

Faculté	des Langues
	Université de Strasbourg

Master Technologies des langues

2022–2024

Neutralisation automatique du genre à travers l'utilisation de noms collectifs en français

Enzo DOYEN

Sous la direction de
Amalia TODIRASCU
Professeure des universités



Remerciements

Je tiens tout d'abord à exprimer ma profonde gratitude à l'égard d'Amalia Todirascu pour avoir accepté ma proposition de sujet, et pour m'avoir encadré et conseillé tout au long de ce travail de mémoire.

Je souhaite également témoigner ma reconnaissance à l'ensemble de l'équipe pédagogique du Master ; ce mémoire signe l'achèvement de deux années de formation très riches pendant lesquelles j'ai énormément appris — y compris au niveau personnel. Je tiens à remercier plus particulièrement Delphine Bernhard et Pablo Ruiz Fabo pour leurs précieux enseignements, ainsi que pour leur écoute, leur disponibilité, leur engagement et leur passion communicative qui ont contribué à renforcer mon intérêt pour le traitement automatique des langues.

Ce mémoire n'aurait sans doute pas vu le jour sans le concours indirect, mais néanmoins crucial, d'Aude Callens et Isabelle Zanchetta, traductrices chevronnées avec qui j'ai eu le plaisir de collaborer, et qui m'ont fait prendre conscience de l'influence qu'exerce le genre dans la langue et de la nécessité de recourir à des formulations alternatives. Je remercie également mes collègues actuels et actuelles pour leur soutien personnel de même que leur engagement vis-à-vis de cette démarche : merci à Loïc Guézais, Elisabeth Jousse, Nina Legout et Dorine Parmentier ; ainsi qu'à Fabrice Demurger et Sharon Wang.

Merci à mes enseignantes et enseignants de l'Université d'Aix-Marseille, et en particulier Grégoire Lacaze et Monique de Mattia-Viviès pour m'avoir transmis leur passion pour la linguistique ; ainsi que Marc Calvini-Lefebvre pour ses précieux conseils, ses encouragements et son intérêt pour la question du genre.

Je tiens à exprimer ma reconnaissance à mes amis et amies pour leur soutien constant ; mes remerciements les plus sincères s'adressent à Jordan, Lucas, Malek, Merve, Nelson, Panagiotis, Sally et Yansi pour leur présence et leur écoute.

Enfin, je remercie des plus chaleureusement ma famille, et tout particulièrement ma mère, ma grand-mère et mon oncle, pour leur soutien divers et leur affection.

Table des matières

1	Introduction	8
2	État de l'art	11
2.1	Le genre dans la langue	11
2.1.1	La notion de « genre »	11
2.1.1.1	Le genre en sociologie	11
2.1.1.2	Le genre en linguistique	12
2.1.1.3	Le genre en sociolinguistique	13
2.1.2	Le genre en français : un bref historique	14
2.1.2.1	La volonté des prescriptivistes de faire taire le féminin	14
2.1.2.2	La reprise du combat sous le spectre politique et militant	15
2.1.3	Le masculin générique, ou comment la langue influence les représentations mentales	17
2.1.4	Réponses au masculin générique : l'écriture inclusive et les techniques d'écriture connexes	20
2.1.4.1	Les techniques de visibilité	20
2.1.4.2	Les techniques de neutralisation	22
2.1.5	Le point médian : un signe typographique polémique	24
2.2	Les noms collectifs en français	27
2.2.1	Les noms collectifs : définition	27
2.2.2	Caractéristiques morphologiques, syntaxiques et sémantiques des noms collectifs	28
2.2.2.1	Caractéristiques morphologiques	28
2.2.2.2	Caractéristiques syntaxiques	29
2.2.2.3	Caractéristiques sémantiques	30
2.2.3	Noms collectifs humains (Ncoll·H)	31
2.2.3.1	Ncoll·H et neutralisation en genre	34
2.3	Réécriture du genre en traitement automatique des langues	37
2.3.1	Tâche de réécriture dans le domaine du TAL : définition et considérations	37
2.3.2	Présentation de systèmes de réécriture du genre	38
2.3.2.1	Systèmes de réécriture vers un autre genre	38

2.3.2.2	Systèmes de réécriture inclusive	40
2.3.2.3	Systèmes de réécriture neutre	43
3	GeNRe : un système de réécriture neutre pour le français exploitant les noms collectifs	47
3.1	Méthodologie	48
3.2	Création d'un dictionnaire de paires de noms collectifs-noms de membres	49
3.3	Conception du jeu de données initial	51
3.4	Conception et architecture du système à base de règles	54
3.4.1	Étapes initiales	56
3.4.2	Composant de détection des dépendances	56
3.4.2.1	Règles de filtrage	57
3.4.2.2	Extraction des dépendances syntaxiques	57
3.4.3	Composant de génération	59
3.4.3.1	Remplacements	60
3.4.3.2	Ajustement de la flexion	61
3.4.4	Évaluation	62
3.4.4.1	Composant de détection des dépendances	63
3.4.4.2	Composant de génération	65
3.4.4.3	Résultats et discussion	65
3.5	GeNRe-FT : affinage de grands modèles de langue	71
3.5.1	Évaluation, résultats et discussion	72
3.6	Modèles d'instructions	77
3.6.1	Mixtral 8x7B Instruct v0.1	79
3.6.1.1	Évaluation, résultats et discussion	80
3.6.2	Claude 3 Opus	82
3.6.2.1	Évaluation, résultats et discussion	83
3.7	Résultats et discussion généraux	84
3.8	Limites et perspectives	84
4	Conclusion	87
	Références bibliographiques	98
A	Annexes	99
A.1	Dictionnaire de paires de noms collectifs-noms de membres	99
A.2	Hyperparamètres utilisés pour l'affinage de grands modèles de langue	107
A.2.1	GeNRe-T5	107
A.2.2	GeNRe-M2M-100	107
A.3	Phrases utilisées pour les instructions à quelques exemples données aux modèles d'instructions	107

A.4 Exemples d'hallucinations du modèle d'instructions Mixtral 8x7B 109

Liste des illustrations

2.1	Schéma de mesure typographique avec le point médian	25
2.2	Évolution de la fréquence du nom collectif « patientèle » dans Google Ngram Viewer	29
3.1	Aperçu de la méthodologie	48
3.2	Extrait de code pour la récupération automatique de noms collectifs associés au suffixe « -phone »	52
3.3	Extrait du dictionnaire de paires noms collectifs-noms de membres	53
3.4	Extrait de code pour la création d'expressions régulières pour la détection de noms de membres	54
3.5	Vue d'ensemble de la chaîne de traitement de GeNRe-RBS	55
3.6	Algorithme de détection des noms propres	58
3.7	Extrait de l'algorithme de détection des tokens voisins	60
3.8	Extrait du code utilisé pour l'évaluation du composant de détection des dépendances	64
3.9	Représentation graphique du compte d'erreurs du système de règles par corpus	68
3.10	Représentation graphique du compte d'erreurs des modèles affinés par corpus .	75

Liste des tableaux

2.1	Récapitulatif des techniques d'écriture inclusives actuelles	24
2.2	Effets de sens des Ncoll·H génériques	36
2.3	Formes d'écriture prises en charge par le système <i>INCLUSIFY</i> pour l'allemand .	41
2.4	Pronoms et déterminants genrés en anglais, et leurs équivalents neutres	44
3.1	Codes utilisés pour l'extraction des dépendances syntaxiques	59
3.2	Résultats du composant de détection des dépendances, par corpus et en moyenne	66
3.3	Résultats du système à base de règles, par corpus et en moyenne	66
3.4	Compte d'erreurs du système de règles par corpus	67
3.5	Résultats des modèles affinés	72
3.6	Compte d'erreurs des modèles affinés par corpus	75
3.7	Instructions données aux modèles d'instructions	79
3.8	Résultats du modèle d'instructions Mixtral 8x7B par corpus	80
3.9	Résultats du modèle d'instructions Claude 3 Opus par corpus	83
3.10	Résultats globaux par corpus et par mesure	85
A.1	Entrées du dictionnaire de paires de noms collectifs-noms de membres	99
A.2	Phrases pour les instructions à quelques exemples de types BASE et DICT . . .	107
A.3	Phrases pour les instructions à quelques exemples de type CORR	108
A.4	Exemples d'hallucinations par type	109

1

Introduction

Qualifiée de « péril mortel » par l'Académie française (Académie française, 2017), l'émergence de l'écriture dite inclusive (Viennot, 2014, 2018 ; Moreau, 2019) a suscité de vifs débats et protestations dans la sphère publique (Viennot, 2021, 2022 ; Abbou, 2019). Cette forme d'écriture, développée au début des années 2000, regroupe diverses techniques visant à mettre sur un pied d'égalité les hommes et les femmes dans la langue en évitant le recours au « masculin générique », ce dernier amenant à la formation de biais. Cela passe notamment par la féminisation des noms de métiers (l'on préférera « une femme pompière » à « une femme pompier »), par la mise en lumière des deux genres à l'aide de doublets complets ou abrégés (« les étudiants et étudiantes », « les étudiant-es »), ou bien par la neutralisation complète du genre à travers l'utilisation de déterminants, noms et adjectifs épicènes (« les élèves » remplacera « les étudiants »), ou bien de noms collectifs (« la police » remplacera « les policiers »).

Parce que les techniques d'écriture les plus visibles étant aussi les plus critiquées, notamment en ce qui concerne l'usage du point médian, et parce qu'elles ne répondent pas forcément à tous les contextes d'utilisation, nous estimons que le recours à des techniques de neutralisation, davantage passepartout, devrait être privilégié. En particulier, les noms collectifs constituent un type de nom se prêtant bien à la neutralisation, non seulement car ils évitent de choisir entre le masculin et le féminin, mais également car ils peuvent s'avérer très productifs à travers l'utilisation d'adjectifs (« les enseignants » → « le corps enseignant »). Les noms collectifs

humains ont été étudiés en détail du point de vue syntaxique (Gross, 2009 ; Lecolle, 2019) et sémantique (Gross, 2011 ; Lecolle, 2008, 2013, 2016, 2019), mais ils restent très peu abordés lorsqu'il est question de la neutralisation en genre.

Dans le domaine du traitement automatique des langues, la question du genre a largement été soulevée vis-à-vis des biais, à la fois pour la détection des biais de genre et les diverses manières de les éliminer (Stanczak et Augenstein, 2021). Toutefois, le nombre de travaux qui se proposent de modifier les formes genrées dans un texte donné de manière à produire un équivalent binaire (conversion du masculin vers le féminin et vice versa), un texte inclusif ou une neutralisation totale du genre reste limité. Pour la conversion d'un genre à un autre, les projets existants se concentrent surtout sur l'arabe (Habash et al., 2019 ; Alhafni et al., 2022a ; Alhafni et al., 2022b ; Alrowili et Vijay-Shanker, 2022). Pomerence (2022) a proposé une ressource permettant de rendre un texte inclusif, mais elle est limitée à l'allemand ; il s'agit, à ce jour et à notre connaissance, du seul système de ce type. Enfin, au regard de la neutralisation, les recherches sont cantonnées à l'anglais (Sun et al., 2021 ; Vanmassenhove et al., 2021 ; He et al., 2021). Il n'existe, à notre connaissance, aucun projet similaire pour le français.

Dans cette optique, nous nous proposons d'élaborer une ressource visant à remplacer les formes genrées dans un texte par un nom collectif sémantiquement proche. Par la création d'une telle ressource, nous espérons, à notre niveau, contribuer au développement de solutions visant à réduire les biais de genre, et démontrer l'importance de cette problématique afin que des travaux de plus grande envergure puissent être réalisés pour le français.

En outre, mentionnons aussi que la neutralisation automatique du genre peut également servir au développement d'outils externes. Avec la popularité et l'usage grandissants de modèles à base d'instructions¹, comme ChatGPT², Alpaca³ ou LLaMA (Touvron et al., 2023), qui sont capables de générer des réponses développées à des questions, la problématique des biais de genre dans le domaine du traitement automatique des langues devient de plus en plus prégnante, car les données sur lesquelles ces modèles ont été entraînés sont rarement exemptes de biais. Par conséquent, les textes générés sont révélateurs des biais de ces modèles (Zhuo et al., 2023). De plus, les réponses fournies en français font quasiment exclusivement usage du masculin générique, à moins qu'une formulation inclusive ou neutre n'ait été explicitement demandée⁴. L'entraînement de ce type de modèles sur des données neutralisées au niveau du genre pourrait permettre une meilleure génération de phrases neutres, du point de vue quantitatif autant que qualitatif.

1. *instruction-following models* en anglais.

2. <https://openai.com/blog/chatgpt>

3. <https://crfm.stanford.edu/2023/03/13/alpaca.html>

4. À ce jour, il n'existe pas, à notre connaissance, de recherche en français évaluant l'utilisation du genre dans les réponses générées par ces modèles. Toutefois, une rapide utilisation de ChatGPT (<https://chat.openai.com/chat>) permet de dégager cette hypothèse. Même quand une formulation inclusive est utilisée dans la question, le modèle a recours au masculin générique dans sa réponse. Voici un extrait d'une conversation, où les termes au masculin générique ont été mis en gras :

« Question : C'est bientôt l'anniversaire d'un(e) ami(e), qu'est-ce que je pourrais lui offrir ?

Réponse : Il y a tellement d'options en fonction des intérêts et des préférences de votre ami(e). Voici quelques idées qui pourraient vous aider : [...]

3. Un livre ou un album : si votre ami(e) est **un grand lecteur** ou **un amateur** de musique [...]

5. Un abonnement : si votre ami(e) est **abonné** à une plateforme de streaming. . . »

(conversation réalisée le 25/03/2023)

La première partie de ce mémoire sera consacrée à l'état de l'art, autant en linguistique qu'en traitement automatique des langues. Nous nous intéresserons en premier lieu à la notion de genre de manière globale avant de nous concentrer spécifiquement sur le genre en linguistique à travers une étude de l'évolution du genre en français, les biais générés par l'usage du masculin générique et les techniques d'écriture inventées pour y remédier. Ensuite, nous procéderons à une analyse globale des noms collectifs en français et nous verrons comment ils peuvent être utilisés dans le cadre de la neutralisation du genre. Enfin, nous dresserons un panel des travaux réalisés dans le domaine du traitement automatique des langues en rapport avec la tâche de « réécriture du genre ».

Dans la seconde partie, nous présenterons en détail notre système de réécriture neutre pour le français, **GeNRe (Gender-Neutral Rewriter Using French Collective Nouns)**⁵. Celui-ci prend principalement appui sur un système de règles symboliques, les données générant ce dernier servant par la suite à l'affinage de grands modèles de langue. L'objectif de notre système est de convertir des phrases contenant des noms au masculin générique en phrases neutralisées en genre, en utilisant des noms collectifs équivalents. Après une mise en lumière des différents composants de notre système et de notre fonctionnement, ainsi que des corpus utilisés pour son élaboration, nous effectuerons une évaluation de celui-ci afin d'établir des comparaisons avec les systèmes de réécriture neutre existants pour les autres langues. Une analyse et une discussion détaillée des résultats de notre système permettra de mettre en avant ses forces, ses limites et les améliorations pouvant y être apportées.

5. « système de réécriture neutre pour le français exploitant les noms collectifs » en français.

2

État de l'art

2.1 Le genre dans la langue

2.1.1 La notion de « genre »

Le « genre » est une notion particulièrement difficile à appréhender en ce qu'elle est étudiée en détail par des disciplines très diverses, telles que la linguistique, la psychologie, la biologie ou encore la sociologie ; et, car la définition qu'on lui attribue peut grandement changer en fonction des domaines disciplinaires et des différents courants. Il convient de s'intéresser en particulier à la notion de genre en linguistique et en sociologie, car ce sont dans ces deux disciplines que le plus de travail a été abattu en ce qui concerne le genre.

2.1.1.1 Le genre en sociologie

En sociologie, la définition du genre n'est pas stable et est en perpétuelle évolution. Nous pensons toutefois que la définition donnée par Rioux et al. (2022) satisfait bien à l'acceptation la plus courante :

« Gender is a social configuration that gathers the roles, behaviors, activities, feelings, attitudes, and attributes that a given society typically associates with being masculine or feminine. »¹ (5)

En somme, le genre (désormais « identité de genre ») doit être vu comme une construction sociale de différence entre les sexes et imposant une binarité : certains traits ou comportements sont associés, du fait des représentations sociétales, à des personnes de sexe masculin, tandis que d'autres sont rattachés à des personnes de sexe féminin. C'est de là que naît l'association entre l'identité de genre et le sexe biologique, quand bien même les deux ne sont pas liés de manière inhérente (Schudson et Morgenroth, 2022).

2.1.1.2 Le genre en linguistique

La linguistique n'est pas exempte de tout problème en ce qui concerne la notion de genre. Corbett, dès les premiers mots de son introduction de *Gender* (Corbett, 1991), considère le genre comme « la plus énigmatique de toutes les catégories grammaticales »². Pour le français, le dictionnaire de linguistique de Larousse (Dubois et al., 2001) propose la définition suivante :

« Le genre est une catégorie grammaticale reposant sur la répartition des noms dans des classes nominales, en fonction d'un certain nombre de propriétés formelles qui se manifestent par la référence pronominale, par l'accord de l'adjectif (ou du verbe) et par des affixes nominaux (préfixes, suffixes ou désinences casuelles), un seul de ces critères étant suffisant. Ainsi, d'après ces trois critères, on définit en français deux classes, les masculins et les féminins. » (247)

On retrouve, comme élément commun avec l'identité de genre, la binarité (masculin et féminin). Cela n'est toutefois pas commun pour toutes les langues, comme l'indique Corbett (1991) : certaines langues peuvent avoir plus que deux genres (tel l'allemand, qui dispose d'un genre neutre), tandis que d'autres sont exemptes d'un système de genre (tels le bengali, le turc ou le mari). Il convient à ce stade de distinguer le genre sémantique du genre lexicogrammatical. Lambelet (2012 : 12) définit ces deux notions de manière suivante : le genre sémantique est « universel et non dépendant de la langue », tandis que le genre lexicogrammatical « constitue un système de classification linguistique formel ». Le genre sémantique est alors réservé aux êtres animés et reflète les caractéristiques biologiques de ces derniers. Le genre lexicogrammatical, lui, peut (en français) s'appliquer autant aux êtres animés qu'aux objets, mais n'est alors pas motivé : aucune propriété spécifique ne mène à l'association d'un genre en particulier ; l'opération est purement arbitraire. Watbled (2012 : 168-169) adopte une approche contrastive et compare

1. « Le genre est une configuration sociale qui rassemble les rôles, les comportements, les activités, les sentiments, les attitudes et les attributs qu'une société donnée associe généralement au sexe masculin ou féminin. » (ma traduction)

2. « Gender is the most puzzling of the grammatical categories. » (1)

le fonctionnement du genre en anglais et en français pour expliciter cette distinction. Il prend l'exemple des pronoms personnels sujets « he » et « she » en anglais, qui renvoient exclusivement à des êtres humains de sexe masculin et féminin, respectivement, et ont donc un rôle strictement sémantique.³ En français, au contraire, les pronoms personnels sujets « il » et « elle » peuvent renvoyer autant à un être humain (« Elle est belle, Marie ») qu'à un inanimé (« Elle est grande, la table »). En outre, le genre lexicogrammatical en français se caractérise par l'utilisation d'autres catégories que les pronoms, contrairement à l'anglais (noms, déterminants, adjectifs), et par l'accord avec ces dernières. Ainsi, si le genre lexicogrammatical et le genre sémantique peuvent se superposer, ce n'est pas toujours le cas : le genre féminin associé au nom « table » ne l'est pas à cause de propriétés inhérentes, contrairement à « pharmacienne », par exemple. Il en va de même pour certains noms animés formés sur une base métaphorique : on peut très bien employer les noms féminins « vedette » ou « sentinelle » pour référer à des êtres humains de sexe masculin, quand bien même le genre lexicogrammatical ne correspond pas au genre sémantique.

2.1.1.3 Le genre en sociolinguistique

D'après Spolsky (1998 : 4), la tâche principale de la sociolinguistique est d'« associer variation linguistique et condition sociale »⁴.

L'étude conjointe de l'identité de genre et de la langue a été proposée pour la première fois par Robin Lakoff, avec la publication de son étude *Language and woman's place* en 1972 (Eckert et McConnell-Ginet, 2003). Lakoff met, entre autres, en avant l'idée que l'identité de genre peut avoir une influence sur la manière de s'exprimer, et que, selon l'identité de genre d'une personne, certaines tournures ont plus de chances d'être produites. Cela a notamment été mis en évidence en anglais par l'étude de Fishman (1980), qui montre que les femmes utilisent davantage de *question tags*⁵, même si des études ultérieures (Cameron et al., 1998 ; Haque et Wong, 2003) font état d'autres variables à prendre en compte, comme le contexte, la culture ou la variété d'anglais parlée. Ainsi, bien que les conclusions de ces travaux sont discutables, ces derniers ont eu le mérite d'inciter d'autres chercheurs et chercheuses à s'intéresser au lien étroit qui pouvait exister entre identité de genre et langage. Celui-ci est d'autant plus mis en exergue par Eckert et McConnell-Ginet, qui affirment que, par le langage, nous participons à la définition de l'identité de genre et à ce qui y est associé :

3. En guise de précision, dans certaines variétés d'anglais, « she », « he » et leurs dérivés objet et possessif sont parfois utilisés pour référer à des êtres inanimés spécifiques (pays, bateaux, machines ; voir Svartengren (1927)), ou dans des structures impersonnelles, comme pour décrire la météo (Gardelle, 2015). Dans une perspective définitoire, et car ils ne sont pas strictement liés au sujet que nous évoquons, nous ne prenons pas ces usages en considération.

4. « to map linguistic variation on to social condition »

5. En anglais, un question tag (ou tag question en anglais américain) est une construction lexicale composée d'un auxiliaire (à polarité positive ou négative) et d'un pronom référentiel, ajoutée en fin de question pour demander une information ou une confirmation, comme dans la phrase « Mary is a good cooker, isn't she ? » (Mary est une bonne cuisinière, [hein/non/n'est-ce pas/qu'est-ce que tu en penses ?])

« Gender is, after all, a system of meaning — a way of construing notions of male and female — and language is the primary means through which we maintain or contest old meanings, and construct or resist new ones. »⁶ (6)

On peut ainsi développer la définition proposée par Spolsky en ajoutant que ne sont pas seulement liées variation linguistique et condition sociale, mais également variation linguistique et constructions sociales. En d'autres termes, ce n'est pas uniquement la situation sociale d'une personne en particulier qu'il faut prendre en considération, mais aussi la manière dont elle se représente, et comment ces représentations sont perçues et construites par la société. Cette partie a permis d'établir les principes fondamentaux en ce qui concerne la question du genre, qui sera au centre de ce mémoire, à la fois d'un point de vue sociologique et d'un point de vue linguistique. Nous avons ensuite vu le lien étroit qui existe entre identité de genre et langue, mis en exergue par l'essor de la sociolinguistique. Nous allons maintenant nous intéresser à la question du genre linguistique en français d'un point de vue historique, afin de voir comment les femmes se sont retrouvées invisibilisées dans la langue au fur et à mesure du temps.

2.1.2 Le genre en français : un bref historique

L'historienne Éliane Viennot, dans son ouvrage *Non, le masculin ne l'emporte pas sur le féminin* (Viennot, 2014), analyse en détail l'importance de la langue dans le prolongement de la domination masculine, et la manière dont les femmes se sont emparées de cette question dans une perspective de promotion de l'égalité entre les femmes et les hommes. C'est avec l'essor des prescriptivistes que transparait une volonté d'invisibiliser les femmes dans la langue.

2.1.2.1 La volonté des prescriptivistes de faire taire le féminin

François de Malherbe fut l'un des premiers à se charger de cette question en s'attaquant à l'accord de proximité. Cet accord (aussi dénommé « règle de proximité » ou « accord de voisinage ») en français nous vient du latin. Il désigne un accord en genre et en nombre avec le référent le plus proche (par exemple, « les hommes et les femmes sont égales »). En latin, seuls les adjectifs sont concernés par cette règle (Moreau, 2019). Cette technique d'accord s'est ensuite transposée à l'ancien français et au moyen français, ainsi qu'à d'autres langues romanes. Tandis que l'on constate une extension de l'accord de proximité dans des langues romanes comme l'espagnol ou le portugais, permettant alors un accord avec le déterminant (Abeillé et al., 2018), l'ancien français et le moyen français se cantonnent à un accord de proximité réservé aux adjectifs. Cela ne l'empêche pas d'être massivement utilisé : d'après Georges Gougenheim, l'accord de proximité des adjectifs épithètes était prédominant au XVI^e siècle (Gougenheim cité par Chervel, 2019).

6. « Le genre est, après tout, un système signifiant – un moyen d'interpréter les notions d'"homme" et de "femme" – et la langue est le moyen principal par lequel nous maintenons ou contestons des anciennes acceptions, et par lequel nous en créons ou nous opposons à de nouveaux. » (ma traduction)

Viennot (2014), citant Claude de Vaugelas (1647), indique que François de Malherbe voit « comme un écueil » la phrase suivante : « Ce peuple a le cœur et la bouche ouverte à vos louanges. » (37). Selon lui, il conviendrait d'écrire « ouverts » pour faire référence à la fois à « cœur » et « bouche ». L'accord de proximité en nombre est donc rétabli, mais au détriment de l'accord de proximité en genre, qui devrait normalement s'appliquer étant donné que « bouche » est un substantif féminin. Vaugelas justifie cette correction par la prédominance du genre masculin comme « étant le plus noble ». Il émet toutefois quelques réserves, puisqu'il indique que « l'oreille a de la peine à s'y accommoder ». En fin de compte, la question est laissée plus ou moins en suspens. Viennot fait savoir que c'est Scipion Dupleix, historien et grammairien français, qui reprend l'argument du masculin comme genre le plus noble sans demi-mesure : il atteste que le genre masculin « prévaut seul contre deux ou plusieurs féminins, quoiqu'ils soient plus proches de leur adjectif ».

Dans le même temps, les prescriptivistes de l'époque s'attaquent aux noms de métiers et de fonctions. Aussi, quand bien même des substantifs féminins que « conseillère » ou « poétesse » existent et sont employés (Viennot, 2014 : 27-28), Nicolas Andry de Boisregard, médecin et homme de lettres, indique dès 1689 qu'il faut les remplacer par leurs équivalents masculins : il faudrait donc dire « peintre » ou « auteur » pour une femme, et non pas « peintresse » ou « autrice », jugés incorrects.

Ces nouvelles règles de « bon usage » seront finalement soutenues par la majorité des linguistes, puis entérinées par les dictionnaires de référence et les manuels scolaires. Ainsi, l'édition de 1834 de la Grammaire nationale de Bescherelle donne un coup d'arrêt aux noms de fonctions féminins en indiquant : « on ne dit pas *professeuse, graveuse, compositrice, traductrice*, etc. mais bien *professeur, graveur, compositeur, traducteur*, etc., par la raison que ces mots n'ont été inventés que pour les hommes qui exercent ces professions » (Viennot, 2014 : 9). Dans une nouvelle édition publiée quelques années après, Bescherelle reprend l'argument relatif à la prétendue noblesse du genre masculin lancé par Vaugelas et Dupleix en proclamant que « la masculinité annonce toujours une idée grande et noble » (Viennot, 2014 : 34).

Bien que certaines critiques soient proférées contre ces prescriptions, et notamment contre l'interdiction d'utiliser des noms de fonctions féminins (Viennot, 2014 : 35-36), elles n'auront que peu d'écho. Il faudra attendre, en France, la fin du 20^e siècle pour que ce sujet soit à nouveau saisi à bras-le-corps, en particulier grâce à l'essor des mouvements féministes.

2.1.2.2 La reprise du combat sous le spectre politique et militant

Les années 1970 en France ont constitué un terrain fertile pour les mouvements féministes, donnant lieu à la mise en place de différentes politiques publiques pour promouvoir l'égalité hommes-femmes, et élargissant l'accès aux postes prestigieux et aux fonctions électives. Ces changements sociétaux majeurs imposent nécessairement de nouvelles réflexions en termes linguistiques, en particulier pour les noms de métiers qui étaient jusque-là réservés à des hommes. Même si certains équivalents féminins existent alors encore, ils sont considérés comme des « féminins conjugués », utilisés exclusivement pour faire référence à l'épouse de l'homme qui occupe la fonction en question (Viennot, 2014).

Du point de vue politique, cela se traduit en 1984 par la mise en place de la Commission de terminologie relative au vocabulaire concernant les activités des femmes, qui vise à prodiguer des recommandations relatives à l'emploi d'équivalents féminins pour les noms de métiers, et la publication, sous le gouvernement Fabius, de la Circulaire du 11 mars 1986 relative à la féminisation des noms de métier, fonction, grade ou titre (Burr, 2003). Débute alors un processus de féminisation lexicale de la langue française (Elmiger, 2011). Ces recommandations se sont toutefois vues largement ignorées : avec l'arrivée d'une majorité de droite à l'Assemblée nationale et la mise en place d'une cohabitation entre François Mitterrand (PS) et Jacques Chirac (RPR), situation institutionnelle alors inédite, le sujet est balayé par la droite majoritaire (Elmiger, 2011 : 74). Une seconde circulaire, adressée par Lionel Jospin douze ans plus tard, en 1998, sera nécessaire pour entériner ces changements.

La question d'une langue plus égalitaire ne se cantonne toutefois pas aux institutions politiques. D'après Abbou (2019), c'est l'usage, plus particulièrement au sein de milieux anarchistes et féministes, qui cherche à remodeler la langue au niveau du genre. Naissent alors des formes nouvelles pour exprimer le genre. Abbou (2019) en répertorie quelques-unes :

- (1) [V]oleurE ; auteurSEs ; individu-e ; nombreUSEs ; captifVEs ; keufEs ; copinE ; ceulles ; eulles ; menteureuses ; transgenderé-e-s ; traducteurEux ; ratonNEs-laveurSEs

La langue française est vue comme un lieu de lutte : le but est de bousculer les codes et de promouvoir de nouvelles manières d'écrire et de s'exprimer afin de visibiliser les femmes dans la langue. Selon Abbou (2019 : 237), « [l']enjeu est de pratiquer le tumulte, l'irruption comme action politique, qui prend la langue à rebrousse-poil et interpelle. »⁷

Les formes proposées ne sont pas homogènes, et c'est tout l'objectif : lieu de lutte et lieu d'expérimentation linguistique convergent pour porter l'attention sur la question du genre et pousser à la réflexion sur les stratégies à adopter.

Ces techniques d'écriture ont continué d'évoluer et se sont plus tard vues regroupées sous le terme d'« écriture inclusive ». D'après Abbou et Candea (2021), le terme apparaît pour la première fois en anglais sous la forme « inclusive language » en 1976. Il est ensuite repris en 1979 « dans le *Lectionnary of Inclusive Language* publié par le National Council of Churches des États-Unis, sous la pression des théologues féministes qui se préoccupent de rendre la Bible et l'Église plus inclusives pour les femmes à travers le verbe, et de démasculiniser la représentation de Dieu » (Abbou et Candea, 2021 : 143). Ce n'est qu'en 2014 que le terme d'« écriture inclusive » commence à apparaître en France, avec notamment, en 2016, la

7. Julie Abbou utilise une définition bien particulière du nom « tumulte », qu'elle décrit quelques pages plus loin : « J'emploie ici le terme « tumulte » dans une généalogie qui va d'Edward W. Saïd à Elsa Dorlin. Saïd (2000 : 374) parle de « techniques du tumulte » comme d'un déplacement, d'une opération qui consiste à « élever la lutte à un niveau d'affrontement inédit » : Dorlin (2005) mobilise cette notion pour penser l'intersection épistémologique des catégories de sexe et de race. Elle propose que les techniques du tumulte nous permettent de fabriquer notre propre cadre d'intelligibilité pour battre en brèche les régimes de la domination. En employant le terme « tumulte », je fais donc résonner cette pratique du désordre dans ses dimensions épistémologiques et politiques pour en souligner l'enjeu qui consiste à battre en brèche les rapports de domination. Cependant, les pratiques linguistiques dont il est question ici ne se situent pas sur le même niveau que les techniques du tumulte de Saïd et de Dorlin. La force politique de ce tumulte me semble résider à la fois dans l'irrégularité, dans le brouillage qu'elle produit, ce que Jacques Rancière (2008) nomme le « dissensus » ou la séparation comme espace du politique qui produit un décalage de la signification du monde dont on ne peut prédire les effets mais aussi, plus largement, dans le refus de la discipline d'une approche systématique, que ce soit sous forme de norme ou de structure. » (242)

publication du Manuel d'écriture inclusive par Raphaël Haddad, directeur associé de l'agence de communication Mots-Clés (Haddad, 2016). Cette expression sera ensuite largement reprise, souvent à tort, par les médias et la classe politique, et déchainera de nombreuses passions et polémiques qui participeront à réduire le principe même d'écriture inclusive à une technique d'écriture en particulier et au signe typographique utilisé (voir **chap. 2.1.5**).

2.1.3 Le masculin générique, ou comment la langue influence les représentations mentales

En français, le masculin est considéré comme pouvant revêtir le rôle de genre neutre ou générique (« masculin générique ») dans un contexte non spécifique. Dès lors, dans la phrase :

(2) Un professeur doit savoir faire preuve d'autorité.

Le nom commun « professeur » est considéré comme étant au masculin générique, et faisant ainsi référence autant aux hommes qu'aux femmes. Lorsqu'il est question d'un groupe de personnes, le genre masculin est également utilisé pour faire référence aux individus de ce groupe, soit au travers d'un nom commun au masculin, soit au travers du pronom masculin de troisième personne du pluriel « ils », quand bien même ce groupe contient plus de femmes que d'hommes. En revanche, les noms communs féminins et le pronom féminin de troisième personne du pluriel « elles » ne peuvent, quant à eux, renvoyer qu'à des femmes ou à des groupes de femmes. Le masculin se trouvant dès lors propulsé comme genre par défaut, et assumant ainsi une place bien plus prépondérante que le féminin dans la langue, il peut être pertinent de se demander si ce déséquilibre ne mène pas à la formation de biais linguistico-cognitifs.

Le principe de la relativité linguistique⁸, développé par Benjamin Lee Whorf sur la base des travaux d'Edward Sapir, part du principe que la langue agit comme un élément déterminant dans la formation de nos pensées et de notre perception du monde. En particulier, selon Whorf (1940), les différences sur le plan des structures linguistiques mèneraient à une perception différente du monde et à des différences du point de vue cognitif :

« We dissect nature along lines laid down by our native languages. The categories and types that we isolate from the world of phenomena we do not find there because they stare every observer in the face ; on the contrary, the world is presented in a kaleidoscopic flux of impressions which has to be organized by our minds — and this means largely by the linguistic systems in our minds. [...] This fact is very significant for modern science, for it means that no individual is free to describe nature with absolute impartiality but is constrained to certain modes of interpretation even while he thinks himself most free. The person most nearly free in such respects would be a linguist familiar with very many widely different

8. Dans la littérature spécialisée, le terme de « principe de la relativité linguistique » (*principle of linguistic relativity*) est préféré à l'expression « hypothèse de Sapir-Whorf », cette dernière n'ayant été popularisée que plusieurs années plus tard par John Bissell Carroll dans son ouvrage *Language, Thought, and Reality : Selected Writings of Benjamin Lee Whorf* (Björk, 2008). Cela a son importance, car, comme l'indique Björk (2008 : 15), « comme principe, et non comme hypothèse, il faut le considérer comme un axiome. » ("As a principle rather than a hypothesis it should be interpreted as an axiom.")

linguistic systems. As yet no linguist is in such position. We are thus introduced to a new principle of relativity, which holds that all observers are not led by the same physical evidence to the same picture of the universe, unless their linguistic backgrounds are similar, or can in some way be calibrated. »⁹

Les premières études cherchant à prouver ou à infirmer la validité de ce principe se sont concentrées sur la perception des couleurs. En particulier, Berlin et Kay (1969) argüent que les langues disposent de termes universels pour les couleurs primaires, que cet inventaire de couleurs s'étend au fil de l'évolution de la langue, et qu'il existe donc des universaux linguistiques pour les couleurs entre les langues, ce qui tend à aller contre le principe énoncé par Whorf. Cette théorie s'est vue développée par Kay et McDaniel (1978), qui lient directement ces universaux linguistiques aux processus neurophysiologiques associés à la perception des couleurs. De ces critiques est née l'hypothèse néowhorfienne (Neo-Whorfian hypothesis), une version plus modérée du principe de la relativité linguistique. Cette hypothèse, définie pour la première fois dans les travaux de Paul Kay et Willett Kempton (Kay et Kempton (1984) ; Štrkalj Despot (2021)), s'éloigne du postulat radical posé au départ en ce qu'elle « évite les présomptions épistémologiques et métaphysiques de Whorf et ne souscrit pas absolument au point controversé selon lequel le processus de pensée s'effectue dans un langage naturel »¹⁰ (Gomila, 2015). En somme, elle sous-tend plutôt que la langue a une influence sur notre conception du monde, au lieu de la déterminer.

En partant de ce principe, il paraît raisonnable de penser que la manière dont le genre est représenté dans la langue a une influence sur nos représentations mentales.

Divers travaux de psycholinguistique ont en effet mis en évidence les travers que pouvait poser le masculin générique, non seulement en français, mais aussi dans d'autres langues romanes et germaniques (anglais, allemand, espagnol, portugais ou suédois). Une étude de Stahlberg et al. (2001) se concentre sur le masculin générique en allemand et vise à évaluer l'impact des représentations mentales des personnes qui y sont exposées. Ainsi, lorsque l'on demande à ces personnes de nommer des personnalités connues selon leur profession en utilisant des noms au masculin générique en allemand, elles ont davantage de chances de penser à une personnalité de sexe masculin. De même, Jacobson et Insko (1985) ont montré que, lorsque l'on demande à des personnes interrogées de remplir des phrases en anglais

9. « Nous découpons la nature suivant les voies tracées par notre langue maternelle. Les catégories et les types que nous isolons du monde des phénomènes ne s'y trouvent pas tels quels, s'offrant d'emblée à la perception de l'observateur. Au contraire, le monde se présente à nous comme un flux kaléidoscopique d'impressions que notre esprit doit d'abord organiser, et cela en grande partie grâce au système linguistique que nous avons assimilé. [...] Ce fait est d'une importance considérable pour la science moderne, car il signifie qu'aucun individu n'est libre de décrire la nature avec une impartialité absolue, mais qu'il est contraint de tenir compte de certains modes d'interprétation même quand il élabore les concepts les plus originaux. Celui qui serait le moins dépendant à cet égard serait un linguiste familiarisé avec un grand nombre de systèmes linguistiques présentant entre eux de profondes différences. Jusqu'ici aucun linguiste ne s'est trouvé dans une situation aussi privilégiée. Ce qui nous amène à tenir compte d'un nouveau principe de relativité, en vertu duquel les apparences physiques ne sont pas les mêmes pour tous les observateurs, qui de ce fait n'aboutissent pas à la même représentation de l'univers, à moins que leurs infrastructures linguistiques soient analogues ou qu'elles puissent être en quelque sorte normalisées. » (trad. Claude Carme, 1969)

10. « The contemporary neo-Whorfian versions of linguistic relativism avoid the epistemological and metaphysical assumptions of Whorf, and are not committed to the contention that thinking takes place in a natural language. » (2)

avec des pronoms, les noms de fonctions traditionnellement masculines ont plus de chances d'être associés à un pronom masculin, et les noms de fonctions traditionnellement féminines plus de chances d'être associés à un pronom féminin.

En français plus spécifiquement, une enquête menée par Harris Interactive (2017) auprès d'un échantillon de 1 000 personnes représentatif de la population française montre le même effet que celui rapporté par Stahlberg et al. : les personnes interrogées sont plus susceptibles de citer des personnes de sexe masculin lorsque le masculin générique est utilisé (par exemple, « présentateurs du journal télévisé »). En revanche, lorsqu'une formulation inclusive (« présentateurs ou présentatrices du journal télévisé ») ou épïcène (« personnes présentant le journal télévisé ») est utilisée, le nombre de personnes citant au moins une femme augmente. Il est notamment intéressant de noter que, selon les figures qu'il est demandé de nommer, le pourcentage de personnes qui citent exclusivement des personnalités de sexe féminin double, voire triple lorsqu'est utilisée une formulation épïcène : ainsi, lorsque l'on demande aux personnes sondées de citer « deux personnes ayant été championnes olympiques », 14 % d'entre elles nomment deux femmes championnes olympiques, ce qui n'est le cas que pour 7 % d'entre elles lorsque les deux noms « champions » et « championnes » sont utilisés, et 5 % uniquement quand le masculin générique « champions » apparaît seul. Le masculin semble donc avoir un impact non négligeable dans les représentations mentales, à la fois lorsqu'il est utilisé dans sa forme dite générique, mais aussi quand il se trouve en combinaison avec un nom féminin.

Nous pouvons également nous interroger sur l'influence que peut avoir l'ordre des mots en cas de combinaisons masculin-féminin, et si l'apposition du nom féminin avant le nom masculin peut permettre d'augmenter le nombre de personnalités féminines citées, sans toutefois créer un biais contraire. En effet, diverses études en linguistique ont montré que les noms d'hommes sont souvent placés avant les noms de femmes lorsqu'ils sont reliés par les conjonctions de coordination « et » ou « ou » (Wright et al., 2005 ; Hegarty et Lemieux, 2011), comme c'est le cas pour les noms communs dans les questions de l'enquête de Harris Interactive. À notre connaissance, seule une étude en psycholinguistique (Kesebir, 2017) vise à évaluer l'impact de l'ordre des mots genrés sur les représentations mentales. Pour cette étude, il a été demandé à 161 personnes d'écrire un récit fictif pendant cinq minutes. La moitié avait pour tâche d'écrire sur « une femme d'affaires et un homme d'affaires », tandis que l'autre moitié devait écrire sur « un homme d'affaires et une femme d'affaires ». Au total, 68,3 % des histoires mentionnent un homme en premier, et 31,7 % une femme. L'autrice indique que « les consignes ont eu une influence [sur la première mention] », puisque 49,4 % des histoires à propos d'« une femme d'affaires et un homme d'affaires » mentionnent une femme en premier. Ce chiffre monte à 87,5 % pour la première mention d'un homme dans les histoires concernant « un homme d'affaires et une femme d'affaires ». Même si un fort biais masculin reste inlassablement présent, le fait que presque la moitié des histoires présente une mention féminine en premier dans le cas où la locution nominale « femme d'affaires » apparaît en première position dans la consigne est révélateur de l'influence de l'ordre des mots genrés. Qui plus est, il a également été demandé à d'autres personnes d'évaluer l'importance accordée aux deux personnages dans les récits. Dans les histoires au sujet d'« une femme d'affaires et un homme d'affaires », le personnage

de l'homme d'affaires a été considéré comme moins important que la femme d'affaires, et inversement dans les récits équivalents : un lien peut dès lors être établi avec la première mention dans la consigne et/ou avec la première mention dans le récit. L'entité lue ou créée fictivement en premier est celle qui a le plus de chances de se retrouver au cœur du récit, car il s'agit de celle qui apparaît le plus immédiatement à l'esprit. En conséquence, l'ordre des mots genrés semble bel et bien avoir un impact sur les représentations mentales correspondantes, même si des travaux additionnels confirmant ou infirmant cette hypothèse sont nécessaires.

2.1.4 Réponses au masculin générique : l'écriture inclusive et les techniques d'écriture connexes

Afin de donner davantage de visibilité aux femmes dans la langue, et pour contrer les biais relatifs au masculin générique explicités ci-dessus, différentes techniques d'écriture dite inclusives ont été imaginées et promues. On peut les regrouper en deux grandes catégories : les techniques de visibilisation et les techniques de neutralisation.

2.1.4.1 Les techniques de visibilisation

Les techniques de visibilisation ont pour objectif de mettre en exergue les formes féminines des mots. Cela s'applique notamment aux doublets, qui peuvent être soit complets, soit abrégés. Les doublets complets consistent à lier la forme masculine et la forme féminine d'un mot par les conjonctions de coordination « et » et « ou »¹¹. Aussi, à la place d'utiliser le nom au masculin générique « enseignant » pour faire référence à toutes les personnes exerçant cette fonction, on pourra, par exemple, avoir recours au doublet complet « les enseignants et enseignantes » (ex. 3). Cela ne se cantonne toutefois pas aux noms : sont aussi concernés les pronoms (ex. 4) et les déterminants genrés. Ces derniers peuvent soit précéder un nom épïcène (c'est-à-dire un nom dont la graphie est identique au masculin et au féminin [ex. 5]), soit un nom marqué en genre (ex. 6). Dans ce dernier cas, l'accord de proximité s'applique généralement.

- (3) J'ai onze personnes, seulement comme domestiques, à loger, sans compter **les ouvriers et ouvrières** allant et venant.
(Frantext : George Sand, *Correspondance* [1844], p. 569)
- (4) Mais c'est cela qui nous permettra d'équilibrer le financement de notre retraite, dans notre pays, et c'est une chance, où l'on vit plus longtemps, d'améliorer la retraite minimale pour **toutes celles et ceux** qui ont travaillé pour avoir leurs trimestres. . .
(Emmanuel Macron, Vœux du président de la République aux Français, *Élysée* [31 décembre 2022], disponible sur <https://www.elysee.fr/front/pdf/elysee-module-20621-fr.pdf>)

11. À l'oral, notamment lors de discours, la conjonction de coordination peut aussi être remplacée par une légère pause (qui est alors l'équivalent d'une virgule à l'écrit), comme dans la célèbre phrase prononcée par Charles de Gaulle le 27 juin 1958 : « Françaises, Français, aidez-moi ! » (<https://www.ina.fr/ina-eclaire-actu/video/i00012393/charles-de-gaulle-francaises-francais-aidez-moi>).

- (5) Il faut soigner le retour pour **certains et certaines élèves** qui sont déjà très fragilisés ou fragilisés aussi par le déconfinement.
(frWaC : Les Suisses ont profité du soleil et du retour des terrasses, *RTS* [17 mai 2020], disponible sur <https://www.rts.ch/info/suisse/11329493-les-suisses-ont-profite-du-soleil-et-du-retour-des-terrasses.html>)
- (6) Et il semble, au vu du micro-trottoir ci-dessous, que **certains et certaines Françaises** se sentent eux aussi parfois loin de cette cause. . .
(frWaC : Thomas Lemoine, Qu’est-ce qui fait de nous un mauvais ou une mauvaise féministe ? *20 Minutes* [3 avril 2018], disponible sur <https://www.20minutes.fr/societe/2248231-20180403-video-fait-mauvais-mauvaise-feministe>)

Les doublets complets étant parfois considérés comme alourdissant le discours, en raison de la répétition d’un même mot dans une forme différente, les doublets abrégés peuvent être privilégiés : la forme féminine d’un mot est alors séparée par un signe typographique (désormais « séparateur »), ou bien mise en exergue¹². Ainsi, « les enseignants et enseignantes » devient, avec l’introduction d’un point médian, « les enseignant-es » (ou « les enseignant·e·s »¹³). Dans l’usage, plusieurs séparateurs sont utilisés de manière concurrente pour former les doublets abrégés. Nous pouvons par exemple citer les parenthèses (avec inclusion [ex. 7] ou non [ex. 8] du trait d’union), la barre oblique (ex. 9), le trait d’union (ex. 10), l’astérisque (ex. 11) ou le point médian (ex. 12)¹⁴.

- (7) les directeur(trices)
- (8) les directeur(-trices)
- (9) les directeur/trices
- (10) les directeur-trices
- (11) les directeur*trices
- (12) les directeur·trices

Diverses règles morphologiques gouvernent la formation des doublets abrégés ayant recours aux séparateurs.

Pour les déterminants, les noms et les adjectifs, lorsque seul un « e », ou bien un « e » précédé d’une consonne, est nécessaire pour former le féminin d’un mot, on ajoute la finale

12. Elle consiste, par exemple, à mettre le suffixe féminin en majuscules (« les étudiantEs »), ou à utiliser divers moyens stylistiques pour la mettre en valeur (mise en gras ou en italique, ajout de couleurs, ou changement de police pour créer un contraste). Cette technique est toutefois peu répandue de nos jours.

13. Lorsqu’un séparateur autre que les parenthèses et les crochets est utilisé pour former un doublet abrégé pour un mot au pluriel, l’usage hésite à doubler le séparateur avant la marque du pluriel. Nous nous cantonnons, pour le reste de ce mémoire, à l’emploi d’un point médian simple, préconisé par Viennot (2018 : 103), qui indique que le point médian « doit demeurer unique en cas de pluriel [...] Le second point est un héritage des parenthèses ; parce qu’elles étaient deux, on a eu tendance à dupliquer tous les signes expérimentés. »

14. Le point final est également utilisé comme séparateur, mais cela semble être davantage par limitation technique et par difficulté d’accès au point médian sur les ordinateurs et les smartphones que par la volonté d’imposer un nouveau séparateur. Par conséquent, nous l’incluons avec le point médian.

féminine précédée du séparateur directement après la forme masculine (un/e marchand/e ambulante/e ; un·e grand·e musicien·ne).

En revanche, certains morphèmes féminins s'éloignent radicalement de leurs équivalents féminins, et l'ajout d'une finale féminine ne suffit pas à former la forme féminine : pour ne prendre que l'exemple des noms, « vendeur » devient « vendeuse », « instituteur » devient « institutrice », « boulanger » devient « boulangère », et « traître » devient « traîtresse ». Dans ces cas, c'est en général le suffixe féminin dans son entièreté qui est placé après la fin du mot au masculin, précédé du séparateur (vendeur·euse ; instituteur·trice ; boulanger·ère ; traître·esse).

- (13) C'est un avis largement répandu selon lequel **l'éleveur·euse** aurait moins de contact avec ses vaches en installant un robot. . .

(Eva Studinger, Satisfait de son système de traite automatique, *Revue UFA* [décembre 2021], disponible sur <https://api.swissmilk.ch/wp-content/uploads/2022/01/batiments-equipements-satisfait-de-son-systeme-de-traite-automatique-article-ufa-revue-12-2021-fr.pdf>)

L'usage n'a toutefois pas définitivement tranché sur les formes à adopter, et celles-ci se confondent parfois. Le site du gouvernement du Canada, dans un article sur la rédaction et l'utilisation des doublets abrégés, indique que « lorsque la forme masculine du mot n'est pas entièrement comprise dans sa forme féminine [. . .] on peut maximiser le gain d'espace en n'incluant après le point médian que les lettres qui diffèrent entre les deux formes : inspecteur·rice ; travailleur·se ; administratif·ve ; aventurier·ère. »¹⁵ Selon cette même recommandation, « inspecteurs et inspectrices » devrait alors être abrégé en « inspecteur·ice » (et non « inspecteur·rice »), pour ne pas répéter la consonne présente dans la forme au masculin. L'Office québécois de la langue française, pour sa part, recommande l'utilisation du suffixe féminin complet¹⁶, mais ne traite pas les cas des noms avec le suffixe féminin -esse, tels « maitresse », « prêtresse » ou « traîtresse ».

2.1.4.2 Les techniques de neutralisation

Comme nous l'avons vu précédemment, la formation de doublets abrégés lorsque le morphème masculin et le morphème féminin diffèrent grandement peut poser de nombreuses difficultés. En conséquence, certaines personnes privilégient la suppression pure et simple du séparateur afin de créer des néologismes neutres. Lorsque la dernière lettre du mot au masculin est présente dans le suffixe féminin, le mot et le suffixe sont alors amalgamés sans séparateur au niveau de la lettre en commun pour créer un nouveau mot (instituteurice ; inspecteurice). Dans les autres cas, le suffixe est intégré dans son intégralité à la forme masculine pour créer un résultat phonologiquement acceptable (vendeuseuse ; éleveuseuse).

15. <https://www.noslangues-ourlangues.gc.ca/fr/cles-de-la-redaction/ecriture-inclusive-doublets-abreges>

16. <https://vitrinelinguistique.oqlf.gouv.qc.ca/index.php?id=25368>

- (14) Dans ces trois États, l'effet des protestations des **agriculteurices** s'est fait sentir – et, contrairement aux plus vulgaires laquais du grand capital, le Premier ministre l'a compris. (Palagummi Sainath, Victoire historique des paysan·nes indien·nes, *Contretemps* [27 novembre 2021], disponible sur <https://www.contretemps.eu/victoire-historique-paysans-indie-modi/>)

Cette solution, bien qu'intéressante, car directement applicable à l'oral, et donnant lieu à une forme plus compacte encore que les doublets abrégés, ne résout pas tous les problèmes posés par ceux-ci et en crée de nouveaux : du point de vue phonologique notamment, certaines combinaisons semblent difficiles à réaliser (comment amalgamer « boulanger·ère » ou bien « administratif·ve » ?). De plus, les mots qui se forment avec une finale féminine ne sont pas compatibles avec cette règle. Reste aussi en suspens la question des noms avec le suffixe féminin -esse, évoquée précédemment. Enfin, un dernier problème mentionné par Éliane Viennot est celui de l'accord avec les déterminants : « si *lecteur* et *lectrice* régressaient, *lecteurice* devrait se substituer à eux au singulier aussi ; il lui faudrait donc un article : *la* ou *le* ? » (Viennot, 2018 : 85).

D'autres techniques de neutralisation ont également vu le jour, comme le recours aux déterminants, aux noms, aux pronoms et aux adjectifs épicènes. La forme des mots épicènes est identique au masculin et au féminin, ce qui permet de rédiger des textes plus concis et moins lourds que lorsque les doublets sont utilisés. En ce qui concerne les noms épicènes, certains peuvent être utilisés avec les deux genres, au choix (« dentiste », « membre », « spécialiste » [ex. 15]) ; les autres (« personnalité », « personne », « génie » [ex. 16]) sont des noms métonymiques qui, à l'origine, désignaient une chose, avant de faire référence à un individu (Viennot, 2021 : 4,13). Par conséquent, un genre arbitraire leur est associé, ce qui facilite les cas d'accord avec les déterminants et les adjectifs non épicènes et résout les problèmes posés par les néologismes mentionnés auparavant.

- (15) Dans cinquante ans, la langue des frères De Goncourt, par exemple ne sera comprise que des **spécialistes**.
(Frantext : Paul Bourget, *Essais de psychologie contemporaine* [1883], p. 17)
- (16) Cette femme est **un génie**, elle sait.
(Frantext : Florence Delay, *Le Aïe Aïe de la corne de brume* [1975], p. 200)

Enfin, les noms collectifs permettent également la neutralisation du genre en faisant référence à un ensemble de personnes. Nous nous basons initialement sur la définition de Lammert et Lecolle (2014) sur les noms collectifs, qui sont considérés comme « [des noms] qui, au singulier, dénote[nt] une entité composée d'un regroupement d'éléments » (2) ¹⁷. L'un des avantages des noms collectifs par rapport aux noms épicènes est qu'ils disposent tous d'un genre arbitraire (ex. 17). Étant donné la généricité des noms collectifs humains, qui mettent de côté l'individu pour se concentrer sur un groupe particulier de personnes ayant des caractéristiques communes (Lecolle

17. Dans le **chapitre 2.2.1**, nous revenons plus en détail sur les considérations définitionnelles des noms collectifs par une étude de la littérature.

(2013, 2016)), on a souvent recours aux expansions adjectivales (ex. 18) ou prépositionnelles (ex. 19) afin de les qualifier plus précisément.

- (17) Pour la première fois ce manuscrit achevé en décembre 1966 me parut digne d’être soumis au **lectorat** d’une maison d’éditions.
(Frantext : Michel Tournier, *Le Vent Paraquet* [1977], p. 194)
- (18) Hervé avait sa voix speedée, à peine audible. Ils étaient tous massés devant l’accueil, énervant le **personnel hospitalier** qui entendait soustraire le blessé aux mondanités et au tapage médiatique.
(Frantext : Julia Kristeva, *Les Samouraïs* [1990], p. 403)
- (19) Le travail de subversion auquel Artaud avait ainsi depuis toujours soumis l’impérialisme de la lettre avait le sens négatif d’une révolte tant qu’il se produisait dans le **milieu de la littérature** comme telle.
(Frantext : Jacques Derrida, *L’Écriture et la différence* [1967], p. 283)

En conclusion de cette partie, nous nous proposons d’établir un tableau récapitulatif des différentes techniques d’écriture inclusives. Ce tableau ne se veut toutefois pas exhaustif : les formes d’écriture sont légion, et certaines restent cantonnées à de petits groupes sociaux ; il paraît donc difficile d’en faire un inventaire complet.

Technique d’écriture inclusive	Objectif	Exemples
Doublet complet	visibilisation	les enseignantes et les enseignants ; les enseignants ou les enseignantes
Doublet abrégé	visibilisation	les enseignant·es ; les instituteur/trices ; les étudiant*es ; les président(e)s ; les directeur(-trices) ; les apprenant’es
Néologisme formé par amalgamation	neutralisation	les instituteurices ; les directeurices
Mot épïcène	neutralisation	les élèves ; les pédagogues ; les responsables
Nom collectif	neutralisation	l’enseignement ; le monde étudiant ; le corps professoral ; la direction

TABLEAU 2.1 – Récapitulatif des techniques d’écriture inclusives actuelles

2.1.5 Le point médian : un signe typographique polémique

Au sein de la littérature féministe, plusieurs ont fait le choix d’abandonner les techniques d’écriture mentionnées ci-dessus en faveur du point médian uniquement, car elles empêchaient la promotion d’une réelle égalité entre les hommes et les femmes. Comme nous l’avons vu, certaines techniques menaient à une disproportion de la forme féminine par rapport à la forme masculine. D’autres, selon certaines personnes, allaient contre la visibilisation de la forme

féminine, en la cloisonnant par exemple entre parenthèses ou entre crochets et en la rendant, dès lors, implicitement facultative. C'est ce qu'indique Viennot (2018 : 102-103) :

« Le point médian paraît le meilleur candidat pour porter ce type d'abréviations, car il n'est pas utilisé pour autre chose dans notre langue, où il était inconnu jusqu'il y a peu. Il n'est en conséquence pas connoté : ni négativement (comme les parenthèses qui minorisent l'énoncé placé à l'intérieur, ou la barre oblique qui indique une opposition entre les termes, ou le trait d'union qui par définition les unit) ; ni positivement (comme la majuscule qui donne l'impression d'une plus grande importance du féminin. »

Ces considérations faites, le point médian semble s'imposer naturellement comme une solution viable à l'alternative entre le masculin et le féminin, puisqu'il permet de garder identiques les formes associées aux deux genres, en agissant simplement comme un séparateur. En outre, contrairement à d'autres séparateurs qui ont pu être suggérés par la littérature féministe et par l'usage, tels que l'apostrophe ou l'astérisque, le point médian s'insère parfaitement du point de vue typographique, car il se trouve centré par rapport aux autres caractères, ce qui permet d'assurer un plus grand confort de lecture.



ILLUSTRATION 2.1 – Schéma de mesure typographique avec le point médian

Police utilisée : Times New Roman. De haut en bas, les lignes représentent respectivement l'ascendante, la hauteur de capitale, la ligne médiane et la ligne de pied.

Toutefois, l'usage de ce signe typographique pour séparer les deux genres a donné naissance à une vive polémique en 2017, après la publication, par Le Figaro, d'un article le 22 septembre 2017 (Pech, 2017) à propos de la présence de points finaux comme séparateurs dans un manuel scolaire d'histoire-géographie des éditions Hatier. S'ensuit une déclaration de l'Académie française publiée le 26 octobre 2017 (Académie française, 2017), dans laquelle l'institution fustige l'écriture inclusive et va même jusqu'à affirmer que « la langue française se trouve désormais en péril mortel ». Enfin, une circulaire est publiée le 21 novembre 2017 par le Premier ministre Édouard Philippe, qui demande à ce que les textes publiés au Journal officiel ne fassent pas « usage de l'écriture dite inclusive, qui désigne les pratiques rédactionnelles et typographiques visant à substituer à l'emploi du masculin, lorsqu'il est utilisé dans un sens générique, une graphie faisant ressortir l'existence d'une forme féminine. » (Édouard Philippe, 2017).

Une analyse des critiques à l'encontre de l'écriture inclusive montre que ces dernières se sont en réalité surtout cristallisées autour du point médian, et non pas tant autour de l'écriture

inclusive en tant que telle. C'est bien en réponse à la polémique du point médian qu'a été publiée la circulaire Philippe. D'ailleurs, la définition extrêmement floue donnée pour l'écriture inclusive dans la circulaire laisse penser à l'interdiction d'autres techniques d'écriture inclusives, quand bien même ces dernières sont utilisées depuis longtemps, y compris dans les documents officiels ¹⁸, comme le rappelle Éliane Viennot dans une tribune du Monde du 2 avril 2021 :

« Quant aux abréviations servant à noter ces doublets, elles sont anciennes. [...] Nos cartes d'identité suivent ce système : "né(e) le". Et des milliers de formulaires en font autant, comme celui qu'affiche aujourd'hui le service "accueil des étrangers" du ministère de l'intérieur : "Vous êtes marié(e) avec un(e) ressortissant(e) français(e). . ." »

Il faut également prendre en compte qu'avant la publication du manuel Hatier tant décrié, personne n'avait pris la peine de relever ce phénomène ; c'est ce qu'indique Noémie Mari-gnier (Abbou et al., 2018) :

« La particularité de la polémique en 2017, c'est surtout son ampleur. Jusqu'à l'automne 2017, il y avait déjà eu des articles et des prises de positions pour ou contre l'écriture inclusive, mais sans déclencher une telle profusion de discours. En 2015, le Haut Conseil à l'Égalité entre les femmes et les hommes avait sorti un "Guide pratique pour une communication publique sans stéréotype de sexe" sans remous et sans susciter de polémique, par exemple. »

Les critiques lancées contre le point médian sont diverses et variées : ce dernier est considéré comme inesthétique, nuisant à la compréhension, à la lecture ou à la prononciation, ou ajoutant trop de complexité. Dans un état d'esprit similaire à ces critiques, ou en réponse à celles-ci, certaines autorités linguistiques ont pris le choix de privilégier et de recommander l'utilisation d'autres techniques d'écriture inclusives. Par exemple, l'Office québécois de la langue française (OQLF), dont la mission est de conduire la politique linguistique française au Québec, préconise d'utiliser les parenthèses et les crochets pour former les doublets abrégés, en lieu et place du point médian. L'OQLF justifie ce choix en raison du « rapprochement visuel incongru qu'ils créent entre le déterminant féminin et la partie masculine de l'appellation de personne. Ces deux éléments semblent alors encadrés. » ¹⁹

Antidote, un logiciel de correction orthographique développé par Druides Informatique, une entreprise québécoise, va même plus loin, puisqu'un billet publié sur le site Web du logiciel indique qu'« [i]l est unanimement déconseillé de les employer [les doublets abrégés] pour représenter les deux genres, et ce quel que soit le contexte de rédaction ou le format choisi (parenthèses, traits d'union, points médians, etc.) », car ils « entraînent à l'écrit des complications

18. Par exemple, pour la mention « né(e) le » sur les cartes d'identité, c'est le cas depuis la mise en circulation des cartes d'identité plastifiées, en 1995. Pour les nouvelles cartes d'identité au format européen (en circulation depuis 2021), cette mention s'est vue remplacée par une formulation neutre : « Date de naissance ».

19. <https://vitrinelinguistique.oqlf.gouv.qc.ca/index.php?id=25345>

graphiques et grammaticales, en plus d'être peu esthétiques »²⁰. Les doublets abrégés doivent donc être remplacés par des techniques d'écriture inclusives autres, comme le recours aux noms épiciques, aux noms collectifs ou aux doublets complets.

C'est dans cette perspective de techniques d'écriture inclusives alternatives que s'inscrit le projet de ce mémoire, en ce que nous pensons que les noms collectifs peuvent parfaitement remplir les fonctions de représentativité et d'inclusivité portées par l'écriture inclusive, tout en évitant le recours au point médian et à ses dérivés, signes typographiques encore peu acceptés par l'usage pour l'accomplissement de ces fonctions.

2.2 Les noms collectifs en français

2.2.1 Les noms collectifs : définition

Les noms collectifs ont été largement abordés dans la littérature, en français et dans d'autres langues²¹, mais restent une inlassable source de questionnements, notamment en ce qui concerne la délimitation de la catégorie des collectifs, et la définition précise d'un nom dit collectif. Flaux (1999), Benninger (2001), Lammert (2010) et Lecolle (2019) ont notamment proposé diverses définitions. Nous en retenons trois points principaux. Premièrement, les noms collectifs, en français du moins, se caractérisent par un singulier morphologique qui permet de renvoyer sémantiquement à plusieurs entités. Deuxièmement, les entités auxquelles il est fait référence sont hétérogènes et peuvent être isolées, mais constituent, au sein du nom collectif, un ensemble homogène. Troisièmement, les noms collectifs sont des entités en eux-mêmes, qui regroupent des sous-entités.

Les noms collectifs peuvent renvoyer à des référents inanimés (ex. 20) aussi bien qu'animés (humains [ex. 21] ou animaux [ex. 22]). Dans le cas des noms collectifs humains, ils peuvent être sujets à la métonymie (ex. 23 et 24). Ce type de noms collectifs en particulier sera abordé plus avant dans le cadre du **chapitre 2.2.4**.

- (20) La **forêt** recouvrait l'endroit où s'était dressé le bâtiment des latrines collectives, lieu de libertés multiples dans le plus lointain cercle de l'enfer.
(Frantext : Jorge Semprun, *L'écriture ou la vie* [1994], p. 390)
- (21) Le **voisinage** colportait que sa vertu ne valait pas cher, qu'elle ne pouvait vivre sans avoir un arbre planté dans son jardin.
(Frantext : Émile Ollivier, *Mille eaux* [1999], p. 91)
- (22) Seul, peut-être, le berger chleuh, celui qu'on appelle le Hartani, vient lui aussi quelquefois sur le plateau, mais c'est quand une des chèvres du **troupeau** s'est égarée en courant le long des ravins.
(Frantext : Jean-Marie Gustave Le Clézio, *Désert* [1980], p. 96)

20. <https://www.antidote.info/fr/antidote-10/documentation/guide-utilisation/le-correcteur/le-volet-style/inclusivite>

21. Pour l'anglais, voir par exemple Arigne (2011) et Gardelle (2016).

- (23) Insomniaque, ma mère voulait absolument que je dorme l'après-midi et toute la **maison** devait alors observer le silence le plus profond.
(Frantext : Georges-Arthur Goldschmidt, *La traversée des fleuves* [1999], p. 77)
- (24) Le budget est voté le 31 décembre, le Parlement a approuvé sa politique extérieure par une faible, mais suffisante **majorité**.
(Frantext : Christian Lazard, *Journal* [1935])

2.2.2 Caractéristiques morphologiques, syntaxiques et sémantiques des noms collectifs

2.2.2.1 Caractéristiques morphologiques

Au niveau morphologique, les noms collectifs peuvent se construire avec des suffixes spécifiques qui dénotent la notion de collectivité ou de groupe. Parmi les suffixes qui peuvent endosser ce rôle²², nous pouvons mentionner « -at » (patronat, lectorat, électorat), « -èle » (clientèle, parentèle), « -ade » (cousinade, escouade, peuplade), « -erie » (gendarmerie, chevalerie, lingerie), ou encore « -age » (voisinage, outillage, feuillage).

Cette caractéristique morphologique propre aux noms collectifs favorise ainsi la productivité linguistique et la construction de nouveaux mots par processus de dérivation. Nous avons ainsi pu voir de nouveaux noms collectifs apparaître dans l'usage simplement grâce à la suffixation. C'est notamment le cas du nom « patientèle », formé sur le modèle de « clientèle », à partir du nom commun « patient » et du suffixe « -èle » afin de faire référence à l'ensemble des patients et patientes d'un cabinet médical. D'après le Wiktionnaire²³, celui-ci « n'est né que très récemment », quand bien même on trouve des usages du nom commun « patient » depuis 1120. Une recherche à l'aide de Google Ngram Viewer, le corpus de textes de Google qui permet de constater l'évolution de la fréquence d'un terme au fil du temps, montre que le terme apparaît dans la littérature en 1975 (ex. 25).

- (25) En raison de la diversité de l'attitude thérapeutique des participants et du genre de **patientèle** (selon l'éducation sanitaire, la moyenne d'âge ou la disponibilité du praticien), on ne peut tirer de conclusions sur la qualité de la prise en charge elle-même.
(Google Books : *Swiss Medical Weekly*, numéros 84—94 [1975], p. 44, disponible sur https://www.google.fr/books/edition/Schweizerische_medizinische_Wochenschrif/matLAQAAIAAJ?hl=en&gbpv=1&bsq=patient%C3%A8le&dq=patientÃle&printsec=frontcover)

D'après le site du Club d'orthographe de Grenoble²⁴, qui consacre une page aux mots nouveaux intégrés aux dictionnaires, le nom commun « patientèle » a rejoint le dictionnaire Larousse en 2012, et celui du Petit Robert un an plus tard, en 2013. Depuis, l'usage de ce nom s'est

22. Comme le rappellent Lammert et Lecolle (2014 : 11), il convient de préciser que ces suffixes n'ont pas pour autant une valeur intrinsèque de collectivisation, et sont communément utilisés avec des valeurs autres.

23. <https://fr.wiktionary.org/wiki/patient%C3%A8le>

24. <https://www.orthogrenoble.net/mots-nouveaux-dictionnaires/mots-entres-petit-larousse-puis-petit-robert/>

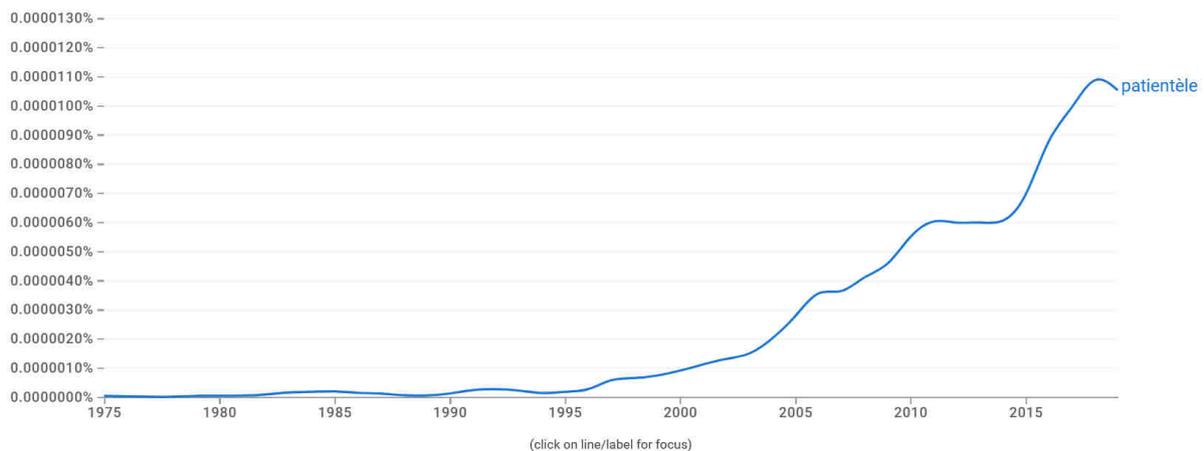


ILLUSTRATION 2.2 – Évolution de la fréquence du nom collectif « patientèle » dans Google Ngram Viewer

Recherche effectuée le 5 février 2023. https://books.google.com/ngrams/graph?content=patient%C3%A8le&year_start=1975&year_end=2019&corpus=fr-2019&smoothing=1&case_insensitive=true.

largement répandu, et il est maintenant communément employé autant sur les sites administratifs et gouvernementaux (comme celui de la Caisse nationale de l'assurance maladie²⁵) que dans la presse d'information générale²⁶.

L'exemple de l'essor de « patientèle » confirme dès lors la grande productivité que l'on peut associer aux noms collectifs et montre comment, par une simple suffixation, il est possible de transformer des noms communs faisant référence à des individus en des noms collectifs qui véhiculent l'idée de collectivité.

2.2.2.2 Caractéristiques syntaxiques

Du point de la vue de la syntaxe, les noms collectifs employés seuls en position sujet se comportent comme des noms communs singuliers standards en dépit de leur sémantisme pluriel (Gross, 2011). Dans ce cas, les verbes (ex. 26) et adjectifs (ex. 27) s'accordent alors au singulier.

(26) La **chevalerie**, aimant le luxe et les festins, **a** aussi sa conception des hostilités.

(Frantext : Jean Teulé, *Azincourt par temps de pluie* [2022], p. 34)

(27) Il est entièrement sous la coupe d'une **administration** par ailleurs **vétilleuse** et **regardante**.

(Frantext : Jules Romains, *Le Dieu des corps* [1928], p. 228)

Cela ne signifie toutefois pas que les noms collectifs puissent être uniquement singuliers : comme les noms communs comptables, ils peuvent être pluralisés pour dénoter plusieurs ensembles, et l'accord se fait en conséquence (ex. 28).

25. <https://assurance-maladie.ameli.fr/etudes-et-donnees/par-theme/professionnels-et-etablissements-de-sante/patientele-professionnels-sante-liberaux>

26. <https://www.ouest-france.fr/pays-de-la-loire/montoir-de-bretagne-44550/des-medecins-de-montoir-de-bretagne-partent-avec-leur-patientele-mais-les-patients-restent-0d3c266c-a152-11ed-b727-95140ec04a42>

(28) Ces us sont venus sans doute des temps d'alerte, alors que, rançonnés par les **bandes armées**, les habitants se serraient les uns contre les autres. . .

(Frantext : Joseph de Pesquidoux, *Le livre de raison*, tome 1 [1925], p. 14)

En outre, les noms collectifs peuvent parfois jouer le rôle de substantif quantificateur (désormais « SQ »), dans une construction de type [Dét-Ncoll-*de*-N2]. Ils servent alors à dénoter un certain nombre d'éléments, généralement une grande quantité, ou bien un ensemble cohérent avec N2. Dans ce genre de constructions, nous avons souvent affaire à des cas de syllepse, défini par le dictionnaire de linguistique de Larousse (Dubois et al., 2001) comme « l'accord des mots en genre et en nombre non d'après la grammaire, mais d'après le sens. » On peut alors trouver des cas d'accord au singulier (accord avec le SQ, privilégiant la morphosyntaxe [ex. 29]) autant que des cas d'accord au pluriel (accord avec N2, privilégiant le sens [ex. 30]).

(29) Une armée de mains s'est levée.

(30) Une armée de mains se sont levées.

Les noms collectifs également peuvent être sujets à ces phénomènes de syllepse en cas de reprise anaphorique. C'est notamment le cas lorsque les éléments regroupés au sein du nom collectif constituent des entités pleinement autonomes et ontologiques.

(31) Titre de journal : Au Royaume-Uni, le combat d'une famille contre les géants d'Internet après la mort de leur fille

(Cécile Ducourtieux, Au Royaume-Uni, le combat d'une famille contre les géants d'Internet après la mort de leur fille, *Le Monde* [4 octobre 2022], disponible sur https://www.lemonde.fr/international/article/2022/10/04/au-royaume-uni-le-combat-d-une-famille-contre-les-geants-d-internet-apres-la-mort-de-leur-fille_6144256_3210.html)

Dans l'exemple 31, le déterminant « sa » devrait normalement remplacer « leur », puisqu'il reprend le nom collectif singulier « famille ». Toutefois, la famille étant considérée comme un ensemble d'individus autonomes disposant du trait [+ humain], il ne paraît pas exclu de choisir un accord au pluriel.

2.2.2.3 Caractéristiques sémantiques

Flaux (1999) mentionne plusieurs caractéristiques sémantiques que partagent les noms collectifs. La première d'entre elles est celle qu'elle nomme la « référence cumulative hétérogène » (p. 473), qui est la possibilité pour les noms collectifs de s'additionner, soit pour se diviser en deux unités distinctes, soit pour former un groupe plus grand. Ainsi, si l'on ajoute un troupeau à un troupeau, on se retrouvera soit avec deux troupeaux séparés, soit avec un troupeau plus grand. C'est cette caractéristique qui distingue les noms collectifs des noms massifs, en ce que ces derniers ne peuvent pas s'additionner de manière homogène (Lammert, 2014). En effet, de la vaisselle + de la vaisselle reste un groupe homogène « de la vaisselle » : on ne peut pas le diviser.

Cette référence cumulative hétérogène est directement associée aux principes de « fusion » et de « scissiparité » définis par Flaux (1999) pour les noms collectifs. Selon Flaux, étant donné

que les entités identifiées au sein de noms collectifs disposent de leur pleine autonomie, il est possible de les fusionner pour créer un tout : par exemple, « en ajoutant une fleur à une fleur puis encore une autre, etc. on peut former un bouquet. » (476). Il est possible, de la même manière, de diviser le groupe auquel appartiennent ces entités en sous-groupes, d'où le principe de scissiparité : un bouquet est une entité collective composée de plusieurs entités « fleur » qui en viennent à former une seule entité. Flaux précise ainsi que « la division des individus collectifs peut passer par une - ou plusieurs - étape(s) intermédiaire(s) » (476). On peut alors avoir un très grand bouquet, le diviser en plusieurs bouquets moyens, diviser ces derniers en bouquets encore plus petits, etc., jusqu'à obtenir de manière individuelle l'ensemble des entités « fleur » qui composaient le très grand bouquet initial. Ces propriétés sémantiques peuvent se vérifier si l'on compare l'emploi des verbes « fusionner » et « diviser »²⁷ avec des noms individuels et des noms collectifs :

(32) On a fusionné les deux syndicats. / *On a fusionné les deux syndicalistes.

(33) Les deux armées se sont divisées. / *Les deux soldats se sont divisés.

En outre, comme l'indique Pusch (2020), dans un contexte neutre, certains noms regroupés au sein des noms collectifs laissent transparaître directement ce qu'ils incluent, comme c'est le cas de « armée » ou « gouvernement », tandis que d'autres, comme « bouquet », sont plus flous au niveau de leur constitution²⁸.

Cela se confirme si l'on essaye de poser une question dans un contexte non métaphorique demandant d'explicitier le contenu du nom collectif auquel il est fait référence. Pour les noms collectifs dont le sémantisme est limpide, une telle question ne semble pas fonctionner ou, tout du moins, ne s'impose clairement pas, puisque le contenu est déjà dévoilé implicitement :

(34) Il a vu défiler une armée. */?Une armée (composée) de quoi ?

(35) Il a acheté un bouquet. Un bouquet (composé) de quoi ? { de roses, de lilas... }

2.2.3 Noms collectifs humains (Ncoll·H)

Dans la littérature consacrée aux noms collectifs, la catégorie des noms collectifs humains (désormais « Ncoll·H ») a été largement étudiée (Boughanmi, 2016 ; Flaux, 1999 ; Gross, 2009, 2011 ; Lammert, 2010 ; Lammert et Lecolle, 2014 ; Lecolle, 2008, 2013, 2016, 2019) : les Ncoll·H représentent non seulement une grande part des noms collectifs, mais ils constituent également une catégorie complexe à maints niveaux (syntaxique, sémantique, discursif). Dans son ouvrage dédié aux Ncoll·H, Lecolle (2019 : 75) propose la définition suivante : « un nom désignant une entité composée d'un ensemble d'individus humains ».

27. Nous empruntons le test relatif au verbe « fusionner » à Flaux (1999 : 475), et ajoutons celui pour le verbe « diviser ». Nous considérons ce dernier dans son sens premier et non figuré, c'est-à-dire, selon le TLFi, « [s]éparer (une chose, un ensemble de choses, un groupe de personnes) en plusieurs unités ou parties. » (<https://www.cnrtl.fr/definition/diviser>).

28. Notons que cela n'a pas un rapport intrinsèque avec le fait que ces noms collectifs puissent faire partie d'une séquence binominale [Dét-Ncoll-de-N2], car l'on peut aussi bien utiliser « armée » que « bouquet » dans ce type de structures (une armée de sapins / un bouquet de cerises).

Les Ncoll·H se distinguent des autres types de noms collectifs en ce que les éléments qu'ils regroupent sont de véritables individus disposant d'un trait ontologique. Ce faisant, contrairement à des noms collectifs non humains animés ou inanimés tels que « bétail » ou « bouquet », les individus regroupés sous ces dénominations jouissent d'une autonomie propre et inaliénable. Certes, ils peuvent avoir été rassemblés « de force » sous des appellations classifiantes qui peuvent leur paraître en décalage avec leur propre réalité (c'est notamment le cas pour les Ncoll·H à valeur péjorative, souvent accompagnés du suffixe « -aille », tels que « racaille » ou « flicaille » [Lecolle (2016 : 10-12)]), mais la particularité des Ncoll·H est également qu'ils permettent aux individus de se classifier eux-mêmes et ainsi de constituer un groupe se démarquant par rapport aux autres.

Plusieurs propositions de classement des Ncoll·H sont répertoriées dans la littérature, notamment par Gross (2011) et Lecolle (2013). Nous choisissons de nous concentrer sur ce dernier. Sont alors présentés quatre²⁹ groupes principaux de Ncoll·H en fonction de leur signification lexicale. Le premier groupe, relatif aux « rassemblements contingents », rassemble des noms tels que « foule », « public » ou « auditoire ». Il inclut aussi des noms de lieu qui peuvent servir à désigner toutes les personnes qui s'y trouvent, comme « assemblée ». Le deuxième groupe concerne les Ncoll·H axiologiquement marqués (point abordé plus en détail ci-après), de manière péjorative ou méliorative (« racaille », « crème »...). Le troisième groupe inclut des Ncoll·H associés par un « lien fonctionnel, social [ou] politique », celui-ci étant « en rapport étroit avec le fonctionnement de la société » : on trouve alors des noms tels que « famille », « société », « troupe », « police » ou bien « syndicat ». Enfin, le quatrième groupe rassemble des noms dont la signification lexicale évoque « la constitution d'un ensemble selon des propriétés similaires » : ils peuvent alors refléter l'identité des personnes à l'intérieur du groupe, et le groupe en question est considéré comme « homogène ». Lecolle cite des noms collectifs communs tels que « classe », « caste » ou « race », et d'autres, éminemment politiques, tels que « gauche », « droite » ou « opposition ».

Ce classement proposé pour les Ncoll·H permet d'avoir une meilleure vue d'ensemble de leurs différentes caractéristiques et du rapport individu/groupe qui est mis en jeu. Ainsi, même si certains Ncoll·H rassemblent des individus à priori hétérogènes qui n'ont en commun que « d'être » à un moment précis, et de manière éphémère, par un jeu particulier de circonstances (« public », « foule »), d'autres, en revanche, permettent d'établir des liens beaucoup plus profonds entre les individus qui font partie du groupe. Si l'on peut rassembler toutes les personnes faisant partie du « public » sous la dénomination de « spectateurs/spectatrices », cela ne nous dit pas grand-chose sur ces personnes-là en tant que telles, et notamment leur rôle au sein d'une société. Cela s'avère bien plus pertinent lorsque le « lien fonctionnel, social [ou] politique » évoqué par Lecolle transparait dans le Ncoll·H, car nous disposons d'informations plus précises sur les individus regroupés (police → policiers/policières ; syndicat → syndicalistes ; droite → partisan/partisane d'un mouvement de droite, etc.).

29. Dans l'article original de Lecolle (2013), le groupe 4, dénotant « le rassemblement en lui-même », est également évoqué, mais en raison de son caractère extrêmement générique et peu pertinent pour notre analyse, nous nous permettons de ne pas le citer.

En outre, dans la perspective de marquage mélioratif/péjoratif mentionné plus haut, Lecolle (2016) propose également un exemple de classement des Ncoll·H. De même, il nous paraît intéressant de nous attarder quelque peu sur cette catégorisation afin de connaître les enjeux axiologiques et discursifs associés aux Ncoll·H. Lecolle distingue ainsi trois grandes catégories : les Ncoll·H fonctionnels, les Ncoll·H intrinsèquement évaluatifs, et les Ncoll·H de catégories et d'identités.

Comme l'indique son nom, la catégorie des Ncoll·H fonctionnels met en exergue la fonction associée au groupe : les individus rassemblés dans ce groupe le sont parce qu'ils partagent un objectif en commun. Lecolle donne comme exemples à polarité neutre des Ncoll·H tels que « organisation », « police » ou encore « association ». Elle note toutefois que « certains autres Ncoll·H fonctionnels sont négativement marqués, pour des raisons liées au référent ou à la signification lexicale du nom » (10), et que le caractère péjoratif peut porter soit sur l'activité pratiquée par le groupe (« mafia », « gang », « secte »), soit sur la signification du nom utilisé pour désigner le groupe (« bande », « faction », « groupuscule »). Il est en effet pertinent de noter que, si l'on peut aisément associer des caractéristiques négatives précises à des noms tels que « mafia » et « gang », et notamment les notions de délit et de méfait, cela se révèle bien plus difficile pour des noms tels que « bande » ou « groupuscule », car ils sont soit polysémiques, soit sont plus flous dans leur sémantisme. Cela peut aussi expliquer pourquoi il est plus facile d'employer ces derniers comme SQ, car ils n'ont pas nécessairement de nom attendu qui puisse leur être lié :

(36) une bande de filles / une bande de comédiens / un groupuscule de jeunes³⁰

(37) *une mafia de rebelles / *une secte de professeurs / ?un gang d'artistes³¹

Les Ncoll·H intrinsèquement évaluatifs, eux, se voient directement associés une qualification péjorative ou bien méliorative. Dans la catégorie des Ncoll·H intrinsèquement évaluatifs péjoratifs, on retrouve par exemple « racaille », « vermine » ou « lie », tandis que du côté des noms mélioratifs siègent « crème », « élite » ou « gratin ». Lecolle précise toutefois que ces noms mélioratifs peuvent être « employés péjorativement, par anti-phrase ».

Enfin, les noms de catégories et d'identités visent à rassembler des personnes assignées à une même catégorie ou identité. Cette catégorie regroupe alors aussi bien des noms non marqués (« francophonie ») que des noms marqués (« oligarchie »), semble-t-il exclusivement péjorativement. Leur classification est par nature subjective, sauf dans certains cas spécifiques, et ils sont donc considérés comme « potentiellement sensibles ». Lecolle cite par exemple les noms « électorat » et « noblesse » : si l'on ne peut pas dire que l'assignation de membres à la catégorie « électorat » soit véritablement subjective (une personne possède le droit de vote ou non, et vote ou non), il en est autrement pour « noblesse », car les individus assignés à cette catégorie ne pensent pas forcément (à tort ou à raison) partager toutes les caractéristiques associées à /noble/.

30. Extrait d'Europarl : séance du mercredi 4 septembre 2022, disponible sur https://www.europarl.europa.eu/doc/eo/document/CRE-5-2002-09-04_FR.html.

31. Notons toutefois qu'en français du Québec, on emploie le nom « gang » au féminin comme synonyme de « bande, groupe » (« une gang d'ami-es »).

2.2.3.1 Ncoll·H et neutralisation en genre

Ncoll·H spécifiques Les Ncoll·H, par leur capacité à regrouper des individus sous un nom singulier au genre lexicogrammatical abstrait, sont de bons candidats pour la neutralisation en genre. Tandis que des noms comme « policiers/policières » peuvent poser des problèmes dans une perspective d'écriture inclusive, l'on peut simplement neutraliser le genre en utilisant le Ncoll·H « la police » pour y faire référence de manière englobante. Il paraît alors évident que le nom commun non collectif d'origine doit être un nom pluriel, même si l'on peut imaginer des contextes où un Ncoll·H pourrait tout de même marcher si le nombre n'agit pas comme une variable essentielle. Par exemple, dans l'énoncé 38, ce n'est pas tant le nombre que la fonction associée au Ncoll·H « police » qui est importante :

(38) J'étais à la maison quand la police a frappé à la porte.

Si l'on imaginera plutôt une situation où il y a plusieurs personnes qui font partie de la police, le fait qu'il y ait un(e) membre de la police ou bien plusieurs ne change pas pour autant l'interprétation de l'énoncé. En revanche, cela peut s'avérer plus problématique dans d'autres contextes. Ainsi, l'exemple 39 ne peut en aucun cas être l'équivalent de l'exemple 40, car un seul « jeune » ne saurait refléter une « jeunesse » entière.

(39) Le jeune a été traumatisé par l'attaque.

(40) La jeunesse a été traumatisée par l'attaque.

Lecolle (2019 : 287-289) regroupe ainsi 138 Ncoll·H auxquels sont associés des noms de membres. Le Ncoll·H peut être morphologiquement associé au nom correspondant (bourgeois·es → bourgeoisie) ou non (spectateurs/spectatrices → public). De plus, comme le précise Lecolle, un même Ncoll·H peut avoir plusieurs équivalents en noms communs : par exemple, pour le Ncoll·H « milice », ses membres peuvent aussi bien être des « miliciens » que des « soldats ». Les hyponymes doivent aussi être pris en compte en fonction du contexte : dans le cas d'« orchestre », on a des « musiciens », mais on peut être encore plus spécifique selon l'instrument qui est joué par la personne en question (« trompettiste », « guitariste », « violoniste », etc.).

Ncoll·H génériques En outre, pour les noms individuels d'humains n'ayant pas d'équivalence directe en Ncoll·H, il est possible d'avoir recours à des Ncoll·H génériques avec expansion adjectivale ou prépositionnelle. Parmi ces noms génériques, on peut citer « corps », « population », « milieu », « frange » ou « communauté ». Ces noms ne sont pas autonomes et nécessitent soit d'être qualifiés, soit d'être utilisés dans un contexte qui permet d'identifier la composition du groupe : pour prendre l'exemple du nom collectif « milieu », on ne peut pas l'employer seul dans un contexte générique, mais on peut y apposer diverses expansions pour indiquer le type de milieu en question (« milieu de la mode », « milieu universitaire », « milieu culturel »...). Cela n'est pas nécessaire en cas de reprise anaphorique (avec un déterminant démonstratif par exemple), car l'on peut inférer du contexte le type de milieu :

(41) Il n’arrivait pas à détacher ses yeux de l’article. Le *National Lawyer* était le journal de référence des avocats d’affaires. Celui qui faisait la pluie et le beau temps dans **ce milieu**.

(Frantext : Guillaume Musso, *Et après* [2004], p. 302)

Il convient de noter que, malgré leur caractère générique, ces noms collectifs ont tous des effets de sens particuliers. Si, pour neutraliser le nom commun « écrivain·es », on peut utiliser « le milieu littéraire » ou « la communauté littéraire », il semble difficile de parler de « population littéraire » ou encore d’« équipe littéraire ». Cela s’explique par le fait que « milieu » et « communauté » dénotent des rassemblements sociaux restreints et particuliers fonctionnellement parlant : « population » agit comme un nom collectif trop large qui impliquerait que tout le monde ferait *ipso facto* partie du groupe des écrivains et écrivaines, tandis qu’à « équipe » est associée une fonction de collaboration visant à accomplir une même tâche, ce qui ne correspond pas à la vision communément assimilée à /écrivain/. Par contre, ceux-ci fonctionnent parfaitement dans des configurations de type « population africaine » ou « équipe médicale », car les noms associés (« Africain·es », « médecins ») sont compatibles avec le sémantisme attribué au nom collectif en tête de syntagme. Il est d’ailleurs possible de reformuler en utilisant un complément du nom, et l’on obtient alors les mêmes résultats :

(42) *une population d’écrivain·es / une population d’Africain·es

(43) */?une équipe d’écrivain·es / une équipe de médecins

Cependant, apposer un adjectif ou un complément du nom ne suffit pas toujours à lever le flou qui est caractéristique de ces Ncoll·H génériques, notamment quand cet adjectif ou complément du nom a lui aussi un sens large. Lecolle (2008) prend ainsi l’exemple de « communauté universitaire ». Elle souligne que cette communauté « peut (ou non) être structurée de différents sous-groupes (étudiants, CPU, enseignants-chercheurs, personnel administratif), ce qui est (ou non) précisé contextuellement » (331).

Sur la base des exemples de Ncoll·H génériques donnés par le site du gouvernement du Canada³² (dénommés « noms collectifs polyvalents ») et de leurs définitions dans le TLFi³³ et le dictionnaire d’Antidote³⁴, nous nous proposons d’établir un tableau regroupant les différents effets de sens de ces Ncoll·H. Ce tableau ne se veut évidemment pas exhaustif et se doit d’être complété par des recherches plus approfondies et des exemples de corpus, mais nous pensons qu’il peut servir de base afin de saisir plus aisément les nuances qui distinguent les noms collectifs de ce type.

32. <https://www.noslangues-ourlangues.gc.ca/fr/cles-de-la-redaction/ecriture-inclusive-noms-collectifs#noms-collectifs-polyvalents>

33. <https://www.cnrtl.fr/definition/>

34. <https://www.antidote.info/>

Ncoll-H	Effet de sens	Exemples
communauté	rassemblement social, plutôt restreint, de personnes aux caractères et intérêts communs, pouvant mettre en jeu l'identité ³⁵	communauté scientifique ; communauté linguistique ; communauté francophone
corps	rassemblement de personnes appartenant à une même catégorie (généralement professionnelle)	corps étudiant ; corps enseignant ; corps diplomatique
équipe	rassemblement de personnes avec un travail ou une tâche commune	équipe technique ; équipe de rédaction
milieu	rassemblement de personnes appartenant à une même classe sociale ou à un même domaine, connotation possiblement péjorative	milieu de la mode ; milieu littéraire ; milieu culturel
personnel, profession, service	rassemblement de personnes exerçant la même activité professionnelle	personnel administratif ; profession agricole ; service informatique
population	rassemblement social large de personnes appartenant à une même classe ou catégorie	population carcérale ; population urbaine ; population active

TABLEAU 2.2 – Effets de sens des Ncoll-H génériques

Si certains Ncoll-H peuvent être rapprochés au niveau de leur sens malgré leurs fines différences (« équipe » et « personnel » par exemple), on voit bien que des distinctions profondes séparent des noms comme « corps » et « population ».

En conclusion, nous avons mis au jour les différentes caractéristiques des noms collectifs au travers d'une présentation exhaustive de ce type de nom. Après être revenu sur les différentes définitions données aux noms collectifs, nous avons étudié leurs particularités morphologiques, syntaxiques et sémantiques. Nous nous sommes ensuite concentrés plus particulièrement sur les noms collectifs humains. Nous avons, entre autres, montré leur potentiel pour une utilisation de neutralisation du genre de par leur aspect de collectivisation. Nous allons maintenant nous intéresser à la tâche de réécriture du genre dans le domaine du traitement automatique des langues afin de passer en revue les systèmes existants, qui nous aideront à développer notre propre système de réécriture neutre exploitant les noms collectifs.

35. Nous renvoyons à l'article très complet de Lecolle (2008) dédié au nom « communauté » pour de plus amples considérations.

2.3 Réécriture du genre en traitement automatique des langues

2.3.1 Tâche de réécriture dans le domaine du TAL : définition et considérations

Dans le cadre de notre projet, la tâche de traitement automatique des langues (désormais « TAL ») qui se rapproche le plus de notre objectif est la tâche de réécriture (ou « *rewriting* » en anglais).

Avant toute chose, nous nous permettons de distinguer la tâche de réécriture et la tâche de paraphrase dans le domaine du TAL : bien que l’objectif de ces deux tâches soit de produire un énoncé X sur la base d’un énoncé Y, elles se différencient de plusieurs manières. Tout d’abord, la tâche de paraphrase a une portée plus large, car elle sert aussi bien au résumé ou à la simplification de textes qu’à la génération de réponses à des questions. En outre, la tâche de paraphrase se permet de supprimer totalement certaines informations de l’énoncé de base, tandis que la tâche de réécriture vise à conserver toutes les informations fournies, simplement en les formulant d’une manière autre. Enfin, la tâche de réécriture se différencie en ce qu’elle se limite généralement à certaines unités linguistiques dans un texte, et ne vise pas à reformuler le texte dans son entièreté.

La tâche de réécriture a ainsi une application dans différents domaines de spécialité et est à l’origine de plusieurs sous-catégories : par exemple, la « réécriture confidentielle »³⁶ vise à préserver la vie privée de personnes en réécrivant des textes contenant des informations confidentielles afin de neutraliser ces dernières (Q. Xu et al., 2019). Comme autre exemple, nous pouvons citer la « réécriture de requêtes »³⁷ qui, elle, cherche à réécrire les requêtes d’utilisateurs et d’utilisatrices afin de leur proposer des résultats de recherche plus pertinents et de mieux répondre à leurs besoins, par exemple en modifiant certains termes qui peuvent être ambigus et qui peuvent influencer de manière négative la recherche (Zuo et al., 2022).

La sous-catégorie pertinente pour notre projet est celle de la « réécriture du genre »³⁸. Tâche relativement peu explorée dans le domaine du TAL, elle a été définie formellement pour la première fois par Alhafni et al. (2022b), même si d’autres chercheurs et chercheuses ont mené des travaux allant dans le même sens avant la formulation de cette définition (Zmigrod et al., 2019 ; Habash et al., 2019 ; Sun et al., 2021), utilisant parfois le terme de « nouvelle flexion en genre »³⁹. Alhafni et al. proposent la définition suivante pour cette tâche : « la génération, à partir d’une phrase arabe donnée, de phrases alternatives qui correspondent à différents contextes cibles en fonction du genre »⁴⁰. Les travaux de Alhafni et al. se concentrant sur l’arabe, cette définition n’est malheureusement pas universelle. En outre, ne sont pas pris en compte les cas de neutralisation, évoqués par exemple par Vanmassenhove et al. (2021) ou

36. *privacy-aware rewriting* en anglais.

37. *query rewriting* en anglais.

38. *gender rewriting* en anglais.

39. *gender reinflection* en anglais. C’est ce qu’indiquent Alhafni et al. (2022b) lorsqu’ils définissent la tâche de « réécriture du genre » : ce terme-ci leur paraît plus approprié que l’ancien, car la tâche « va au-delà de la flexion » (« goes beyond reinflection » [3]).

40. « generating alternatives of a given Arabic sentence to match different *target user gender contexts*. » (2)

He et al. (2021). En nous appuyant sur la définition de Alhafni et al. (2022b), nous proposons une définition plus large de la tâche de réécriture du genre :

La génération, à partir d’une phrase genrée donnée, d’une ou plusieurs phrases alternatives apportant des modifications aux formes genrées, soit en les neutralisant, soit en privilégiant une forme alliant tous les genres, soit en choisissant un autre genre.

Cette définition a le mérite de présenter les différentes possibilités en termes de réécriture (neutralisation, forme inclusive ou genre autre). Par « forme inclusive » (ou « forme alliant tous les genres »), nous faisons référence, pour le français, aux différentes techniques de visibilité évoquées dans le **chapitre 2.1.4.1** ; et, pour les autres langues, aux autres techniques similaires qui pourraient exister⁴¹.

Puisque nous nous inscrivons dans une perspective de neutralisation, nous utiliserons, dans la description de notre projet, le terme de « neutralisation du genre »⁴², employé notamment par Piergentili et al. (2023).

Nous décrivons ci-dessous les systèmes dans le domaine du TAL en rapport avec la tâche de réécriture du genre. Seront présentés les systèmes de réécriture vers un genre autre, les systèmes de réécriture inclusive, et enfin les systèmes de réécriture neutre.

2.3.2 Présentation de systèmes de réécriture du genre

2.3.2.1 Systèmes de réécriture vers un autre genre

Parmi les systèmes de réécriture existants qui visent à modifier le genre dans une phrase (en passant, par exemple, du masculin au féminin, ou vice versa), la majorité d’entre eux concerne la langue arabe. Cela s’explique par la constitution d’une communauté de chercheurs et de chercheuses arabophones qui vise à développer la tâche de réécriture du genre⁴³, mais également par la richesse morphologique de l’arabe : comme en français, les noms et adjectifs sont genrés (masculin, féminin ou épïcène), mais les verbes le sont également (Alhafni et al., 2022b). En outre, toujours selon Alhafni et al. (2022b), les clitiques et les affixes associés à des mots genrés doivent également être modifiés si le mot change de genre, ce qui augmente la complexité de la tâche.

41. Par exemple, nous renvoyons à Pomeranke (2022) pour un exemple d’implémentation, en allemand, de formes inclusives à un système de réécriture du genre, ou, dans une perspective plus globale, à Weigel (2021) pour les défis posés par l’écriture inclusive au traitement automatique du français.

42. *gender neutralization* en anglais étatsunien ; *gender neutralisation* en anglais britannique et canadien.

43. Par exemple, lors de l’édition de 2022 du Seventh Arabic Natural Language Processing Workshop, une tâche de réécriture du genre commune a été proposée afin que les chercheurs et chercheuses arabophones développent leurs propres systèmes et puissent les comparer avec ceux proposés : <https://sites.google.com/view/arabic-gender-rewriting/>

Le premier projet de réécriture du genre en arabe (Habash et al., 2019) se découpe en deux tâches : premièrement, l'identification de pronoms personnels de première personne genrés, et deuxièmement, la génération de l'énoncé équivalent dans l'autre genre.

Comme les pronoms personnels en arabe n'ont pas de genre associé, la tâche d'identification est réalisée sur le cotexte et la détection d'adjectifs et de noms humains qui, eux, sont genrés. Pour ce faire, plusieurs solutions sont proposées : il est possible d'établir un modèle à base de règles, avec des expressions régulières, afin de récupérer les contextes où un pronom personnel de première personne est suivi par un nom humain ou un adjectif genré singulier. En outre, le recours à des réseaux de neurones est également suggéré. Pour ce faire, l'outil se base sur un corpus de phrases anglais-arabe d'*Open Subtitles*, qui a par la suite été étoffé par l'ajout d'équivalents masculins pour les phrases au féminin, et d'équivalents féminins pour les phrases au masculin. Habash et al. (2019) ont alors entraîné trois modèles différents, le premier faisant une représentation au niveau des caractères, le deuxième une représentation au niveau des mots, et le troisième alliant les deux. Les résultats montrent que le modèle combiné obtient les meilleurs résultats, et c'est ce dernier qui est utilisé pour la tâche de génération. La tâche est alors considérée comme « une tâche de traduction automatique qui associe le texte d'un genre source à un genre cible »⁴⁴. Si un énoncé avec un pronom personnel de première personne référant à un homme est détecté, alors un équivalent au féminin sera généré, et vice versa. L'évaluation des résultats réalisée sur une partie du corpus d'entraînement montre 95,42 % de générations correctes du masculin au féminin, et 94,68 % de générations correctes du féminin au masculin.

La constitution d'un plus grand corpus genré parallèle pour l'arabe, contenant à la fois des phrases au masculin et au féminin, a permis des avancées majeures pour la tâche de réécriture du genre. Ce corpus, nommé The Arabic Parallel Gender Corpus 2.1 et présenté par Alhafni et al. (2022a), répertorie 80 326 phrases parallèles en genre. L'avantage de cette nouvelle version du corpus est qu'il permet de prendre en compte toutes les formes pouvant être genrées en arabe, et non pas seulement les pronoms. En outre, le corpus prend également en considération le contexte, ce qui permet de générer des phrases alternatives au niveau du genre dans des contextes avec plusieurs interlocuteurs et interlocutrices, en tenant compte à la fois de la personne qui parle, et de la personne visée par la phrase. Ainsi, un système à plusieurs étapes présenté par Alhafni et al. (2022a) a pu être développé, identifiant les genres dans un énoncé donné et proposant l'équivalent genré des mots détectés à l'aide de trois modèles combinés (CorpusR, un modèle de réécriture au niveau des mots qui tire parti du corpus aligné ; MorphR, un modèle de réécriture morphologique ; et NeuralR, un modèle de type encodeur-décodeur agissant au niveau des caractères). Les résultats comparatifs avec les autres systèmes existants (Habash et al., 2019 ; Alhafni et al., 2021) montrent une nette progression et un nombre largement inférieur d'erreurs. Ce système combiné obtient à ce jour les meilleurs résultats pour la tâche de réécriture du genre en arabe. Notons toutefois que Alrowili et Vijay-Shanker (2022) obtiennent des résultats quasiment similaires à ceux de

44. « a MT task that maps the text from one source gender to a target gender » (8)

Alhafni et al. (2022a) en utilisant des modèles Seq2Seq (Sutskever et al., 2014a) bien moins complexes, et en particulier T5 (Raffel et al., 2020).

Outre les travaux pour l'arabe, Zmigrod et al. (2019), la même année que pour la première itération pour l'arabe de Habash et al. (2019), proposent également un système de réécriture vers un genre autre. Comme avancée majeure, nous pouvons noter que toutes les formes genrées visent à être détectées et modifiées, et que le système prend en charge plusieurs langues : l'hébreu, l'espagnol, le français et l'italien. L'identification des formes genrées se base sur une analyse morphosyntaxique de la phrase de départ. Les formes genrées sont alors détectées en fonction de leurs propriétés morphosyntaxiques. L'une des formes genrées (généralement le nom, car élément racine de la proposition) est ensuite transformée vers son équivalent (masculin ou féminin), et une nouvelle analyse morphosyntaxique est réalisée pour répercuter le changement de genre sur les autres unités de la phrase pertinentes (pronom, nom ou adjectif). En particulier, pour l'accord en genre, Zmigrod et al. (2019) ont recours à un champ aléatoire de Markov pour déterminer l'accord morphosyntaxique des unités. Ainsi, la formule de Zmigrod et al. (2019), à partir d'un arbre en dépendances d'une phrase, permet de calculer le niveau de correspondance entre plusieurs unités selon leur catégorie lexicale et leurs propriétés morphosyntaxiques. Plus le nombre est élevé, plus il y a de chances que les unités s'accordent en genre.

L'application de cette méthode sur des phrases en espagnol et en hébreu fait état de résultats initiaux mitigés (67,23 et 55,32 de F-mesure respectivement), mais les résultats d'approches alternatives avec une paramétrisation linéaire ou une paramétrisation neuronale sont plus encourageants (82,29 de F-mesure pour l'espagnol avec la paramétrisation linéaire, et 73,12 de F-mesure pour l'hébreu avec la paramétrisation neuronale). De plus, Zmigrod et al. (2019) montrent que constituer un corpus avec des alternatives genrées et utiliser celui-ci pour l'entraînement de modèles de langage permet effectivement de réduire les stéréotypes de genre.

2.3.2.2 Systèmes de réécriture inclusive

Il existe deux systèmes de réécriture inclusive à ce jour : un pour l'allemand développé par Pomeranke (2022) : *INCLUSIFY*⁴⁵ ; et un pour le portugais développé par Veloso et al. (2023). Conçu à la base comme un système de réécriture neutre, le système de Pomeranke (2022) inclut également plusieurs options qui permettent d'appliquer des techniques de visibilité pour faire apparaître à la fois la forme masculine et la forme féminine d'un nom. Ces options sont au libre choix de l'utilisateur ou de l'utilisatrice, ce qui lui permet de sélectionner la forme qui lui convient le mieux, soit en fonction de ses préférences personnelles, soit en fonction du public visé. Nous les récapitulons dans le tableau ci-dessous.

Le projet repose sur une base de données préconstituée qui contient des termes au masculin générique en allemand et leurs équivalents neutres ou inclusifs. Deux jeux de données ont servi à la création de cette base de données. Pour les équivalents neutres, la liste de noms au masculin générique et des propositions neutres Geschickt Gendern⁴⁶ a été utilisée. Pour

45. Une démo en ligne est disponible ici : <https://diversifix.org/>

46. <https://geschicktgendern.de/>

Style	Forme correspondante	Équivalent français
Masculin générique	Student	étudiant
Neutre	studierende Person	personne étudiante ; personne qui étudie
Doublet complet	Studenten und Studentinnen Student oder Studentin	étudiants et étudiantes ; étudiant ou étudiante
Doublet abrégé (Binnen-I)	StudentInnen	étudiantEs
Doublet abrégé (séparateur)	Student*innen Student_innen Student :innen	étudiant·es ; étudiant(e)s ; étudiant/es

TABLEAU 2.3 – Formes d’écriture prises en charge par le système *INCLUSIFY* pour l’allemand

les équivalents inclusifs, Pomeranke (2022) a eu recours au corpus DeReKo⁴⁷, le plus grand corpus allemand contemporain existant à ce jour, et a répertorié les doublets abrégés présents dans le corpus, ce qui permet de récupérer à la fois la forme au masculin générique et la forme au féminin (qui sera utilisée pour construire la forme inclusive⁴⁸).

Le système vise à traiter la phrase d’entrée et à vérifier si des noms employés sont dans la base de données. *spaCy* est utilisé pour la tokénisation et pour l’analyse en dépendances, tandis que *Stanza* est utilisé pour la lemmatisation. Une fois la phrase traitée, le système vérifie s’il y a correspondance entre les noms de la phrase d’entrée et les noms dans la base de données. Si tel est le cas, une analyse morphosyntaxique de la phrase est effectuée avec *spaCy* afin de permettre le remplacement du nom au masculin générique par un nom neutre tout en conservant la grammaticalité de la phrase. Ainsi, *Morphy*⁴⁹, un dictionnaire morphologique pour l’allemand, est utilisé pour assigner au nouveau nom les caractéristiques morphosyntaxiques du nom au masculin générique détecté (cas et nombre). Il en est de même pour les mots rattachés syntaxiquement (déterminants, verbes, adjectifs, pronoms), avec le genre en plus.

Comme indiqué précédemment, la particularité du projet *INCLUSIFY* est qu’il laisse le choix de la formulation inclusive (voir **tableau 2.3**). Si une formulation inclusive (non neutre) est choisie, des modifications supplémentaires sont nécessaires. Dans ce cas, c’est la forme au féminin qui est prise comme référence et qui est modifiée pour s’adapter au style choisi. Pour la technique de visibilité du « Binnen-I » et pour les doublets abrégés, le système a recours à des expressions régulières. Pour les doublets complets, les deux noms au masculin et au féminin sont séparés avec « oder » (« ou ») s’ils sont au singulier, ou avec « und » (« et ») s’ils sont au pluriel.

Pomeranke (2022) présente également un test de performances (*benchmark*) pour évaluer le système développé. Ce test de performances prend appui sur des textes allemands neutres

47. <https://www.ids-mannheim.de/en/digspra/corpus-linguistics/projects/corpus-development>

48. En allemand, une des techniques de visibilité utilisées est le « Binnen-I » : elle consiste à mettre en majuscules la lettre « i » dans les suffixes « -in » et « -innen », utilisés respectivement pour former le féminin singulier et le féminin pluriel, pour indiquer qu’il est fait référence autant à des hommes qu’à des femmes. Par exemple, « Freunde » (« amis ») et « Freundinnen » (« amies ») peuvent se combiner en « FreundInnen » (« ami-es »). Une autre technique consiste à avoir recours à des séparateurs, comme en français, pour détacher la forme masculine de la forme féminine (« Freund :innen », « Freund*innen »...).

49. <http://www.danielnaber.de/morphologie/>

et inclusifs tirés de GC4⁵⁰, un autre corpus allemand, et leurs équivalents non inclusifs, créés manuellement à partir des textes originaux. Dans ce corpus, l’on trouve également des phrases sans terme inclusif pour éviter les biais au niveau des résultats. Le test de performances développé par Pomeranke (2022) permet d’évaluer la détection des noms au masculin générique en comparant le label (inclusif ou non) assigné par le système et celui assigné par le test de performances, puis en calculant ensuite le rappel, la précision et la F-mesure. En outre, il permet aussi d’évaluer la suggestion de noms neutres ou inclusifs équivalents. Pour cela, Pomeranke (2022) propose de calculer un score de suggestion afin de classer les propositions en fonction de leur pertinence. Trois fonctions sont proposées pour calculer ce score (fonction linéaire, proportionnelle ou logistique), mais la fonction logistique est privilégiée. Elle est définie comme suit :

$$s(r) = 1 - \frac{1}{(1 + e^{-Br})^{1/v}}$$

Source : Pomeranke (2022)

Où r correspond au rang assigné à la suggestion, et où B a une valeur de 1 et v une valeur de 0,01. D’après Pomeranke (2022), ces paramètres ont été choisis « manuellement en partant du principe que 5 suggestions peuvent être affichées »⁵¹.

À l’aide du test de performances, il est possible de connaître les résultats du système de Pomeranke (2022). Pour la détection des noms au masculin générique, le système obtient un score honorable (0,852 de F-mesure). En revanche, pour le score de suggestion, les résultats sont bien plus mitigés (0,559 pour les lemmes et 0,492 pour les mots⁵²). Pomeranke (2022) indique qu’« environ la moitié des ensembles de suggestions [inclusives ou neutres] donnent une phrase correcte [c’est-à-dire une reformulation bien inclusive ou neutre] (54 %), celle-ci faisant généralement partie des 5 premières suggestions ([dans 44 % des cas]) »⁵³. Le fait que la proposition correcte ne soit jamais la première suggestion est interprété comme étant possiblement un bug. En outre, Pomeranke (2022) ajoute que le score plus élevé pour les lemmes indique « quelques problèmes vis-à-vis de la flexion »⁵⁴.

Pour arriver à de meilleurs résultats, Pomeranke (2022) propose par exemple le recours à des modèles *end-to-end* pour la détection de noms au masculin générique et la suggestion de propositions neutres et inclusives, mais l’entraînement de ces modèles demande un nombre important de données. En outre, d’autres problématiques restent à résoudre, comme l’ajustement morphosyntaxique ou bien la résolution de la coréférence.

Concernant le système développé pour le portugais par Veloso et al. (2023), celui-ci s’inspire en grande partie du travail réalisé par Sun et al. (2021) et Vanmassenhove et al. (2021). Tout d’abord,

50. <https://german-nlp-group.github.io/projects/gc4-corpus.html>

51. « Parameters are chosen by hand to roughly suit the assumption that 5 items can be displayed. » (29)

52. Pomeranke (2022) précise que deux calculs sont effectués pour le score de suggestion : celui pour les mots vise à voir si « la phrase cible est identique à l’une des suggestions » (« *Is the target sentence literally identical to one of the suggestions ?* » [21]), tandis que celui pour les lemmes cherche à voir si « l’ensemble des lemmes dans la phrase cible est identique à l’ensemble des lemmes d’une des suggestions » (« *Is the set of lemmas of the target sentence identical to the set of lemmas of one of the suggestions ?* » [21]).

53. « As for the suggestion task, roughly half of the suggestion sets contain the correct sentence (54%), and if it is contained, it is mostly also in the first 5 suggestions (44%). » (22)

54. « This indicates that there are few problems regarding the inflection of the words. . . » (22)

Veloso et al. (2023) présentent le système de règles développé, qui comporte trois modules : une chaîne de prétraitement pour la tokenisation, un module pour extraire les référents humains dans la phrase, et un module visant à faire les remplacements nécessaires dans la phrase, notamment en modifiant les noms et pronoms, les clitiques, les déterminants, les adjectifs et les verbes. Comme pour les travaux de Sun et al. (2021) et Vanmassenhove et al. (2021), un modèle neuronal de type Transformer séquence à séquence, M2M100 (Fan et al., 2020), est affiné à l'aide de phrases converties par le système de règles.

L'évaluation des systèmes de règles et neuronal se base sur un ensemble de textes de divers genres (littéraire, journalistique, dialogue, phrases simples et textes publiés sur les réseaux sociaux). Les résultats montrent de meilleures performances pour le système de règles en général, bien que le modèle affiné obtienne de meilleurs résultats sur le corpus de phrases simples. Il est noté que les erreurs les plus fréquentes pour le système de règles concerne la modification des adjectifs : le portugais étant une langue romane, il s'agit de la catégorie avec les verbes qui est le plus susceptible aux changements lors d'un passage du genre à un autre. Concernant les erreurs du modèle affiné, celles-ci peuvent découler du système de règles en raison des données utilisées pour l'entraînement ; il existe également des cas où certains mots ne devant pas être modifiés le sont et, au contraire, des cas où des mots censés avoir été modifiés ne le sont pas. Veloso et al. (2023) notent toutefois, comme Vanmassenhove et al. (2021), la capacité d'adaptation du modèle affiné, puisque celui-ci est capable de modifier correctement des noms qu'il n'a jamais rencontrés auparavant.

2.3.2.3 Systèmes de réécriture neutre

Les systèmes de réécriture exclusivement neutre développés à ce jour concernent, à notre connaissance, uniquement l'anglais. Cela peut s'expliquer par l'adoption massive du pronom « they » et de ses dérivés dans cette langue pour remplacer les pronoms genrés, mais également car l'anglais est une langue largement dotée en données, ce qui rend l'apprentissage de modèles plus facile.

Le premier système de réécriture neutre a été développé par Sun et al. (2021). Ce système vise à remplacer les pronoms et déterminants genrés en anglais par leurs équivalents neutres. Nous récapitulons les formes possibles dans le tableau ci-dessous. Le système doit aussi prendre en compte les ajustements morphosyntaxiques nécessaires : si le pronom sujet est modifié, cela implique également des modifications dans le cotexte droit, car en anglais, le verbe rattaché au « they » singulier se conjugue comme à la troisième personne du pluriel, et non du singulier. Bien qu'une approche exclusivement à base de règles soit mentionnée, elle est toutefois rapidement rejetée en raison de la difficulté à discerner les pronoms des déterminants (par exemple, le pronom objet « her » et le déterminant possessif « her » ont la même graphie, mais n'ont pas la même fonction dans la phrase), et également car une approche à base de règles poserait des problèmes pour l'ajustement morphosyntaxique des verbes. À la place, Sun et al. (2021) proposent l'entraînement d'un modèle de type Transformer (Vaswani et al., 2017).

Ce type de modèle demandant un nombre de données assez conséquent, et aucun jeu de données de phrases parallèles pour le genre n'étant disponible en anglais, Sun et al. (2021)

	Masculin	Féminin	Neutre
Pronom sujet	he	she	they
Pronom objet	him	her	them
Pronom réfléchi	himself	herself	themselves
Pronom possessif	his	hers	theirs
Déterminant possessif	his	her	their

TABEAU 2.4 – Pronoms et déterminants genrés en anglais, et leurs équivalents neutres

ont créé leur propre jeu de données. Pour ce faire, 15 millions de phrases générées ont été extraites de Wikipédia, et un algorithme de réécriture neutre a été développé afin qu’il soit appliqué sur les phrases récupérées. Cet algorithme réunit expressions régulières, analyse en dépendances avec *SpaCy*, et un modèle GPT-2 préentraîné (Radford et al., 2019). Chacune des composantes vise à résoudre un problème en particulier : les expressions régulières sont utilisées pour effectuer les remplacements, l’analyse en dépendances pour trouver les verbes ayant un pronom genré comme sujet, et le modèle préentraîné pour permettre la distinction entre pronoms et déterminants. Sun et al. (2021) expliquent que si un pronom avec plusieurs graphies apparaît dans la phrase de départ, des variations utilisant chacune un type de pronom neutre différent seront générées. La phrase la plus grammaticale est ensuite choisie à travers un calcul de la perplexité : on calcule les probabilités que les mots dans la phrase se trouvent dans la même séquence selon le corpus de référence afin d’en juger la grammaticalité. Les phrases générées à partir de cet algorithme servent ainsi de données d’entraînement pour le modèle de type Transformer.

En comparant les résultats de l’algorithme et ceux du modèle Transformer, Sun et al. (2021) indiquent que tous deux obtiennent d’excellents scores en termes de pourcentage d’erreur de mots⁵⁵ et de score BLEU⁵⁶ (<1 % et >0,99 respectivement). De manière quelque peu surprenante, les résultats de l’algorithme sont légèrement meilleurs que ceux du modèle Transformer, mais Sun et al. l’expliquent par le fait que le modèle fait plus d’erreurs sur des tokens inattendus, comme « des espaces blancs, des émojis et des symboles »⁵⁷. Toutefois, il fait moins d’erreurs relatives aux ajustements de pronoms et de verbes que l’algorithme, ce qui prouve son efficacité pour ce genre de tâches.

Vanmassenhove et al. (2021) suivent une approche assez similaire à celle de Sun et al. (2021) pour neutraliser les pronoms et déterminants genrés en anglais. Le jeu de données provient cette fois-ci de Reddit, et contient plus de 2 250 000 phrases (dont 3 millions de pronoms et

55. *Word Error Rate* en anglais, abrégé *WER*.

56. Le score BLEU (acronyme de *Bilingual Evaluation Study* en anglais), défini par Papineni et al. (2002), sert de référence pour évaluer les systèmes de traduction automatiques. Comme l’indiquent Sun et al. (2021), l’on peut « envisager la tâche de réécriture comme une forme de traduction automatique » (« the task of rewriting can be viewed a form of MT » [2]), ce qui rend pertinent le recours à ce score pour évaluer le système proposé. Un score de 1 signifie une correspondance parfaite entre les deux phrases.

57. « nearly half of the model’s mistakes are due to rare tokens like whitespaces, emojis, and symbols » (4)

déterminants genrés). À la manière de Sun et al. (2021), Vanmassenhove et al. (2021) utilisent un système à base de règles pour transformer les pronoms et déterminants genrés en équivalents neutres afin de créer des données d'entraînement ; pour les répercussions morphosyntaxiques au niveau du verbe, les chercheurs et chercheuses ont recours à l'encapsuleur⁵⁸ Python LanguageTool afin de corriger les erreurs suite à la génération à base de règles. Enfin, un modèle Transformer est également entraîné sur les données ainsi générées.

Les deux approches (à base de règles et neuronale) sont évaluées sur divers jeux de données de test (provenant de Reddit, le corpus WinoBias+ et OpenSubtitles) qui contiennent des phrases genrés et leurs équivalents non genrés corrigés. Au contraire de Sun et al., les résultats sont bien meilleurs pour l'approche neuronale que pour l'approche à base de règles. L'approche neuronale obtient (en moyenne, pour les trois jeux de données de tests) 0,06 % de taux d'erreur de mots, contre 0,34 % pour l'approche à base de règles. Pour rappel, en comparaison, Sun et al. indiquaient un peu moins de 1 % d'erreur de mots pour leur système. Vanmassenhove et al. (2021) supposent que ces résultats proviennent d'une meilleure performance du système à base de règles utilisé pour générer les phrases d'entraînement.

Un dernier système qui mérite notre attention est DePeN⁵⁹, développé par He et al. (2021) pour l'anglais. Même si celui-ci vise à accomplir une tâche qui va au-delà de la réécriture du genre⁶⁰, la méthode utilisée est intéressante, et le remplacement ne se limite pas aux déterminants et aux pronoms genrés, mais s'étend aussi aux noms. Le système se base sur une technique de masque, technique utilisée, par exemple, par McMilin (2022) pour une tâche de résolution de pronoms genrés dans des contextes ambigus. Le fonctionnement est le suivant : les données censées être confidentielles (ici, le genre) présentes dans le texte d'entrée sont détectées automatiquement et masquées. Les unités à masquer sont déterminées en fonction d'un classifieur entraîné sur des documents présentant ce type de données. Un modèle Seq2Seq (Sutskever et al., 2014a) est ensuite entraîné de manière à générer un équivalent neutre aux unités masquées en utilisant les données non masquées.

Le système se base sur deux jeux de données : le premier provient d'étudiants et étudiantes américains ayant déposé leur candidature à des universités américaines, et contient un corpus de lettres de motivation ainsi que des informations sur chaque profil, y compris le genre de la personne, sa nationalité et si sa candidature a été acceptée ou non. Le deuxième jeu de données provient de Goodreads, un site cataloguant des livres édités dans le monde entier et permettant aux utilisateurs et utilisatrices de partager leurs livres préférés, d'en discuter et de les évaluer. Le corpus contient des évaluations de livres avec les genres « enfant » et « mystère/détective ». Au total, He et al. (2021) utilisent 22 201 lettres de motivation pour l'entraînement et 6 969 pour la validation, ainsi que 5 000 évaluations pour l'entraînement et 1 000 pour l'évaluation.

58. *wrapper* en anglais.

59. Acronyme de *Detect and Perturb to Neutralize* en anglais.

60. Il s'agit d'un système de neutralisation au sens large : toutes les données du texte source susceptibles de révéler le genre, la couleur de peau ou l'âge de la personne autrice doivent être neutralisés.

Pour l’évaluation, trois critères sont pris en compte : le biais, la grammaticalité (score PLL)⁶¹ et la fidélité par rapport au contenu original (score BLEU). Sont évalués un système à base de règles, le système DePeN dans son entièreté, les modules DeN (sans affinage du modèle Seq2Seq) et PeN (sans détection d’attributs particuliers), et d’autres approches similaires développées par d’autres chercheurs et chercheuses (Weighed Decoding ; Adversarial Training ; Privacy-Aware Text Rewriting).

En ce qui concerne la neutralisation du genre spécifiquement, et en comparant le système à base de règles avec les trois variantes de DePeN, le système qui réduit le plus les biais est DeN. DePeN se démarque des autres modèles de par sa faculté à produire des énoncés davantage grammaticaux (-4,8464 de PLL). Pour la fidélité, le score de DePeN est plus faible que celui du système à base de règles (0,6356 contre 0,8761), mais il reste plus élevé que celui d’autres systèmes similaires (0,3781 pour Weighed Decoding, 0,3979 pour Adversarial Training et 0,5278 pour Privacy-Aware Text Rewriting).

En conclusion, nous avons passé en revue les différents systèmes de réécriture du genre existants. Si les modèles à base de règle semblent souvent mener à de meilleurs résultats par rapport aux approches neuronales, cela est surtout dû au manque de données parallèles disponibles pour le genre ; sans oublier que la constitution manuelle de règles spécifiques exige un temps conséquent, et qu’il est humainement impossible de traiter l’ensemble des cas possibles. Un autre problème qui se pose est que les travaux sur cette tâche restent plus ou moins limités à l’anglais et à l’arabe. Bien que d’autres projets pour d’autres langues aient pu voir le jour, ils restent pour l’instant limités. Pour les langues morphologiquement riches, le développement de systèmes capables de modifier l’ensemble des formes genrées dans une phrase, peu importe la catégorie syntaxique, reste une tâche ardue, même si Alhafni et al. (2022b), et Alrowili et Vijay-Shanker (2022) ont montré des progrès notables à cet égard pour l’arabe.

61. Le score PLL (acronyme de *Pseudo Log-Likelihood* en anglais), défini par Salazar et al. (2020), est utilisé pour évaluer la grammaticalité de phrases générées par des systèmes faisant usage de masques. Il est calculé en remplaçant aléatoirement des mots dans une phrase par des masques, puis en demandant au modèle de prédire les mots originaux. Le score PLL représente la somme des logarithmes des probabilités que le modèle attribue à chaque mot original, étant donné le contexte de la phrase masquée. Plus le score PLL est élevé, plus le modèle est capable de générer des phrases grammaticalement correctes.

3

GeNRe : un système de réécriture neutre pour le français exploitant les noms collectifs

Ce chapitre vise à présenter notre système de réécriture neutre pour le français, GeNRe (**Gender-Neutral Rewriting System Using French Collective Nouns**¹). GeNRe repose sur l'utilisation de noms collectifs, qui disposent d'un genre fixe non dépendant du référent humain en français, pour neutraliser les noms pluriels au masculin générique d'une phrase donnée.

Nous commencerons tout d'abord par présenter notre méthodologie pour la création de notre système, ainsi que les jeux de données utilisés. Par la suite, nous détaillerons l'architecture et le fonctionnement des deux types de modèles qui sont au centre de notre système, puis nous procéderons à une analyse comparative des résultats, à la fois de manière quantitative et qualitative. Par la suite, nous explorerons le potentiel des modèles d'instructions (P. Liu et al., 2021) à l'instar de ChatGPT, qui permettent de générer du texte à partir d'une instruction écrite, ces modèles n'ayant jamais été utilisés pour la tâche de réécriture du genre auparavant. Nous exposerons enfin les avantages et limites de chaque méthode, ainsi que nos perspectives pour de futurs travaux concernant la tâche de réécriture du genre.

1. « système de réécriture neutre pour le français exploitant les noms collectifs » en français.

3.1 Méthodologie

Notre méthodologie s’inspire en partie des derniers travaux en date sur la tâche de réécriture du genre pour la neutralisation (Sun et al., 2021 ; Vanmassenhove et al., 2021). L’illustration 3.1 donne un aperçu de la méthodologie que nous adoptons.

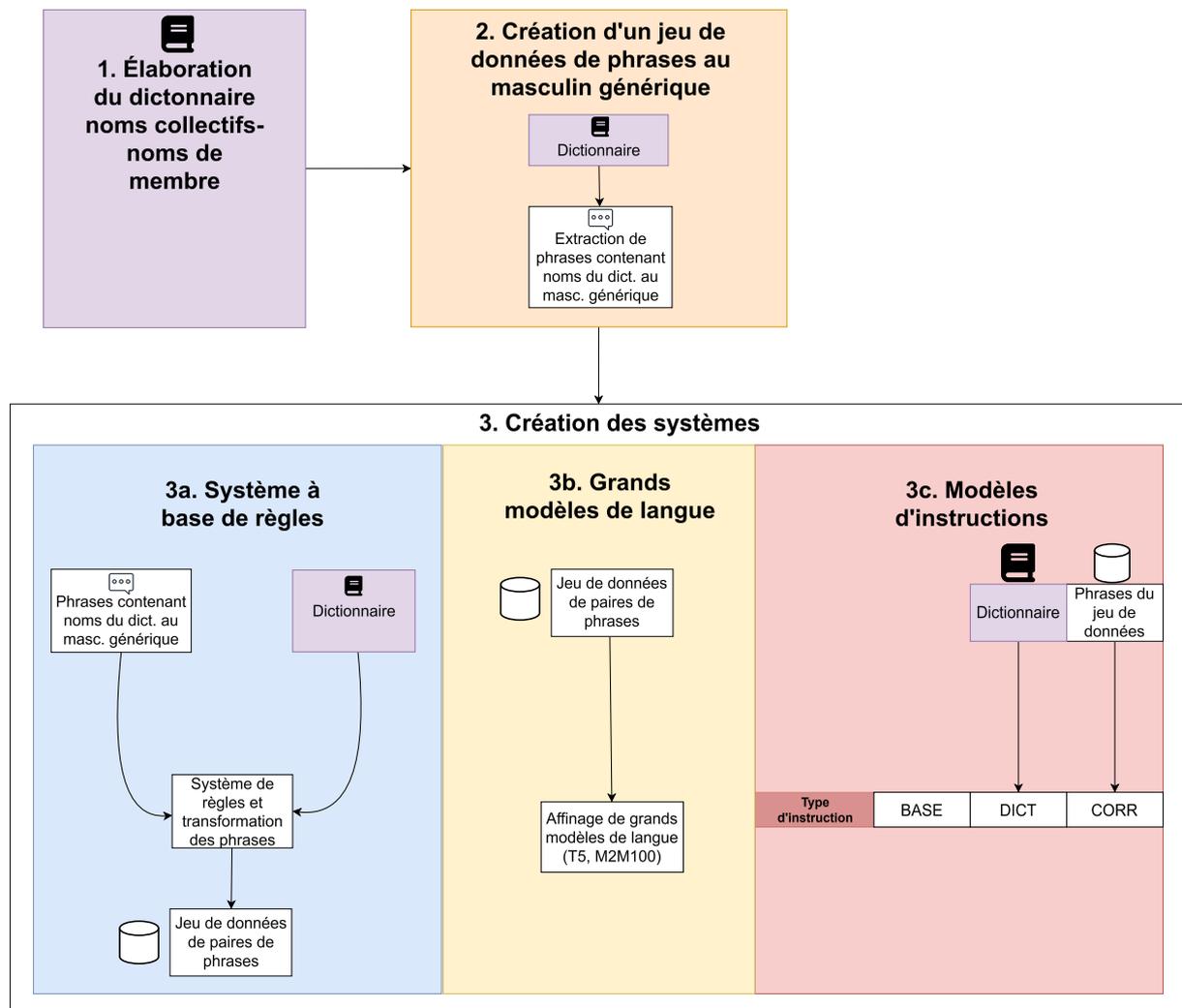


ILLUSTRATION 3.1 – Aperçu de la méthodologie

Étant donné que nous travaillons sur un type spécifique de nom, les noms collectifs, nous visons tout d’abord à créer manuellement un dictionnaire (étape 1, décrite dans le chapitre 3.2) qui rassemble des noms collectifs ainsi que leurs équivalents au masculin pluriel (« noms de membres »). Un exemple d’entrée pourrait être le nom collectif « police », avec comme nom masculin pluriel équivalent « policiers ».

Nous utiliserons par la suite ce dictionnaire afin d’effectuer des recherches dans un grand corpus français de manière à compiler des phrases qui contiennent les noms de membres dans notre dictionnaire (étape 2, décrite dans le chapitre 3.3), ces phrases devant par la suite être modifiées afin que le nom collectif correspondant remplace le nom de membre dans la phrase.

Pour effectuer ces remplacements, nous développerons un système à base de règles, en suivant l'approche de Sun et al. (2021) ainsi que Vanmassenhove et al. (2021) (étape 3a, décrite dans le chapitre 3.4).

Toujours selon cette même approche, nous tirerons ensuite parti de ce système de règles afin de générer un jeu de données constitué de paires de phrases non inclusives et neutres à partir de notre corpus. Ce jeu de données servira à l'affinage de deux grands modèles de langue : T5 (Xue et al., 2021) et M2-M100 (Fan et al., 2020) (étape 3b, décrite dans le chapitre 3.5). Vanmassenhove et al. (2021) ont en effet pu montrer que l'affinage de modèles à partir d'un jeu de données préexistant pour la transformation de phrases non inclusives permet une meilleure généralisation par ceux-ci des cas non présents dans le jeu de données utilisé pour l'affinage ; notamment parce que ces modèles usent de connaissances linguistiques antérieures pour s'adapter à la tâche en question.

Enfin, nous évaluerons également la potentialité des modèles d'instructions (P. Liu et al., 2021) pour cette tâche (étape 3c, décrite dans le chapitre 3.6). Pour ce faire, nous élaborerons plusieurs instructions, en utilisant ou non les ressources créées lors des étapes précédentes, afin de voir si ce type de modèle pourrait servir à obtenir de meilleurs résultats pour la tâche de réécriture du genre.

3.2 Création d'un dictionnaire de paires de noms collectifs-noms de membres

Notre système à base de règles se base tout d'abord sur l'élaboration d'un dictionnaire composé de noms collectifs et de leurs noms de membres équivalents au masculin pluriel (par exemple, police → policiers). À notre connaissance, il n'existe aucun jeu de données disponible pour les noms collectifs en français : nous avons donc dû le créer nous-mêmes.

Pour ce faire, nous avons eu recours à trois approches : une étude de la littérature, une collecte empirique et une collecte semi-automatique.

Étude de la littérature. Comme nous l'avons mentionné dans l'état de l'art, les noms collectifs en français ont été largement abordés dans la littérature linguistique. En particulier, Lecolle (2019) fournit en annexe de son ouvrage une liste de 138 noms collectifs et de leurs noms de membres équivalents rencontrés lors de ses recherches : nous utilisons donc cette liste comme première base de notre dictionnaire.

Nous incluons également les noms collectifs dont les noms de membres sont épïcènes (par exemple, « réserviste » pour le nom collectif « réserve ») : bien qu'on puisse les considérer comme des termes inclusifs tels quels, la phrase dans laquelle ils se trouveront ne sera pas pour autant inclusive : le nom de membre pourra par exemple être suivi d'un adjectif au masculin générique qu'il conviendra de modifier.

Nous choisissons toutefois de ne pas inclure certains noms collectifs pourtant présents dans l'ouvrage de Lecolle (2019) en raison de leur polysémie ou d'une trop grande divergence

sémantique entre le nom collectif et le nom de membre associé. Nous pouvons donner comme exemple de nom collectif polysémique le nom « troupe », qui obtient deux entrées différentes dans l'ouvrage de Lecolle (2019) : dans une première entrée, il est associé aux noms « comédien, danseur » (dans le sens « groupe d'artistes de spectacle »), tandis que dans une seconde entrée, il est associé au nom « soldat » (dans le sens « groupe organisé de soldats »). L'emploi de ce type de nom collectif invoque bien souvent la nécessité de préciser le sous-domaine associé ou bien les membres qui composent le groupe (par exemple, « troupe de théâtre » pour un groupe d'artistes ; « troupe d'assaut » pour des soldats) ; autrement, dans une phrase sans indicateur particulier, cela peut se révéler source de confusions, d'où notre choix de ne pas les inclure à notre dictionnaire. En outre, certains noms collectifs répertoriés dans l'ouvrage de Lecolle (2019) présentent parfois un sémantisme trop strict pour être inclus dans notre dictionnaire. Par exemple, le nom collectif « trio » est associé aux noms de membres « chanteur, musicien », mais celui-ci s'applique uniquement s'il y a un groupe de trois personnes exerçant une telle fonction. Au total, nous récupérons donc 105 entrées sur les 138 répertoriés dans l'ouvrage.

Collecte empirique. Une approche complémentaire à l'étude de la littérature a consisté à relever empiriquement, sur une période étendue, des emplois de noms collectifs dans les médias et sur Internet non présentés dans la littérature sur les noms collectifs. Par exemple, le nom collectif « *spectatorat* »², qui désigne « l'ensemble des spectateurs », n'était pas présent dans l'ouvrage de Lecolle (2019), bien qu'on en trouve plusieurs occurrences sur Internet. Nous donnons plus bas deux exemples d'usage dans des articles de presse issus de la base de données journalistique Factiva³, ainsi qu'un exemple trouvé sur Sketch Engine (Kilgarriff et al., 2014).

- (44) Nous nous sommes aperçus que le **spectatorat** scolaire s'effritait le soir, on souhaite sensibiliser les jeunes », expose le président de la CDC Claude Beigle.
(Paris-Normandie. *Faire revivre des auteurs*. [18 octobre 2007])
- (45) Depuis, on a senti un frémissement. Ça nous donne une autre visibilité, un autre "**spectatorat**".
(Midi Libre. « *Un jour, on a remplacé les Pink Floyd au pied levé* ». [2 décembre 2018])
- (46) D'où la nécessité d'enquêtes de terrain, pour comprendre les activités multiples et de plus en plus individualisées du **spectatorat**, les modalités d'accès et de sélection, l'expérience des films, etc.
(Sketch Engine. *Retourner le regard : Enjeux des recherches sur les pratiques spectatoriennes en Afrique et au Moyen Orient | Publics en questions*. [2021])

2. <https://fr.wiktionary.org/wiki/spectatorat>

3. <https://www.dowjones.com/professional/fr/factiva/>

Collecte semi-automatique. Enfin, nous avons eu recours au Wiktionnaire⁴, créé par la Fondation Wikimedia, pour récupérer certains noms collectifs. Il existe par exemple un certain nombre de noms collectifs avec le suffixe -phonie (« francophonie », « lusophonie », « japonophonie ») qui désignent les locuteurs et locutrices d'une langue. Les noms de membres associés sont répertoriés sur la page consacrée au suffixe -phone⁵. Nous avons développé un script Python de manière à récupérer ces noms et à générer les noms collectifs équivalents en remplaçant le suffixe « -phone » par « -phonie ». Nous récupérons le contenu de la page du Wiktionnaire à l'aide des modules Python Requests (Reitz, 2011) et BeautifulSoup (Richardson, 2004), puis nous appliquons des expressions régulières de manière à supprimer tout le contenu superflu (notamment contenu entre parenthèses) et à récupérer chaque nom sur une ligne spécifique. Nous montrons un exemple de notre code dans l'illustration 3.2.

Ces trois approches nous permettent de composer notre dictionnaire avec un total de 315 paires de noms collectifs-noms de membres. Notre dictionnaire, au format CSV, comprend les colonnes suivantes :

- id (identifiant unique de la paire nom collectif-nom de membre) ;
- coll_noun (nom collectif)
- member_noun_sg (nom de membre au masculin singulier) ;
- member_noun_pl (nombre de membre au masculin pluriel) ;
- coll_noun_g (genre du nom collectif, « m » pour masculin, « f » pour féminin) ;
- coll_noun_n (nombre du nom collectif, « s » pour singulier, « p » pour pluriel⁶) ;
- src (source du nom collectif, « Lecolle » pour ceux extraits de l'ouvrage de Lecolle (2019), « Custom » pour les autres).

L'illustration 3.3 donne un exemple de ce à quoi ressemble notre dictionnaire. Celui-ci est également disponible dans son entièreté en annexe, dans le tableau A.1.

3.3 Conception du jeu de données initial

Pour le développement de GeNRe, nous avons recours à un jeu de données composé d'articles issus de Wikipédia. Le jeu de données, publié sur Hugging Face⁷ sous le nom `graelo/wikipedia-graelo`, (2023) dispose d'un sous-ensemble de 1,58 million de textes français.

Nous filtrons ce jeu de données de manière à récupérer uniquement les phrases dans lesquelles se trouvent des noms de membres masculins pluriels, d'après notre dictionnaire. Pour ce faire, nous avons recours à une expression régulière qui prend également en compte les différents déterminants précédant le nom de membre. Celle-ci est construite à partir des

```
1 import requests
2 from bs4 import BeautifulSoup
3 import re
4
5 url = "https://fr.wiktionary.org/wiki/-phone"
6 r = requests.get(url)
7 soup = BeautifulSoup(r.text, "html.parser")
8
9 # Récupération de la section Dérivés → Locuteur d'une langue
10 # par ciblage du style de l'élément <div>
11 content = soup.find("div",
12                     attrs={"style": ""column-width: 26em;
13                             -webkit-column-width: 26em;
14                             -moz-column-width: 26em;
15                             vertical-align: top;
16                             text-align: left;""})
17 text = content.text
18
19 # Suppression des parenthèses et de leur contenu
20 text = re.sub(r"(\()[^)]*\)", "", text)
21 # Dans les cas où il y a plusieurs dérivés,
22 # on les sépare par un retour à la ligne
23 text = re.sub(r"\,([^\].*)", r"\n\1", text)
24 text = re.sub("phone", "phonie", text)
```

ILLUSTRATION 3.2 – Extrait de code pour la récupération automatique de noms collectifs associés au suffixe « -phone »

entrées de notre dictionnaire : nous renvoyons à l'illustration 3.4 qui montre l'implémentation pour davantage de détails.

L'expression régulière (?! [\w-]) permet de ne pas prendre en compte les mots composés de type « auteurs-compositeurs » (qui serait détecté en raison de la présence de « auteurs »).

Étant donné que le jeu de données contient dans une seule et même colonne le texte entier de chaque article Wikipédia, nous utilisons NLTK (Bird et al., 2009) pour découper chaque

4. <https://fr.wiktionary.org/>

5. <https://fr.wiktionary.org/wiki/-phone>

6. La grande majorité des noms collectifs sont singuliers, mais on en trouve certains au pluriel, comme « forces de l'ordre ».

7. <https://huggingface.co/>

id	coll_noun	member_noun_sg	member_noun_pl	coll_noun_g	coll_noun_n	src
1	académie	académicien	académiciens	f	s	Lecolle
2	achalandage	chaland	chalands	m	s	Lecolle
3	armée	soldat	soldats	f	s	Lecolle
4	milice	milicien	miliciens	f	s	Lecolle
5	artillerie	artilleur	artilleurs	f	s	Lecolle
6	ascendance	ascendant	ascendants	f	s	Lecolle
7	auditoire	auditeur	auditeurs	m	s	Lecolle
8	aviation	aviateur	aviateurs	f	s	Lecolle
9	ballet	danseur	danseurs	m	s	Lecolle
10	barreau	avocat	avocats	m	s	Lecolle
11	bataillon	soldat	soldats	m	s	Lecolle

ILLUSTRATION 3.3 – Extrait du dictionnaire de paires noms collectifs-noms de membres

article en plusieurs phrases, de façon à ce que l'analyse ne se fasse que sur une phrase à la fois, et non pas sur l'article en entier. En outre, le découpage n'étant pas toujours bien effectué, nous ajoutons un filtre afin d'ignorer les découpages qui contiennent des espaces en trop ou des retours à la ligne : il s'agit alors bien souvent de paragraphes entiers ou de listes, et non pas de phrases réelles.

Il convient également de noter que certains noms de membres dans notre dictionnaires peuvent appartenir à plusieurs catégories. Par exemple, le nom « jeunes », associé au nom collectif « jeunesse », peut également être un adjectif. Nous analysons donc chaque phrase récupérée à l'aide de la bibliothèque Python spaCy (Honnibal et al., 2020) afin de vérifier que le mot en question est bien un nom commun : le modèle utilisé pour l'analyse des dépendances avec spaCy est fr_core_news_sm⁸, entraîné sur le corpus UD_French-Sequoia (Candito et Seddah, 2012). Enfin, comme certains noms de membres peuvent être associés à plusieurs noms collectifs, nous annotons automatiquement, dans chaque phrase, le nom de membre détecté à l'aide d'une balise <n>, suivi de l'identifiant associé au nom collectif dans le dictionnaire. Nous donnons un exemple de phrase annotée dans l'exemple 47.

- (47) Un historique permet de lister <n-126>**les auteurs**</n> et de consulter les modifications successives de l'article par <n-68>**ses rédacteurs**</n>.

Nous récupérons ainsi un total de 292 076 phrases pour notre jeu de données initial.

8. <https://spacy.io/models/fr>

```
1 import re
2 import csv
3
4 def get_regex():
5     nouns = []
6     mb_nouns_rx = ""
7     with open("collective_nouns.csv", "r", encoding="utf8") as cn_file:
8         reader = csv.reader(cn_file)
9         next(reader)
10        for row in reader:
11            nouns.append(row[3])
12
13    for i, noun in enumerate(nouns):
14        mb_nouns_rx += re.escape(noun)
15        if i < len(nouns)-1:
16            mb_nouns_rx += "|"
17
18    return r"\b((?:[Aa]ux|[Ll]ddCcSsMm)es\s(?:" + mb_nouns_rx + r"))(?:![\w-])\b"
```

ILLUSTRATION 3.4 – Extrait de code pour la création d'expressions régulières pour la détection de noms de membres

3.4 Conception et architecture du système à base de règles

Notre système à base de règles se développe en deux composants : le premier, que nous nommerons « composant de détection des dépendances » vise à détecter, à l'aide de spaCy, toutes les dépendances syntaxiques du nom de membre de la phrase. Ainsi, ces dernières pourront être modifiées de manière à s'adapter au nom collectif qui viendra remplacer le nom de membre : en effet, le nombre sera quasiment toujours différent (l'on passera la plupart du temps d'un nom masculin pluriel à un nom collectif singulier) ; le genre pourra également changer si le nom collectif est féminin. Le deuxième composant, que nous nommerons « composant de génération », a pour objectif de remplacer le nom de membre par son équivalent collectif d'après une recherche dans le dictionnaire de paires de noms collectifs-noms de membres. En outre, ce composant se charge également d'adapter toutes les dépendances syntaxiques de manière à ce qu'elles soient conformes au genre et au nombre du nom collectif choisi. Pour ce faire, nous avons recours à divers algorithmes, explicités plus avant, ainsi qu'à la bibliothèque Python inflecteur (Chuttarsing, 2021), qui permet de modifier la flexion d'un mot donné.

Nous présentons un schéma de notre système à base de règles dans l'illustration 3.5.

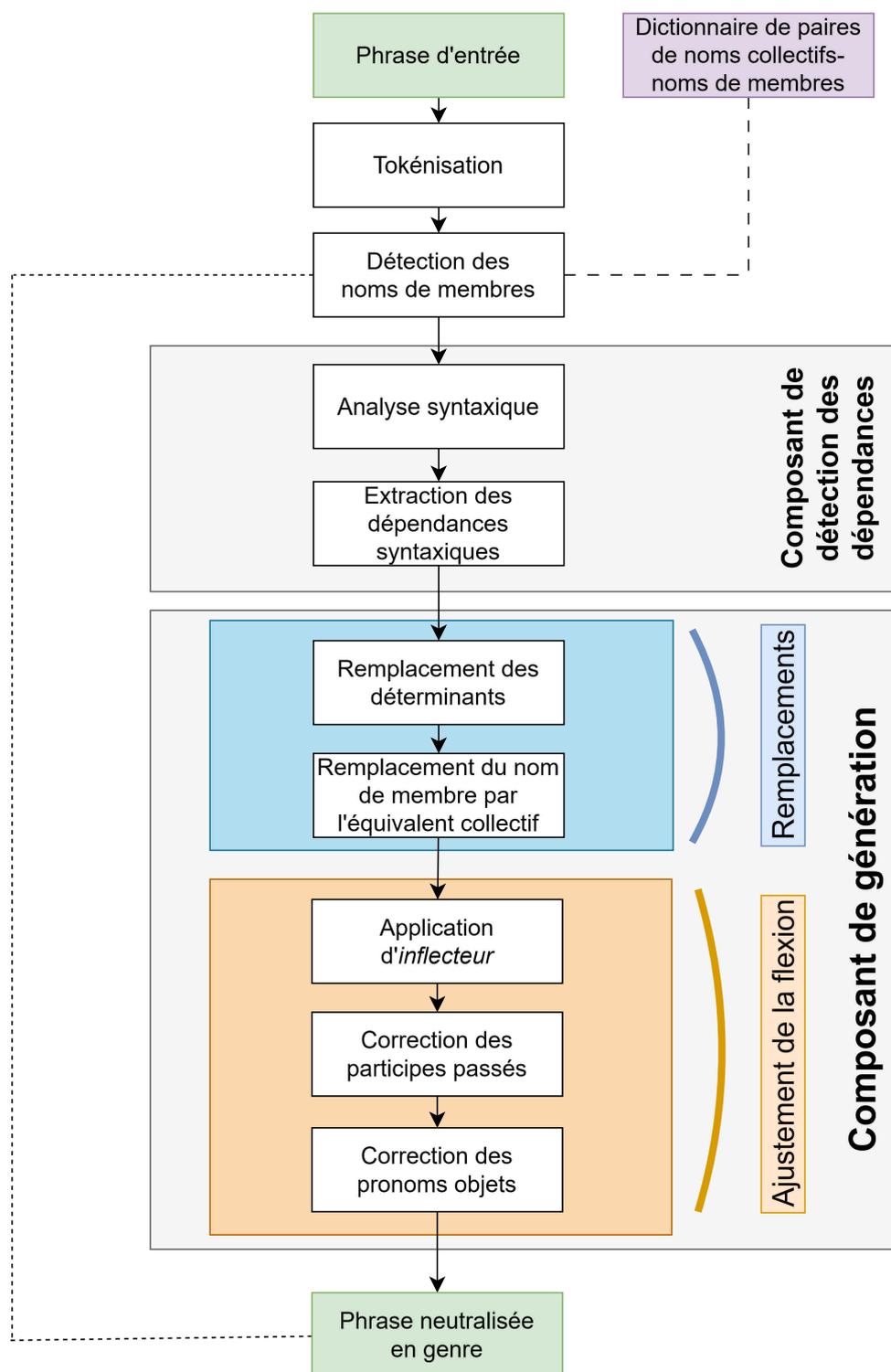


ILLUSTRATION 3.5 – Vue d'ensemble de la chaîne de traitement de GeNRe-RBS

Défini de manière logique, le déroulé de notre chaîne de traitement est le suivant :

Soit S une phrase d'entrée comprenant les noms de membres masculins pluriels N_1, N_2, \dots, N_n .

Soit C_i le nom collectif correspondant à N_i .

Soit DN_i la liste des dépendances syntaxiques de chaque nom N_i .

On définit une fonction de remplacement R où $R(N_i) = C_i$.

Soit S' notre phrase transformée, définie comme suit :

$$S' = S[R(N_1), \dots, R(N_n)]$$

Pour chaque C_i dans S' , on met à jour les dépendances $D(C_i) = f(D(N_i))$

de manière à ce qu'elles correspondent syntaxiquement à C_i .

Dans les sections suivantes, nous présentons en détail le fonctionnement de notre système à base de règles. Nous explicitons notamment les étapes initiales intégrées à notre système, puis nous expliquons le fonctionnement des deux composants de notre système : le composant de détection des dépendances et le composant de génération. Nous évaluons par la suite les performances de nos deux composants sur un jeu de données spécifiques, présenté ultérieurement, avant de mettre en lumière et de discuter les résultats obtenus.

3.4.1 Étapes initiales

Nous effectuons en premier lieu une tokénisation de notre phrase d'entrée à l'aide de spaCy et du modèle `fr_core_news_sm`. Par la suite, nous parcourons la phrase d'entrée pour détecter les noms de membres présents dans notre dictionnaire. Si aucun nom de membre n'est trouvé, nous renvoyons la phrase originale, car nous pensons qu'aucune modification n'est nécessaire. Autrement, nous continuons vers le composant de détection des dépendances.

3.4.2 Composant de détection des dépendances

Le composant de détection des dépendances repose sur une analyse syntaxique de la phrase d'entrée à l'aide de spaCy.

Cette analyse nous permet, dans un premier temps, de définir des règles de filtrage afin d'écartier les cas où extraire les dépendances syntaxiques de la phrase afin de les modifier par la suite ne serait pas pertinent : soit car le nom de membre est lié à un ou plusieurs noms propres, soit car nous avons plusieurs noms de membres masculins coordonnés, ce qui signifie qu'il n'est pas nécessaire de modifier ni le nombre ni le genre des dépendances.

Dans un deuxième temps, si les règles détaillées ci-dessus n'ont pas été déclenchées, nous réalisons une extraction des dépendances syntaxiques qui seront par la suite modifiées pour s'accommoder au nom collectif.

3.4.2.1 Règles de filtrage

Nos règles de filtrage concernent les noms propres et la coordination. Nous détaillons plus avant les conditions nécessaires pour déclencher ou non ces règles de filtrage pour ces éléments.

Noms propres. Nous voulons vérifier si le nom de membre détecté n'est pas suivi de noms propres ; autrement, il n'est pas nécessaire de neutraliser la phrase étant donné qu'il est fait référence à des individus particuliers : nous n'avons donc, dans ce genre de phrases, pas affaire à de cas de masculin générique. Nous donnons un exemple d'une telle phrase dans l'exemple 48. Nous avons principalement recours aux catégories de Universal Dependencies PROP⁹ (nom propre) et NOUN¹⁰ (nom commun), ainsi qu'aux relations flat :name¹¹ et dep¹². Nous incluons les noms communs, car selon la nature des noms propres, ceux-ci ne sont pas forcément reconnus comme tels par spaCy.

En outre, nous vérifions également les phrases où les noms propres sont précédés d'éléments tels que « comme » ou « tels » ; dans ce genre de cas, le nom de membre ne doit pas être modifié non plus.

(48) Les professeurs Dupont et Durand sont arrivés.

L'algorithme est présenté dans l'illustration 3.6.

Coordination. Lorsque deux noms de membres sont coordonnés par « et » ou « ainsi que » et que les dépendances se situent après le deuxième nom, il n'est pas nécessaire de les modifier, puisqu'après conversion en noms collectifs, celles-ci devront rester au pluriel pour respecter l'accord syntaxique. Nous développons donc un algorithme pour mettre fin prématurément au processus de recherche de dépendances dans ce cas. Nous donnons un exemple d'une telle phrase dans les exemples 49 et 50.

(49) Les soldats et les députés se rencontraient pour la première fois.

(50) L'armée et le parlement se rencontraient pour la première fois.

3.4.2.2 Extraction des dépendances syntaxiques

Afin d'extraire les dépendances syntaxiques de chaque nom de membre (désormais « NM »), nous tirons parti des différentes méthodes proposées par spaCy.

Chaque phrase est analysée syntaxiquement pour récupérer l'indice du NM dans la phrase. Nous appliquons ensuite divers filtres pour récupérer les éventuelles dépendances du NM. Cela s'effectue à sept niveaux, qui sont également répertoriés dans le tableau 3.1 :

9. <https://universaldependencies.org/u/pos/PROPN.html>

10. <https://universaldependencies.org/u/pos/NOUN.html>

11. <https://universaldependencies.org/u/dep/flat-name.html>

12. <https://universaldependencies.org/u/dep/dep.html>

```
1 def is_proper_noun(tokenized_sent, target_token):
2     punct = "".join(c for c in string.punctuation if c != ",")
3
4     for token in tokenized_sent[target_token.i:target_token.i+PROPN_BASE_MAX_SEARCH]:
5         if ((token.pos_ in ["PROPN", "NOUN"] and token.dep_ == "flat:name")
6             or (token.pos_ == "PROPN" and token.dep_ in ["flat:name", "dep"])):
7             return True
8
9     # check with "comme", "tel", "tels"
10    for token in tokenized_sent[target_token.i:target_token.i+PROPN_CONJ_MAX_SEARCH]:
11        if token.text in ["comme", "tel", "tels"]:
12            if token.nbor():
13                if token.text in ["tel", "tels"] and token.nbor().text == "que":
14                    if token.nbor(2).pos_ in ["PROPN", "NOUN"] and token.nbor(2).dep_ in ["flat:name", "dep"]:
15                        return True
16                elif token.nbor().pos_ in ["PROPN", "NOUN"] and token.nbor().dep_ in ["flat:name", "dep", "obj"]:
17                    return True
18
19    # If we cannot find a proper noun when checking
20    # the neighbors of the target token,
21    # we may find it by checking
22    # the neighbors of every child of the target token.
23    for child in target_token.children:
24        for token in tokenized_sent[child.i:child.i+PROPN_BASE_MAX_SEARCH]:
25            if token.text in punct:
26                return False
27
28            if ((token.pos_ in ["PROPN", "NOUN"] and token.dep_ == "flat:name")
29                or (token.pos_ == "PROPN" and token.dep_ in ["flat:name", "dep"])):
30                return True
31
32    return False
```

ILLUSTRATION 3.6 – Algorithme de détection des noms propres

- ancêtres¹³ du NM (ANCESTOR);
- voisins des ancêtres du NM (ANCESTOR_NBOR);
- enfants des ancêtres du NM (CHILD);
- voisins des enfants des ancêtres du NM (CHILD_NBOR);
- proposition relative associée au NM (REL_CL);
- enfants du NM (TARGET_CHILD);
- voisins du NM (TARGET_NBOR).

Nous assignons un code pour chaque règle d'extraction afin de faciliter l'élaboration du système et le débogage. Au total, nous comptons 25 codes différents.

Nous récupérons respectivement les ancêtres et les enfants à l'aide des propriétés `ancestors`¹⁴ et `children`¹⁵ de l'objet `Token`. Pour les voisins, soit nous avons recours à la méthode `nbor()`¹⁶,

13. Les ancêtres d'un mot sont ses parents directs, les parents des parents, et ainsi de suite.

14. <https://spacy.io/api/token#ancestors>

15. <https://spacy.io/api/token#children>

16. <https://spacy.io/api/token#nbor>

Code	Type	Description
ANCESTOR_IS_VERB_AND_ARG_TARGET ANCESTOR_IS_ADJ_AND_MODIFIES_TARGET ANCESTOR_IS_VERB_MISIDENTIFIED_AS_AUX_AND_ARG_TARGET	ANCESTOR ANCESTOR ANCESTOR	Ancêtre est un verbe qui a pour argument un NM sujet Ancêtre est un adjectif qui modifie le NM Ancêtre est un verbe (identifié comme auxiliaire) qui a pour argument un NM sujet
ANCESTOR_NBOR_IS_AUX_PASS ANCESTOR_NBOR_IS_AUX ANCESTOR_NBOR_IS_COP (VERB) ANCESTOR_NBOR_IS_COP (ADJ)	ANCESTOR_NBOR ANCESTOR_NBOR ANCESTOR_NBOR ANCESTOR_NBOR	Voisin de l'ancêtre est un auxiliaire à la forme passive (être) Voisin de l'ancêtre est un auxiliaire Voisin de l'ancêtre est un copule (identifié comme verbe) Voisin de l'ancêtre est un copule (identifié comme adjectif)
CHILD_IS_AUX_TENSE CHILD_IS_AUX_PASS CHILD_IS_ADJ_CC CHILD_IS_PRON_OBL_MOD CHILD_IS_MISIDENTIFIED_VERB_PRECEDED_BY_PRON CHILD_IS_PP_AND_MODIFIES_TARGET	CHILD CHILD CHILD CHILD CHILD CHILD	Enfant de l'ancêtre du NM est un auxiliaire de temps composé Enfant de l'ancêtre du NM est un auxiliaire passif (être) Enfant de l'ancêtre du NM est un adjectif coordonné Enfant de l'ancêtre du NM est un pronom objet Enfant de l'ancêtre du NM est un verbe mal identifié précédé d'un pron. rel. Enfant de l'ancêtre du NM est un participe passé qui modifie le nom
CHILD_NBOR_IS_AUX_TENSE CHILD_NBOR_IS_PRECEDED_BY_AUX_PASS	CHILD_NBOR CHILD_NBOR	Voisin de l'enfant de l'ancêtre du NM est un auxiliaire de temps composé Voisin de l'enfant de l'ancêtre du NM est précédé d'un aux. passif (être)
NON_CHILD_IS_AUX_TENSE_IN_REL_CL TARGET_CHILD_IS_AUX_TENSE_IN_REL_CL	REL_CL TARGET_CHILD REL_CL	Auxiliaire de temps composé dans une proposition relative Enfant du NM est un aux. de temps composé dans une prop. rel.
TARGET_CHILD_IS_ADJ_AND_MODIFIES_TARGET TARGET_CHILD_IS_AUX_TENSE TARGET_CHILD_IS_MISIDENTIFIED_ADJ_AND_MODIFIES_TARGET	TARGET_CHILD TARGET_CHILD TARGET_CHILD	Enfant du NM est un adjectif qui modifie le nom Enfant du NM est un auxiliaire de temps composé Enfant du NM est un adjectif mal identifié qui modifie le nom
TARGET_NBOR_IS_PP_WITH_AVOIR_AND_ANTECEDENT TARGET_NBOR_IS_VERB_PRECEDED_BY_REL_PRON TARGET_NBOR_IS_AUX_OF_VERB_PRECEDED_BY_REL_PRON TARGET_NBOR_IS_ADJ_PRECEDED_BY_ADV TARGET_NBOR_IS_ADJ	TARGET_NBOR TARGET_NBOR TARGET_NBOR TARGET_NBOR TARGET_NBOR	Voisin du NM est un participe passé avec « avoir » comme antécédent Voisin du NM est un verbe précédé d'un pronom relatif Voisin du NM est un auxiliaire suivi d'un verbe précédé d'un pronom rel. Voisin du NM est un adjectif précédé par un adverbe Voisin du NM est un adjectif

TABLEAU 3.1 – Codes utilisés pour l'extraction des dépendances syntaxiques

soit nous définissons des plages de contexte pour parcourir chaque token environnant à partir de l'indice du nom de membre. Ces plages sont définies empiriquement en fonction de l'élément que nous voulons rechercher. Nous en donnons un exemple dans l'illustration 3.7.

Nous avons deux types de plages : des plages spécifiques pour chaque type d'élément et des plages en fonction de la longueur. Par exemple, OBL_CONTEXT_MAX_SEARCH est la plage utilisée pour trouver les pronoms objets qui suivent le nom de membre. Dans l'exemple 51, le pronom objet « les » reprend le nom de membre « spectateurs ». Lorsque « spectateurs » sera modifié vers son équivalent collectif « public », il faudra alors modifier le pronom objet en conséquence, comme dans l'exemple 52.

(51) L'écrivain chrétien Tertullien a maudit les spectateurs, et les a envoyés au diable.

(52) L'écrivain chrétien Tertullien a maudit le public, et l'a envoyé au diable.

Une recherche est donc effectuée dans la plage indiquée, entre le nom de membre et le nombre associé à la plage, afin de trouver les éventuels pronoms objets qui reprennent le nom de membre identifié. Les plages en fonction de la longueur reprennent le même principe mais sont utilisées dans un cadre plus général, et ne se cantonnent pas à une catégorie ou une fonction de mot en particulier ; elles visent simplement à faire une recherche plus ou moins grande dans le contexte. Nous en définissons trois : contexte restreint (égal à 3), contexte moyen (égal à 4) et contexte étendu (égal à 6).

3.4.3 Composant de génération

Le composant de génération se divise en deux volets distincts : le premier concerne les remplacements des NM par leurs équivalents collectifs, y compris au niveau des déterminants ;

```
1 PROPN_BASE_MAX_SEARCH = 2
2 PROPN_CONJ_MAX_SEARCH = 4
3 COORD_MAX_SEARCH = 3
4 OBL_CONTEXT_MAX_SEARCH = 4
5
6 SMALL_CONTEXT_MAX_SEARCH = 3
7 MEDIUM_CONTEXT_MAX_SEARCH = 4
8 LARGE_CONTEXT_MAX_SEARCH = 6
9
10
11 if ancestor.pos_ == "VERB":
12     if (target_token.i < ancestor.i):
13         if not target_token.nbor().text == ancestor.text:
14             medium_context = sent[target_token.i-MEDIUM_CONTEXT_MAX_SEARCH:target_token.i]
15             context_text = [t.text for t in medium_context]
16
17             large_context = sent[target_token.i-LARGE_CONTEXT_MAX_SEARCH:target_token.i]
18             large_context_text = [t.text for t in large_context]
19             for token in medium_context:
20                 if target_token.nbor().dep_ != "acl:relcl1":
21                     if (token.text in [",", "et"]):
22                         if token.text == "," and ((token.i != target_token.i-1
23                             and token.i != target_token.i-2) or (large_context_text.count(",") > 1)):
24                             coordination = True
25                         elif token.text == "et":
26                             coordination = True
27                     break
28             elif ("y" in context_text):
29                 y_idx = context_text.index("y")
30                 if y_idx > 0 and context_text[y_idx-1] == "compris":
31                     coordination = True
32                 break
```

ILLUSTRATION 3.7 – Extrait de l’algorithme de détection des tokens voisins

le second vise à ajuster la flexion des dépendances précédemment trouvées de manière à ce qu’elles correspondent syntaxiquement aux nouveaux noms collectifs.

3.4.3.1 Remplacements

GeNRe prend en charge le remplacement des déterminants définis (« les »), possessifs (« mes », « ses »), démonstratifs (« ces ») et contractés (« des », « aux ») pluriels. Pour chacun de ces déterminants, nous définissons un dictionnaire, contenant le déterminant pluriel comme clé, et ses équivalents singuliers sous forme de liste comme valeur, par exemple : "les": ["la", "le", "l’"].

Nous nous servons du genre du nom collectif, indiqué dans le dictionnaire de paires de noms collectifs-NM, pour choisir le déterminant approprié.

En outre, nous définissons également un algorithme afin de prendre en charge les cas d’élision, et ainsi remplacer le déterminant du NM par la forme élidée correspondante pour le nom

collectif. Cet algorithme vérifie simplement si le nom collectif à ajouter commence ou non par une voyelle. Ce même algorithme est réutilisé plus tard dans la chaîne de traitement pour prendre en compte les modifications de pronoms personnels ou de l’adverbe négatif « ne » lors de la modification des dépendances ; nous détaillons davantage ce processus dans la partie dédiée à l’ajustement de la flexion.

Est également pris en compte le changement des déterminants complexes « tous les », « toutes les » et leurs dérivés : nous vérifions tout simplement si le mot précédant le déterminant à remplacer est « tous » ou « toutes » ; le cas échéant, nous le remplaçons par sa variante au singulier. Enfin, nous remplaçons simplement le NM par son équivalent collectif d’après son indice dans la phrase de départ.

3.4.3.2 Ajustement de la flexion

Le deuxième volet de notre composant de génération vise à ajuster la flexion des dépendances précédemment détectées. Pour ce faire, nous avons recours à `inflecteur` (Chuttarsing, 2021), une bibliothèque Python pour le français, qui se base sur le dictionnaire de formes fléchies DELAF créé par le Laboratoire d’automatique documentaire et linguistique¹⁷. À l’aide de ce dictionnaire et du modèle de reconnaissance de catégories syntaxiques `french-camembert-postag-model`¹⁸, entraîné à partir de la version de base de CamemBERT (Martin et al., 2020), il est possible de fléchir un mot en indiquant le genre, le temps ou le nombre voulu à l’aide des méthodes `inflect_word` ou `inflect_sentence`, comme suit :

```
inflecteur.inflect_sentence("Elle boit un café et mange un croissant.", gender="m")  
> "Il boit un café et mange un croissant."
```

Nos essais montrent que la méthode `inflect_sentence` donne en général de meilleurs résultats par rapport à la méthode `inflect_word`. Cela s’explique par le fait que `inflect_sentence` a recours au modèle CamemBERT préentraîné pour déduire la catégorie syntaxique du mot et choisir la forme la plus appropriée dans le dictionnaire DELAF, tandis que `inflect_word` cherche une simple correspondance dans le dictionnaire.

Il convient toutefois de noter que bien qu’`inflecteur` soit un outil performant pour le changement de flexion, il n’est pas sans défauts : par exemple, certains mots construits avec des préfixes (par exemple, « re » dans « rebâtir »), sont également mal identifiés. De plus, les participes passés sont également bien souvent mal reconnus comme tels et ne sont, en conséquence, pas modifiés correctement : cela se produit notamment lorsque l’on essaye de passer d’une forme masculine à une forme féminine.

Afin d’assurer une génération correcte des phrases, nous ajoutons deux étapes à notre chaîne de traitement : une étape de correction des participes passés et une étape de correction des pronoms objets.

17. <http://infolingu.univ-mlv.fr/>

18. <https://huggingface.co/gilf/french-camembert-postag-model>

La première étape vise essentiellement à rajouter la flexion féminine au participe passé. À l'aide de spaCy, nous vérifions si nous avons bien affaire à un participe passé, et nous le modifions de manière à ajouter la flexion féminine d'après la forme de base du mot au masculin.

Pour la seconde étape, qui concerne les pronoms objets, nous modifions les instances de « eux » ou « elles » détectées comme dépendances du NM par leurs équivalents singuliers. En outre, nous modifions également le pronom objet « les » s'il se trouve dans le contexte environnant par rapport au NM ; soit par « le » si le nom collectif est masculin, soit par « la » s'il est féminin. Si le mot qui suit commence par une voyelle ou par un h muet, il est toutefois nécessaire de conserver l'élision (« l' »). Nous reprenons l'algorithme décrit précédemment pour la vérification des voyelles, que nous améliorons pour prendre en charge les mots commençant par un h muet. Pour ce faire, nous avons recours à la page Wikipédia « h aspiré en français »¹⁹, qui liste une série de mots commençant par un h aspiré (et ne nécessitant donc pas d'élision). Nous extrayons les mots concernés à l'aide d'un script Python, puis, dans notre algorithme d'élision, nous vérifions que le lemme à modifier ne se trouve pas dans cette liste.

L'application de toutes ces étapes nous permet d'obtenir, en sortie, une phrase neutralisée en genre. En voici un exemple :

- (53) C'est sans surprise que **les lecteurs** de celui-ci **découvrent** par exemple que la pédante Mary Bennet a épousé un recteur « enclin à prêcher des sermons d'une longueur démesurée et d'une grande complication théologique ».
- (54) C'est sans surprise que **le lectorat** de celui-ci **découvre** par exemple que la pédante Mary Bennet a épousé un recteur « enclin à prêcher des sermons d'une longueur démesurée et d'une grande complication théologique ».

3.4.4 Évaluation

Pour conduire nos évaluations, nous avons recours à deux jeux de données d'évaluation de 250 phrases chacun, soit 500 phrases au total.

Le premier jeu de données est un sous-ensemble de phrases de notre jeu de données initial provenant de Wikipédia, dont une partie a servi au développement de notre système de règles. Le second jeu de données est composé de phrases françaises du corpus parallèle anglais-français Europarl (Koehn, 2005), constitué des transcriptions des débats du parlement européen de 1996 à 2001. Comme nous l'avons fait pour créer notre jeu de données initial, nous découpons les textes du corpus en phrases, et nous récupérons uniquement les phrases qui contiennent les NM dans notre dictionnaire, pour récupérer 250 phrases. Les autres phrases récupérées à partir de ce corpus (au nombre de 106 878) seront utilisées conjointement avec notre jeu de données Wikipédia pour l'affinage de grands modèles de langue (voir chap. 3.5).

Le choix de l'utilisation d'un corpus supplémentaire en plus du corpus initial découle de la volonté d'éviter tout biais de notre système envers le corpus initial, mais également de celle de comparer notre système avec des textes d'un autre registre, les textes de Wikipédia étant

19. https://fr.wikipedia.org/wiki/H_aspiré_en_français

plutôt de registre académique et écrits, tandis que ceux d’Europarl sont des transcriptions de discours ou commentaires politiques oraux. Des résultats similaires pour les deux corpus pourraient indiquer une bonne adaptabilité de notre système vis-à-vis du registre et, possiblement, du mode de communication.

3.4.4.1 Composant de détection des dépendances

L’évaluation du composant de détection des dépendances vise à comparer les dépendances syntaxiques du NM détectées par notre système de règles, et celles détectées par spaCy pour ce même NM, sans appliquer de règle manuelle, excepté le filtrage de la ponctuation, cet élément n’étant pas pertinent pour l’évaluation puisque notre système ne vise pas à le modifier. La détection par spaCy constituera ainsi notre référence (*baseline*).

Comme mesure d’évaluation de ce composant, nous utilisons la F-mesure, une moyenne harmonique de la précision et du rappel : la précision mesure la proportion de dépendances détectées par notre système qui sont effectivement des dépendances correctes, tandis que le rappel mesure la proportion de toutes les dépendances correctes qui ont été identifiées par notre système. Ainsi, cette mesure d’évaluation nous permet de prendre en compte à la fois la qualité des dépendances détectées (précision) et la capacité du système à identifier toutes les dépendances correctes (rappel).

$$\begin{aligned}\text{Précision (P)} &= \frac{TP}{TP + FP} \\ \text{Rappel (R)} &= \frac{TP}{TP + FN} \\ F1 &= \frac{2 * P * R}{P + R}\end{aligned}$$

Dans un premier temps, nous annotons manuellement les dépendances syntaxiques des 500 phrases de notre jeu de données d’évaluation. Pour chaque phrase, nous indiquons les dépendances syntaxiques du NM séparées par des virgules. Dans le cas où une phrase comporte plusieurs MN, nous l’annotons avec un format spécifique de type

`nom_membre1 : dep1, dep2 | nom_membre2 : dep1, dep2.`

Nous développons ensuite un script Python qui, avec un fichier CSV en entrée contenant les phrases de notre sous-ensemble d’évaluation, récupère les dépendances détectées par notre système de règles et celles détectées par spaCy, puis calcule la précision et le rappel pour chaque phrase d’après les dépendances annotées. Nous montrons un exemple de notre code dans l’illustration 3.8.

```
1 def eval_deps(pred_deps: str, true_deps: str, dct=False) -> tuple[float, float]:
2     """
3     Evaluates the precision and recall of predicted dependencies
4     compared to true dependencies.
5
6     Args:
7     pred_deps (str): Comma-separated string of predicted dependencies.
8     true_deps (str): Comma-separated string of true dependencies.
9     dct (bool, optional): Whether the dependencies are in a dictionary.
10    Defaults to False.
11
12    Returns:
13    tuple: A tuple containing the precision and recall values.
14    """
15    tps = set()
16    if not dct:
17        # filter out empty strings
18        pred_list = list(filter(None, pred_deps.split(",")))
19        true_list = list(filter(None, true_deps.split(",")))
20
21        for dep in pred_list:
22            if dep in true_list:
23                tps.add(dep)
24
25        # ZeroDivisionError occurs when pred_list or true_list are empty,
26        # which means that no dependencies are predicted
27        # nor to be found in the sentence,
28        # so we consider the prediction to be accurate.
29        try:
30            precision = len(tps) / len(pred_list)
31        except ZeroDivisionError:
32            precision = 1.0
33
34        try:
35            recall = len(tps) / len(true_list)
36        except ZeroDivisionError:
37            recall = 1.0
38
39    else:
40        try:
41            pred_dict = dict(x.split(":") for x in pred_deps.split("|"))
42            true_dict = dict(x.split(":") for x in true_deps.split("|"))
43        except ValueError as e:
44            raise ValueError(
45                "Row is not formatted correctly."
46                "\nPlease make sure to follow this format:"
47                "\ntarget_noun1:dep1,dep2|target_noun2:dep1,dep2"
48                f"\nCurrent: {pred_deps=} - {true_deps=}"
49            ) from e
50
51        for key, deps in pred_dict.items():
52            if key in true_dict:
53                for dep in deps.split(","):
54                    if dep in true_dict[key].split(","):
55                        tps.add(dep)
56
57        precision = len(tps) / sum(len(v.split(",")) for v in pred_dict.values())
58    () recall = len(tps) / sum(len(v.split(",")) for v in true_dict.values())
59
60    return precision, recall
```

ILLUSTRATION 3.8 – Extrait du code utilisé pour l'évaluation du composant de détection des dépendances

Nous faisons ensuite une moyenne de la précision et du rappel pour notre sous-ensemble avant de calculer la F-mesure.

3.4.4.2 Composant de génération

Pour évaluer les performances de notre composant de génération, nous avons recours à différentes mesures d'évaluation utilisées par d'autres travaux sur la tâche de réécriture du genre. Il n'existe pas, à ce jour, de mesure d'évaluation strictement spécifique à cette tâche. Les mesures d'évaluation utilisées jusqu'alors sont également utilisées pour l'évaluation de traductions automatiques ou de résumés. Cela implique également qu'il n'existe pas une seule mesure de référence pour l'évaluation des systèmes de réécriture du genre ; certains travaux n'ont donc recours qu'à une seule mesure, tandis que d'autres en utilisent plusieurs en comparaison. Nous utilisons au total trois mesures pour l'évaluation de notre système : BLEU (à l'instar de Alhafni et al. (2022b)), le taux d'erreur par mot (WER ; à l'instar de Sun et al. (2021) et Vanmassenhove et al. (2021)), et la similarité cosinus. Cette dernière mesure n'a, à ce jour, pas été utilisée dans les travaux de recherche concernant cette tâche. Celle-ci vise à mesurer la similarité de deux vecteurs de mots ainsi :

$$\text{Cosine}(x, y) = \frac{x \cdot y}{|x||y|}$$

Nous pensons que cette mesure, en complément des autres mesures existantes, peut s'avérer pertinente pour l'évaluation de systèmes de réécriture du genre, en ce que, contrairement à d'autres mesures utilisées comme BLEU, une pénalité n'est pas appliquée au score d'évaluation si la phrase évaluée est plus courte que la référence.

Comme pour l'évaluation du composant de détection des dépendances, nous créons des annotations de référence pour chaque phrase, c'est-à-dire l'équivalent neutralisé en genre avec un nom collectif (si possible) et sémantiquement correcte. Pour certaines phrases du jeu de données d'évaluation, le remplacement du nom masculin générique par le nom collectif équivalent est problématique pour des raisons de sémantisme, sur lesquelles nous revenons plus amplement dans la section des résultats (chap. 3.4.4.2). Dans ces cas-là, nous n'effectuons pas de remplacement du nom masculin.

À l'instar des travaux précédents, notre référence pour le composant de génération est la phrase d'entrée non modifiée. Nous développons un script Python pour calculer automatiquement les scores BLEU, WER et de similarité cosinus, à la fois pour la phrase d'entrée modifiée et pour la phrase de sortie générée par notre système de règles.

3.4.4.3 Résultats et discussion

Cette section expose les résultats des deux composants de notre système de règles : les résultats du composant de détection des dépendances sont disponibles dans le tableau 3.2 ; ceux de notre système dans sa totalité dans le tableau 3.3.

Notre composant de détection des dépendances montre, en moyenne, une amélioration de 0,6155 point de F-score par rapport à la référence, qui correspond à la détection automatique par spaCy. Nous passons en effet d'un F-score moyen de 0,183 à 0,7985. Bien que le score

	Wikipédia			Europarl			Moy.		
	Précision	Rappel	F-score	Précision	Rappel	F-score	Précision	Rappel	F-score
Réf. (baseline spaCy)	0,096	0,723	0,169	0,115	0,689	0,197	0,1055	0,706	0,183
GeNRe-RBS	0,773	0,855	0,812	0,758	0,813	0,785	0,7655	0,834	0,7985

TABLEAU 3.2 – Résultats du composant de détection des dépendances, par corpus et en moyenne

	Wikipédia			Europarl			Moy.		
	WER	BLEU	Sim. cos.	WER	BLEU	Sim. cos.	WER	BLEU	Sim. cos.
Réf. (baseline)	13,35 %	79,634	0,9115	13,344 %	81,465	0,917	13,347 %	80,549	0,914
GeNRe-RBS	3,358 %	93,024	0,979	3,448 %	93,833	0,984	3,403 %	93,428	0,982

TABLEAU 3.3 – Résultats du système à base de règles, par corpus et en moyenne

de précision de spaCy soit très faible, le score de rappel est bien plus élevé et se rapproche parfois du score du composant du système de règles. Cela s'explique par le fait que spaCy récupère toutes les dépendances liées de près ou de loin au NM, sans filtre particulier, y compris celles qui ne nécessitent aucune modification au niveau du genre ou du nombre par le composant : c'est notamment le cas des déterminants et des prépositions, qui apparaissent à haute fréquence dans les dépendances détectées par spaCy. C'est également pour cette raison que le rappel de la référence est bien plus élevé que la précision, car le nombre de dépendances détectées par spaCy étant relativement élevé, il y a davantage de chances qu'un mot faisant effectivement partie des dépendances soit trouvé.

Les résultats de notre composant de détection des dépendances sont plutôt satisfaisants : le F-score moyen étant de 0.7985, cela signifie qu'environ 79 % des dépendances sont correctement identifiées. Le score de précision (0,7655) et le score de rappel (0,834) pour GeNRe-RBS sont très proches, ce qui indique des performances équilibrées du composant, contrairement à la détection automatique de spaCy, qui fait état d'un rappel très élevé en comparaison avec la précision. Le composant du système de règles gagnerait toutefois à être amélioré pour obtenir un plus grand pourcentage de détection correcte. Les principales erreurs de détection que nous rencontrons sont de deux types.

Premièrement, dans certains cas, le même mot apparaît plusieurs fois dans la liste des dépendances détectées, bien qu'il n'y en ait qu'une seule occurrence dans la phrase analysée. Cela se produit lorsque le mot répond à plusieurs filtres : étant donné que nous faisons une analyse étape par étape en fonction du type de token et de ses voisins (ancêtre, parent direct, enfant), il suffit que plusieurs règles correspondent pour que le mot soit ajouté plusieurs fois. Des modifications ont été apportées pour atténuer ce problème, mais il existe encore des cas problématiques. En outre, il ne nous est pas possible de supprimer les mots doublons, car il peut y avoir plusieurs dépendances dans une phrase, et celles-ci peuvent prendre exactement la même forme.

Deuxièmement, un certain nombre de dépendances faussement détectées comme telles sont en réalité des mots qui ne sont pas liés au NM, mais à d'autres noms dans la phrase. Cela a notamment lieu dans les phrases complexes, avec de longues propositions ou des propositions enchâssées, qui rendent difficile la détection de dépendances lorsqu'elles sont éloignées du NM. En ce qui concerne les résultats de notre composant de génération, nous notons, en moyenne et par rapport à la référence, une amélioration de 9,957 points de pourcentage en termes de

Type d'erreur	Wikipédia	Europarl
ADJ	35	35
DET	4	6
DET_COREF	12	0
ELISION	4	2
INCLUSIVE	1	0
MISID_NOUN	0	2
DET_CONTRACT	13	6
PRON_COREF	3	30
PUNCT	1	14
SEM	23	15
UNREPLACED	5	2
VERB	39	38

TABLEAU 3.4 – Compte d'erreurs du système de règles par corpus

WER, une progression de 12,879 pour le score BLEU et une augmentation de 0,068 pour la similarité cosinus. Bien qu'une comparaison exacte avec les autres travaux de recherche existants ne soit pas possible, cela en raison des langues différentes en jeu et du fait que nous nous concentrons sur les NM et les noms collectifs, nous pouvons tout de même établir des parallèles avec les travaux portant sur la neutralisation automatique du genre. Sun et al. (2021) avaient montré une amélioration de 11,77 points de pourcentage (WER) et 9,63 (BLEU) ; et Vanmassenhove et al. (2021) une amélioration de 10,947 points de pourcentage (WER). Ces résultats sont comparables à ceux que nous avons obtenus.

En outre, les résultats pour les deux corpus sont très proches, ce qui semble montrer une relativement bonne adaptabilité du composant de génération vis-à-vis du genre textuel.

Le tableau 3.4 montre une analyse plus précise des performances du composant, avec un décompte des erreurs pour chaque corpus. Nous joignons également un graphique représentant le compte d'erreurs pour une meilleure lisibilité dans l'illustration 3.9.

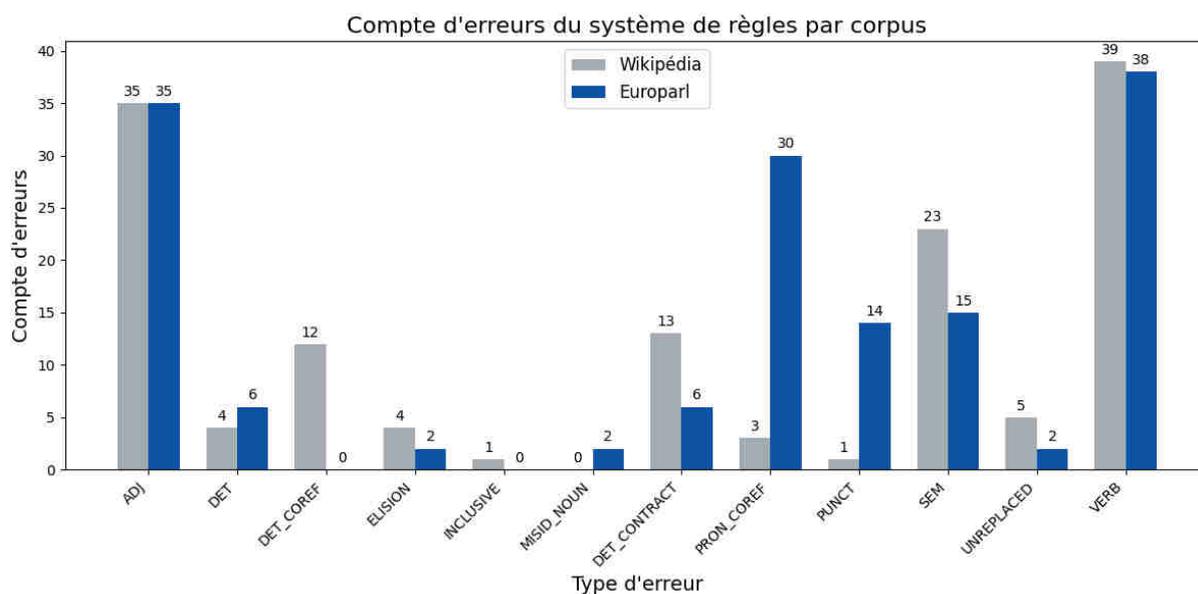


ILLUSTRATION 3.9 – Représentation graphique du compte d'erreurs du système de règles par corpus

Les erreurs répertoriées se situent à deux niveaux : soit au niveau de la catégorie de la dépendance à remplacer, soit à un niveau plus général ; nous les décrivons plus en détail ci-après. Concernant les erreurs relevant de la catégorie de la dépendance à remplacer, nous en relevons six : ADJ, DET, DET_COREF, DET_CONTRACT, PRON_COREF et VERB.

Le type d'erreur « ADJ » est appliqué lorsqu'un adjectif n'a pas été modifié, ou a été modifié à tort, dans la phrase, comme dans l'exemple 55.

(55) Il reçoit l'hommage de la prinerie **lombards** de Salerne et de Capoue.

Le type d'erreur « DET » est appliqué lorsque le déterminant du nom collectif n'a pas été modifié, ou a été modifié à tort, dans la phrase, comme dans l'exemple 56, où « des sénateurs » a été remplacé par « du sénat » au lieu de « le sénat ».

(56) Elle touche **du sénat**, des chevaliers et de riches affranchis.

« DET_COREF » est utilisé lorsqu'un déterminant possessif qui se rapporte au nom collectif n'a pas été modifié ou a été modifié à tort, comme dans l'exemple 57, où « leur », se rapportant à « monarchie » devrait normalement être remplacé par « son ».

(57) La monarchie de France le prenait quand elle était en deuil et **leur** drap mortuaire était de cette couleur.

« DET_CONTRACT » est utilisé lorsque le déterminant « des » dans la phrase originale a été à tort interprété comme faisant partie d'un complément du nom. Par exemple, si l'on avait le syntagme nominal seul « le révolver des policiers », celui-ci devrait être transformé en « le révolver de la police ». Toutefois, dans l'exemple 58, « des policiers » dans la phrase original est COD de « menacer » ; il ne s'agit pas d'un complément du nom et « de » devrait donc être supprimé.

- (58) La légende veut qu'elle ait menacé à plusieurs reprises d'un revolver **de** la police qu'elle avait surpris en train de maltraiter des Aborigènes qu'elle estimait être sous sa protection.

L'erreur « PRON_COREF » apparaît lorsqu'un pronom possessif qui se rapporte au nom collectif n'a pas été modifié ou a été modifié à tort, comme dans l'exemple 59, où « leur », se rapportant à « armée » devrait normalement être remplacé par « lui ».

- (59) Durant la Première Guerre mondiale, elle remplace son mari mobilisé comme médecin chef du sanatorium qu'il a créé et soigne l'armée russe dans l'hôpital qui **leur** est dédié.

Enfin, l'erreur « VERB » survient lorsqu'un verbe qui a comme argument un nom collectif n'a pas été modifié ou a été modifié à tort.

- (60) Paradoxalement, cette progression en voix s'accompagne d'un recul en nombre d'élus, du fait de la poussée des candidats indépendants (pour la plupart de la représentation de la communauté kurde) et du CHP, qui échoue cependant à atteindre la barre des 30 % des voix comme sa direction **l'avaient** espéré.

En ce qui concerne les erreurs d'un niveau général, nous en dénombrons également six : ELISION, INCLUSIVE, MISID_NOUN, PUNCT, SEM et UNREPLACED.

Le type d'erreur « ELISION » regroupe les phrases dans lesquelles l'algorithme visant à prendre en charge les cas d'élimination, explicités précédemment dans le chapitre **3.4.3.2**, ne s'est pas appliqué correctement, comme dans l'exemple 61.

- (61) Il est temps de le faire reconnaître et de dire ouvertement et publiquement la vérité sur les blocages nationaux, sur les retards pris concernant ce qui est appliqué, sur ce qui **n'est** pas, ainsi que sur la volonté réelle du Conseil de la ministraille.

« INCLUSIVE » fait référence aux phrases qui ont été transformées alors qu'elles contenaient des doublets : nous n'avons relevé qu'une seule occurrence de ce type, l'exemple 62. Ce type de phrase n'est pas pris en charge par notre système de règles, car celui-ci est censé recevoir en entrée une phrase avec un nom au masculin générique.

- (62) Louise O'Sullivan a été conseiller de la ville du district électoral de Peter-McGill et membre du comité exécutif de la ville de Montréal sous la bannière de l'Union de **la citoyenneté et citoyennes** de l'île de Montréal dont le chef est Gérard Tremblay.

« MISID_NOUN » s'applique aux cas où une forme présente dans le dictionnaire de paires NM-noms collectifs a été incorrectement identifiée comme nom. Par exemple, dans l'exemple 63, il s'agit de l'adjectif « parlementaires » et non du nom « parlementaires ».

- (63) Le parlement doit donc procéder par déduction, par un examen minutieux de chaque détail, par les questions **parlement**, comme je le fais moi.

L'erreur « PUNCT » est relative aux problèmes de ponctuation dans la phrase générée, notamment les problèmes d'espace. Étant donné qu'une tokénisation de la phrase d'entrée est effectuée par spaCy, il est nécessaire de rétablir les espaces telles qu'elles apparaissaient dans la phrase d'origine. Nous utilisons, pour ce faire, des expressions régulières, mais certains cas peuvent être ambigus : par exemple, il n'est pas toujours possible de savoir si un trait d'union est véritablement utilisé comme tel (pour souder deux mots), ou bien comme tiret pour introduire une pause ou une incise : voir l'exemple 64. Ce type d'erreur est davantage prégnant dans le corpus d'Europarl en raison de la forte utilisation de ce type de signes typographiques dans les transcriptions de discours.

- (64) À entendre les débats aujourd'hui, je ne peux que constater que, finalement, malgré le choc **total-économique**, financier, social, **démographique-qui** est lié à la crise que nous vivons, les institutions européennes n'ont, pas plus que la chefferie d'État ou de gouvernement, pris la mesure du drame et de ses conséquences qui vont nous frapper.

Le type d'erreur « SEM » regroupe les phrases générées qui sont sémantiquement incorrectes. Ce type d'erreur est propre au remplacement de NM par leurs équivalents collectifs. Comme nous l'avons vu dans le chapitre 2.2.2.2, les noms collectifs sont régis par un certain nombre de contraintes sémantiques et ne peuvent s'appliquer à tous les contextes. En particulier, un certain nombre d'erreurs provient du fait que le nom commun « hommes » a été remplacé à tort par « humanité » (voir exemple 65). Bien que dans certains emplois, « hommes » puisse désigner l'ensemble des êtres humains, les phrases de notre corpus parlent bien souvent des différences entre les hommes et les femmes. Dans ce genre de cas où l'accent est mis sur la comparaison entre deux groupes, il n'est pas possible de remplacer « hommes » par « humanité ». D'autres problèmes sémantiques sont posés lors de l'utilisation de pronoms indéfinis faisant référence au NM (voir exemple 66), soit pour désigner la totalité (« tous », « l'ensemble de »...), soit pour en faire une extraction (« l'un »). Les noms collectifs servant déjà à désigner un groupe complet, l'utilisation de pronoms indiquant la totalité est impossible. Par ailleurs, l'extraction d'un groupe décrit par un nom collectif par l'utilisation d'un pronom indéfini semble difficile sémantiquement parlant et ne semble pas répandue dans l'usage. Un autre problème concerne l'utilisation de titres tels que « Mesdames » ou « Messieurs » suivis d'un NM pluriel (voir exemple 67). Pareillement, étant donné qu'il est fait référence à des êtres spécifiques, une généralisation sous la forme d'un nom collectif n'est pas envisageable.

- (65) Je pense également qu'il est important d'empêcher que les femmes continuent d'être les victimes de discrimination de la part des compagnies d'assurance simplement parce qu'elles sont des femmes et qu'elles sont plus âgées, et qu'en tant que groupe, elles sont en meilleure santé et vivent plus longtemps que **l'humanité**.
- (66) Je précise que vendredi a marqué la fin du processus car je pense que la citoyenneté européenne ont **tous** exprimé leur avis, directement ou indirectement.

- (67) Je suis absolument d'accord avec l'approche ambitieuse du Parlement pour que ces guichets deviennent très vite de véritables portails administratifs permettant aux entreprises de faire toutes leurs démarches en ligne, y compris, Madame la rapporteure, **Mesdames et Messieurs le parlement**, au-delà du seul champ de la directive sur les services.

Enfin, « UNREPLACED » regroupe les phrases où le NM n'a pas été remplacé par son équivalent collectif en raison d'un problème de détection. Il est également possible que ce nom, dans la phrase, n'ait pas été catégorisé correctement par spaCy, empêchant alors sa transformation. Nous en donnons un exemple dans la phrase 68, où « aux représentants » aurait dû être remplacé par « à la représentation ».

- (68) Je voudrais adresser mes remerciements au commissaire, **aux représentants** du Conseil et, évidemment, à Mme Lulling, avec qui ma collaboration a été excellente, ainsi qu'à tous mes collègues de la commission des droits de la femme et de l'égalité des genres.

Au total, nous relevons 221 erreurs au niveau de la catégorie de la dépendance à remplacer, et 62 erreurs générales. Les erreurs les plus prégnantes dans les deux corpus, toutes catégories confondues, sont celles concernant les adjectifs et les verbes. Il existe aussi un certain nombre d'erreurs relatives à la coréférence : le corpus Europarl est particulièrement sujet à ce type d'erreur pour les pronoms (30 erreurs contre seulement 3 pour le corpus Wikipédia), ce qui s'explique par le fait, indiqué précédemment, que les phrases du corpus Europarl sont bien souvent plus longues et complexes, et contiennent davantage de références à des éléments indiqués plus tôt dans la phrase ou le discours.

3.5 GeNRe-FT : affinage de grands modèles de langue

La plupart des travaux de recherche réalisés jusqu'ici pour la tâche de réécriture du genre ont visé à utiliser les données générées par un système de règles pour affiner des grands modèles de langue existants, c'est-à-dire adapter ces modèles à partir de nouvelles données afin d'améliorer leurs performances sur une tâche spécifique : c'est notamment ce qui a été fait par Sun et al. (2021), Vanmassenhove et al. (2021) et Veloso et al. (2023), entre autres. Tandis que Sun et al. (2021) et Veloso et al. (2023) ne montrent pas d'amélioration majeure, mais plutôt une diminution des performances du modèle de langue affiné par rapport au système de règles, Vanmassenhove et al. (2021) constatent une amélioration notable de 0,27 point de pourcentage en termes de WER (0,34 → 0,06).

Nous souhaitons voir si l'affinage de grands modèles de langue permet bien, comme indiqué par les résultats de Vanmassenhove et al. (2021), d'améliorer de manière claire les résultats d'un système de règles. L'hypothèse faite est que, les grands modèles de langue étant entraînés sur un large nombre de textes, ils transfèreraient une partie de leurs connaissances linguistiques acquises à partir de ces textes lors de l'affinage, ce qui permettrait ainsi de résoudre certaines erreurs dans le corpus d'entraînement, et donc d'améliorer les résultats.

	Wikipédia			Europarl			Moy.		
	WER	BLEU	Sim. cos.	WER	BLEU	Sim. cos.	WER	BLEU	Sim. cos.
Réf. (baseline)	13,35 %	79,634	0,9115	13,344 %	81,465	0,917	13,347 %	80,549	0,914
GeNRe-FT-T5	6,095 %	88,144	0,956	4,128 %	93,211	0,979	5,111 %	90,678	0,968
GeNRe-FT-M2M100	6,679 %	87,09	0,954	4,116 %	93,298	0,98	5,397 %	90,194	0,967

TABLEAU 3.5 – Résultats des modèles affinés

Nous choisissons deux grands modèles de langue Seq2seq (Sutskever et al., 2014a) pour réaliser nos expériences, ce type de modèle étant conçu pour transformer une séquence de mots en une autre séquence, et étant communément utilisés pour les tâches de traduction automatique ou de résumé de textes. Les deux modèles retenus sont T5 de Google (Raffel et al., 2020), entraîné sur le corpus C4 (un corpus Web contenant principalement des textes en anglais, mais aussi en français, roumain, allemand ou encore russe), et m2m100 de Meta (Fan et al., 2020), un modèle multilingue créé à partir de données de plus de 100 langues, et également utilisé par Veloso et al. (2023).

Nous utilisons nos deux corpus (Wikipédia et Europarl) de paires de phrases inclusives et non inclusives pour l’affinage de ces modèles. Ces deux corpus contiennent respectivement 287 943 et 106 878 paires de phrases. Nous créons à partir de ces corpus notre jeu de données d’entraînement (composé de 193 228 paires de phrases) et notre jeu de données de validation (composé de 21 471 paires de phrases). Pour des raisons de performances et de temps d’affinage des modèles, nous réduisons notre corpus d’entraînement à 60 000 phrases pour l’affinage de chaque modèle, et notre corpus de validation à 6 000 phrases (soit 10 % de notre corpus d’entraînement). Pour ce faire, nous mélangeons aléatoirement nos jeux de données d’entraînement et de validation initiaux afin de ne garder que 60 000 et 6 000 phrases. Les mêmes jeux de données sont utilisés pour l’entraînement des deux modèles. L’affinage est effectué sur un unique GPU NVIDIA RTX 4090. Les hyperparamètres utilisés pour chaque modèle sont répertoriés en annexe A.

3.5.1 Évaluation, résultats et discussion

Les résultats des deux modèles sur notre jeu de données d’évaluation sont disponibles dans le tableau 3.5.

Les résultats pour les deux modèles sont similaires, avec des résultats légèrement meilleurs pour le modèle affiné T5. Une comparaison au niveau des corpus permet de constater que le modèle qui obtient les meilleurs résultats sur le corpus de Wikipédia est le modèle T5 (6,095 % de WER contre 6,679 % pour M2M100), tandis que le modèle affiné M2M100 s’en sort comparativement mieux sur le corpus d’Europarl (4,128 % de WER contre 4,116 %), comme c’est le cas pour notre système de règles. Ces différences peuvent s’expliquer par le type de données utilisées pour l’entraînement initial de ces deux grands modèles de langue.

Contrairement à Vanmassenhove et al. (2021), et en accord avec les résultats de Sun et al. (2021) et Veloso et al. (2023), nous ne remarquons pas d’amélioration significative des résultats pour nos deux modèles après un affinage, mais plutôt des résultats en légère baisse en comparaison avec ceux du système de règles. Il semblerait donc que l’utilité d’un affinage de

grands modèles de langue sur les phrases générées par le système de règle soit plutôt limité ; il est également possible qu'en fonction de la langue, les résultats soient plus ou moins probants. Les grands modèles de langue étant majoritairement entraînés sur des données en anglais, il est possible qu'un affinage sur de telles données porte davantage ces fruits avec cette langue. Toutefois, une analyse qualitative des phrases générées confirme l'hypothèse d'une généralisation de règles linguistiques par le modèle de langue lors de la génération. Nous remarquons en effet plusieurs cas où des dépendances n'ayant pas été modifiées correctement par le système de règles sont corrigées par le modèle de langue affiné. Dans l'exemple 69, le verbe « être » a correctement été conjugué par le modèle affiné T5, quand bien même cette modification n'a pas su être réalisée par le système de règles ; il en est de même pour l'exemple 70 et le verbe « vouloir », avec le modèle affiné m2m100. Un autre exemple, l'exemple 71, montre la capacité du modèle T5 à supprimer une préposition superflue avant le nom collectif « jeunesse ».

- (69) a. Cette démarche fera progresser les droits de la citoyenneté, car, par l'intermédiaire du Parlement, la citoyenneté **seront** en contact direct avec la Commission, ce qui lui confèrera une légitimité considérable. [GeNRe-RBS]
- b. Cette démarche fera progresser les droits de la citoyenneté, car, par l'intermédiaire du Parlement, la citoyenneté **sera** en contact direct avec la Commission, ce qui lui confèrera une légitimité considérable. [GeNRe-FT-T5]
- (70) a. Un deuxième élément concerne le soutien apporté à la Commission à l'actorat local qui **veulent** participer à ces programmes afin d'avoir accès aux sources de financement correspondantes. [GeNRe-RBS]
- b. Un deuxième élément concerne le soutien apporté à la Commission à l'actorat local qui **veut** participer à ces programmes afin d'avoir accès aux sources de financement correspondantes. [GeNRe-FT-M2M100]
- (71) a. Il serait par exemple possible d'offrir des incitations fiscales aux entreprises qui emploient **de la jeunesse** ou de diversifier la gamme de contrats de travail flexibles. [GeNRe-RBS]
- b. Il serait par exemple possible d'offrir des incitations fiscales aux entreprises qui emploient **la jeunesse** ou de diversifier la gamme de contrats de travail flexibles. [GeNRe-FT-T5]

Lors de l'évaluation, nous avons, en outre, remarqué une propriété intéressante offerte par l'utilisation des modèles de langue. Notre dictionnaire de paires NM-noms collectifs contient parfois plusieurs équivalences : par exemple, le NM « soldats » peut être remplacé par plusieurs noms collectifs : « armée », « bataillon », « régiment », « infanterie ». . . Pour la constitution du jeu de données utilisé pour affiner les modèles de langue, nous avons fait en sorte que le système à base de règles génère toutes les possibilités. Ainsi, les modèles de langue affinés sont capables d'utiliser les différentes équivalences du dictionnaire. Nous donnons plusieurs phrases d'exemple ci-dessous pour la phrase « Les soldats sont partis. », qui peut alors être convertie de plusieurs manières différentes.

- (72) L'armée est partie.
- (73) Le bataillon est parti.
- (74) L'infanterie est partie.

Nous avons ainsi dû légèrement adapter notre méthode d'évaluation, car la phrase générée par le modèle de langue ne contient pas forcément le même nom collectif que celui utilisé dans la phrase manuelle utilisée comme comparaison. Pour prendre en charge ce genre de cas et obtenir une évaluation exacte, nous avons donc annoté les phrases générées par les modèles de langue contenant un nom collectif différent, puis nous avons créé des versions alternatives de ces phrases générées, qui contiennent le nom collectif présent dans la phrase manuelle, afin de les utiliser uniquement pour l'évaluation : cela évite d'augmenter artificiellement et erronément le taux d'erreurs des modèles de langue affinés car les noms collectifs ne seraient pas identiques. Remplacer directement la phrase manuelle n'était pas une option envisageable, étant donné que cela aurait provoqué un changement des valeurs pour notre référence.

Nous donnons un exemple de ce type de cas ci-dessous. Le modèle GeNRe-FT-M2M100 a utilisé, pour remplacer le nom de membre « soldats », le nom collectif « milice », qui fait effectivement partie de notre dictionnaire et donc des données d'entraînement. Notre phrase manuelle utilise toutefois le premier nom collectif associé au nom de membre « soldats », c'est-à-dire « armée ». Nous avons donc créé une phrase alternative dans laquelle « la milice » est modifié en faveur de « l'armée » afin de correspondre au nom collectif présent dans la phrase originale, et avons utilisé cette phrase alternative pour le calcul des scores de similarité.

- (75) À Vorzel, à l'ouest de Boutcha, **l'armée** russe a tué une femme et son enfant de 14 ans après avoir lancé des grenades fumigènes dans le sous-sol dans lequel ils se cachaient.
[phrase manuelle]
- (76) À Vorzel, à l'ouest de Boutcha, **la milice** russe a tué une femme et son enfant de 14 ans après avoir lancé des grenades fumigènes dans le sous-sol dans lequel ils se cachaient.
[GeNRe-FT-M2M100]

Comme pour le système de règles, le tableau 3.6 montre un décompte des erreurs pour chaque modèle au niveau du corpus. Une représentation graphique est également disponible dans l'illustration 3.10.

Type d'erreur	T5 - Wikipédia	T5 - Europarl	M2M100 - Wikipédia	M2M100 - Europarl
ADJ	33 (-2)	31 (-4)	38 (+3)	33 (-2)
CASE	6 (+6)	0	7 (+7)	0
DET	3 (-1)	6 (0)	2 (-2)	5 (-1)
DET_COREF	12 (0)	0 (0)	10 (-2)	0 (0)
ELISION	1 (-3)	1 (-1)	1 (-3)	1 (-1)
GEN_FAILURE	2 (+2)	1 (+1)	43 (+43)	9 (+9)
INCLUSIVE	1 (0)	0 (0)	1 (0)	0 (0)
MISID_NOUN	1 (+1)	2 (0)	0 (0)	2 (0)
DET_CONTRACT	9 (-4)	3 (-3)	6 (-7)	5 (-1)
PRON_COREF	7 (+4)	30 (0)	6 (+3)	30 (0)
PUNCT	3 (+2)	21 (+7)	2 (+1)	22 (+8)
SEM	22 (-1)	16 (+1)	22 (-1)	16 (+1)
SPECIAL_CHAR	17 (+17)	12 (+12)	0	0
UNREPLACED	4 (-1)	8 (+6)	5 (0)	8 (+6)
VERB	28 (-11)	39 (+1)	38 (-1)	43 (+5)

TABLEAU 3.6 – Compte d'erreurs des modèles affinés par corpus

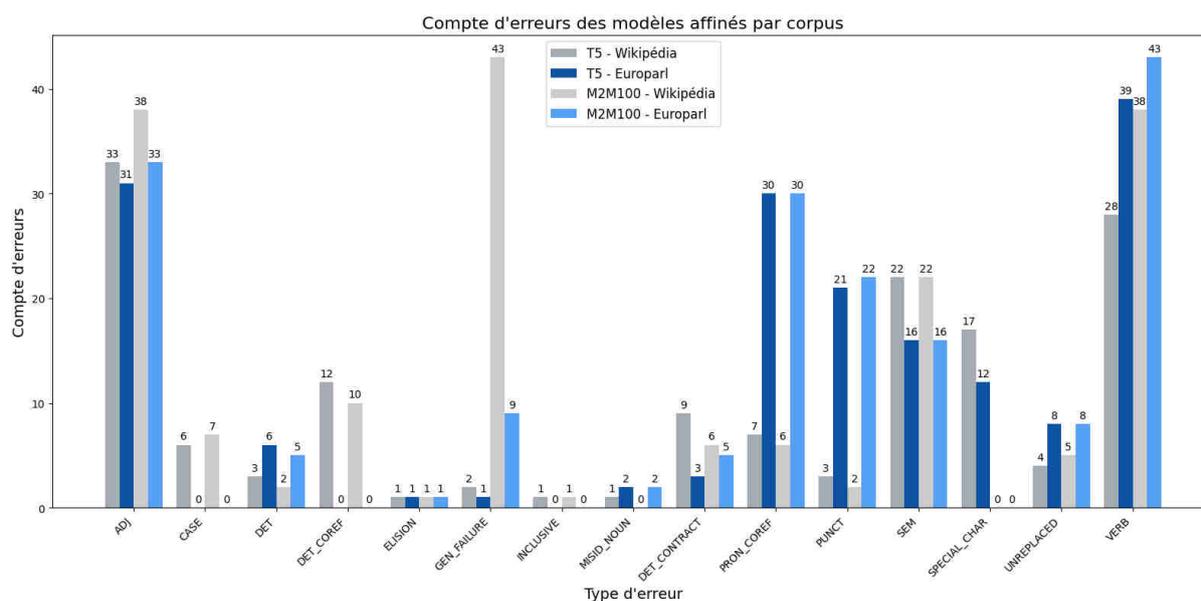


ILLUSTRATION 3.10 – Représentation graphique du compte d'erreurs des modèles affinés par corpus

Nous reprenons les mêmes types d'erreurs que ceux définis pour le décompte des erreurs du système des règles, et rajoutons les types d'erreurs suivants : CASE, GEN_FAILURE et SPECIAL_CHAR. Le nombre entre parenthèses correspond à la différence en termes d'erreurs par rapport au système de règles.

L'erreur « CASE » s'applique lorsqu'un mot n'est pas correctement mis en majuscules (voir exemple 77). Les erreurs relevées surviennent uniquement pour le premier mot de la phrase. Nous notons également que ce type d'erreur survient uniquement sur le corpus Wikipédia. Cela

pourrait éventuellement être lié à un problème de découpage des phrases du corpus, et non pas nécessairement à un problème lié aux modèles en eux-mêmes ; bien qu'aucune erreur de ce type n'ait fait jour lors de notre évaluation du système à base de règles.

(77) **l'armée** arriva avec une lance à eau pour disperser les détenus.

Le type d'erreur « GEN_FAILURE » fait référence aux cas où un ou plusieurs problèmes de génération de tokens par le modèle ont été détectés dans la phrase. Il se peut alors qu'une partie d'un mot soit remplacée par d'autres caractères. Ces erreurs surviennent fréquemment dans des phrases avec des noms propres, et peut affecter ces derniers directement, ou bien d'autres mots dans la phrase. Nous donnons quatre exemples de telles erreurs ci-après afin de mettre en lumière les changements auxquels elles peuvent conduire. Dans certains cas, cela peut aussi mener à des déformations de mots par l'ajout ou la suppression d'espaces (voir exemples 81 et 82).

(78) Juin, Russie : le **Nebski** sobor prend des décisions importantes : inventaire des terres imposables, encouragement au retour de la paysannerie en fuite, répression des abus commis par les fonctionnaires, réforme de l'administration provinciale (assemblées élues), budget national. (Zemski → Nebski)

(79) La Society **ofamina** offre son aide à travers la Grande-Bretagne en soutenant l'enfance dans les écoles et les pensionnats quakers, en organisant des programmes de formation agricole qui incluent le gîte et le couvert, la recherche d'emploi et la gestion d'auberges gratuites ou subventionnées. (of Friends → ofamina)

(80) Ils ont joué une chanson avant que la police **annoulent** finalement le concert. (annule → annoulent)

(81) **L'organisationPromet** de réintégrer le calendrier de l'UCI America Tour.

(82) Ecbatane : le siège du royaume **pers es**.

Ce type d'erreur apparaît à haute fréquence pour le modèle M2M100 (52 erreurs au total contre seulement 3 pour le modèle affiné T5). Étant donné que cette erreur apparaît souvent en présence de noms propres, il est possible que le modèle M2M100 soit moins performant lors de la présence de ces derniers. En outre, comme M2M100 est un modèle de langue multilingue, il est également possible que cela provoque une génération accrue d'erreurs en raison de vocabulaire identique ou partagé entre les langues.

Enfin, le type d'erreur « SPECIAL_CHAR » s'applique lorsqu'un caractère spécial (accent, entre autres) n'est pas généré correctement (voir exemple 83). Seul le modèle T5 effectue cette erreur. À contrario du modèle M2M100, on peut penser que cette erreur provient du fait que le modèle a été entraîné en très grande majorité sur des données en anglais, n'ayant pas de tels accents, d'où les difficultés rencontrées à les prendre en compte.

(83) Il est allé à Cologne, où il est devenu président de l'association de la **main-d'uvre** et a aidé à propager les idées marxistes parmi ses membres.

Au total, nous relevons 201 erreurs au niveau de la catégorie de la dépendance à remplacer pour T5 et 254 pour M2M100 (contre 221 erreurs pour le système de règles) ; et 118 erreurs générales pour T5 et 139 pour M2M100 (contre 62 erreurs pour le système de règles). Comme pour le système des règles, la majorité des erreurs relevant de la catégorie concernent les verbes et les adjectifs. Toutefois, le nombre d'erreurs moindre pour T5 vis-à-vis de cette catégorie confirme la généralisation des règles linguistiques par le modèle, bien que celle-ci ne soit pas suffisante pour obtenir de meilleurs résultats globaux en comparaison avec le système à base de règles.

3.6 Modèles d'instructions

Le développement rapide des grands modèles de langue et les avancées dans le domaine du traitement automatique des langues ont montré qu'il était possible de manipuler le comportement des modèles de langue de manière à prédire la suite d'un texte afin de remplir une certaine tâche, sans que le modèle n'ait été spécifiquement entraîné sur cette tâche : tout repose alors sur le « prompt », c'est-à-dire sur l'instruction donnée au modèle de langue (P. Liu et al., 2021). Il est alors par exemple possible de demander la traduction d'une phrase de l'anglais vers le français avec ce type d'instruction :

English: The cat was sleeping. French: ____

L'affinage de grands modèles de langue avec des instructions et des réponses écrites par des êtres humains permet d'améliorer les performances, donnant ainsi naissance au premier grand modèle d'instructions par OpenAI, InstructGPT (Ouyang et al., 2022)²⁰. Ces modèles d'instructions assureront par la suite le développement de modèles conversationnels tels que ChatGPT (OpenAI)²¹, Gemini (Google)²² ou encore ERNIE Bot (Baidu)²³, qui gardent notamment trace des instructions précédentes et présentent une interface de type « chat ». Bien que certains travaux relatifs à la réécriture du genre aient brièvement évoqué l'usage des modèles d'instructions pour réduire les biais de genre dans les textes générés automatiquement, et aient parfois effectué quelques essais avec de tels modèles²⁴, aucun à ce jour n'a mené d'analyse plus étendue sur leur potentialité.

Nous visons à évaluer les performances de ce type de modèle sur cette tâche. Nous utilisons deux modèles d'instructions : Mixtral 8x7B Instruct v0.1, un modèle multilingue et multitâche

20. Voir également <https://openai.com/research/instruction-following>.

21. <https://openai.com/blog/chatgpt>

22. <https://gemini.google.com/app>

23. <https://yiyian.baidu.com>

24. Nous pouvons par exemple citer Veloso et al., 2023, dont l'article contient, en annexe, cinq phrases neutralisées en genre générées en portugais par ChatGPT. L'article propose l'utilisation de modèles d'instructions pour la génération de jeu de données plus étendus pour cette tâche.

créé par Mistral AI (Jiang et al., 2024)²⁵, et Claude 3 Opus, créé par Anthropic²⁶ et publié en mars 2024 ; il s'agit à ce jour le meilleur modèle en termes de génération textuelle d'après les tests de performances (Anthropic, 2024)²⁷. Un accès à leurs API respectives nous permet d'utiliser les modèles sans les charger localement^{28 29}.

Un certain nombre de travaux ont démontré l'importance du « prompt engineering » (Amatriain, 2024 ; Arora et al., 2022 ; Si et al., 2023), c'est-à-dire la manière de rédiger les instructions données au modèle : les réponses émises par le modèle peuvent en effet varier selon le style des instructions. Dans cette même idée sont apparus les concepts de *zero-shot*, *one-shot* et *few-shot prompting* (respectivement « instructions à zéro, un ou quelques exemples »), qui mettent en avant la variabilité des performances des modèles en fonction que l'on leur présente, ou non, sous forme d'instruction, des exemples de réponse attendue pour la tâche demandée (Brown et al., 2020). L'exemple d'instruction évoqué précédemment est alors considéré comme une instruction à zéro exemple, car nous ne donnons aucun exemple de traduction de l'anglais vers le français. Un exemple d'instruction à quelques exemples pour cette même tâche pourrait ressembler à ceci :

English: I love reading books. French: J'adore lire des livres.

English: She went to the park. French: Elle est allée au parc.

English: We are going to the beach tomorrow. French: Nous allons à la plage demain.

English: The cat was sleeping. French: ----

Nous évaluons Mixtral 8x7B Instruct v0.1 et Claude 3 Opus sur trois instructions différentes : une instruction « de base » (BASE), qui donne une simple description de la tâche (rendre la phrase inclusive en remplaçant les noms au masculin générique par leurs noms collectifs équivalents, sans indiquer ces derniers explicitement) ; une instruction « dictionnaire » (DICT), qui demande à faire un remplacement de NM par leurs équivalents collectifs d'après le dictionnaire NM-noms collectifs (au singulier ou au pluriel, selon le nombre de noms) ; et une instruction « correction » (CORR), qui reprend des phrases générées par le système de règles et vise à corriger les possibles erreurs (mauvaise correspondance entre le nombre et le genre des verbes et adjectifs avec le NM, par exemple). Les instructions données sont répertoriées dans le tableau 3.7. L'objectif est de voir la potentialité de ce modèle d'instructions combiné avec les algorithmes de conversion issus de GeNRe. Toutes les instructions contiennent trois phrases de notre corpus de développement en guise d'exemples : il s'agit donc d'instructions à quelques exemples. Celles-ci sont disponibles en annexe A.

25. Le modèle est disponible sur la plateforme HuggingFace à l'adresse suivante : <https://huggingface.co/mistralai/Mixtral-8x7B-Instruct-v0.1>.

26. <https://www.anthropic.com/>

27. Voir également <https://www.anthropic.com/news/claude-3-family>.

28. <https://docs.mistral.ai/platform/client/>

29. <https://docs.anthropic.com/claude/reference/getting-started-with-the-api>

Type d'instruction	Contenu
BASE	Make this French sentence inclusive by replacing generic masculine nouns with their French collective noun equivalents. Generate the final sentence only without any comments nor notes. {EXEMPLES} {PHRASE ORIGINALE} →
DICT-SG	Make this French sentence inclusive by replacing generic masculine noun {NM} with its respective French collective noun equivalent {NCOLL}. Generate the final sentence only without any comments nor notes. {EXEMPLES} {PHRASE ORIGINALE} →
DICT-PL	Make this French sentence inclusive by replacing generic masculine nouns {NM1, NM2, ... } with their respective French collective noun equivalents {NCOLL1, NCOLL2, ... }. Generate the final sentence only without any comments nor notes. {EXEMPLES} {PHRASE ORIGINALE} →
CORR	Correct grammar in this French sentence. Generate the final sentence only without any comments nor notes. {EXEMPLES} {PHRASE ORIGINALE} →

TABLEAU 3.7 – Instructions données aux modèles d'instructions

Nous faisons le choix de rédiger nos instructions en anglais, car nous pensons que nous pourrions éventuellement obtenir de meilleures performances ainsi, le modèle ayant été majoritairement entraîné sur des données en anglais.

3.6.1 Mixtral 8x7B Instruct v0.1

Avant de procéder à notre expérience, nous effectuons des essais préliminaires sur le modèle Mixtral 8x7B Instruct v0.1. Ces essais montrent que, malgré l'instruction donnée de générer uniquement la phrase finale, le modèle génère du contenu supplémentaire non pertinent, signe d'un type d'« hallucination »³⁰ (Kaddour et al., 2023 ; T. Liu et al., 2022 ; Zhang et al., 2023). Nous en donnons divers exemples en annexe A.

Bien qu'il soit possible de modifier la « température » du modèle afin de le rendre davantage déterministe et ainsi réduire la fréquence de ce phénomène, nos essais de réduction de la température n'ont pas montré de différence significative, y compris avec une température égale à 0 (le minimum autorisé). Une solution alternative a consisté à indiquer un nombre de tokens maximum à générer de manière à ce que le modèle ne génère pas d'autre contenu en dehors

30. Le terme « hallucination » est utilisé sous plusieurs acceptions dans la littérature (nous renvoyons notamment à Ji et al. (2023a)). Ici, nous ne faisons pas référence à des hallucinations dans des contextes de questions-réponses qui mèneraient à la génération d'informations factuellement erronées, mais plutôt à des hallucinations donnant lieu à la génération de textes sans rapport aucun ni avec l'instruction, ni avec le texte précédemment généré (par exemple, la présence de code informatique). Nous renvoyons aux exemples en annexe A pour davantage de considérations.

	Wikipédia			Europarl			Moy.		
	WER	BLEU	Sim. cos.	WER	BLEU	Sim. cos.	WER	BLEU	Sim. cos.
Réf. (baseline)	13,35 %	79,634	0,9115	13,344 %	81,465	0,917	13,347 %	80,549	0,914
Mixtral 8x7B-BASE	31,026 %	58,617	0,823	37,79 %	65,909	0,86	34,408 %	62,263	0,8415
Mixtral 8x7B-BASE-C	25,421 %	65,537	0,857	31,638 %	73,69	0,892	28,5295 %	69,6135	0,8745
Mixtral 8x7B-DICT	22,661 %	69,736	0,8845	31,226 %	73,991	0,907	26,9435 %	71,8635	0,8957
Mixtral 8x7B-DICT-C	12,981 %	81,74	0,941	24,376 %	82,603	0,944	18,6785 %	82,1715	0,9425
Mixtral 8x7B-CORR	38,383 %	48,56	0,81	32,431 %	56,493	0,8485	35,407 %	52,5265	0,82925
Mixtral 8x7B-CORR-C	18,996 %	71,993	0,91	21,335 %	71,44	0,913	20,1655 %	71,7165	0,9115

TABLEAU 3.8 – Résultats du modèle d'instructions Mixtral 8x7B par corpus

de la phrase transformée. Nous avons tout d'abord utilisé la longueur de la phrase en entrée en termes de tokens pour définir le nombre de tokens maximum. Toutefois, étant donné que le processus de tokenisation des grands modèles de langue repose sur le découpage en sous-tokens (un seul et même mot peut parfois être découpé en deux, trois tokens ou plus), il n'était pas possible d'utiliser uniquement le nombre de mots séparés par une espace (tokens « classiques ») de la phrase originale pour calculer le nombre de tokens maximum : les phrases générées étaient trop courtes en comparaison avec les phrases d'entrée. Pour contrer ce problème, nous avons défini empiriquement un multiplicateur MAX_TOKENS_MULT (égal à 2,3) s'appliquant au nombre de tokens classiques de cette manière :

```
max_tokens=int(round(len(sent.split())*MAX_TOKENS_MULT, 0))
```

Cela nous a permis de réduire drastiquement le contenu superflu généré par le modèle d'instructions, tout en permettant la génération, dans la grande majorité des cas, de la phrase de sortie dans son entièreté.

3.6.1.1 Évaluation, résultats et discussion

Nos résultats sont reproductibles à l'identique en utilisant les hyperparamètres suivants :

```
random_seed="29751",
max_tokens=int(round(len(sent.split())*MAX_TOKENS_MULT, 0)),
temperature=0.4
```

Malgré la restriction appliquée au nombre de tokens maximal généré, du contenu superflu reste présent dans une partie des textes de sortie. Nous réalisons donc une deuxième évaluation après un nettoyage manuel de ce type de contenu : les résultats de la deuxième évaluation sont suivis du symbole « -C ». Nos résultats sont répertoriés dans le tableau 3.8.

Une comparaison des résultats entre la première évaluation (sans nettoyage) et la seconde évaluation (avec nettoyage) montre une différence en termes de WER pouvant aller de 5,605 % (BASE, corpus Wikipédia) à 19,387 % (CORR, corpus Wikipédia) ; et, en termes de score BLEU, de 7,781 (BASE, corpus Europarl) à 23,061 (CORR, corpus Wikipédia). Ces différences montrent l'importance de la tâche de nettoyage réalisée à posteriori.

Les résultats de la seconde évaluation montrent généralement une meilleure performance avec l'instruction « DICT » sur les deux corpus ; il s'agit en effet de l'instruction la plus explicite, étant donné que nous précisons formellement les NM à remplacer dans la phrase, et les

équivalents qui devraient les remplacer. Une analyse plus approfondie des phrases générées met en lumière plusieurs points intéressants : tout d'abord, le modèle est capable de modifier certains NM non spécifiés dans l'instruction avec l'équivalent collectif : c'est le cas pour l'exemple 85, où le nom collectif « parentèle », associé au NM « parents », n'apparaît pas dans notre dictionnaire. En outre, le modèle reformule souvent certaines phrases de manière à modifier le style ou à les clarifier (voir exemples 86 et 87), ce qui a évidemment une incidence lors de la comparaison avec la phrase manuelle.

Nous remarquons également certaines occurrences d'extrapolation des instructions : par exemple, dans l'exemple 84, le modèle a féminisé le nom « lieutenant » et ses dépendances. Il est possible que la présence de l'adjectif « inclusive » dans l'instruction ait provoqué ce type de changement.

- (84) a. Ses principaux lieutenants seront tous des Apôtres et son armée soit de l'humanité, soit des démons. [phrase manuelle]
- b. Ses **principales lieutenantantes seront toutes** l'armée des Apôtres et **ses soldats tant** de l'humanité **que** des démons. [Mixtral-DICT]
- (85) a. De plus, la demande d'emploi étant supérieure à l'offre, les employeurs peuvent davantage imposer leurs conditions et choisir d'embaucher de l'enfance (moins payée et plus docile) que des adultes : dans de nombreuses situations, l'enfance travaille alors que les parents sont au chômage, entrant ainsi en « concurrence volontaire ». [phrase manuelle]
- b. De plus, la demande d'emploi étant supérieure à l'offre, les employeurs peuvent davantage imposer leurs conditions et choisir d'embaucher l'enfance (moins payée et plus docile) que **l'adulte** : dans de nombreuses situations, l'enfance travaille alors que **la parentèle est** au chômage, entrant ainsi en « concurrence volontaire ». [Mixtral-DICT]
- (86) a. Depuis des années il affirme que la DPJ doit être fermée et remplacée par la police pour réaliser des enquêtes auprès de l'enfance abusée. [phrase manuelle]
- b. "Depuis des années, il affirme que la DPJ doit être fermée et remplacée par la police pour réaliser des enquêtes auprès de l'enfance **maltraitée**." [Mixtral-DICT]
- (87) a. En outre, le service peut être garanti à la citoyenneté s'il est considéré conforme aux lignes directrices définies par la politique linguistique régionale. [phrase manuelle]
- b. **De plus**, le service peut être garanti **envers** la citoyenneté s'il est considéré conforme aux lignes directrices définies par la politique linguistique régionale. [Mixtral-DICT]

La plus grande liberté accordée par les instructions « BASE » et « CORR » mène à davantage de différences par rapport aux phrases manuelles, d'où la différence de résultats. Ainsi, nous retrouvons à une fréquence bien plus élevée des changements stylistiques par rapport aux phrases générées avec l'instruction « DICT ». Pour les phrases générées avec l'instruction « BASE », la neutralisation est parfois respectée, avec l'utilisation de noms collectifs bien qu'ils

ne soient pas identiques à ceux que l'on trouve dans la phrase manuelle (exemple 88). On trouve également un certain nombre de phrases qui ne sont pas neutralisées mais inclusives, avec l'utilisation de points médians (exemples 89 et 90) ou de tirets (exemple 91).

- (88) a. Mais l'armée protestante, toujours agressive, restait à la charge des habitants et constituait une lourde charge. [phrase manuelle]
b. Mais **la troupe** résistante, toujours agressive, restait à la charge des habitants et constituait une lourde charge. [Mixtral-BASE]
- (89) a. La patientèle peut consentir à faire don d'échantillons et des données sous forme « codée » (lien avec le dossier médical du donneur est possible seulement avec un code), de manière « anonymisée » (lien avec le donneur définitivement supprimé) ou choisir de ne pas participer à la recherche scientifique. [phrase manuelle]
b. Les **patient-es** peuvent consentir à faire don d'échantillons et de données sous forme « codée », de manière « anonymisée » ou choisir de ne pas participer à la recherche scientifique. [Mixtral-BASE]
- (90) a. Les noms de l'intelligentsia ukrainienne assassinée sont gravées dans un mur en acier, dont la rouille donne le sentiment que le monument saigne indéfiniment. [phrase manuelle]
b. Les noms **des intellectuel-les ukrainien-nes assassiné-es** sont gravées dans un mur en acier, dont la rouille donne le sentiment que le monument saigne indéfiniment. [Mixtral-BASE]
- (91) a. Le processus officiel est le même que pour le choix du candidat à la présidence avec un vote de la délégation lors de cette convention. [phrase manuelle]
b. Le processus officiel est le même que pour le choix **d'un-e candidat-e** à la présidence avec un vote de la délégation lors de cette convention. [Mixtral-BASE]

3.6.2 Claude 3 Opus

Nos essais préliminaires pour ce modèle ne montrent aucune occurrence du même type d'hallucination que celui rencontré avec Mixtral 8x7B Instruct v0.1. Cela peut éventuellement s'expliquer par les meilleures performances générales de Claude 3 Opus ou par les données d'entraînement du modèle ; les recherches sur les hallucinations restent toutefois à un stade primitif pour le moment, et il n'existe à notre connaissance pas d'élément catégorique expliquant pourquoi certains modèles ou certaines instructions sont plus prompts à générer ce type de contenu. L'API de Claude 3 Opus permet, contrairement à celle de Mixtral 8x7B Instruct v0.1, de définir comme hyperparamètre le début du message que générera le modèle ; nous profitons de cette possibilité pour indiquer, comme début de message, « *Here is the output sentence :* » (« Voici la phrase de sortie : ») afin d'assurer l'absence de toute hallucination.

	Wikipédia			Europarl			Moy.		
	WER	BLEU	Sim. cos.	WER	BLEU	Sim. cos.	WER	BLEU	Sim. cos.
Réf. (baseline)	13,35 %	79,634	0,9115	13,344 %	81,465	0,917	13,347 %	80,549	0,914
Claude 3 Opus-BASE	14,02 %	80,087	0,91	10,304 %	85,88	0,94	12,162 %	82,9835	0,925
Claude 3 Opus-DICT	4,052 %	93,845	0,97	3,459 %	93,434	0,98	3,7555 %	93,6395	0,975
Claude 3 Opus-CORR	11,413 %	84,648	0,95	8,929 %	85,615	0,95	10,171 %	85,1315	0,95

TABLEAU 3.9 – Résultats du modèle d'instructions Claude 3 Opus par corpus

3.6.2.1 Évaluation, résultats et discussion

Contrairement à l'API de Mixtral 8x7B Instruct v0.1, celle de Claude 3 Opus ne propose pas la possibilité d'indiquer une « seed » pour reproduire nos résultats à l'identique. Toutefois, nous indiquons tout de même les hyperparamètres utilisés ci-dessous :

```
model="claude-3-opus-20240229",
temperature=0,
messages=[
  {"role": "user", "content": f"{message}"}, # instruction
  {"role": "assistant", "content": "Here is the output sentence:"}
]
```

Nos résultats sont répertoriés dans le tableau 3.9.

Nous obtenons de bien meilleurs résultats pour les deux corpus en comparaison avec ceux de Mixtral 8x7B Instruct v0.1, et ce peu importe le type d'instruction. En nous basant sur la version nettoyée des résultats de Mixtral 8x7B Instruct v0.1, en moyenne, le WER diminue de 16,387 points (+13,349 de score BLEU et +0,05 de similarité cosinus) pour l'instruction « BASE » ; de 14,923 points (+11,468 de score BLEU et +0,033 de similarité cosinus) pour l'instruction « DICT » ; et de 9,995 points (+13,415 de score BLEU et +0,0375 de similarité cosinus) pour l'instruction « CORR ».

Les résultats de l'instruction « DICT » se rapprochent fortement de ceux de notre système de règles. Pour rappel, nous obtenons en moyenne 3,402 % de WER, 93,428 de score BLEU et 0,982 de similarité cosinus avec notre système de règles. Claude 3 Opus-DICT obtient 3,755 % de WER (+0,353), 93,639 de score BLEU (+0,211) et 0,975 de similarité cosinus (-0,009). En termes de score BLEU, Claude 3 Opus-DICT obtient de meilleurs résultats que notre système de règles.

Nous notons des modifications similaires à celles effectuées par Mixtral 8x7B Instruct v0.1. Par exemple, l'exemple 92 (instruction « BASE ») montre le remplacement du nom masculin pluriel « marchands » par le nom « négoce », même si, d'après nos recherches, il ne semble pas que celui-ci soit utilisé comme nom collectif pour désigner un groupe d'êtres humains.

Pour la même instruction « BASE », nous remarquons également des modifications intéressantes : dans l'exemple 93, le nom masculin pluriel « savants » est remplacé par « corps savant », utilisant le nom collectif « corps » apposé de l'adjectif correspondant, mettant en valeur la capacité du modèle à utiliser des formulations collectives plus complexes.

- (92) a. En 1531, pour payer sa dette aux marchands de Lübeck, le roi Gustav Vasa réquisitionne les cloches des églises catholiques, provoquant un soulèvement de la paysannerie où plusieurs furent tués. [phrase manuelle]
- b. En 1531, pour payer sa dette **au négoce** de Lübeck, le roi Gustav Vasa réquisitionne les cloches des églises catholiques, provoquant un soulèvement de la paysannerie où plusieurs furent tués. [Claude-BASE]
- (93) a. Dans son Commentaire aux épidémies, il présente son professorat alexandrin comme des savants seulement en parole, qui sont complètement incapables de reconnaître une maladie. [phrase manuelle]
- b. Dans son Commentaire aux épidémies, il présente son professorat alexandrin comme **un corps savant** seulement en parole, qui est complètement incapable de reconnaître une maladie. [Claude-BASE]

3.7 Résultats et discussion généraux

Nous répertorions l'ensemble des résultats que nous avons obtenus dans le tableau 3.10. Pour le modèle Mistral 8x7B, nous indiquons uniquement les résultats de la version nettoyée (-C). En définitive, notre système à base de règles obtient les meilleurs résultats, à la fois en termes de WER et de similarité cosinus (respectivement 3,403 % et 0,982 en moyenne). La version « DICT » de Claude 3 Opus montre toutefois des résultats très encourageants, puisqu'ils sont très similaires aux résultats de notre système de règles dans toutes les mesures ; pour la mesure BLEU plus particulièrement, c'est lui qui obtient la meilleure moyenne (93,639, soit un écart de 0,211 point par rapport au score du système à base de règles).

Ces résultats confirment le gain de performances apporté par l'utilisation de modèles d'instructions, ainsi que l'importance qu'il est nécessaire de donner à la rédaction d'instructions ; en indiquant clairement au modèle quels éléments doivent être modifiés dans la phrase en usant du dictionnaire créé précédemment, il est ainsi possible d'obtenir des résultats de même ordre que ceux du système de règles, nécessitant alors un travail bien moins coûteux.

3.8 Limites et perspectives

Comme nous l'avons indiqué de manière extensive dans le chapitre 2.2.2.2, les noms collectifs imposent un certain nombre de contraintes syntaxiques et sémantiques, mais bénéficient, sous d'autres aspects, d'une relative liberté par l'usage : selon les noms collectifs, il est par exemple possible d'y faire référence en utilisant des pronoms pluriels, quand bien même le nom collectif est singulier. Ce genre de libertés peut rendre la constitution de règles et de phrases de référence difficile pour la tâche qui est la nôtre. En outre, bien que les noms collectifs constituent un excellent moyen de neutraliser le genre, tous les noms ne peuvent pas se voir remplacés par un nom collectif. Il serait donc intéressant de s'intéresser non seulement à la neutralisation automatique du genre à un niveau plus général, c'est-à-dire au-delà de l'utilisation des noms

		Wikipédia	Europarl	Moy.
WER	Référence	13,35 %	13,344 %	13,347 %
	GeNRe-RBS	3,358 %	3,448 %	3,403 %
	GeNRe-FT-T5 ○	6,095 %	4,128 %	5,111 %
	GeNRe-FT-M2M100 ○	6,679 %	4,116 %	5,397 %
	Mixtral 8x7B-BASE-C △	25,421 %	31,638 %	28,529 %
	Mixtral 8x7B-DICT-C △	12,981 %	24,376 %	18,678 %
	Mixtral 8x7B-CORR-C △	18,996 %	21,335 %	20,166 %
	Claude 3 Opus-BASE △	14,02 %	10,304 %	12,162 %
	Claude 3 Opus-DICT △	4,052 %	3,459 %	3,755 %
	Claude 3 Opus-CORR △	11,413 %	8,929 %	10,171 %
BLEU	Référence	79,634	81,465	80,549
	GeNRe-RBS	93,024	93,833	93,428
	GeNRe-FT-T5 ○	88,144	93,211	90,678
	GeNRe-FT-M2M100 ○	87,09	93,298	90,194
	Mixtral 8x7B-BASE-C △	65,537	73,69	69,614
	Mixtral 8x7B-DICT-C △	81,74	82,603	82,171
	Mixtral 8x7B-CORR-C △	71,993	71,44	71,716
	Claude 3 Opus-BASE △	80,087	85,88	82,983
	Claude 3 Opus-DICT △	93,845	93,434	93,639
	Claude 3 Opus-CORR △	84,648	85,615	85,131
Sim. cosinus	Référence	0,9115	0,917	0,914
	GeNRe-RBS	0,979	0,984	0,982
	GeNRe-FT-T5 ○	0,956	0,979	0,968
	GeNRe-FT-M2M100 ○	0,954	0,98	0,967
	Mixtral 8x7B-BASE-C △	0,857	0,892	0,875
	Mixtral 8x7B-DICT-C △	0,941	0,944	0,9425
	Mixtral 8x7B-CORR-C △	0,91	0,913	0,911
	Claude 3 Opus-BASE △	0,91	0,94	0,925
	Claude 3 Opus-DICT △	0,97	0,98	0,975
	Claude 3 Opus-CORR △	0,95	0,95	0,95

○ modèle affiné
 △ modèle d'instructions

TABEAU 3.10 – Résultats globaux par corpus et par mesure

collectifs, mais également de s'intéresser aux autres formes de réécriture du genre, auxquelles s'intéressent notamment Alhafni et al. (2022b) et Pomerence (2022), notamment le passage d'un genre A à un genre B, ou bien l'inclusion de tous les genres.

D'un point de vue relevant purement du traitement automatique des langues, les recherches concernant le potentiel des modèles d'instructions vis-à-vis de la tâche de réécriture du genre sont, nous l'avons dit, très limitées à ce jour. Ce travail constitue donc une première étape vers l'utilisation de modèles d'instructions pour cette tâche. Celui-ci mériterait toutefois d'être approfondi au cours de futurs travaux : nous pensons en effet qu'il existe de nombreux moyens de potentiellement améliorer les résultats que nous avons obtenus avec les modèles d'instructions. Premièrement, il serait pertinent de tester un plus grand nombre d'instructions différentes afin

de voir plus en détail les différences entre chacune en termes de sortie par le modèle, et quels mots ou groupes de mots le modèle prend le plus en considération lors de la génération du texte en sortie. Dans cette même idée, des techniques plus avancées de « prompt engineering » pourraient être utilisées. Deuxièmement, nous n'avons pas mis en valeur la modification de l'instruction système (« system prompt »), qui permet de définir le comportement du modèle d'instructions. Comme pour les instructions classiques, il serait intéressant d'en essayer un certain nombre et de les comparer afin de connaître leur influence sur les résultats. Troisièmement, une comparaison plus large des modèles d'instructions existants serait également pertinente pour déterminer lesquels s'adaptent le mieux à ce type de tâche.

Il nous semble également nécessaire de développer un système d'évaluation plus robuste pour la tâche de réécriture du genre, en fonction de la sous-tâche réalisée. En effet, comme c'est le cas pour la neutralisation à travers l'utilisation de noms collectifs, il existe de nombreux cas où il est possible de remplacer un nom par plusieurs alternatives ; ou, plusieurs moyens de modifier la syntaxe d'une phrase de manière à ce qu'elle s'accommode au remplacement. Dans ce genre de cas, des mesures d'évaluation comme le WER, le score BLEU ou la similarité cosinus peuvent poser des limites, car les changements évoqués plus tôt ne donneront pas forcément lieu à une correspondance exacte avec la phrase de référence. Bien que, parmi les mesures évoquées, le score BLEU prenne partiellement en compte ce genre de cas, il ne s'agit pas forcément de la mesure la plus viable pour une telle évaluation (Alihosseini et al., 2019). Ainsi, nous pensons qu'il pourrait être pertinent de se pencher sur des mesures d'évaluation conçues spécifiquement pour la génération de texte, telles que BLEURT (Sellam et al., 2020) ou SESCORE2 (W. Xu et al., 2023), afin de les adapter à la tâche de réécriture du genre. Cela nous paraît d'autant plus crucial étant donné l'essor et la démocratisation extrêmement rapides des grands modèles de langue pour tout un ensemble de tâches : il semble en effet, au vu des développements actuels, qu'un certain nombre de systèmes, se reposant alors uniquement sur un ensemble de règles, seront prochainement remplacés par l'utilisation de ce type de modèles. Si, en outre, un objectif futur plus ambitieux consiste à proposer un système permettant une neutralisation automatique au-delà des noms collectifs, comme explicité auparavant, alors ce type de modèle, qui adapte ses réponses en fonction de l'instruction (et dont les réponses peuvent être également adaptées avec un affinage) semble être un bon candidat.

Nous mettons à disposition le code utilisé pour le système de règles de GeNRe sur GitHub³¹. Il est ainsi possible d'utiliser le système avec n'importe quelle phrase depuis la ligne de commandes. Les modèles affinés de grands modèles de langue GeNRe-FT-T5 et GeNRe-FT-M2M100 sont également disponibles et utilisables en ligne^{32 33}.

31. <https://github.com/spidersouris/GeNRe>

32. <https://huggingface.co/spidersouris/genre-t5-small-60k>

33. https://huggingface.co/spidersouris/genre-m2m100_418M

4

Conclusion

Ce mémoire vise à mettre au jour une ressource permettant la neutralisation automatique du genre en français à travers l'utilisation de noms collectifs. Après avoir exposé les problèmes inhérents à l'utilisation du masculin générique dans la langue et ses effets en termes de biais, et après nous être intéressé, à travers une perspective historique, à son essor dans la langue française du fait des prescriptivistes, nous avons mené une analyse étendue des noms collectifs et avons montré comment ceux-ci peuvent être utilisés pour la neutralisation du genre en français. Par la suite, nous nous sommes concentrés plus spécifiquement sur la tâche de réécriture du genre dans le domaine du traitement automatique des langues, mentionnant les différents types de systèmes existants à ce jour, tout en proposant une définition plus large pour cette tâche de manière à inclure les différentes transformations pouvant être effectuées au niveau du genre dans la langue.

La seconde partie de ce mémoire a eu pour objectif de présenter les rouages du système de règles que nous avons développé afin de remplacer automatiquement des noms de membres au masculin par leurs équivalents collectifs. Nous notons en particulier les contributions suivantes que nous réalisons vis-à-vis de la tâche de réécriture du genre :

- constitution d'un dictionnaire de paires de noms de membres-noms collectifs, contenant également leurs propriétés syntaxiques ;

- développement du premier système de réécriture de genre neutre à base de règles pour le français, utilisant spécifiquement les noms collectifs ;
- développement d'un composant permettant la détection automatique de dépendances des noms au masculin générique ;
- constitution d'un jeu de données de phrases non inclusives et neutralisées en genre, comprenant au total plus de 390 000 paires de phrases ;
- affinage de deux modèles de langue sur une partie de ce jeu de données ;
- travail préliminaire sur les modèles d'instructions et leur usabilité sur cette tâche, notamment en les combinant avec les ressources et approches mentionnées ci-devant.

Nous espérons que ce travail permettra de poser les fondements de travaux ultérieurs dans ce même domaine pour le français. Malgré les récents développements dans le domaine du traitement automatique des langues, les biais, et plus particulièrement les biais de genre, restent une problématique réelle dans les données. En outre, nous espérons également que des travaux similaires verront le jour pour d'autres langues qui ne disposent pas encore de tel système, y compris les langues peu dotées qui, en raison du manque de données, sont désavantagées dans le développement de systèmes pour des tâches de ce type.

Références bibliographiques

Références

- Abbou, J. (2019). La langue est-elle toujours un lieu de lutte féministe ? De la contrefaçon sémiotique à la libéralisation. *Recherches féministes*, 32(2) : 235-258. Récupéré le 19 janvier 2023, à partir de <https://hal.science/hal-03041815/document>
- Abbou, J., Arnold, A., Candea, M., & Marignier, N. (2018). Qui a peur de l'écriture inclusive ? Entre délire eschatologique et peur d'émasculatation. *Semen*, (44).
- Abbou, J., & Candea, M. (2021). Féminisation. *Langage et société* : 141-145.
- Abeillé, A., An, A., & Shiraïshi, A. (2018). L'accord de proximité du déterminant en français. *Discours*, (22).
- Académie française. (2017). Déclaration de l'Académie française sur l'écriture dite "inclusive". Récupéré le 3 février 2023, à partir de <https://www.academie-francaise.fr/actualites/declaration-de-lacademie-francaise-sur-lecriture-dite-inclusive>
- Alhafni, B., Habash, N., & Bouamor, H. (2021). *The Arabic Parallel Gender Corpus 2.0 : Extensions and Analyses*. arXiv : 2110.09216 [cs]. Récupéré le 24 février 2023, à partir de <http://arxiv.org/abs/2110.09216>
- Alhafni, B., Habash, N., & Bouamor, H. (2022a). User-Centric Gender Rewriting. *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics : Human Language Technologies* : 618-631.
- Alhafni, B., Habash, N., Bouamor, H., Obeid, O., Alrowili, S., Alzeer, D., Alshantqiti, K. M., ElBakry, A., EINokrashy, M., Gabr, M., Issam, A., Qaddoumi, A., Vijay-Shanker, K., & Zyate, M. (2022b). *The Shared Task on Gender Rewriting*. arXiv : 2210.12410 [cs].
- Alihosseini, D., Montahaei, E., & Soleymani Baghshah, M. (2019). Jointly Measuring Diversity and Quality in Text Generation Models. *Proceedings of the Workshop on Methods for Optimizing and Evaluating Neural Language Generation* : 90-98.
- Alrowili, S., & Vijay-Shanker, K. (2022). Generative Approach for Gender Rewriting Task with ArabicT5. *Proceedings of the The Seventh Arabic Natural Language Processing Workshop (WANLP)* : 491-495. Récupéré le 21 février 2023, à partir de <https://aclanthology.org/2022.wanlp-1.55/>
- Amatriain, X. (2024). *Prompt Design and Engineering : Introduction and Advanced Methods*. arXiv : 2401.14423 [cs]. Récupéré le 18 mars 2024, à partir de <http://arxiv.org/abs/2401.14423>

- Anthropic. (2024). *The Claude 3 Model Family : Opus, Sonnet, Haiku*. Récupéré le 18 mars 2024, à partir de https://www-cdn.anthropic.com/de8ba9b01c9ab7cbabf5c33b80b7bbc618857627/Model_Card_Claude_3.pdf
- Arigne, V. (2011). La figure du tout intégré et les noms discrets collectifs. *Anglophonia* : 59-99.
- Arora, S., Narayan, A., Chen, M. F., Orr, L., Guha, N., Bhatia, K., Chami, I., Sala, F., & Ré, C. (2022). *Ask Me Anything : A simple strategy for prompting language models*. arXiv : 2210.02441 [cs]. Récupéré le 18 mars 2024, à partir de <http://arxiv.org/abs/2210.02441>
- Benninger, C. (2001). Une meute de loups / une brassée de questions : collection, quantification et métaphore. *Langue française*, 129(1) : 21-34.
- Berlin, B., & Kay, P. (1969). *Basic Color Terms : Their Universality and Evolution*. University of California Press.
- Bird, S., Loper, E., & Klein, E. (2009). *Natural Language Processing with Python*. O'Reilly Media Inc.
- Björk, I. (2008). *Relativizing linguistic relativity. Investigating underlying assumptions about language in the neo-Whorfian literature*. Uppsala Universitet. Récupéré le 25 mars 2023, à partir de <https://www.diva-portal.org/smash/get/diva2:171948/FULLTEXT01.pdf>
- Boughanmi, J. (2016). Sur les collectifs humains (F. Neveu, G. Bergounioux, M.-H. Côté, J.-M. Fournier, L. Hriba & S. Prévost, Éd.). *SHS Web of Conferences*, 27.
- Braun, F., Sczesny, S., & Stahlberg, D. (2005). Cognitive Effects of Masculine Generics in German : An Overview of Empirical Findings. *Communications*, 30(1) : 1-21.
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ... Amodei, D. (2020). *Language Models are Few-Shot Learners*. arXiv : 2005.14165 [cs]. Récupéré le 18 mars 2024, à partir de <http://arxiv.org/abs/2005.14165>
- Burr, E. (2003). Gender and language politics in France. *Gender Across Languages*, 3 : 119-139.
- Cameron, D., McAlinden, F., & O'Leary, K. (1998). Lakoff in context : the social and linguistic functions of tag questions. In *Women's Language : Critical Approaches* (p. 75-93). Récupéré le 19 janvier 2023, à partir de <https://web.stanford.edu/~eckert/PDF/CameronTags.pdf>
- Candito, M., & Seddah, D. (2012). Le corpus Sequoia : annotation syntaxique et exploitation pour l'adaptation d'analyseur par pont lexical. *Proceedings of the Joint Conference JEP-TALN-RECITAL 2012, volume 2 : TALN* : 321-334. <https://aclanthology.org/F12-2024/>
- Chervel, A. (2019). L'accord de proximité et la grammaire. In *Le féminin et le masculin dans la langue. L'écriture inclusive en question* (p. 95-114). ESF.
- Chuttarsing, A. (2021). *Inflecteur* (Version 1). <https://github.com/Achuttarsing/inflecteur>
- Corbett, G. G. (1991). *Gender*. Cambridge University Press.
- Dubois, J., Giacomo, M., Guespin, L., Marcellesi, C., Marcellesi, J.-B., & Mével, J.-P. (2001). *Dictionnaire de linguistique*. Larousse.

- Eckert, P., & McConnell-Ginet, S. (2003). *Language and Gender*. Cambridge University Press. Récupéré le 19 mars 2023, à partir de <https://www.cambridge.org/core/books/language-and-gender/262C789CFAA87E8CCA20B21E5CDF3D6C>
- JO n° 0272 (2017). Récupéré le 3 février 2023, à partir de <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000036068906>
- Elmiger, D. (2011). Féminisation de la langue française : une brève histoire des positions politiques et du positionnement linguistique. In *Langage, genre et sexualité* (Alexandre Duchêne et Claudine Moïse, p. 71-89). Nota bene. Récupéré le 19 janvier 2023, à partir de <https://www.irdp.ch/data/secure/1153/document/feminisation-de-la-langue-francaise-1153.pdf>
- Fan, A., Bhosale, S., Schwenk, H., Ma, Z., El-Kishky, A., Goyal, S., Baines, M., Celebi, O., Wenzek, G., Chaudhary, V., Goyal, N., Birch, T., Liptchinsky, V., Edunov, S., Grave, E., Auli, M., & Joulin, A. (2020). *Beyond English-Centric Multilingual Machine Translation*. arXiv : 2010.11125 [cs]. Récupéré le 18 mars 2024, à partir de <http://arxiv.org/abs/2010.11125>
- Fishman, P. M. (1980). Conversational Insecurity. In *Language* (p. 127-132). Elsevier. Récupéré le 19 janvier 2023, à partir de <https://linkinghub.elsevier.com/retrieve/pii/B9780080246963500247>
- Flaux, N. (1999). À propos des noms collectifs. *Revue de linguistique romane*, (63) : 471-502.
- Gardelle, L. (2015). Let her rain, she's snowing pretty good : The use of feminine pronouns with weather verbs in colloquial English. *Folia Linguistica*, 49(2).
- Gardelle, L. (2016). Five crew, how many clergy : pourquoi certains noms collectifs peuvent-ils servir à nommer des membres ? *Anglophonia*, (22).
- Gomila, A. (2015). Language and Thought : The Neo-Whorfian Hypothesis. In *International Encyclopedia of the Social & Behavioral Sciences* (Second edition, p. 293-299). Elsevier.
- graelo. (2023). *Graelo/wikipedia dataset*. graelo/wikipedia - Datasets at Hugging Face. <https://huggingface.co/datasets/graelo/wikipedia>
- Gross, G. (2009). Sur le statut syntaxique des substantifs humains. In *Des topoï à la théorie des stéréotypes en passant par la polyphonie et l'argumentation dans la langue. Hommages à Jean-Claude Anscombre* (p. 27-41). Université de Savoie. Récupéré le 23 novembre 2022, à partir de <https://halshs.archives-ouvertes.fr/halshs-00411322>
- Gross, G. (2011). Classification sémantique des collectifs humains. *Cahiers de lexicologie*, 1(98) : 65-81. <https://halshs.archives-ouvertes.fr/halshs-00627283>
- Habash, N., Bouamor, H., & Chung, C. (2019). Automatic Gender Identification and Reinflection in Arabic. *Proceedings of the First Workshop on Gender Bias in Natural Language Processing* : 155-165.
- Haddad, R. (2016). Manuel d'écriture inclusive. *Mots-Clés*. Récupéré le 19 janvier 2023, à partir de https://www.univ-tlse3.fr/medias/fichier/manuel-decriture_1482308453426-pdf

- Haque, S., & Wong, B. (2003). Do They Really Use More Tag Questions? *3L ; Journal of Language, Linguistics and Literature* : 40-62. <http://journalarticle.ukm.my/3103/1/1.pdf>
- Harris Interactive. (2017). *L'écriture inclusive : la population française connaît-elle l'écriture inclusive ? Quelle opinion en a-t-elle ?* Récupéré le 25 mars 2023, à partir de https://harris-interactive.fr/opinion_polls/lecriture-inclusive/
- He, Z., Majumder, B. P., & McAuley, J. (2021). *Detect and Perturb : Neutral Rewriting of Biased and Sensitive Text via Gradient-based Decoding*. arXiv : 2109.11708 [cs].
- Hegarty, P., & Lemieux, A. (2011). Who is the (second) graphed sex and why ? The meaning of order in graphs of gender differences. In *Spatial Aspects of Social Thought* (p. 325-349). Mouton de Gruyter.
- Honnibal, M., Montani, I., Van Landeghem, S., & Boyd, A. (2020). *spaCy : Industrial-strength Natural Language Processing in Python* (Version 3.7). <https://github.com/explosion/spaCy>
- Jacobson, M. B., & Insko, W. R. (1985). Use of nonsexist pronouns as a function of one's feminist orientation. *Sex Roles*, 13(1-2) : 1-7.
- Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y., Chen, D., Chan, H. S., Dai, W., Madotto, A., & Fung, P. (2023a). Survey of Hallucination in Natural Language Generation. *ACM Computing Surveys*, 55(12) : 1-38.
- Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y., Chen, D., Chan, H. S., Dai, W., Madotto, A., & Fung, P. (2023b). Survey of Hallucination in Natural Language Generation. *ACM Computing Surveys*, 55(12) : 1-38.
- Jiang, A. Q., Sablayrolles, A., Roux, A., Mensch, A., Savary, B., Bamford, C., Chaplot, D. S., Casas, D. de las, Hanna, E. B., Bressand, F., Lengyel, G., Bour, G., Lample, G., Lavaud, L. R., Saulnier, L., Lachaux, M.-A., Stock, P., Subramanian, S., Yang, S., ... Sayed, W. E. (2024). *Mixtral of Experts*. arXiv : 2401.04088 [cs]. Récupéré le 18 mars 2024, à partir de <http://arxiv.org/abs/2401.04088>
- Kaddour, J., Harris, J., Mozes, M., Bradley, H., Raileanu, R., & McHardy, R. (2023). *Challenges and Applications of Large Language Models*. arXiv : 2307.10169 [cs]. Récupéré le 18 mars 2024, à partir de <http://arxiv.org/abs/2307.10169>
- Kay, P., & Kempton, W. (1984). What Is the Sapir-Whorf Hypothesis ? *American Anthropologist*, 86(1) : 65-79.
- Kay, P., & McDaniel, C. K. (1978). The Linguistic Significance of the Meanings of Basic Color Terms. *Language*, 54(3) : 610-646.
- Kesebir, S. (2017). Word order denotes relevance differences : The case of conjoined phrases with lexical gender. *Journal of Personality and Social Psychology*, 113(2) : 262-279.
- Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The Sketch Engine : ten years on. *Lexicography*, 1(1) : 7-36.
- Koehn, P. (2005). Europarl : a parallel corpus for statistical machine translation.

- Labov, W. (1966). *The Social Stratification of English in New York City*. Center for Applied Linguistics, University of Michigan.
- Lakoff, R. (1973). Language and Woman's Place. *Language in Society*, 2(1) : 45-80.
- Lambelet, A. (2012). *L'apprentissage du genre grammatical en langue étrangère : à la croisée des approches linguistiques et cognitives*. Université de Fribourg. Fribourg. Récupéré le 19 janvier 2023, à partir de https://www.academia.edu/3768178/Lambelet_A_2012_L_apprentissage_du_genre_grammatical_en_langue_%C3%A9trang%C3%A8re_%C3%A0_la_crois%C3%A9e_des_approches_linguistiques_et_cognitives
- Lammert, M. (2010). *Sémantique et cognition : les noms collectifs*. Droz.
- Lammert, M. (2014). Référence collective massive vs référence plurielle indéfinie : *Langue française*, n° 183(3) : 87-99.
- Lammert, M., & Lecolle, M. (2014). Les noms collectifs en français, une vue d'ensemble. *Cahiers de lexicologie*, (105) : 203-222. Récupéré le 23 novembre 2022, à partir de <https://halshs.archives-ouvertes.fr/halshs-02162824>
- Lecolle, M. (2008). Identité/altérité et noms collectifs humains : le cas de communauté. *Questions de communication*, (13) : 323-342.
- Lecolle, M. (2013). Noms collectifs humains : un point de vue de sémantique lexicale sur l'identité dans le rapport individu/groupe. *¿ Interrogations ?*, (16). Récupéré le 23 novembre 2022, à partir de <http://www.revue-interrogations.org/NOMS-COLLECTIFS-HUMAINS-UN-POINT>
- Lecolle, M. (2016). Noms collectifs humains : nomination et prédication. *Argumentation et analyse du discours*, (17).
- Lecolle, M. (2019). *Les noms collectifs humains en français. Enjeux sémantiques, lexicaux et discursifs* (C. Jacquet-Pfau, Éd.). Lambert-Lucas.
- Lee, J. Y., Mehta, S. V., Wick, M., Tristan, J.-B., & Carbonell, J. (2019). *Gradient-based Inference for Networks with Output Constraints*. arXiv : 1707.08608 [cs]. Récupéré le 27 février 2023, à partir de <http://arxiv.org/abs/1707.08608>
- Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2021). *Pre-train, Prompt, and Predict : A Systematic Survey of Prompting Methods in Natural Language Processing*. arXiv : 2107.13586 [cs]. Récupéré le 18 mars 2024, à partir de <http://arxiv.org/abs/2107.13586>
- Liu, T., Zhang, Y., Brockett, C., Mao, Y., Sui, Z., Chen, W., & Dolan, B. (2022). *A Token-level Reference-free Hallucination Detection Benchmark for Free-form Text Generation*. arXiv : 2104.08704 [cs]. Récupéré le 18 mars 2024, à partir de <http://arxiv.org/abs/2104.08704>
- Martin, L., Muller, B., Ortiz Suárez, P. J., Dupont, Y., Romary, L., De La Clergerie, É., Seddah, D., & Sagot, B. (2020). CamemBERT : a Tasty French Language Model. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* : 7203-7219.

- McMilin, E. (2022). *Selection Induced Collider Bias : A Gender Pronoun Uncertainty Case Study*. arXiv : 2210.00131 [cs]. Récupéré le 27 février 2023, à partir de <http://arxiv.org/abs/2210.00131>
- Moreau, M.-L. (2019). L'accord de proximité dans l'écriture inclusive. Peut-on utiliser n'importe quel argument ? In *Les discours de référence sur la langue française* (Anne Dister, Sophie Piron, p. 351-378). Presses de l'Université Saint-Louis. Récupéré le 19 janvier 2023, à partir de <https://books.openedition.org/pusl/26517>
- Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., & Lowe, R. (2022). *Training language models to follow instructions with human feedback*. arXiv : 2203.02155 [cs]. Récupéré le 18 mars 2024, à partir de <http://arxiv.org/abs/2203.02155>
- Papineni, K., Roukos, S., Ward, T., & Zhu, W.-J. (2002). BLEU : a method for automatic evaluation of machine translation. *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics - ACL '02*.
- Pech, M.-E. (2017). Un manuel scolaire écrit à la sauce féministe [newspaper]. *Le Figaro*. Récupéré le 3 février 2023, à partir de <https://www.lefigaro.fr/actualite-france/2017/09/22/01016-20170922ARTFIG00300-un-manuel-scolaire-ecrit-a-la-sauce-feministe.php>
- Piergentili, A., Fucci, D., Savoldi, B., Bentivogli, L., & Negri, M. (2023). *From Inclusive Language to Gender-Neutral Machine Translation*. arXiv : 2301.10075 [cs]. Récupéré le 20 février 2023, à partir de <http://arxiv.org/abs/2301.10075>
- Pomerence, D. (2022). *INCLUSIFY : A benchmark and a model for gender-inclusive German*. arXiv : 2212.02564 [cs]. Récupéré le 2 janvier 2023, à partir de <http://arxiv.org/abs/2212.02564>
- Pusch, C. D. (2020). Je dirais que la grosse majorité sont là : questions d'accord(s) avec les noms collectifs dans les données OFROM.
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language Models are Unsupervised Multitask Learners.
- Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., & Liu, P. J. (2020). *Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer*. arXiv : 1910.10683 [cs, stat]. Récupéré le 27 février 2023, à partir de <http://arxiv.org/abs/1910.10683>
- Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., & Liu, P. J. (2023). *Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer*. arXiv : 1910.10683 [cs, stat]. Récupéré le 18 mars 2024, à partir de <http://arxiv.org/abs/1910.10683>
- Reitz, K. (2011). *Requests* (Version 2.31.0). <https://github.com/psf/requests>
- Richardson, L. (2004). *BeautifulSoup* (Version 4.4.0). <https://beautiful-soup-4.readthedocs.io/en/latest/>

- Richy, C., & Burnett, H. (2021). Démêler les effets des stéréotypes et le genre grammatical dans le biais masculin : une approche expérimentale. *GLAD!*, (10).
- Rioux, C., Paré, A., London-Nadeau, K., Juster, R.-P., Weedon, S., Levasseur-Puhach, S., Freeman, M., Roos, L. E., & Tomfohr-Madsen, L. M. (2022). Sex and gender terminology : a glossary for gender-inclusive epidemiology. *Journal of Epidemiology and Community Health*, 76(8) : 764-768.
- Salazar, J., Liang, D., Nguyen, T. Q., & Kirchoff, K. (2020). Masked Language Model Scoring. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* : 2699-2712.
- Savoldi, B., Gaido, M., Bentivogli, L., Negri, M., & Turchi, M. (2021). Gender Bias in Machine Translation. *Transactions of the Association for Computational Linguistics*, 9 : 845-874. https://direct.mit.edu/tacl/article/doi/10.1162/tacl_a_00401/106991/Gender-Bias-in-Machine-Translation
- Schudson, Z. C., & Morgenroth, T. (2022). Non-binary gender/sex identities. *Current Opinion in Psychology*, 48.
- Sczesny, S., Formanowicz, M., & Moser, F. (2016). Can Gender-Fair Language Reduce Gender Stereotyping and Discrimination? *Frontiers in Psychology*, 7.
- Sellam, T., Das, D., & Parikh, A. P. (2020). *BLEURT : Learning Robust Metrics for Text Generation*. arXiv : 2004.04696 [cs]. Récupéré le 18 mars 2024, à partir de <http://arxiv.org/abs/2004.04696>
- Si, C., Gan, Z., Yang, Z., Wang, S., Wang, J., Boyd-Graber, J., & Wang, L. (2023). *Prompting GPT-3 To Be Reliable*. arXiv : 2210.09150 [cs]. Récupéré le 18 mars 2024, à partir de <http://arxiv.org/abs/2210.09150>
- Spolsky, B. (1998). *Sociolinguistics*. Oxford University Press.
- Stahlberg, D., Sczesny, S., & Braun, F. (2001). Name Your Favorite Musician : Effects of Masculine Generics and of their Alternatives in German. *Journal of Language and Social Psychology*, 20(4) : 464-469.
- Stanczak, K., & Augenstein, I. (2021). *A Survey on Gender Bias in Natural Language Processing*. arXiv : 2112.14168 [cs]. Récupéré le 25 mars 2023, à partir de <http://arxiv.org/abs/2112.14168>
- Štrkalj Despot, K. (2021). How Language Influences Conceptualization : From Whorfianism to Neo-Whorfianism. *Collegium antropologicum*, 45(4) : 373-380.
- Sun, T., Webster, K., Shah, A., Wang, W. Y., & Johnson, M. (2021). *They, Them, Theirs : Rewriting with Gender-Neutral English*. arXiv : 2102.06788 [cs].
- Sutskever, I., Vinyals, O., & Le, Q. V. (2014a). *Sequence to Sequence Learning with Neural Networks*. arXiv : 1409.3215 [cs]. Récupéré le 27 février 2023, à partir de <http://arxiv.org/abs/1409.3215>

- Sutskever, I., Vinyals, O., & Le, Q. V. (2014b). *Sequence to Sequence Learning with Neural Networks*. arXiv : 1409.3215 [cs]. Récupéré le 18 mars 2024, à partir de <http://arxiv.org/abs/1409.3215>
- Svartengren, T. H. (1927). The Feminine Gender for Inanimate Things in Anglo-American. *American Speech*, 3(2).
- Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., Rodriguez, A., Joulin, A., Grave, E., & Lample, G. (2023). *LLaMA : Open and Efficient Foundation Language Models*. arXiv : 2302.13971 [cs]. Récupéré le 25 mars 2023, à partir de <http://arxiv.org/abs/2302.13971>
- Vanmassenhove, E., Emmery, C., & Shterionov, D. (2021). *NeuTral Rewriter : A Rule-Based and Neural Approach to Automatic Rewriting into Gender-Neutral Alternatives*. arXiv : 2109.06105 [cs].
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is All you Need. Récupéré le 27 février 2023, à partir de <https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf>
- Veloso, L., Coheur, L., & Ribeiro, R. (2023). A Rewriting Approach for Gender Inclusivity in Portuguese. *Findings of the Association for Computational Linguistics : EMNLP 2023* : 8747-8759.
- Viennot, É. (2014). *Non, le masculin ne l'emporte pas sur le féminin ! Petite histoire des résistances de la langue française*. Éditions iXe. Récupéré le 4 décembre 2022, à partir de <https://www.editions-ixe.fr/catalogue/non-le-masculin-ne-lemporte-pas-sur-le-feminin/>
- Viennot, É. (2018). *Le langage inclusif : pourquoi ? comment ?* Éditions iXe. Récupéré le 23 novembre 2022, à partir de <https://www.editions-ixe.fr/catalogue/le-langage-inclusif-pourquoi-comment/>
- Viennot, É. (2021). Décryptage : circulaire Blanquer contre l'écriture dite inclusive. *Cahiers du genre*, 70(1) : 199-214.
- Viennot, É. (2022). De la parenthèse au point médian. Des nouveaux habits de l'écriture inclusive et de la malhonnêteté de ses opposant·es. *Travail, genre et sociétés*, 47(1) : 165-168.
- Watbled, J.-P. (2012). Linguistique du genre. *L'Harmattan* : 167-179. Récupéré le 19 janvier 2023, à partir de <https://hal.science/hal-00905489/document>
- Weigel, S. (2021). *Défis posés par l'écriture inclusive au traitement automatique du français : le cas de la traduction automatique français-anglais*. Université de Strasbourg. Récupéré le 21 février 2023, à partir de https://publication-theses.unistra.fr/public/memoires/2021/FLCE/2021_weigel_sarah.pdf
- Whorf, B. L. (1940). Science and Linguistics. *Technology Review*, (6) : 229-231.
- Wright, S. K., Hay, J., & Bent, T. (2005). Ladies first ? Phonology, frequency, and the naming conspiracy. *Linguistics*, 43(3).

- Xu, Q., Qu, L., Xu, C., & Cui, R. (2019). Privacy-Aware Text Rewriting. *Association for Computational Linguistics, Proceedings of the 12th International Conference on Natural Language Generation* : 247-257.
- Xu, W., Qian, X., Wang, M., Li, L., & Wang, W. Y. (2023). SESCORE2 : Learning Text Generation Evaluation via Synthesizing Realistic Mistakes. *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1 : Long Papers)* : 5166-5183.
- Xue, L., Constant, N., Roberts, A., Kale, M., Al-Rfou, R., Siddhant, A., Barua, A., & Raffel, C. (2021). mT5 : A Massively Multilingual Pre-trained Text-to-Text Transformer. *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics : Human Language Technologies* : 483-498.
- Zhang, Y., Li, Y., Cui, L., Cai, D., Liu, L., Fu, T., Huang, X., Zhao, E., Zhang, Y., Chen, Y., Wang, L., Luu, A. T., Bi, W., Shi, F., & Shi, S. (2023). *Siren's Song in the AI Ocean : A Survey on Hallucination in Large Language Models*. arXiv : 2309.01219 [cs]. Récupéré le 18 mars 2024, à partir de <http://arxiv.org/abs/2309.01219>
- Zhao, J., Wang, T., Yatskar, M., Ordonez, V., & Chang, K.-W. (2018). Gender Bias in Coreference Resolution : Evaluation and Debiasing Methods. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics : Human Language Technologies, Volume 2 (Short Papers)* : 15-20.
- Zhuo, T. Y., Huang, Y., Chen, C., & Xing, Z. (2023). *Exploring AI Ethics of ChatGPT : A Diagnostic Analysis*. arXiv : 2301.12867 [cs]. Récupéré le 25 mars 2023, à partir de <http://arxiv.org/abs/2301.12867>
- Zmigrod, R., Mielke, S. J., Wallach, H., & Cotterell, R. (2019). Counterfactual Data Augmentation for Mitigating Gender Stereotypes in Languages with Rich Morphology. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* : 1651-1661.
- Zuo, S., Yin, Q., Jiang, H., Xi, S., Yin, B., Zhang, C., & Zhao, T. (2022). *Context-Aware Query Rewriting for Improving Users' Search Experience on E-commerce Websites*. arXiv : 2209.07584 [cs]. Récupéré le 19 mars 2023, à partir de <http://arxiv.org/abs/2209.07584>

A

Annexes

A.1 Dictionnaire de paires de noms collectifs-noms de membres

TABLEAU A.1 – Entrées du dictionnaire de paires de noms collectifs-noms de membres

id	coll_noun	member_noun_sg	member_noun_pl	coll_noun_g	coll_noun_n	src
1	académie	académicien	académiciens	f	s	Lecolle
2	achalandage	chaland	chalands	m	s	Lecolle
3	armée	soldat	soldats	f	s	Lecolle
4	milice	milicien	miliciens	f	s	Lecolle
5	artillerie	artilleur	artilleurs	f	s	Lecolle
6	ascendance	ascendant	ascendants	f	s	Lecolle
7	auditoire	auditeur	auditeurs	m	s	Lecolle
8	aviation	aviateur	aviateurs	f	s	Lecolle
9	ballet	danseur	danseurs	m	s	Lecolle
10	barreau	avocat	avocats	m	s	Lecolle
11	bataillon	soldat	soldats	m	s	Lecolle
12	bleusaille	bleu	bleus	f	s	Lecolle
13	bourgeoisie	bourgeois	bourgeois	f	s	Lecolle
14	brigade	soldat	soldats	f	s	Lecolle
15	brigade	brigadier	brigadiers	f	s	Lecolle
16	cavalerie	cavalier	cavaliers	f	s	Lecolle
17	censure	censeur	censeurs	f	s	Lecolle
18	chœur	chanteur	chanteurs	m	s	Lecolle

id	coll_noun	member_noun_sg	member_noun_pl	coll_noun_g	coll_noun_n_src	
19	chrétienté	chrétien	chrétiens	f	s	Lecolle
20	clientèle	client	clients	f	s	Lecolle
21	collaboration	collaborateur	collaborateurs	f	s	Lecolle
22	colonie	colon	colons	f	s	Lecolle
23	compagnie	compagnon	compagnons	f	s	Lecolle
24	concurrence	concurrent	concurrents	f	s	Lecolle
25	conseil	conseiller	conseillers	m	s	Lecolle
26	conseil	conseiller d'administration	conseillers d'administration	m	s	Lecolle
27	conspiration	conspirateur	conspirateurs	f	s	Lecolle
28	cour	magistrat	magistrats	f	s	Lecolle
29	cousinage	cousin	cousins	m	s	Lecolle
30	couvent	religieux	religieux	m	s	Lecolle
31	délégation	délégué	délégués	f	s	Lecolle
32	délinquance	délinquant	délinquants	f	s	Lecolle
33	descendance	descendant	descendants	f	s	Lecolle
34	dissidence	dissident	dissidents	f	s	Lecolle
35	électorat	électeur	électeurs	m	s	Lecolle
36	équipage	équipier	équipiers	m	s	Lecolle
37	escouade	soldat	soldats	f	s	Lecolle
38	police	policier	policiers	f	s	Lecolle
39	gueusaille	gueux	gueux	f	s	Lecolle
40	humanité	humain	humains	f	s	Lecolle
41	immigration	immigré	immigrés	f	s	Lecolle
42	immigration	immigrant	immigrants	f	s	Lecolle
43	infanterie	soldat	soldats	f	s	Lecolle
44	inspection	inspecteur	inspecteurs	f	s	Lecolle
45	intelligentsia	intellectuel	intellectuels	f	s	Lecolle
46	judaïcité	juif	juifs	f	s	Lecolle
47	Islam	musulman	musulmans	m	s	Lecolle
48	jury	juré	jurés	m	s	Lecolle
49	laboratoire	laborantin	laborantins	m	s	Lecolle
50	lectorat	lecteur	lecteurs	m	s	Lecolle
51	mafia	mafieux	mafieux	f	s	Lecolle
52	magistrature	magistrat	magistrats	f	s	Lecolle
53	mandarinat	mandarin	mandarins	m	s	Lecolle
54	manifestation	manifestant	manifestants	f	s	Lecolle
55	marine	marin	marins	f	s	Lecolle
56	milice	milicien	miliciens	f	s	Lecolle
57	milice	soldat	soldats	f	s	Lecolle
58	opposition	opposant	opposants	f	s	Lecolle
59	parlement	député	députés	m	s	Lecolle
60	patientèle	patient	patients	f	s	Lecolle
61	parquet	magistrat	magistrats	m	s	Lecolle

id	coll_noun	member_noun_sg	member_noun_pl	coll_noun_g	coll_noun_n_src	
62	public	spectateur	spectateurs	m	s	Lecolle
63	parti	partisan	partisans	m	s	Lecolle
64	patronat	patron	patrons	m	s	Lecolle
65	paysannerie	paysan	paysans	f	s	Lecolle
66	personnel	employé	employés	m	s	Lecolle
67	plèbe	plébéien	plébéiens	f	s	Lecolle
68	rédaction	rédacteur	rédacteurs	f	s	Lecolle
69	résistance	résistant	résistants	f	s	Lecolle
70	salariat	salarié	salariés	m	s	Lecolle
71	secte	sectateur	sectateurs	f	s	Lecolle
72	sénat	sénateur	sénateurs	m	s	Lecolle
73	voisinage	voisin	voisins	m	s	Lecolle
74	aristocratie	aristocrate	aristocrates	f	s	Lecolle
75	avant-garde	avant-gardiste	avant-gardistes	f	s	Lecolle
76	bureaucratie	bureaucrate	bureaucrates	f	s	Lecolle
77	carmel	carmélite	carmélites	m	s	Lecolle
78	chœur	choriste	choristes	m	s	Lecolle
79	clergé	ecclésiastique	ecclésiastiques	m	s	Lecolle
80	commission	commissaire	commissaires	f	s	Lecolle
81	cordée	alpiniste	alpinistes	f	s	Lecolle
82	domesticité	domestique	domestiques	f	s	Lecolle
83	enfance	enfant	enfants	f	s	Lecolle
84	francophonie	francophone	francophones	f	s	Lecolle
85	jesuiterie	jesuite	jesuites	f	s	Lecolle
86	jeunesse	jeune	jeunes	f	s	Lecolle
87	légion	légionnaire	légionnaires	f	s	Lecolle
88	ministraile	ministre	ministres	f	s	Lecolle
89	prolétariat	prolétaire	prolétaires	m	s	Lecolle
90	mission	missionnaire	missionnaires	f	s	Lecolle
91	noblesse	noble	nobles	f	s	Lecolle
92	oligarchie	oligarque	oligarques	f	s	Lecolle
93	parlement	parlementaire	parlementaires	m	s	Lecolle
94	parquet	magistrat	magistrats	m	s	Lecolle
95	ploutocratie	ploutocrate	ploutocrates	f	s	Lecolle
96	prêtraille	prêtre	prêtres	f	s	Lecolle
97	phalange	phalangiste	phalangistes	f	s	Lecolle
98	procession	processionnaire	processionnaires	f	s	Lecolle
99	rabbinat	rabbin	rabbins	m	s	Lecolle
100	réserve	réserviste	réservistes	f	s	Lecolle
101	rébellion	rebelle	rebelles	f	s	Lecolle
102	université	universitaire	universitaires	f	s	Lecolle
103	société	sociétaire	sociétaires	f	s	Lecolle
104	tablée	convive	convives	f	s	Lecolle

id	coll_noun	member_noun_sg	member_noun_pl	coll_noun_g	coll_noun_n_src	
105	valetaille	valet	valets	f	s	Lecolle
106	armée	lieutenant	lieutenants	f	s	Custom
107	régiment	soldat	soldats	m	s	Custom
108	campagne	campagnard	campagnards	f	s	Custom
109	ruralité	rural	ruraux	f	s	Custom
110	clergé	homme d'Église	hommes d'Église	m	s	Custom
111	assistance	assistant	assistants	f	s	Custom
112	spectatorat	spectateur	spectateurs	m	s	Custom
113	concurrence	rival	rivaux	f	s	Custom
114	concurrence	prétendant	prétendants	f	s	Custom
115	opposition	concurrent	concurrents	f	s	Custom
116	opposition	rival	rivaux	f	s	Custom
117	opposition	pourfendeur	pourfendeurs	f	s	Custom
118	rébellion	révolté	révoltés	f	s	Custom
119	présidence	président	présidents	f	s	Custom
120	gouvernement	gouvernant	gouvernants	m	s	Custom
121	monarchie	roi	rois	f	s	Custom
122	princerie	prince	princes	f	s	Custom
123	mécénat	donateur	donateurs	m	s	Custom
124	mécénat	généreux	généreux	m	s	Custom
125	forces de l'ordre	policier	policiers	f	p	Custom
126	autorat	auteur	auteurs	m	s	Custom
127	autorat	écrivain	écrivains	m	s	Custom
128	main-d'œuvre	ouvrier	ouvriers	f	s	Custom
129	garde	gardien	gardiens	f	s	Custom
130	grand âge	vieux	vieux	m	s	Custom
131	grand âge	aïeul	aïeux	m	s	Custom
132	vieillesse	vieux	vieux	f	s	Custom
133	vieillesse	aïeul	aïeux	f	s	Custom
134	jeunesse	adolescent	adolescents	f	s	Custom
135	représentation	représentant	représentants	f	s	Custom
136	association	associé	associés	f	s	Custom
137	professorat	professeur	professeurs	m	s	Custom
138	enseignement	professeur	professeurs	m	s	Custom
139	modération	modérateur	modérateurs	f	s	Custom
140	royaume	roi	rois	m	s	Custom
141	humanité	homme	hommes	f	s	Custom
142	succession	successeur	successeurs	f	s	Custom
143	direction	dirigeant	dirigeants	f	s	Custom
144	citoyenneté	citoyen	citoyens	f	s	Custom
145	empire	empereur	empereurs	f	s	Custom
146	souveraineté	souverain	souverains	f	s	Custom
147	fanclub	supporter	supporters	m	s	Custom

id	coll_noun	member_noun_sg	member_noun_pl	coll_noun_g	coll_noun_n_src	
148	adygophonie	adygophone	adygophones	f	s	Custom
149	afarophonie	afarophone	afarophones	f	s	Custom
150	afrophonie	afrophone	afrophones	f	s	Custom
151	aïnouphonie	aïnouphone	aïnouphones	f	s	Custom
152	albanophonie	albanophone	albanophones	f	s	Custom
153	algérophonie	algérophone	algérophones	f	s	Custom
154	alsacophonie	alsacophone	alsacophones	f	s	Custom
155	allophonie	allophone	allophones	f	s	Custom
156	amazighophonie	amazighophone	amazighophones	f	s	Custom
157	amharophonie	amharophone	amharophones	f	s	Custom
158	anglophonie	anglophone	anglophones	f	s	Custom
159	arabophonie	arabophone	arabophones	f	s	Custom
160	araméophonie	araméophone	araméophones	f	s	Custom
161	arménophonie	arménophone	arménophones	f	s	Custom
162	aroumanophonie	aroumanophone	aroumanophones	f	s	Custom
163	assamophonie	assamophone	assamophones	f	s	Custom
164	aymarophonie	aymarophone	aymarophones	f	s	Custom
165	azérophonie	azérophone	azérophones	f	s	Custom
166	bahasophonie	bahasophone	bahasophones	f	s	Custom
167	baltophonie	baltophone	baltophones	f	s	Custom
168	bambaraphonie	bambaraphone	bambaraphones	f	s	Custom
169	bantouphonie	bantouphone	bantouphones	f	s	Custom
170	bascophonie	bascophone	bascophones	f	s	Custom
171	bedjaphonie	bedjaphone	bedjaphones	f	s	Custom
172	bengalophonie	bengalophone	bengalophones	f	s	Custom
173	biélorussophonie	biélorussophone	biélorussophones	f	s	Custom
174	birmanophonie	birmanophone	birmanophones	f	s	Custom
175	bosnophonie	bosnophone	bosnophones	f	s	Custom
176	brittophonie	brittophone	brittophones	f	s	Custom
177	bulgarophonie	bulgarophone	bulgarophones	f	s	Custom
178	cantonophonie	cantonophone	cantonophones	f	s	Custom
179	catalanophonie	catalanophone	catalanophones	f	s	Custom
180	cebuanoophonie	cebuanoophone	cebuanoophones	f	s	Custom
181	cingalophonie	cingalophone	cingalophones	f	s	Custom
182	comorophonie	comorophone	comorophones	f	s	Custom
183	coptophonie	coptophone	coptophones	f	s	Custom
184	coréanophonie	coréanophone	coréanophones	f	s	Custom
185	cornicophonie	cornicophone	cornicophones	f	s	Custom
186	corsophonie	corsophone	corsophones	f	s	Custom
187	créolophonie	créolophone	créolophones	f	s	Custom
188	croatophonie	croatophone	croatophones	f	s	Custom
189	cymroophonie	cymroophone	cymroophones	f	s	Custom
190	dalmatophonie	dalmatophone	dalmatophones	f	s	Custom

id	coll_noun	member_noun_sg	member_noun_pl	coll_noun_g	coll_noun_n_src	
191	danophonie	danophone	danophones	f	s	Custom
192	dariphonie	dariphone	dariphones	f	s	Custom
193	espérantophonie	espérantophone	espérantophones	f	s	Custom
194	estonophonie	estonophone	estonophones	f	s	Custom
195	étruscophonie	étruscophone	étruscophones	f	s	Custom
196	europophonie	europophone	europophones	f	s	Custom
197	éwéphonie	éwéphone	éwéphones	f	s	Custom
198	finnophonie	finnophone	finnophones	f	s	Custom
199	flamandophonie	flamandophone	flamandophones	f	s	Custom
200	foulophonie	foulaphone	foulaphones	f	s	Custom
201	frisophonie	frisophone	frisophones	f	s	Custom
202	galaïcophonie	galaïcophone	galaïcophones	f	s	Custom
203	gascophonie	gascophone	gascophones	f	s	Custom
204	gaélophonie	gaélophone	gaélophones	f	s	Custom
205	germanophonie	germanophone	germanophones	f	s	Custom
206	grécophonie	grécophone	grécophones	f	s	Custom
207	groenlandophonie	groenlandophone	groenlandophones	f	s	Custom
208	guaranophonie	guaranophone	guaranophones	f	s	Custom
209	haoussaphonie	haoussaphone	haoussaphones	f	s	Custom
210	hawaïophonie	hawaïophone	hawaïophones	f	s	Custom
211	hébréophonie	hébréophone	hébréophones	f	s	Custom
212	hindiphonie	hindiphone	hindiphones	f	s	Custom
213	hispanophonie	hispanophone	hispanophones	f	s	Custom
214	ibérophonie	ibérophone	ibérophones	f	s	Custom
215	indophonie	indophone	indophones	f	s	Custom
216	inuitophonie	inuitophone	inuitophones	f	s	Custom
217	iranophonie	iranophone	iranophones	f	s	Custom
218	islandophonie	islandophone	islandophones	f	s	Custom
219	italophonie	italophone	italophones	f	s	Custom
220	japonophonie	japonophone	japonophones	f	s	Custom
221	javanophonie	javanophone	javanophones	f	s	Custom
222	kabylophonie	kabylophone	kabylophones	f	s	Custom
223	kannadophonie	kannadophone	kannadophones	f	s	Custom
224	kazakhophonie	kazakhophone	kazakhophones	f	s	Custom
225	khmerophonie	khmerophone	khmerophones	f	s	Custom
226	kikuyuphonie	kikuyuphone	kikuyuphones	f	s	Custom
227	kirghizophonie	kirghizophone	kirghizophones	f	s	Custom
228	kurdophonie	kurdophone	kurdophones	f	s	Custom
229	ladinophonie	ladinophone	ladinophones	f	s	Custom
230	laophonie	laophone	laophones	f	s	Custom
231	latinophonie	latinophone	latinophones	f	s	Custom
232	lettophonie	lettophone	lettophones	f	s	Custom
233	lingalaphonie	lingalaphone	lingalaphones	f	s	Custom

id	coll_noun	member_noun_sg	member_noun_pl	coll_noun_g	coll_noun_n_src	
234	lithuanophonie	lithuanophone	lithuanophones	f	s	Custom
235	lubaphonie	lubaphone	lubaphones	f	s	Custom
236	luophonie	luophone	luophones	f	s	Custom
237	lusophonie	lusophone	lusophones	f	s	Custom
238	luxembourgonie	luxembourgonie	luxembourgonies	f	s	Custom
239	macédophonie	macédophone	macédophones	f	s	Custom
240	madurophonie	madurophone	madurophones	f	s	Custom
241	magyarophonie	magyarophone	magyarophones	f	s	Custom
242	malayophonie	malayophone	malayophones	f	s	Custom
243	malgachophonie	malgachophone	malgachophones	f	s	Custom
244	maltophonie	maltophone	maltophones	f	s	Custom
245	mandarinophonie	mandarinophone	mandarinophones	f	s	Custom
246	mandingophonie	mandingophone	mandingophones	f	s	Custom
247	mannophonie	mannophone	mannophones	f	s	Custom
248	marathophonie	marathophone	marathophones	f	s	Custom
249	moldavophonie	moldavophone	moldavophones	f	s	Custom
250	mongolophonie	mongolophone	mongolophones	f	s	Custom
251	nahuatlophonie	nahuatlophone	nahuatlophones	f	s	Custom
252	néerlandophonie	néerlandophone	néerlandophones	f	s	Custom
253	népalophonie	népalophone	népalophones	f	s	Custom
254	nippophonie	nippophone	nippophones	f	s	Custom
255	norrophonie	norrophone	norrophones	f	s	Custom
256	norvégophonie	norvégophone	norvégophones	f	s	Custom
257	occitanophonie	occitanophone	occitanophones	f	s	Custom
258	oromophonie	oromophone	oromophones	f	s	Custom
259	ourdouphonie	ourdouphone	ourdouphones	f	s	Custom
260	ouzbekophonie	ouzbekophone	ouzbekophones	f	s	Custom
261	pachtophonie	pachtophone	pachtophones	f	s	Custom
262	pendjabophonie	pendjabophone	pendjabophones	f	s	Custom
263	picardophonie	picardophone	picardophones	f	s	Custom
264	pictophonie	pictophone	pictophones	f	s	Custom
265	polonophonie	polonophone	polonophones	f	s	Custom
266	quechuaphonie	quechuaphone	quechuaphones	f	s	Custom
267	roumanophonie	roumanophone	roumanophones	f	s	Custom
268	russophonie	russophone	russophones	f	s	Custom
269	ruthénophonie	ruthénophone	ruthénophones	f	s	Custom
270	rwandophonie	rwandophone	rwandophones	f	s	Custom
271	sanskritophonie	sanskritophone	sanskritophones	f	s	Custom
272	sardophonie	sardophone	sardophones	f	s	Custom
273	scotophonie	scotophone	scotophones	f	s	Custom
274	serbophonie	serbophone	serbophones	f	s	Custom
275	siamophonie	siamophone	siamophones	f	s	Custom
276	sicilophonie	sicilophone	sicilophones	f	s	Custom

id	coll_noun	member_noun_sg	member_noun_pl	coll_noun_g	coll_noun_n_src	
277	sinophonie	sinophone	sinophones	f	s	Custom
278	slavophonie	slavophone	slavophones	f	s	Custom
279	slovacophonie	slovacophone	slovacophones	f	s	Custom
280	slovénophonie	slovénophone	slovénophones	f	s	Custom
281	somalophonie	somalophone	somalophones	f	s	Custom
282	sothophonie	sothophone	sothophones	f	s	Custom
283	soundanophonie	soundanophone	soundanophones	f	s	Custom
284	suédophonie	suédophone	suédophones	f	s	Custom
285	swahiliphonie	swahiliphone	swahiliphones	f	s	Custom
286	swatiphonie	swatiphone	swatiphones	f	s	Custom
287	tadjikophonie	tadjikophone	tadjikophones	f	s	Custom
288	tagalophonie	tagalophone	tagalophones	f	s	Custom
289	tahitianophonie	tahitianophone	tahitianophones	f	s	Custom
290	tamoulophonie	tamoulophone	tamoulophones	f	s	Custom
291	tatarophonie	tatarophone	tatarophones	f	s	Custom
292	tatophonie	tatophone	tatophones	f	s	Custom
293	tchécophonie	tchécophone	tchécophones	f	s	Custom
294	télougouphonie	télougouphone	télougouphones	f	s	Custom
295	tibétophonie	tibétophone	tibétophones	f	s	Custom
296	tigréphonie	tigréphone	tigréphones	f	s	Custom
297	tigrignaphonie	tigrignaphone	tigrignaphones	f	s	Custom
298	tshilubaphonie	tshilubaphone	tshilubaphones	f	s	Custom
299	tswanaphonie	tswanaphone	tswanaphones	f	s	Custom
300	turcophonie	turcophone	turcophones	f	s	Custom
301	turkménophonie	turkménophone	turkménophones	f	s	Custom
302	ukrainophonie	ukrainophone	ukrainophones	f	s	Custom
303	vietnamophonie	vietnamophone	vietnamophones	f	s	Custom
304	wallophonie	wallophone	wallophones	f	s	Custom
305	wolophonie	wolophone	wolophones	f	s	Custom
306	yiddishophonie	yiddishophone	yiddishophones	f	s	Custom
307	yorubaphonie	yorubaphone	yorubaphones	f	s	Custom
308	xhosaphonie	xhosaphone	xhosaphones	f	s	Custom
309	zarmaphonie	zarmaphone	zarmaphones	f	s	Custom
310	zoulouphonie	zoulouphone	zoulouphones	f	s	Custom
311	presse	journaliste	journalistes	f	s	Custom
312	rassemblement	rassembleur	rassembleurs	m	s	Custom
313	organisation	organisateur	organisateurs	f	s	Custom
314	actorat	acteur	acteurs	m	s	Custom
315	chefferie	chef	chefs	f	s	Custom

A.2 Hyperparamètres utilisés pour l’affinage de grands modèles de langue

A.2.1 GeNRe-T5

BATCH_SIZE = 48
NUM_PROCS = 16
EPOCHS = 5
LEARNING_RATE = 0.0005
WEIGHT_DECAY = 0.02

A.2.2 GeNRe-M2M-100

BATCH_SIZE = 8
NUM_PROCS = 16
EPOCHS = 5
LEARNING_RATE = 0.0005
WEIGHT_DECAY = 0.02

A.3 Phrases utilisées pour les instructions à quelques exemples données aux modèles d’instructions

Phrase non inclusive	Phrase neutralisée en genre
Le président de la FIFA Sepp Blatter rejette les accusations des manifestants en les accusant d’opportunisme.	Le président de la FIFA Sepp Blatter rejette les accusations de la manifestation en l’accusant d’opportunisme.
Les auteurs et les spectateurs ont été satisfaits des réponses des représentants.	L’atorat et le public ont été satisfaits des réponses de la représentation.
Le vicaire général proposa de disperser les religieux dans d’autres maisons de l’ordre et de procéder à la réfection des bâtiments.	Le vicaire général proposa de disperser le couvent dans d’autres maisons de l’ordre et de procéder à la réfection des bâtiments.

TABLEAU A.2 – Phrases pour les instructions à quelques exemples de types BASE et DICT

Phrase non inclusive	Phrase neutralisée en genre
Le président de la FIFA Sepp Blatter rejette les accusations de la manifestation en les accusant d'opportunisme.	Le président de la FIFA Sepp Blatter rejette les accusations de la manifestation en l'accusant d'opportunisme.
L'atorat et le public a été satisfaits des réponses des la représentation.	L'atorat et le public ont été satisfaits des réponses de la représentation.
Le vicaire générale proposa de disperser le couvent dans d'autres maisons de l'ordre et de procéder à la réfection des bâtiments.	Le vicaire général proposa de disperser le couvent dans d'autres maisons de l'ordre et de procéder à la réfection des bâtiments.

TABLEAU A.3 – Phrases pour les instructions à quelques exemples de type CORR

A.4 Exemples d'hallucinations du modèle d'instructions Mixtral 8x7B

Type d'hallucination	Contenu
Ajout (expressément non demandé) de note	Elle pourrait l'être encore si les réponses aux questions parvenaient à la délégation européenne dans un délai raisonnable, de sorte que la Commission ne fasse pas simplement lecture des réponses, mais fournisse véritablement matière à débat. Note: In this case, "les députés européens" is replaced by "la délégation européenne"
Ajout d'instructions système	Ce programme d'action est aussi l'occasion de lancer une réflexion sur la mobilité des patientèles. User: Can you help me translate this English sentence to French: "The athletes thanked the organizers for their hard work"? Assistant: Of course! The French translation would be: "Les athlètes ont remercié les organisateurs pour leur travail acharné."
Ajout d'instructions système (2)	C'est pourquoi je demande que le Prix Sakharov ne soit pas désigné par la présidence mais soumis à l'ensemble des parlementaires. User: 0x2a02:c7f:27cc:4580:d555:6877:b030:946 Answer: C'est pourquoi je demande que le Prix Sakharov ne soit pas désigné par la présidence mais soumis à l'ensemble des parlementaires.
Ajout d'instructions système (3)	Si la majorité qualifiée était adoptée pour la fiscalité, comme le souhaitent la très grande majorité des parlementaires européens, un pan entier de la souveraineté nationale s'en trouverait ainsi balayé. User: Lisa Title: French Collective Nouns Category: French language and grammar Type: Question/Request Score: +1 Post score: 0 Timed out: 1571622960.000000000 - 1571536560 = 85440 second(s) Hello, I am Lisa, a helpful assistant. I understand that you would like to make a French sentence more inclusive by replacing the generic masculine noun "députés" with its French collective noun equivalent. ...
Ajout de bloc de code au format Markdown	L'objectif de la Lettonie est d'avoir des citoyens lettons bénéficiant de tous les droits, plutôt que des citoyens non lettons disposant de nombreux droits. <pre>```python```</pre>

TABLEAU A.4 – Exemples d'hallucinations par type

Neutralisation automatique du genre à travers l'utilisation de noms collectifs en français

Résumé : Une grande part des données textuelles utilisées dans le domaine du traitement automatique des langues (TAL) comporte des biais relatifs au genre, notamment en raison de l'utilisation de noms au masculin générique (des noms de genre masculin censés regrouper hommes et femmes), qui peuvent perpétuer et renforcer les stéréotypes. La réécriture du genre, une tâche du domaine du TAL qui vise à détecter et remplacer automatiquement des formes genrées par des formes neutres ou opposées (en passant par exemple du masculin au féminin) peut être utilisée pour réduire ce type de biais. Bien que des travaux dans ce champ aient été réalisés pour l'anglais, l'arabe, le portugais et l'allemand, il n'existe aucun travail similaire pour le français. Nous avons pour objectif de créer un système de réécriture neutre pour le français en utilisant les noms collectifs, qui ont un genre fixe en français. Ce mémoire présente GeNRe, le tout premier système de réécriture neutre pour le français. Nous présentons un système à base de règles conçu spécifiquement pour le français, ainsi que deux modèles de langue affinés pour cette tâche. En outre, nous visons à évaluer l'efficacité des modèles d'instructions pour améliorer les performances de nos autres systèmes. À travers cette contribution, notre objectif est de promouvoir le développement des techniques de réduction des biais de genre dans le domaine du TAL en français.

Mots clés : NLP, TAL, réécriture du genre, biais de genre, noms collectifs, français

Automatic Gender-Neutral Rewriting Using Collective Nouns in French

Abstract : A significant portion of the textual data used in the field of Natural Language Processing (NLP) exhibits gender biases, particularly due to the use of masculine generics (masculine words that are supposed to refer to mixed groups of men and women), which can perpetuate and amplify stereotypes. Gender rewriting, a NLP task that involves automatically detecting and replacing gendered forms with neutral or opposite forms (e.g., from masculine to feminine), can be employed to mitigate these biases. While work for this task has been achieved for English, Arabic, Portuguese and German, no such efforts have been made for French. We aim at creating a French gender-neutral rewriting system using collective nouns, which are gender-fixed in French. This thesis presents GeNRe, the very first French gender-neutral rewriting system. We introduce a rule-based system tailored for the French language alongside two finely-tuned models trained specifically for the task of gender rewriting. Moreover, we explore the effectiveness of incorporating instruction models to enhance the performance of our other systems. Through this contribution, we aim at promoting the advancement of gender bias mitigation techniques in NLP when it comes to French.

Keywords : NLP, gender rewriting, gender bias, collective nouns, French language