragin



Édition électronique

URL : <http://journals.openedition.org/discours/10522>

DOI : [10.4000/discours.10522](https://doi.org/10.4000/discours.10522)

ISSN : 1963-1723

Éditeur :

Laboratoire LATTICE, Presses universitaires de Caen

Référence électronique

Lucie Rousier-Vercruyssen et Frédéric Landragin, « Interdistance et instabilité au sein des chaînes de référence : indices textuels ? », *Discours* [En ligne], 25 | 2019, mis en ligne le 30 décembre 2019, consulté le 04 avril 2020. URL : <http://journals.openedition.org/discours/10522> ; DOI : <https://doi.org/10.4000/discours.10522>



Revue de linguistique, psycholinguistique et informatique

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

Interdistance et instabilité au sein des chaînes de référence : indices textuels ?

Lucie Rousier-Vercruyssen

Lattice (Langues, textes, traitements informatiques, cognition), CNRS, ENS Paris,
Université PSL (Paris sciences et lettres), Université Sorbonne Nouvelle

Frédéric Landragin

Lattice (Langues, textes, traitements informatiques, cognition), CNRS, ENS Paris,
Université PSL (Paris sciences et lettres), Université Sorbonne Nouvelle

.....
Lucie Rousier-Vercruyssen, Frédéric Landragin, « Interdistance et instabilité au sein des chaînes de référence :
indices textuels ? », *Discours* [En ligne], 25 | 2019, mis en ligne le 30 décembre 2019.

.....
URL : <http://journals.openedition.org/discours/10522>

.....
Titre du numéro : *Varia*

Coordination : Laure Sarda & Denis Vigier

Date de réception de l'article : 22/07/2019

Date d'acceptation de l'article : 13/12/2019



Presses
universitaires
de Caen 

Interdistance et instabilité au sein des chaînes de référence : indices textuels ?

Lucie Rousier-Vercruyssen

Lattice (Langues, textes, traitements informatiques, cognition), CNRS, ENS Paris,
Université PSL (Paris sciences et lettres), Université Sorbonne Nouvelle

Frédéric Landragin

Lattice (Langues, textes, traitements informatiques, cognition), CNRS, ENS Paris,
Université PSL (Paris sciences et lettres), Université Sorbonne Nouvelle

.....

De nombreux facteurs entrent en compte lors du choix des mentions référentielles. D'autant plus qu'il peut y avoir compétition entre les référents mis en jeu. Réalisée sur un extrait du corpus DEMOCRAT, cette étude a pour objectif d'examiner si la distance entre les mentions d'une chaîne de référence a un effet sur l'hétérogénéité lexicale de celle-ci. En ce qui concerne ces mesures, nos résultats montrent une indépendance entre le genre narratif et le genre non narratif. Il y a moins d'interdistance et plus de stabilité dans les textes narratifs que dans les textes non narratifs. Nous montrons également qu'il existe une corrélation inversée entre l'interdistance et l'instabilité référentielle selon le métagenre textuel. Dans les narrations, plus l'interdistance est grande, plus l'instabilité référentielle est présente. Inversement, plus l'interdistance est grande, moins l'instabilité référentielle est présente dans les autres textes. Cette étude constitue un premier pas vers une meilleure caractérisation des chaînes de référence en considérant le métagenre textuel et l'ensemble des référents.

Mots clés : chaîne de référence, expression référentielle, corrélation, interdistance, lexicalisation

.....

Many factors are taken into account when choosing the referential terms. Especially since there may be competition between the referents. Carried out on an extract from the corpus DEMOCRAT, this study aims to examine whether the distance between referring expressions in a reference chain has an effect on the lexical heterogeneity of this reference chain. With regard to these measures, our results show the independence between the narrative genre and the non narrative genre. There is less interdistance and more stability in narrative texts than in non narrative texts. We also show that there is a reverse correlation between interdistance and referential instability according to the textual metagenre. In narrations, the greater the interdistance, the more referential instability is present. Conversely, the greater the interdistance, the less referential instability is present in other texts. This study constitutes a first step towards a better characterisation of the reference chains with consideration of the textual metagenre and of all the referents.

Keywords: reference chain, referring expression, correlation, interdistance, lexicalization

.....

Nous tenons à remercier les membres du projet DEMOCRAT (« Description et modélisation des chaînes de référence : outils pour l'annotation de corpus et le traitement automatique », projet ANR-15-CE38-0008) pour les discussions enrichissantes lors des réunions et pour l'entraide technique. Sans ces réunions, ce travail n'aurait peut-être pas eu lieu car il émane de discussions qui ont suscité notre intérêt scientifique. Un grand merci à Frédérique Mélanie pour l'aide à la prise en main de l'outil TXM et à toutes les personnes qui nous ont aidés à réaliser cette étude. Enfin, un grand merci aux relecteurs anonymes pour leurs remarques et conseils, ainsi qu'à Denis Vigier et Laure Sarda pour leur relecture minutieuse et leurs suggestions pertinentes.

1. Introduction

1 Dans cet article, nous étudions les relations et la composition des expressions qui réfèrent aux objets et personnes. L'exemple [1] montre plusieurs expressions référentielles. Cet exemple est extrait du corpus DEMOCRAT (Landragin, 2018), réalisé dans le cadre d'un projet ANR. Le projet DEMOCRAT vise à développer les recherches sur la langue et la structuration textuelle du français *via* l'analyse détaillée et contrastive des chaînes de référence dans un corpus diachronique, avec des genres textuels variés et une annotation manuelle (Landragin, 2016). Dans l'exemple [1], quatre expressions réfèrent à la société («Une société», «qui» [×2] et «ses», dans la mesure où nous considérons que le possessif réfère). Ces quatre expressions référentielles – ou mentions – constituent une chaîne de référence car elles réfèrent à la même entité, à savoir «la société». Ces expressions sont proches les unes des autres, et pourtant diversifiées au niveau lexical. L'objectif de cet article est d'analyser statistiquement la notion de distance et de lexique sur différents genres textuels.

[1] [Une société] [qui] a évolué et [qui] a aussi fait évoluer [[ses] employés], notamment [Michel], passé par [différents types de poste], pour terminer [préparateur de [profilé de [sol]]].

2 Pour développer les problèmes de référence qui sont en jeu, il nous semble nécessaire de revenir sur la notion d'accessibilité référentielle (Ariel, 1988, 1990, 2001). Chaque expression encode de manière exclusive un degré spécifique d'accessibilité référentielle, allant du nom complet avec modifieur pour le référent d'accessibilité réduite au pronom clitique pour le référent d'accessibilité haute.

3 L'échelle d'accessibilité énonce quatre facteurs qui influencent le degré d'accessibilité d'un référent :

- la distance entre l'expression référentielle et son antécédent : plus la distance entre l'expression référentielle et son antécédent est grande, moins le référent est accessible ;
- la compétition : plus les antécédents sont nombreux, plus le degré d'accessibilité pour chacun sera bas ;
- la saillance de l'antécédent : un référent qui correspond au topique du discours est plus accessible ;
- l'unité : plus les antécédents font partie d'un script d'arrière-plan, de point de vue, de segment, plus le degré d'accessibilité pour chacun sera élevé.

4 Ainsi, selon cette échelle, aucun des facteurs qui affectent l'accessibilité (par exemple la distance, la compétition) ne peut expliquer à lui seul l'utilisation variée des expressions référentielles s'il est pris isolément. Il est nécessaire de considérer l'ensemble de ces facteurs pour comprendre cette variation. Un des points faisant débat pour cette échelle est l'encodage exclusif des expressions référentielles à un degré d'accessibilité d'un référent : chaque expression référentielle encode spécifiquement

un degré d'accessibilité d'un référent. Cela donne un caractère rigide à l'échelle d'accessibilité, puisqu'il n'est pas attendu, par exemple, qu'une expression référentielle introduite par un déterminant défini (codant un degré d'accessibilité intermédiaire) soit utilisée pour récupérer un référent cognitivement saillant (très accessible).

- 5 Le contexte référentiel, c'est-à-dire le contexte visuel partagé par les locuteurs ou le cotexte dans le cas d'un récit écrit, peut également impacter la composition de l'expression référant à une entité. Rousier-Vercruyssen (2017) a montré l'influence du contexte référentiel sur l'emploi des expressions référentielles lors de narrations d'histoires séquentielles. Le principe était de tester la production de sujets placés face à une série de six images décrivant – visuellement – une histoire. Selon la condition expérimentale retenue, ces images étaient présentées dans l'ordre naturel ou dans un ordre volontairement bouleversé. La tâche de narration d'histoires séquentielles a permis d'évaluer la façon dont un participant planifie son discours et, notamment, de déterminer le type d'information discriminante qu'il produit pour permettre à un destinataire d'identifier et d'ordonner les six images qui constituent la séquence de l'histoire. Les résultats montrent que les locuteurs tendent à employer davantage de syntagmes indéfinis et d'expansions lorsqu'ils introduisent un référent d'accessibilité cognitive basse. Lorsqu'ils mentionnent un référent d'accessibilité cognitive moindre – comme lors d'un changement de personnage en focus –, ils utilisent davantage de syntagmes définis et d'expansions. Les pronoms sont favorisés lors du maintien d'un référent d'accessibilité cognitive élevée. Lorsque les images ne se succèdent pas comme attendu pour un schéma d'action canonique, les locuteurs utilisent davantage de syntagmes définis et sur-spécifient lexicalement les syntagmes lorsqu'ils réfèrent aux personnages. De plus, la simple présence de compétiteurs – même de sexe différent – amène les locuteurs à davantage employer des syntagmes définis ou plus informatifs, manifestant une réduction de l'accessibilité cognitive des référents. Enfin, en présence de deux référents de sexe identique, une stratégie émerge chez les jeunes adultes (19-39 ans) lorsqu'ils mentionnent un référent d'accessibilité cognitive moindre lors du marquage référentiel : l'emploi des syntagmes définis.
- 6 L'emploi des expressions référentielles dépend également des (méta)genres textuels dans lesquels elles sont produites. Dans cet article, métagenre est employé comme hyperonyme de genre. Ces dernières sont moins nombreuses et davantage hyperonymiques dans les textes procéduraux par rapport aux romans (Tutin, 2002 ; Condamines, 2005 ; Schnedecker, 2017).
- 7 Actuellement, les expressions mentionnant un même référent sont souvent étudiées sur un empan plus ou moins large, isolément, et généralement limitées aux référents animés. Or, chacune de ces expressions constitue un « maillon » de la chaîne de référence, « elles construisent une relation d'identité référentielle » (Corblin, 1995 : 123) (Schnedecker, 1997 ; Schnedecker, 2017). À propos des référents non animés, nous pouvons citer les études d'Amsili *et al.* (2005) et d'Asher (1993) axées sur la représentation et l'interprétation sémantique des anaphores dénotant des entités abstraites telles que les événements, les situations, les propositions et les faits. L'analyse sur un empan large et sur différents types de référents serait une avancée.

- 8 Les chaînes de référence dépendent également des genres textuels (Schneidecker, 2014 et 2017) et de la langue (Baumer, 2015). Leur distribution et leur utilisation sont aussi influencées par les genres textuels (Boudreau et Kittredge, 2005). Les genres influencent également le nombre des chaînes de référence et de leurs maillons, et leur composition (Schneidecker et Longo, 2012). Tout comme l'entité référée, les chaînes de référence sont très hétérogènes. Certaines chaînes contiennent essentiellement des expressions référentielles avec une tête lexicale identique. À l'inverse, d'autres chaînes contiennent des expressions référentielles avec des têtes lexicales variées (Schneidecker, 2014). Cette variation lexicale peut être influencée par les genres textuels (Schneidecker, 2017). Au plan fonctionnel, les expressions référentielles produites au sein des contes de fées sont plus rapprochées en termes de mots par rapport à celles produites dans des faits divers (Schneidecker, 2017). Au plan lexical, les chaînes de référence produites au sein des contes de fées sont moins sujettes aux variations lexicales par rapport à celles produites dans des faits divers (Schneidecker, 2017). Cette faible variation lexicale – stabilité référentielle – faciliterait l'identification des personnages pour le lectorat jeune des contes (Schneidecker, 2017). La considération des lexèmes et de leur interdistance semble intéressante pour comprendre la composition des mentions au sein des chaînes de référence dans les métagenres.
- 9 Cet article est structuré de la manière suivante. En section 2, nous présentons le corpus dont sont extraits les textes étudiés, leur composition, la méthodologie d'extraction des données analysées et les tests statistiques utilisés. En section 3, nous exposons les résultats statistiques obtenus pour les notions d'interdistance et d'instabilité référentielle provenant des textes présentés précédemment. En section 4, nous discutons ces résultats et leurs apports théoriques.

2. Méthodologie

2.1. Données

2.1.1. *Corpus de départ*

- 10 L'étude s'est effectuée sur des extraits de textes provenant du corpus DEMOCRAT (2019). Les textes narratifs du XXI^e siècle sont absents de ce corpus pour des raisons de droits. Or, nous voulions malgré tout constituer un corpus avec des écrits récents, avec une représentation de la langue française contemporaine, ce qui explique la présence de textes non narratifs récents. Le corpus DEMOCRAT est un corpus textuel annoté en références. Les expressions référentielles sont repérées et annotées avec l'identifiant du référent, ce qui permet de construire les chaînes de référence. Ce corpus est composé de 57 fichiers, soit 688851 mots. L'annotation des expressions référentielles a été effectuée manuellement par les membres du projet et des stagiaires de niveau master. Pour ce faire, ils ont suivi un protocole d'annotation interne au projet DEMOCRAT. Toutes les expressions linguistiques ayant un pouvoir de référence (renvoyant à une entité du réel, de la fiction, etc.) ont été annotées, indépendamment de la question de coréférence (Landragin, 2018).

Fichier	Auteur	Titre	Époque	Métagenre	Nombre de mots	Nombre de chaînes de référence (nombre de mentions)
DIABLAUCORPS	Raymond Radiguet	<i>Le diable au corps</i>	XX ^e	Narratif	14 596	1 881 (4427)
DOUCELUMIERE	Marguerite Audoux	<i>Douce lumière</i>	XX ^e	Narratif	12 268	1 359 (3260)
JEANCHRISTOPHE-2	Romain Rolland	<i>Jean-Christophe (2)</i>	XX ^e	Narratif	10 303	1 189 (3037)
ARTICLEWIKI	NaN	Articles encyclopédiques « Zèbre », « Girafe », « Singe »	XXI ^e	Non narratif	10 628	1 447 (2907)
CONVENTIONTROPIC	NaN	« Convention relative au renforcement de la Commission interaméricaine du thon tropical établie par la convention de 1949 entre les États-Unis d'Amérique et la République du Costa Rica »	XXI ^e	Non narratif	11 173	1 765 (3391)
ESTREPUBLICAIN-2	NaN	« Est Républicain_2 » (collection de faits divers)	XX ^e	Non narratif	11 605	2 018 (2868)

Tableau 1 – Composition du corpus étudié

11 À la suite de cette annotation manuelle, les chaînes de référence ont été créées à partir de la macro «CreationChaine» du logiciel TXM développée pour le projet DEMOCRAT. Au total, 198 000 mentions ont été annotées, représentant 20 000 chaînes de coréférence qui comportent au minimum deux mentions, dont 9 000 chaînes comportant 3 mentions et plus. Au sein de ce corpus, nous avons extrait six textes datant du XX^e et du XXI^e siècle, trois textes narratifs et trois non narratifs.

12 Si l'on se fonde sur le nombre de mentions, la distribution des chaînes de référence semble différente entre les métagenres. Dans les textes narratifs (figures 1, 2, 3), la chaîne de référence la plus élevée comporte 895 mentions. Ce nombre est réduit à 211 mentions dans les textes non narratifs (figures 4, 5, 6).

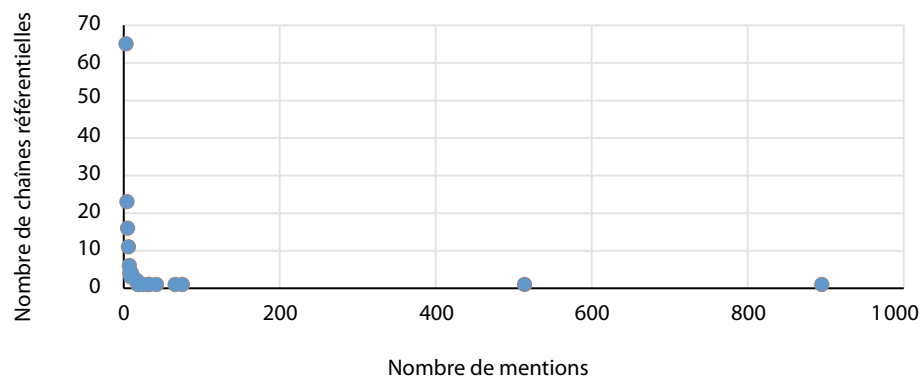


Figure 1 – Distribution des chaînes de référence dans «DIABLEAUCORPS»

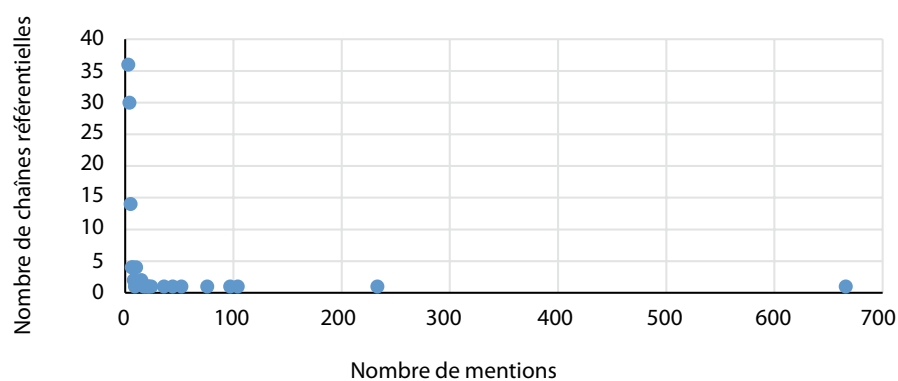


Figure 2 – Distribution des chaînes de référence dans «DOUCELUMIERE»

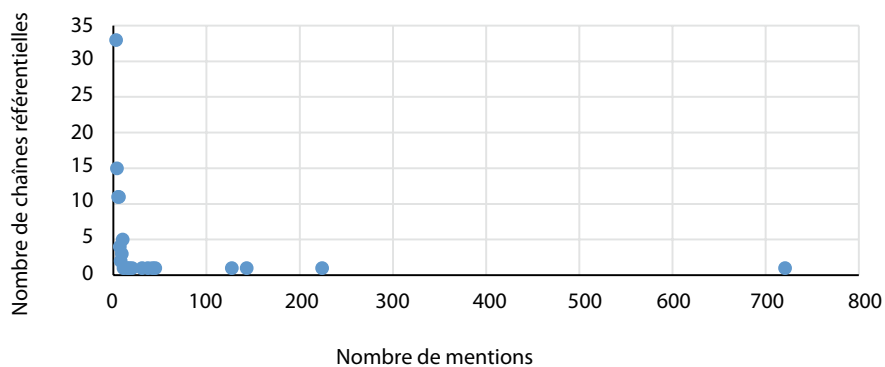


Figure 3 – Distribution des chaînes de référence dans «JEANCHRISTOPHE-2»

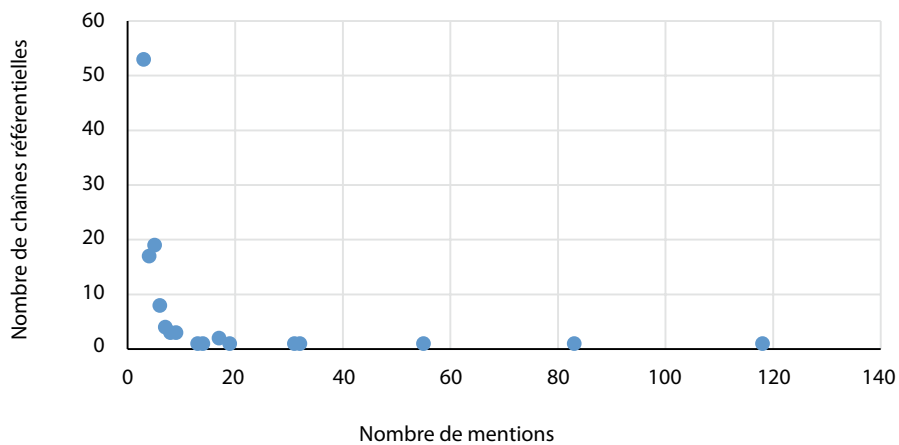


Figure 4 – Distribution des chaînes de référence dans «ARTICLEWIKI»

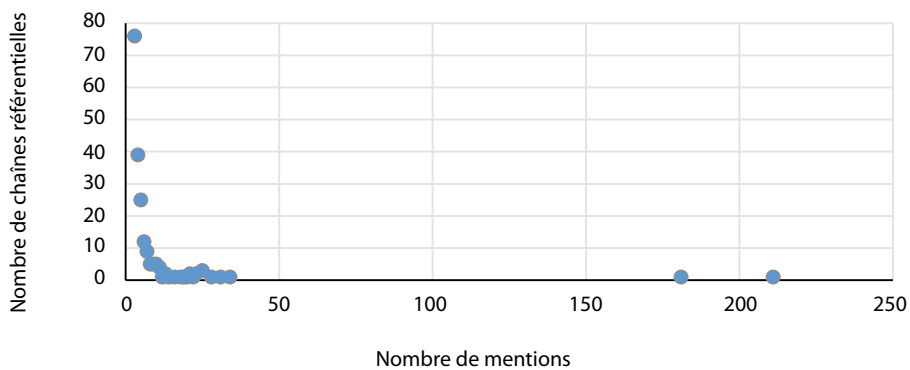


Figure 5 – Distribution des chaînes de référence dans «CONVENTIONTROPIC»

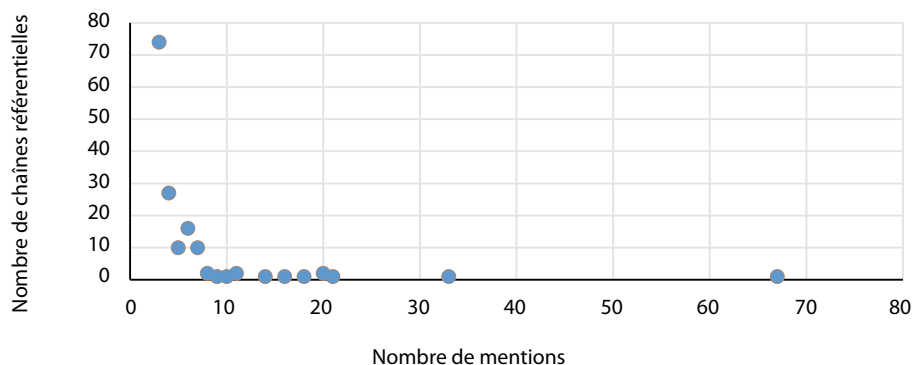


Figure 6 – Distribution des chaînes de référence dans «ESTREPUBLICAIN-2»

2.1.2. Données extraites

13 Le nombre de chaînes de référence varie fortement selon les textes. Pour obtenir une certaine comparabilité entre les fichiers, nous avons extrait les 10 chaînes de référence qui contiennent le plus de mentions pour chacun des textes. Au total, 60 chaînes de référence sont analysées par le biais du logiciel TXM (Heiden *et al.*, 2010) avec l'extension «Annotation URS (Unité-Relation-Schéma)». Cette extension permet d'annoter interactivement des unités au sein des éditions de textes de TXM, d'enrichir l'annotation par des commandes ou des macros, de vérifier sa cohérence et de procéder à diverses extractions pour l'affichage ou des décomptes. La macro «Schemas Lengths» nous a permis de trouver rapidement les dix chaînes contenant le plus de mentions.

14 Dans les textes narratifs, les 30 chaînes de référence analysées comportent entre 17 et 895 mentions. L'empan du nombre de mentions, compris entre 11 et 211 mentions, est beaucoup plus petit dans les textes non narratifs. Les chaînes de référence sont donc plus longues dans les textes narratifs par rapport aux non narratifs.

Fichier	Métagenre	Moyenne des mentions au sein des chaînes référentielles (nombre de chaînes de référence)	Étendue
DIABLEAUCORPS	Narratif	939 (77)	20-895
DOUCELUMIERE	Narratif	739 (110)	22-660
JEANCHRISTOPHE-2	Narratif	742,5 (96)	17-721
ARTICLEWIKI	Non narratif	212 (54)	13-118
CONVENTIONTROPIC	Non narratif	375,5 (201)	23-211
ESTREPUBLICAIN-2	Non narratif	125 (151)	11-67

Tableau 2 – Présentation globale des textes

15 Nous intéressant au rapport entre l'interdistance des mentions et l'instabilité lexicale des chaînes de référence, nous souhaitions connaître la répartition de ces chaînes au sein de chaque fichier. Par le biais de la macro «Schema Progression» (Quignard *et al.*, 2018), nous avons pu visualiser leur répartition (voir figures 7 à 12).

16 Deux grandes tendances se dégagent de ces visualisations. Au sein des textes narratifs, les chaînes de référence progressent continuellement (figures 7 à 9), c'est-à-dire que les référents sont mentionnés à différents moments. En revanche, la progression des chaînes de référence au sein des textes non narratifs ressemble davantage à un escalier (figures 10 à 12), c'est-à-dire qu'un référent n'est mentionné qu'à un moment et plus du tout par la suite ou bien est repris plus loin (effet palier).

17 L'interdistance référentielle correspond à la distance en nombre de mots entre les mentions d'une chaîne de référence. Dans l'exemple [2], il y a deux mots et une ponctuation entre les deux mentions de la chaîne de référence notée *i*. L'interdistance pour le référent *i* est donc de deux.

[2] [Julien]_i est fatigué. [Il]_i n'a pas dormi de [la nuit]_i.

18 Pour calculer cette interdistance, nous avons utilisé la macro «Units Interdistance» en spécifiant les chaînes de référence souhaitées. Les résultats montrent une interdistance plus courte pour les textes narratifs (moyenne = 108,5 mots) comparativement aux textes non narratifs (moyenne = 205 mots) (voir tableau 3).

19 Dans une chaîne de référence, certaines mentions peuvent être identiques. Lorsqu'une chaîne contient de nombreuses mentions identiques, elle est dite «stable». La notion de stabilité référentielle a été calculée pour chaque chaîne à partir de la macro «Units Referential Stability». Cette macro permet d'obtenir un coefficient de stabilité (Perret, 2000; Schnedecker, 2017). Ce coefficient correspond au quotient du nombre total d'anaphores nominales divisé par le nombre de désignations différentes pour chaque référent. Plus le coefficient est élevé, moins les désignations sont différentes, et donc plus la stabilité référentielle est grande. Bien que ce coefficient soit très intéressant, il est difficilement utilisable pour des tests de comparaison. En effet, les nombres obtenus ont chacun leurs propres limites. Par exemple, une chaîne de référence, notée *k*, comportant dix mentions et un lexème aura un coefficient de dix. Une autre, notée *l*, comportant quatre mentions et un lexème aura un coefficient de quatre. La chaîne *k* a un coefficient plus élevé, mais elle contient autant de lexèmes «différents» (1) au regard du nombre de mentions comparativement à la chaîne *l*. Il semble donc difficile de pouvoir comparer ces chaînes entre elles en considérant uniquement le coefficient de stabilité. Nous avons donc choisi de calculer un pourcentage d'instabilité. Ce pourcentage correspond au quotient du nombre de désignations différentes divisé par le nombre total d'anaphores nominales pour chaque référent et multiplié par cent. Ce pourcentage permet une lecture rapide de la stabilité référentielle et une comparabilité entre les chaînes de référence. Plus le pourcentage est élevé, plus il y a de désignations différentes, et donc moins la chaîne de référence est stable. Reprenons les chaînes *k* et *l*, leur pourcentage d'instabilité est respectivement de 10 % ($1/10 \times 100$) et de 25 % ($1/4 \times 100$). La chaîne *k* est donc moins instable que la chaîne *l*.

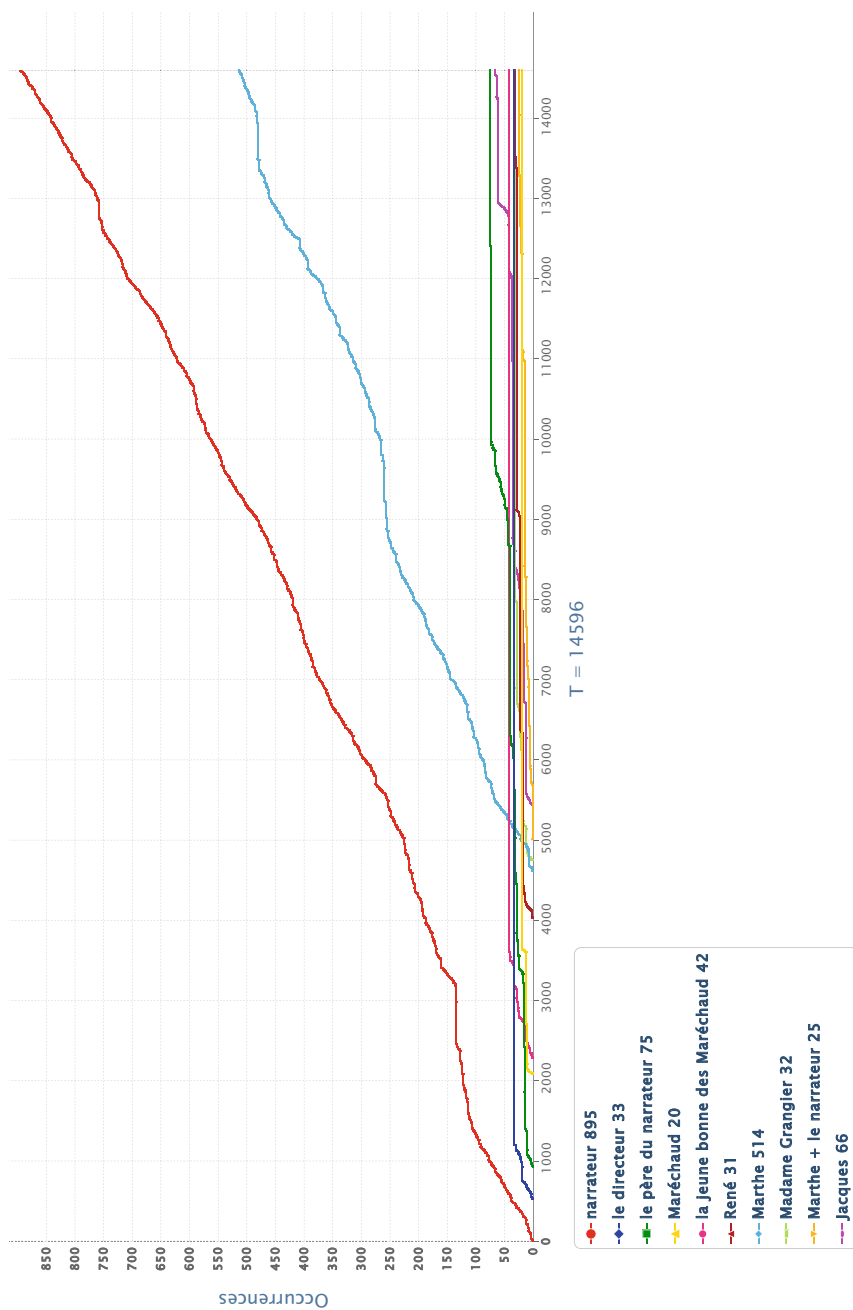


Figure 7 – Schéma de progression dans « DIABLEAUCORPS »

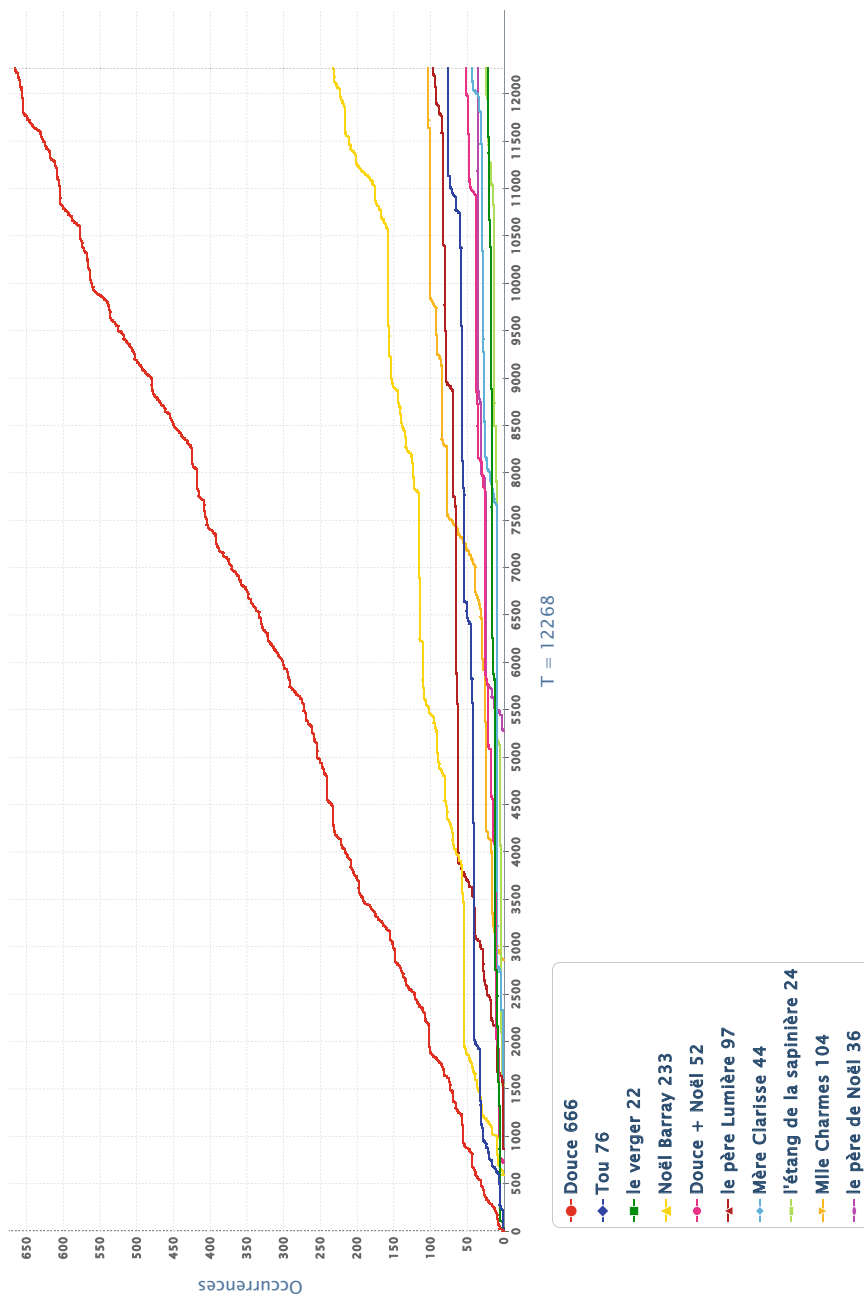


Figure 8 – Schéma de progression dans « DOUCELUMIERE »

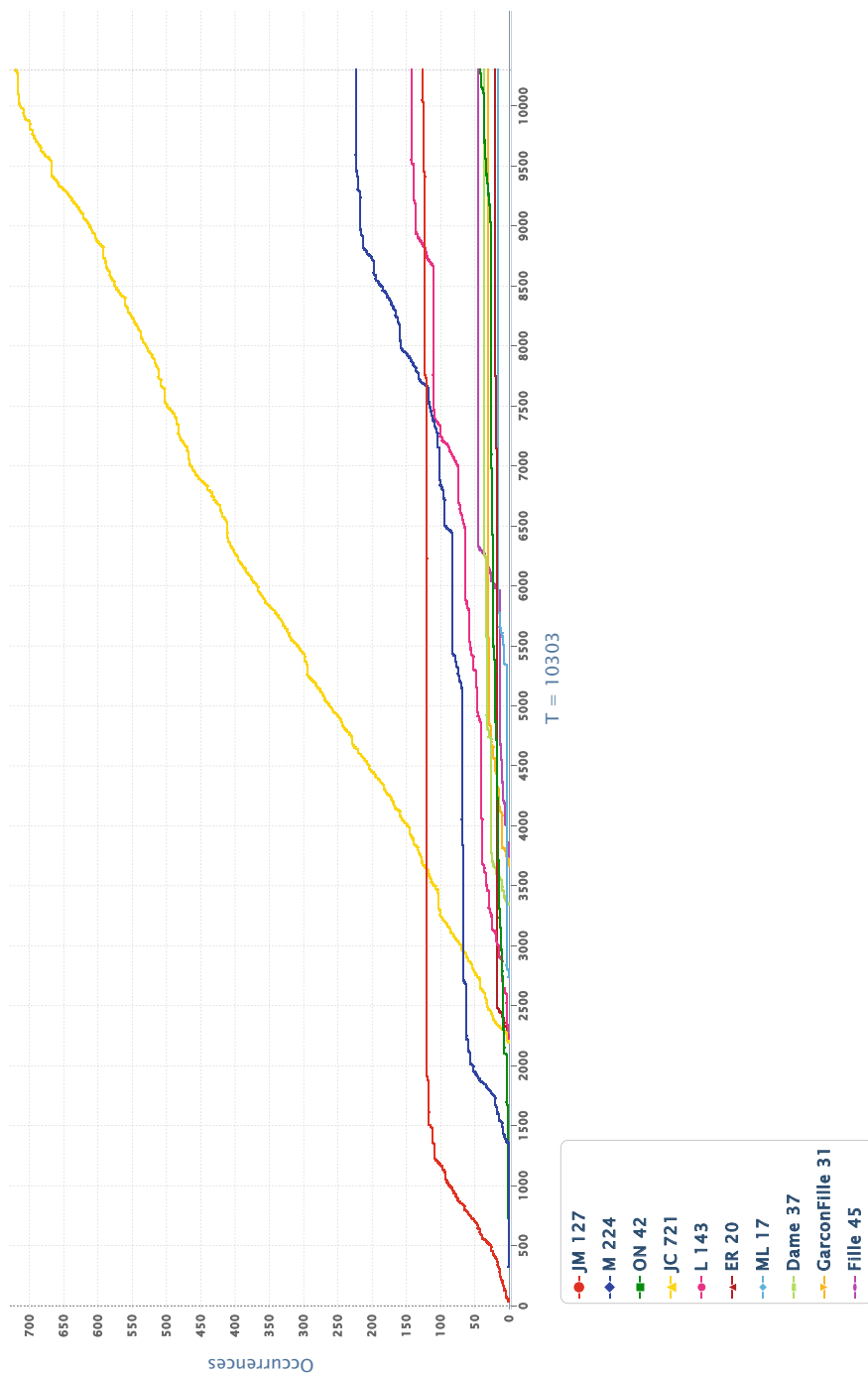


Figure 9 – Schéma de progression dans «JEANCHRISTOPHE-2»

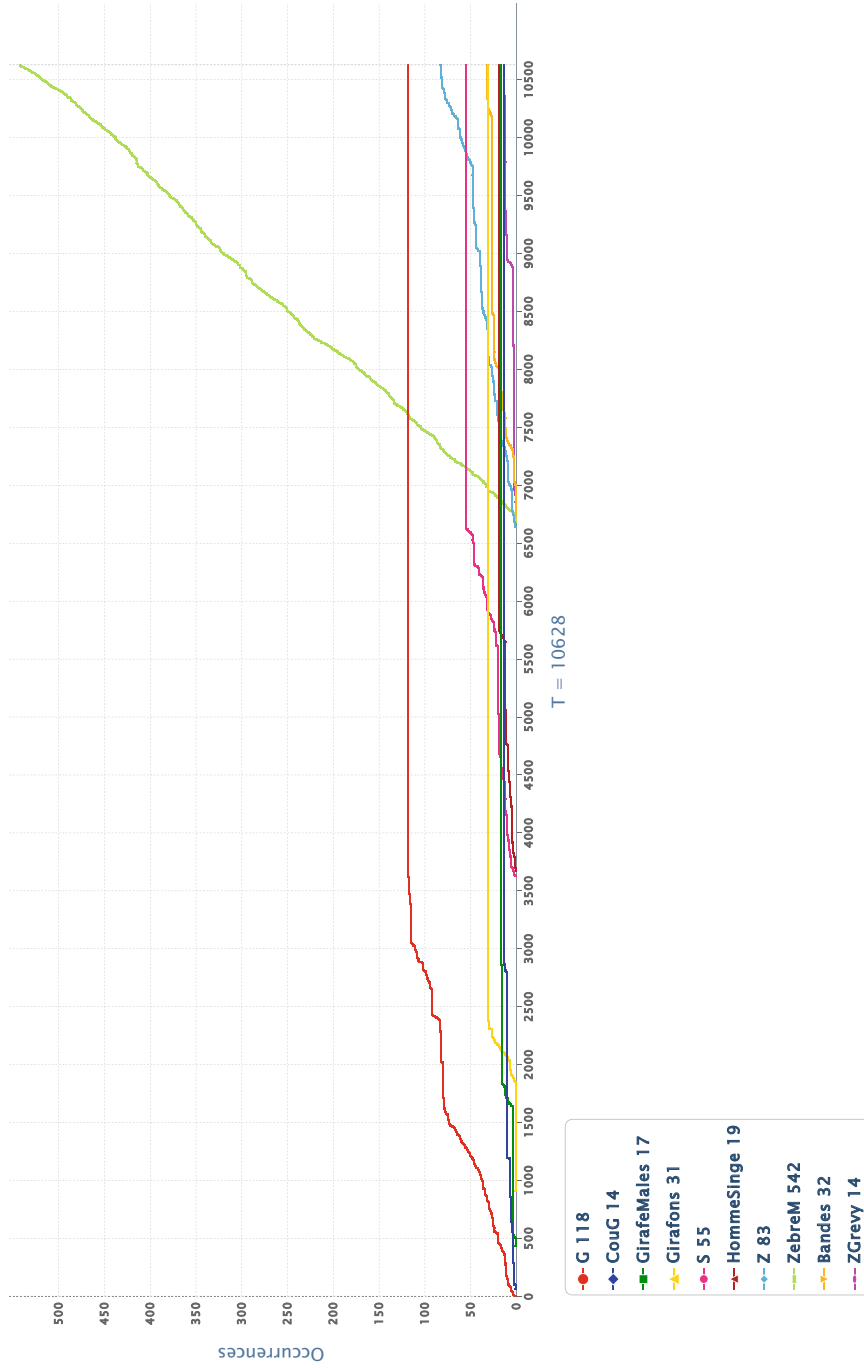


Figure 10 – Schéma de progression dans «ARTICLEWIKI»

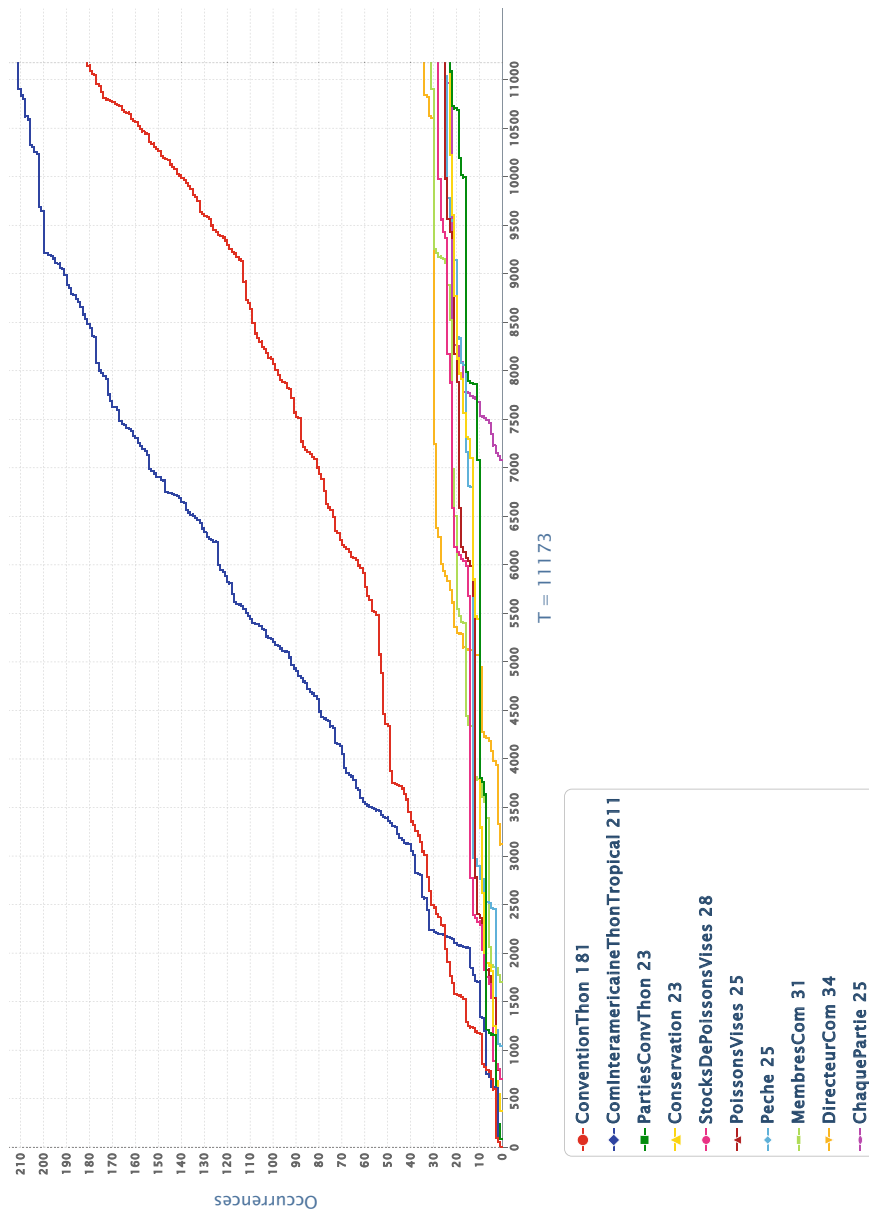


Figure 11 – Schéma de progression dans «CONVENTIONTROPIC»

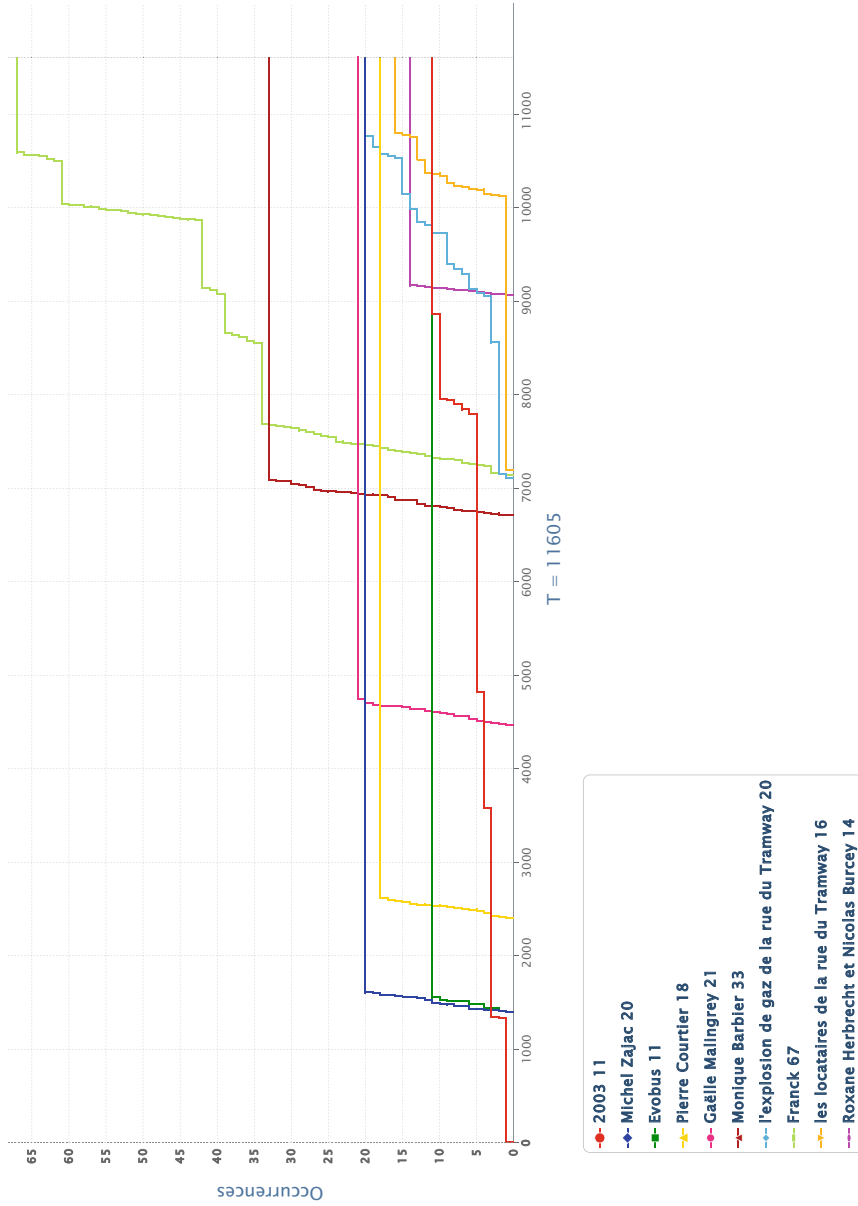


Figure 12 – Schéma de progression dans « ESTREPUBLICAIN-2 »

Texte	Chaîne de référence (nombre de mentions)	Interdistance en nombre de mots
DIABLEAUCORPS	narrateur (895)	16,28
	Marthe (514)	19,39
	le père du narrateur (75)	152,58
	Jacques (66)	137,45
	la jeune bonne des Maréchaud (42)	31
	le directeur (33)	20,39
	Madame Grangier (32)	243,31
	René (31)	305,93
	Marthe + le narrateur (25)	360,72
	Maréchaud (20)	76,95
DOUCELUMIERE	Douce (666)	18,25
	Noël Barray (233)	49,91
	Mlle Charmes (104)	84,83
	le père Lumière (97)	116,38
	Tou (76)	145,86
	Douce + Noël (52)	216,13
	Mère Clarisse (44)	253,88
	le père de Noël (36)	163,66
	l'étang de la sapinière (24)	434,54
	le verger (22)	522,68
JEANCHRISTOPHE-2	JC (721)	11,17
	M (224)	41,27
	L (143)	51,05
	JM (127)	78,68
	Fille (45)	57,31
	ON (42)	227,07
	Dame (37)	78,43
	GarçonFille (31)	60,48
	ER (20)	275,15
	ML (17)	178,52

ARTICLEWIKI	ZGrevy (14)	115,64
	G (118)	30,07
	Z (83)	47,37
	S (55)	53,12
	Bandes (32)	112,34
	Girafons (31)	47,12
	HommeSinge (19)	108,05
	GirafeMales (17)	141,88
	CouG (14)	199
	GirafeMere (13)	15,53
CONVENTIONTROPIC	ComInteramericaineThonTropical (211)	50,44
	ConventionThon (181)	59,48
	DirecteurCom (34)	226,20
	MembresCom (31)	294,25
	StocksDePoissonsVises (28)	323,75
	PoissonsVises (25)	366,48
	Peche (25)	399,72
	ChaquePartie (25)	155,16
	PartiesConvThon (23)	476,26
	Conservation (23)	428,13
ESTREPUBLICAIN-2	Franck (67)	50,85
	Monique Barbier (33)	11,21
	Gaëlle Malingrey (21)	12,95
	Michel Zajac (20)	9,85
	l'explosion de gaz de la rue du Tramway (20)	181
	Pierre Courtier (18)	11,33
	les locataires de la rue du Tramway (16)	224,37
	Roxane Herbrecht et Nicolas Burcey (14)	6,71
	Evobus	13,54
	2003	804,18

Tableau 3 – Récapitulatif de l'interdistance référentielle pour les chaînes de référence étudiées

20 Les résultats montrent une instabilité plus forte pour les textes narratifs (moyenne = 40 %) comparativement aux textes non narratifs (moyenne = 25 %).

21 Dans l'exemple présent en annexe, la chaîne de référence «Evobus» (mentionnées par des crochets et en gras) comporte onze mentions. Le nombre de mentions pourrait être plus élevé si nous acceptions des références non strictes (Delaborde et Landragin, 2018). Par exemple, le pronom « nous » dans « le bienvenu chez nous » n'est pas pris en compte car il peut référer à une partie d'«Evobus», il n'est pas strict. Les onze mentions sélectionnées réfèrent à l'entreprise comme une personne morale. En ce sens, elles sont des références «strictes». Cette chaîne côtoie 29 autres chaînes de référence (mentionnées par de simples crochets). Son interdistance est de 13,54 ($149/11 = 13,54$) et son pourcentage d'instabilité est de 72,72 % ($8/11 \times 100 = 72,72$).

2.2. Tests statistiques utilisés

22 Cette section présente les analyses statistiques utilisées pour cette étude.

2.2.1. Comparaison de moyennes

23 Avant d'examiner le lien entre l'interdistance et l'instabilité des chaînes de référence, nous examinons la distribution des données au sein des textes narratifs et non narratifs. Les tests de comparaison des moyennes permettent de savoir si la distribution des données est identique au sein des groupes. Lorsque le résultat est significatif, l'hypothèse nulle est rejetée, ce qui veut dire que la différence entre les groupes analysés est confirmée. La valeur d'un groupe est donc supérieure à la valeur de l'autre groupe. Le test de Mann-Whitney (Mann et Whitney, 1947) est utilisé car il permet de comparer des moyennes de données indépendantes, à savoir une valeur par chaîne de référence.

2.2.2. Test de corrélation

24 Les coefficients de corrélation ont été exécutés pour examiner les relations entre l'interdistance et l'instabilité. Les données étant non paramétriques, nous avons utilisé le coefficient de corrélation de Spearman «Rs» (Dodge, 2008).

25 Une corrélation est considérée comme parfaite si la valeur du test est de +1 ou -1 et nulle si la valeur est de 0. Lorsque la valeur est supérieure à 0, la corrélation est positive. Et inversement, lorsque la valeur est inférieure à 0, la corrélation est négative. Le coefficient de corrélation est approprié aux variables quantitatives (variables contenant des valeurs mesurables) et ordinales (variables qualitatives numériques classifiant un ensemble). Les corrélations significatives ont permis d'examiner le lien entre l'interdistance et l'instabilité référentielle.

2.2.3. Régression linéaire

26 La régression permet d'examiner si l'interdistance et l'instabilité référentielle ont une relation de causalité, c'est-à-dire si l'interdistance détermine l'instabilité référentielle.

27 La régression linéaire permet d'expliquer une variable Y à l'aide d'une variable X (James *et al.*, 2013). La variable Y est la variable dépendante, celle à expliquer et la variable X est la variable indépendante, la variable explicative. Il faut néanmoins faire attention aux relations de causalité formulées, l'existence d'une relation entre X et Y n'implique pas nécessairement une relation de causalité entre elles.

3. Résultats

28 L'objectif de cette étude est d'examiner si l'interdistance d'une chaîne de référence a un effet sur l'hétérogénéité lexicale de celle-ci.

29 Si c'est le cas, une interdistance faible entraînerait une hétérogénéité. Le référent étant accessible, le locuteur peut se permettre une diversité lexicale.

30 À l'inverse, une interdistance élevée entraînerait davantage une homogénéité. Le référent étant moins accessible, l'emploi de lexèmes identiques faciliterait la compréhension.

3.1. Impact des métagenres

31 Dans les figures 13 et 14, nous remarquons que la plupart des mentions référentielles, représentées par la médiane (trait horizontal central à l'intérieur des boîtes, scindant de manière égale les valeurs supérieures et inférieures), sont moins rapprochées et diversifiées dans les textes non narratifs comparativement aux textes narratifs. Les autres traits horizontaux présents dans les figures correspondent aux quartiles (par exemple, pour le trait le plus bas, un quart des observations ont des valeurs plus petites et trois quarts ont des valeurs plus grandes).

32 Le test de Mann-whitney (Mann et Whitney, 1947) révèle une différence statistique entre les deux métagenres: la distribution de l'interdistance et celle de l'instabilité référentielle sont différentes entre les deux métagenres ($U = 652$, $p < 0,05$ pour l'interdistance, voir figure 13; $U = 669$, $p < 0,05$ pour l'instabilité référentielle, voir figure 14). L'interdistance est statistiquement plus élevée dans les textes non narratifs comparativement aux textes narratifs. À l'inverse, l'instabilité référentielle est statistiquement plus élevée dans les textes narratifs comparativement aux textes non narratifs.

33 Dans la figure 14, plus le pourcentage est élevé, moins les mentions sont identiques, donc plus la chaîne de référence est instable. Les métagenres étant significativement différents, les textes narratifs et non narratifs sont étudiés indépendamment.

3.2. Métagenre narratif

34 Le test de corrélation révèle une corrélation positive entre l'interdistance et l'instabilité référentielle ($R_s = 0,37$, $p < 0,05$, voir figure 15). Plus la chaîne de référence est instable, plus l'interdistance est élevée. Le résultat de la régression linéaire n'est statistiquement pas concluant. L'interdistance n'a pas d'impact sur l'instabilité référentielle.

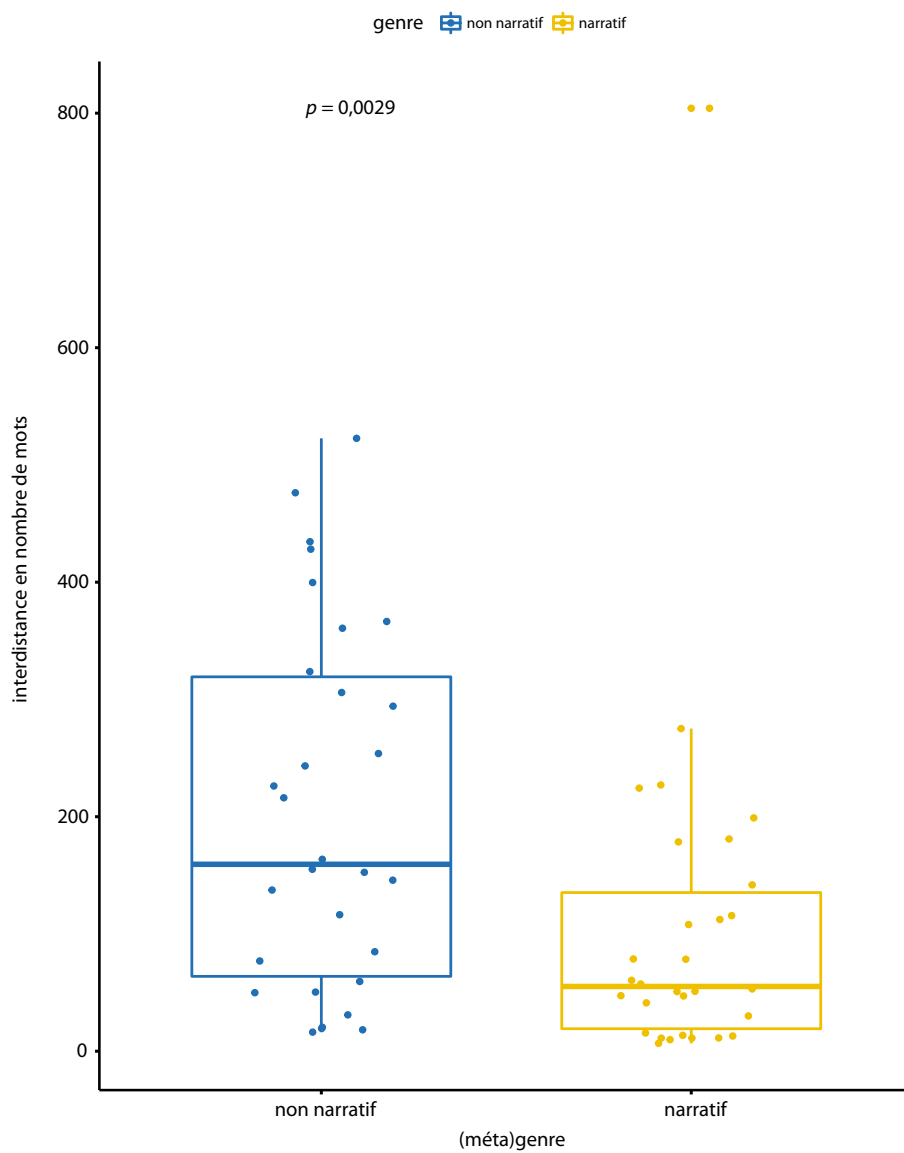


Figure 13 – Interdistance référentielle selon le métagenre

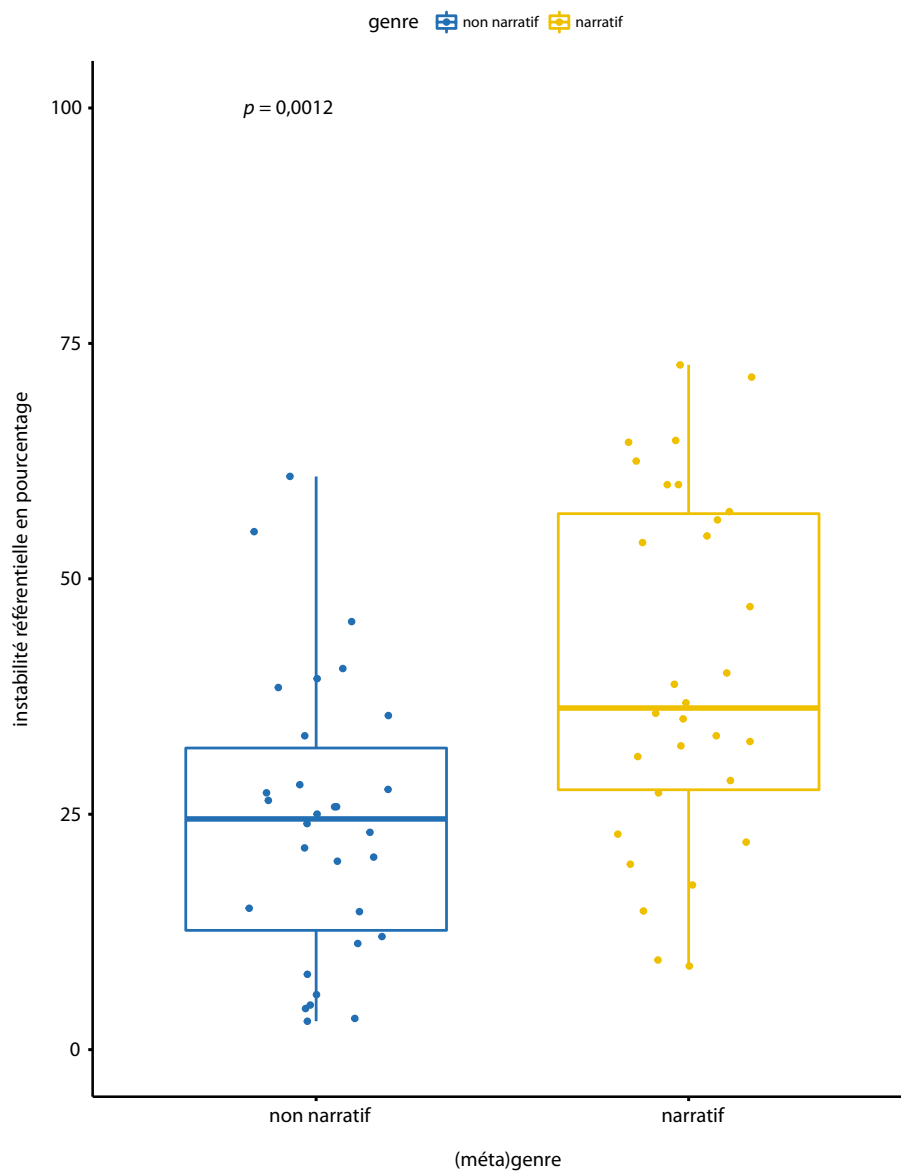


Figure 14 – Instabilité référentielle selon le méta-genre

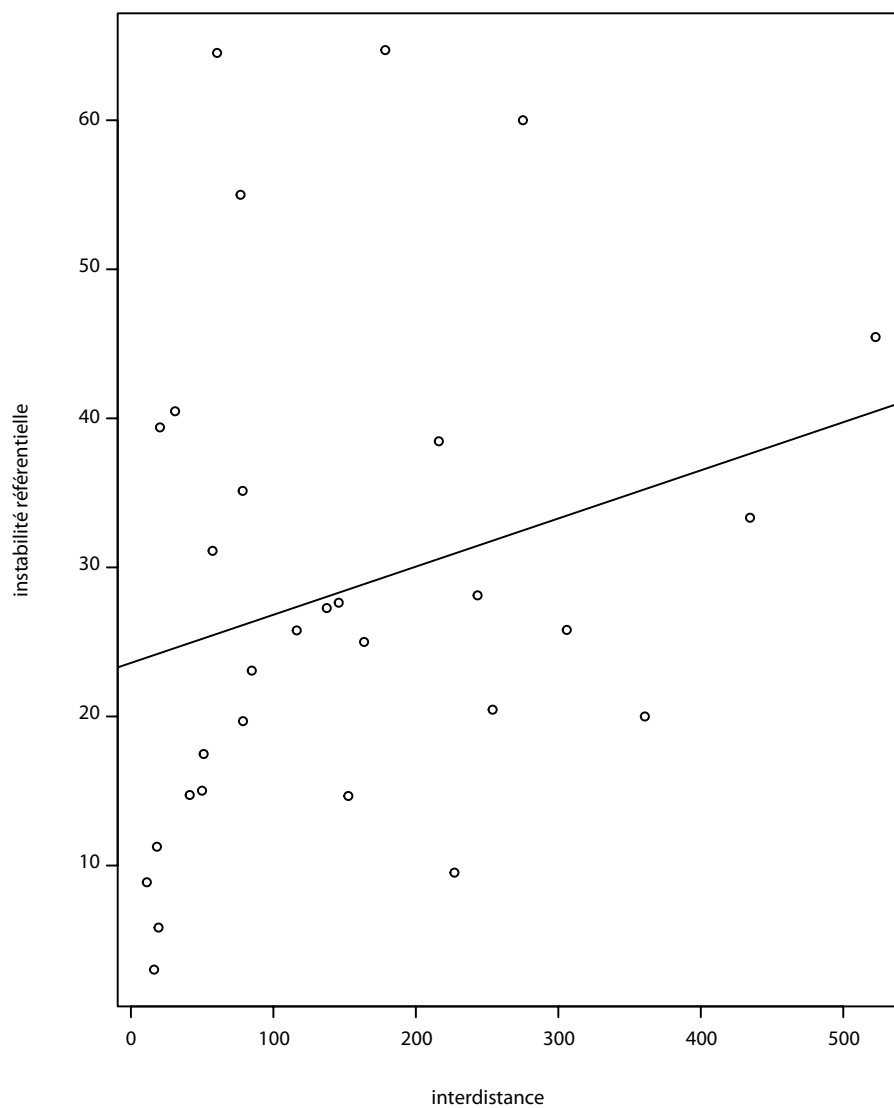


Figure 15 – Corrélation entre interdistance et instabilité référentielle dans les textes narratifs

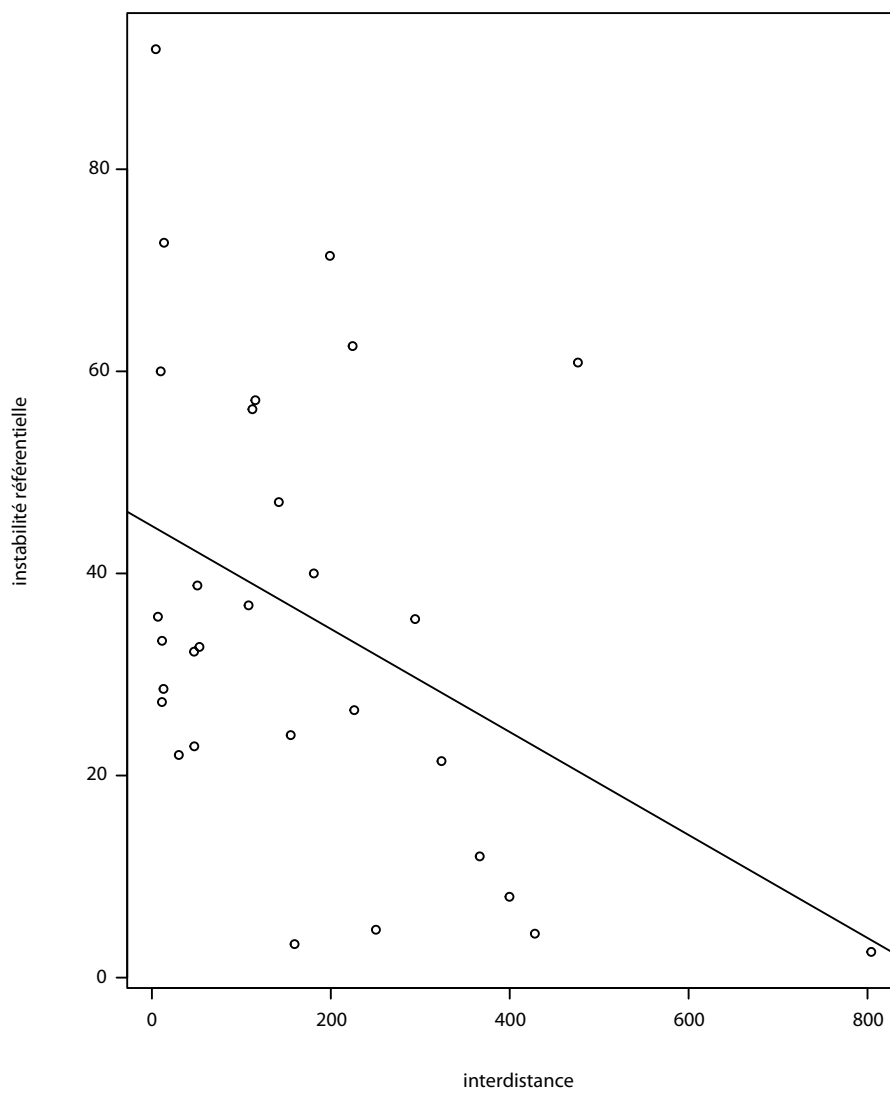


Figure 16 – Corrélation entre interdistance et instabilité référentielle dans les textes non narratifs

3.3. Métagenre non narratif

35 L'interdistance et l'instabilité référentielle ont une corrélation négative au sein des textes non narratifs ($R_s = -0,39$, $p < 0,05$, voir figure 16). Plus la chaîne de référence est instable, moins l'interdistance est élevée.

36 Le résultat de la régression linéaire montre un effet significatif de l'interdistance sur l'instabilité d'une chaîne de référence. Les chaînes de référence ont moins de probabilités d'être instables lorsque l'interdistance devient de plus en plus grande (tableau 4).

Effets fixes	P	Écart type	Valeur de p
(Intercept)	44,69935	5,40	< 0,05
Interdistance	-0,05097	0,02	< 0,05

Tableau 4 – Régression linéaire entre interdistance et instabilité référentielle dans les textes non narratifs

4. Discussion

37 Cette étude avait pour objectif d'examiner le lien entre interdistance et instabilité référentielle au sein de textes narratifs et non narratifs. Pour rappel, elle se base non seulement sur les chaînes de référence contenant le plus de mentions au sein des textes (les chaînes de référence ne sont pas isolées) mais aussi sur l'ensemble des référents (au-delà du genre animé). Les études portant sur les chaînes de référence ont tendance à les analyser comme des listes de maillons sans tenir compte du texte dont elles proviennent et sans mémoriser les expressions référentielles qui évoquent d'autres référents en concurrence. La plupart de ces études sont des études de quantification où les statistiques servent à un décompte de maillons. Cela ne fait que quelques années que la quantification et la visualisation sont appliquées. En effet, les diagrammes de progression (*d*), issus du projet ANR DEMOCRAT, sont très utiles pour visualiser les expressions de chaque chaîne de référence au sein du texte.

38 Ce travail résulte de besoins exprimés par les linguistes de ce projet lors de discussions englobant différents points de vue au cours des journées d'étude. L'un des besoins est notamment une meilleure caractérisation des chaînes de référence. Cette étude constitue un premier pas vers cette caractérisation en s'appuyant sur nos expertises (statistiques applicables aux expressions référentielles et mise en regard avec des habiletés sociocognitives).

4.1. Interdistance, instabilité référentielle et (méta)genres

39 Notre premier résultat présente une différence statistique entre les deux métagenres – narratif et non narratif – pour les mesures d'interdistance et d'instabilité référentielle. L'interdistance et la stabilité référentielle sont plus faibles dans les textes narratifs

comparativement à celles dans les textes non narratifs. Ces résultats ne rejoignent pas les résultats de Schnedecker (2017) qui avait remarqué une stabilité référentielle plus élevée au sein des contes de fées par rapport aux faits divers sans indiquer leur valeur. Ces différences peuvent être occasionnées par l'utilisation d'indices différents : coefficient de stabilité pour Schnedecker (2017) et pourcentage d'instabilité pour nous. Contrairement au coefficient de stabilité, le pourcentage d'instabilité fixe une limite d'instabilité (100 %) qui permet de rendre les chaînes de référence comparables entre elles.

40 Le genre discursif semble donc avoir un impact non seulement sur la répartition des chaînes de référence mais aussi sur la composition des mentions au sein de ces chaînes. Le genre textuel devrait donc être considéré dans les facteurs qui influencent le degré d'accessibilité d'un référent.

41 L'étude de la corrélation révèle un rapport inversé entre ces deux métagenres. Cette corrélation est positive pour le métagenre narratif. Plus la chaîne de référence connaît une hétérogénéité lexicale, plus ses mentions sont espacées. Cette instabilité peut être permise dans un contexte narratif car un référent principal est «ancré». L'auteur se permet cette liberté lexicale car il a précédemment instancié le référent dont il est question. Néanmoins, cette instabilité référentielle peut rendre l'interprétation des chaînes de référence plus difficile (Schnedecker et Landragin, 2014).

42 *A contrario*, la corrélation est négative pour le métagenre non narratif. De plus, l'interdistance influence l'instabilité des chaînes de référence. Un rapprochement des mentions d'un même référent prédit une hétérogénéité lexicale au sein de la chaîne de référence. Dans ce métagenre, les chaînes de référence s'entrecroisent très peu, il n'y a donc potentiellement qu'un référent accessible simultanément. Ce résultat confirme les facteurs de compétition et de distance de l'échelle d'accessibilité (Ariel, 1988, 1990, 2001). Selon le facteur de compétition, plus les antécédents sont nombreux, plus le degré d'accessibilité pour chacun sera bas. Selon le facteur de distance, plus la distance entre l'expression référentielle et son antécédent est grande, moins le référent est accessible. Ainsi, en présence de peu de compétiteurs, les auteurs «se permettent» cette liberté lexicale car le degré d'accessibilité du référent n'est pas bas et réduisent cette liberté lorsque l'écart entre les mentions devient trop grand.

4.2. Expressions référentielles et habiletés (socio)cognitives : collaboration – égocentrisme

43 Les liens entre interdistance et stabilité référentielle peuvent manifester un acte collaboratif ou coopératif. Lorsque les chaînes de référence s'entrecroisent, la conception d'une certaine stabilité référentielle peut manifester un acte collaboratif ou coopératif car cela aide le lecteur à suivre. En effet, l'utilisation de lexèmes identiques (comme «La société Evobus» et «la société» dans l'exemple présenté en annexe) pour mentionner un référent augmente la rapidité de résolution de la référence car ce traitement a précédemment été effectué. Ce lien entre habiletés sociocognitives (ensemble des processus mentaux – empathie et théorie de l'esprit – guidant

les interactions sociales [Achim *et al.*, 2012] et correspondant à la perception, l'interprétation et la production de réponses face aux intentions et comportements d'autrui [Green *et al.*, 2008]) et expressions référentielles peut être illustré par Achim *et al.* (2015). Par le biais du paradigme de communication référentielle qui permet de recréer une situation de communication avec une interaction verbale impliquant la collaboration entre deux locuteurs (Clark et Wilkes-Gibbs, 1986), les auteurs ont demandé aux participants de mentionner des personnages de films présents sur des cartes afin de permettre à leurs interlocuteurs de les replacer dans le même ordre que le leur. Leurs résultats ont montré que les locuteurs qui ont une meilleure capacité de théorie de l'esprit (définie comme étant la capacité du participant à attribuer des états mentaux, des croyances, des connaissances, des intentions, des désirs et des émotions [Achim *et al.*, 2012]) ajustent davantage leurs expressions référentielles en considérant leurs interlocuteurs. Ils ajoutent des informations lexicales au sein des marqueurs syntaxiques pour évoquer des référents supposés non connus par leurs interlocuteurs. Une aptitude collaborative se manifeste à travers ces résultats car les locuteurs utilisent leur savoir partagé construit au cours des interactions pour choisir le marqueur adéquat.

44 Néanmoins, le choix d'une stabilité référentielle peut également manifester un acte égocentrique ou une économie d'énergie. La production de la mention référentielle est un processus complexe et influencé par de multiples facteurs linguistiques (pragmatiques, syntaxiques) et extralinguistiques (contexte référentiel, métagenre, compétences sociocognitives et cognitives correspondant aux processus mentaux et mécanismes qui nous permettent d'effectuer des activités diverses et complexes à la base de la pensée, de l'action et de la communication [Baciu, 2011]). En effet, Rousier-Vercruyssen (2017) a montré l'influence de la flexibilité (définie comme étant la «capacité à réorienter les contenus de pensée et l'action afin d'être à même de percevoir, traiter et réagir aux situations de différentes manières» [Leclercq et Zimmermann, 2014: 151]) sur l'emploi des expressions référentielles lors de narrations d'histoires séquentielles. La flexibilité est notamment impliquée lorsque les locuteurs sont censés «désactiver» un référent accessible pour en mentionner un moins accessible. L'implication de cette capacité indique la difficulté de mettre de côté un référent maintenu pour un autre d'accessibilité cognitive moindre. Les productions spontanées peuvent rendre plus difficiles les reprises et corrections. Ce qui n'est pas nécessairement le cas pour l'écrit, l'auteur peut plus aisément reprendre son texte et le corriger. Néanmoins, l'emploi d'une mention quasi-identique pour un même référent peut permettre à l'auteur de consacrer cette énergie économisée à :

- la planification de la suite de son texte ;
- la gestion de la flexibilité référentielle lorsque les chaînes de référence s'entrecroisent (par exemple, la mention de multiples référents dans un court empan peut facilement engendrer une ambiguïté référentielle et devenir d'autant plus difficile si l'on varie leur tête lexicale) ;
- la gestion de son style d'écriture.

5. Conclusion et perspectives

- 45 Cette étude a permis d'analyser la relation entre interdistance et instabilité référentielle, relation dépendante du métagenre auquel appartient le texte. En effet, l'étude de la corrélation révèle un rapport inversé entre ces deux métagenres. La corrélation est positive dans les textes narratifs, et négative dans les textes non narratifs. De plus, la relation entre ces deux notions est causale au sein des textes non narratifs : plus l'interdistance augmente, moins les chaînes de référence deviennent instables. Ces relations manifestent des particularités selon les métagenres. Ces particularités sont spécifiées par l'indépendance statistique de l'interdistance et l'instabilité référentielle entre les deux métagenres. Il semble donc nécessaire de considérer le métagenre textuel pour en connaître le fonctionnement référentiel.
- 46 Il serait intéressant d'étendre ces analyses en considérant d'autres genres textuels et d'autres périodes du français. L'étude des *n*-grammes semble également intéressante pour mieux appréhender la caractérisation des chaînes de référence. En effet, une étude des prototypes référentiels par le biais des *n*-grammes a montré que les référents principaux sont davantage présents dans les enchaînements de trigrammes et bigrammes (Landragin *et al.*, 2016). Cette approche permet de considérer non seulement les relations entre les expressions référentielles d'une chaîne de référence mais aussi, et surtout, leur relation avec d'autres chaînes de référence.

Références bibliographiques

- ACHIM, A. M., FOSSARD, M., COUTURE, S. et ACHIM, A. 2015. Adjustment of Speaker's Referential Expressions to an Addressee's Likely Knowledge and Link with Theory of Mind Abilities. *Frontiers in Psychology* 6 (823): 1-11. En ligne à l'adresse suivante : <https://www.frontiersin.org/articles/10.3389/fpsyg.2015.00823/pdf>.
- ACHIM, A. M., OUELLET, R., ROY, M. A. et JACKSON, P. L. 2012. Mentalizing in First-Episode Psychosis. *Psychiatry Research* 196 (2-3): 207-213.
- AMSILI, P., DENIS, P. et ROUSSARIE, L. 2005. Anaphores abstraites en français : représentation formelle. *Traitement automatique des langues* 46 (1): 15-40.
- ARIEL, M. 1988. Referring and Accessibility. *Journal of Linguistics* 24 (1): 65-87.
- ARIEL, M. 1990. *Accessing Noun-Phrase Antecedents*. Londres – New York : Routledge.
- ARIEL, M. 2001. Accessibility Theory: An Overview. In T. SANDERS, J. SCHILPEROORD et W. SPOOREN (éd.), *Text Representation: Linguistic and Psycholinguistic Aspects*. Human cognitive processing 8. Amsterdam – Philadelphie : J. Benjamins : 29-87.
- ASHER, N. 1993. *Reference to Abstract Objects in Discourse*. Dordrecht – Boston – Londres : Kluwer Academic Publishers.
- BACIU, M. 2011. *Bases de neurosciences. Neuroanatomie fonctionnelle*. Bruxelles : De Boeck.
- BAUMER, E. 2015. *Noms propres et anaphores nominales en anglais et en français : étude comparée des chaînes de référence*. Paris : L'Harmattan.

- BOUDREAU, S. et KITTREDGE, R. 2005. Résolution des anaphores et détermination des chaînes de coréférences. Différences entre variétés de textes. *Traitement automatique des langues* 46 (1): 41-70.
- CHAROLLES, M. 2002. *La référence et les expressions référentielles en français*. Gap – Paris: Ophrys.
- CLARK, H. H. et WILKES-GIBBS, D. 1986. Referring as a Collaborative Process. *Cognition* 22 (1): 1-39.
- CONDAMINES, A. 2005. Anaphore nominale infidèle et hyperonymie: le rôle du genre textuel. *Revue de sémantique et pragmatique* 18: 23-42.
- CORBLIN, F. 1995. *Les formes de reprise dans le discours. Anaphores et chaînes de référence*. Rennes: Presses universitaires de Rennes.
- DELABORDE, M. et LANDRAGIN, F. 2018. Traitement *good-enough* du pronom *on*: vers une modélisation de la coréférence floue. In *Linguistic and Psycholinguistic Approaches to Text Structuring – LPTS 2018 (Paris, janvier 2018)*. Présentation en ligne sur HAL: <https://halshs.archives-ouvertes.fr/halshs-01795228>.
- DODGE, Y. 2008. *The Concise Encyclopedia of Statistics*. New York: Springer.
- GREEN, M. F., PENN, D. L., BENTALL, R., CARPENTER, W. T., GAEBEL, W., GUR, R. C., KRING, A. M., PARK, S., SILVERSTEIN, S. M. et HEINSEN, R. 2008. Social Cognition in Schizophrenia: An NIMH Workshop on Definitions, Assessment, and Research Opportunities. *Schizophrenia Bulletin* 34 (6): 1211-1220.
- GRICE, H. P. 1975. Logic and Conversation. In J. P. KIMBALL, J. L. MORGAN et P. COLE (éd.), *Syntax and Semantics*. New York – San Francisco – Londres: Academic Press. Vol. 3: *Speech Acts*: 41-58.
- HEIDEN, S., MAGUÉ, J.-P. et PINCEMIN, B. 2010. TXM: une plateforme logicielle open-source pour la textométrie – conception et développement. In S. BOLASCO, I. CHIARI et L. GIULIANO (éd.), *JADT 2010 – Statistical Analysis of Textual Data. Proceedings of the 10th International Conference: 9-11 June 2010, Sapienza University of Rome*. Milan: Edizioni Universitarie di Lettere Economia Diritto. Vol. 2: 1021-1032.
- JAMES, G., WITTEN, D., HASTIE, T. et TIBSHIRANI, R. 2013. *An Introduction to Statistical Learning: With Applications in R*. New York: Springer.
- LANDRAGIN, F. 2016. Description, modélisation et détection automatique des chaînes de référence (DEMOCRAT). *Bulletin de l'Association française pour l'Intelligence Artificielle* 92: 11-15.
- LANDRAGIN, F. 2018. Corpus DEMOCRAT. Corpus disponible en ligne sur la plateforme ORTOLANG (Open Resources and Tools for Language): <https://hdl.handle.net/11403/democrat/v1.1>.
- LANDRAGIN, F., POIBEAU, T. et VICTORRI, B. 2016. Analyse des références et des transitions référentielles: apport de la linguistique outillée. In L. SARDA, D. VIGIER et B. COMBETTES (éd.), *Connexion et indexation. Ces liens qui tissent le texte*. Lyon: ENS Éditions: 123-135.
- LECLERCQ, M. et ZIMMERMANN, P. 2014. L'évaluation des fonctions attentionnelles. In X. SERON et M. VAN DER LINDEN (éd.), *Traité de neuropsychologie clinique de l'adulte*. Louvain-la-Neuve: De Boeck – Solal. T. 1: *Évaluation*: 143-162 [2^e éd.].

- MANN, H. B. et WHITNEY, D. R. 1947. On a Test of Whether One of Two Random Variables Is Stochastically Larger than the Other. *The Annals of Mathematical Statistics* 18 (1): 50-60.
- PERRET, M. 2000. Quelques remarques sur l'anaphore nominale aux XIV^e et XV^e siècles. *L'information grammaticale* 87 : 17-23.
- QUIGNARD, M., HEIDEN, S., LANDRAGIN, F. et DECORDE, M. 2018. Textometric Exploitation of Coreference-Annotated Corpora with TXM: Methodological Choices and First Outcomes. In *JADT' 18: Proceedings of the 14th International Conference on Statistical Analysis of Textual Data (Rome, 12-15 June 2018)*. Rome : UniversItalia. Vol. 1 : 610-615. En ligne à l'adresse suivante : <http://lexicometrica.univ-paris3.fr/jadt/JADT2018/actes-jadt18.pdf>.
- ROUSIER-VERCRUYSSSEN, L. 2017. *Prosodie, Syntaxe, et Référence : processus cognitifs et marqueurs linguistiques*. Thèse de doctorat. Université de Neuchâtel – Université Paris Nanterre.
- SCHNEDECKER, C. 1997. *Nom propre et chaînes de référence*. Paris – Metz : Klincksieck – Université de Metz.
- 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., GLIKMAN, J. et LANDRAGIN, F. 2017. Les chaînes de référence : annotation, application et questions théoriques. *Langue française* 195 : 5-16.
- 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.
- TUTIN, A. 2002. A Corpus-Based Study of Pronominal Anaphoric Expressions in French. In A. BRANCO, R. MITKOV et T. MCENERY (éd.), *DAARC 2002: Proceedings of the 4th Discourse Anaphora and Anaphor Resolution Colloquium*. Lisbonne : Edições Colibri.

Annexe

Exemple illustrant le propos de la section 2, dans son intégralité. La chaîne de référence est composée de onze mentions : [La société Evobus], [ses] employés, [la société], [son] directeur, [une entreprise], [qu’], [Une société], [qui], [qui] et [[notre] entreprise]. Dans cette chaîne, il y a huit mentions différentes : [La société Evobus], [ses] employés, [la société], [son] directeur, [une entreprise], [Une société] et [[notre] entreprise].

[**La société Evobus**] a fêté [le départ en [retraite] d’[un de [ses] plus fidèles employés], [Michel Zajac]]. Entouré de [[ses] collègues et amis], accompagné de [[sa] femme], [il] a reçu [les honorables remerciements de [**la société**]] par [l’entremise de [[son] directeur], [M. Muller]]. « [J]’ai souvent pu apprécier [[votre] travail], [vous] serez toujours [le bienvenu] chez [nous] », expliquait-[il], avant [la remise [des cadeaux]]. [Michel Zajac] quitte donc [**une entreprise**] [**qu’**][il] a connue sous [différents noms], d’abord [Rouméas], à [[son] entrée], puis [Kassbohrer] jusqu’en [1995], et enfin [Evobus France]. [**Une société**] [**qui**] a évolué et [**qui**] a aussi fait évoluer [[ses] employés], notamment [Michel], passé par [différents types de poste], pour terminer [préparateur de [profilé de [sol]]]. « [Vous] avez contribué à [la qualité de fabrication de [[notre] entreprise]] par [[votre] passion du travail] et [[votre] rigueur] », concluait [le directeur] devant [l’équipe] [où] a travaillé [Michel Zajac].