



ISSN 1961-9359

ISSN en ligne 2260-6513

Distinguer des genres textuels en français langue étrangère : étude textométrique de marqueurs subjectifs

Aylin Pamuksaç

Université de Neuchâtel, Suisse
aylin.pamuksac@unine.ch

<https://orcid.org/0000-0002-7043-4133>

Linda Sanvido

Université de Neuchâtel, Suisse
linda.sanvido@unine.ch

<https://orcid.org/0000-0002-2192-999X>

Reçu le 30-12-2022 / Évalué le 15-02-2023 / Accepté le 15-06-2023

Résumé

Une méthodologie textométrique pour différencier quatre genres textuels est utilisée dans la recherche : les discours de presse, les discours encyclopédiques, les discours présidentiels et des écrits d'étudiants universitaires. Ces divergences ou ressemblances sont étudiées à l'aide de marqueurs subjectifs, traces de la locutrice dans son discours : les pronoms personnels de première personne et certains verbes subjectifs (verbes modaux et verbes d'attitude propositionnelle). L'étude de représentativité de ces formes dans des genres différents permet de montrer que des discours à visée informative sont plus imperméables aux marques de la locutrice que des discours à visée argumentative. Cette méthodologie, ainsi que les résultats obtenus, permettent une application didactique en classe de français langue étrangère développée dans la dernière section à l'aide de plusieurs propositions d'exercices.

Mots-clés : textométrie, genres, linguistique de corpus, français langue étrangère

Distinción de géneros textuales en francés como lengua extranjera: un estudio textométrico de marcadores subjetivos

Resumen

En la investigación se utiliza una metodología textométrica para diferenciar cuatro géneros textuales: discursos de prensa, discursos enciclopédicos, discursos presidenciales y escritos de estudiantes universitarios. Investigamos estas diferencias o similitudes mediante el análisis de algunos marcadores subjetivos, ya que todos ellos atestiguan la presencia de la locutora en su discurso. Prestamos especial atención a los pronombres personales de primera persona y a algunos verbos subjetivos (verbos modales y verbos de actitud proposicional). El estudio de la representatividad de estas formas en distintos géneros enseña que el discurso informativo es más

hermético a las marcas de la locutora que el discurso argumentativo. Esta metodología y sus resultados permiten una aplicación didáctica en francés lengua extranjera desarrollada en el último apartado mediante la proposición de varios ejercicios.

Palabras clave: textometría, géneros, lingüística de corpus, FLE

Distinguishing textual genres in French as a foreign language: a textometric study of subjective markers

Abstract

A textometric methodology to differentiate four textual genres is used in the research: press discourse, encyclopaedic discourse, presidential discourse and university students' essays. The differences or similarities are investigated through the analysis of some subjective markers, since they all attest the presence of the speaker in her speech. A particular attention is given to first person personal pronouns and to some subjective verbs (modal verbs and propositional attitude verbs). The study of the representativeness of these forms in different genres shows that the informative discourse is more hermetic to the speaker's marks than the argumentative discourse. This methodology and its results allow for a didactic application in FFL, which we develop in the last section through the proposition of several exercises.

Keywords: textometry, genres, corpus linguistics, French as a foreign language

Introduction¹

La question des genres de textes² est abordée dans les classes de français langue première depuis les années 1970 (Vrydaghs, Cordeiro, 2016 : 13) mais elle l'est beaucoup moins dans les classes de français langue étrangère (désormais FLE). En prenant comme point de départ ce manque, nous proposons une méthodologie inspirée de la textométrie afin de faire émerger des différences notables entre genres textuels différents. Ces observations permettront aux enseignants et aux concepteurs de manuels d'élaborer des applications didactiques ayant pour but la sensibilisation des apprenants à la question du genre textuel.

La connaissance des types de textes est « indispensable en ce qui concerne l'enseignement/apprentissage des langues, car les genres de discours sont des formes de communication verbale propres à des communautés discursives c'est-à-dire, d'une certaine manière et indirectement, à des langues » (Beacco, 2013 : 192). Cette maîtrise permet une « économie cognitive », à savoir une « compréhension anticipée de l'intention, du contenu, du format, de la structure et de la globalité de l'énoncé (Bakhtine, 1978 et 1984). » (Ringoot, 2014 : 13). Les manuels de FLE proposent, pour presque tous les niveaux, des pistes de compréhension, d'analyse

et d'écriture des différents genres textuels, mais ces contributions restent minimes et lacunaires, et souvent cantonnées aux mêmes types de texte (la lettre pour les niveaux inférieurs, l'essai argumenté pour la préparation du DALF C1, pour en citer quelques-uns).

Au lieu d'esquisser une description qui tienne compte des caractéristiques génériques propres à chaque genre, nous proposons d'analyser des types différents de discours au moyen de l'étude des marques de subjectivité. Ainsi, nous nous concentrons sur des procédés linguistiques mettant en jeu un point de vue propre aux locutrices, à travers lequel elles s'inscrivent dans leur discours et se situent par rapport au message communiqué (Kerbrat-Orecchioni, 1980 : 32). Nous nous focalisons en particulier sur des formes telles que les pronoms personnels, les verbes modaux et quelques verbes d'attitude propositionnelle, formes reconnues comme faisant apparaître les locutrices dans leur énoncé (Kerbrat-Orecchioni, 1980 : 40-41, 100-101). Nous avançons l'hypothèse que des discours à visée informative, comme la presse et l'encyclopédie, sont plus imperméables aux traces des locutrices, alors que des discours à visée argumentative, comme les discours politiques, laisseraient libre cours à ces procédés³.

Les premières observations menées à partir de la liste des vingt verbes les plus fréquents montrent déjà une nette différence de représentation des verbes dénotant une subjectivité entre un corpus à visée informative et un corpus à visée argumentative. En effet, si dans le premier seulement deux verbes modaux apparaissent (*pouvoir*, *devoir*) parmi les verbes les plus fréquents, dans le second on observe quatre verbes modaux (*pouvoir*, *falloir*, *devoir*, *vouloir*) et quatre verbes d'attitude propositionnelle (*dire*, *savoir*, *croire*, *penser*).

Ces observations préliminaires seront approfondies dans notre article à l'aide d'une méthodologie fondée sur des calculs statistiques de sous- et surreprésentation des formes, qui permettra de rendre visibles des tendances qui différencient les discours choisis par rapport aux marques de subjectivité. Les mêmes recherches sont menées en italien, afin de vérifier si les tendances observées pour le français se retrouvent également dans une autre langue romane.

1. Genres, discours, types de textes : point terminologique

Avant de présenter nos résultats, nous souhaitons faire un point terminologique concernant la question même du genre textuel. Malgré son importance cruciale, cette notion ne fait pas l'unanimité chez de nombreux auteurs par sa définition même, car « [...] elle est souvent confondue avec celle de type de texte, et tantôt définie à partir de « fonctions du langage » (Biber, 1988 ; 1992) tantôt assimilée avec le domaine sémantique du discours (Illouz et al., 1999). » (Malrieu, Rastier,

2001 : 548). D'ailleurs, le terme *genre* est aujourd'hui plus souvent utilisé dans le cadre des études littéraires (Branca-Rosoff, 1999 : 5). Dans les études linguistiques francophones, on préfère alors adopter la terminologie *types de textes* suite aux études d'Adam (2017) ou même de manière plus récurrente encore, la notion de *discours* depuis les années 1980 (Ringoot, 2014 : 11).

Pour cet article, nous nous appuyons sur les niveaux de classification de Malrieu et Rastier (2001)⁴ pour définir notre cadre théorique et ainsi notre terminologie. Sur la base de quatre niveaux, les auteurs proposent de classer de manière hiérarchique les différents termes qui ont pu être qualifiés de *genres*. En raison de la nature de nos corpus, notre étude se situe au niveau des champs génériques. Nous étudions quatre types de textes qui représentent dans notre travail les quatre corpus que nous investiguons : les discours présidentiels, la presse écrite, l'encyclopédie collaborative et des écrits d'étudiants universitaires. Nous pouvons ainsi⁵ adapter le schéma de Malrieu et Rastier (2001 : 549) à nos corpus (Voir note 5, Figure 1 : *Niveaux de classification de nos corpus, inspirés de Malrieu et Rastier*).

À partir de ce schéma, nous pouvons observer que tous nos corpus se situent sur le même niveau, celui des champs génériques, et qu'ils sont ainsi tous des prototypes d'un certain type de discours. Le niveau des champs génériques est ici défini par le type de locutrice ou de scriptrice en charge du texte. Dans le cadre des discours journalistiques, nous pensons, à l'instar de Ringoot (2014 : 22-23), que le type de média montre également l'importance accordée aux locuteurs car

L'audiovisuel fait disparaître le nom du journal qui identifie les publications de la presse écrite, et ramène la qualification du journal à celle de la chaîne ou de la station. Il s'agit alors du journal de TF1 ou de France-Inter ; pour autant, la personnalisation s'effectue par le biais d'un ou d'une présentatrice qui, selon sa longévité, devient une figure emblématique (Ringoot, 2014 : 22-23).

La presse écrite, par excellence, se base « sur la construction d'une parole collective autour du nom du journal. » (Ringoot, 2014 : 25) et ce malgré le nombre important d'individus qui s'investissent dans une rédaction. De ce fait, les champs génériques permettent de montrer ici quel type de locutrice ou autrice se trouve derrière les textes, allant d'une parole individualisée à une parole plus collaborative.

2. La question du genre textuel en FLE

Dans la recherche en FLE, la question du genre textuel ne semble pas encore faire beaucoup de remous. Ceci peut être dû au fait que dans le *Cadre européen commun de référence pour les langues* (désormais CECR), cette question n'est que peu abordée, même si des connaissances dans ce domaine semblent être un élément primordial pour faciliter l'apprentissage pour l'apprenant,

car « la charge cognitive est réduite et le succès de l'exécution d'une tâche facilité par la plus ou moins grande familiarité de l'apprenant avec [...] le type de texte, le genre [...] » (CECR, 2001 : 122). Lorsque la notion est mentionnée explicitement, elle n'apparaît qu'à partir du niveau B2, voire au niveau C1, qui « [...] semble être caractérisé par le bon accès à *une large gamme de discours*⁶ qui permet une communication aisée et spontanée » (*ibid.*).

Tous les auteurs ayant travaillé sur le sujet reconnaissent l'importance de ce type de connaissance mais ils se centrent sur la compréhension des apprenants et les capacités des enseignants à sélectionner le texte adéquat selon le sujet enseigné. Tous proposent également un panorama de la notion et surtout les difficultés définitoires qui y sont liées. Les premières recherches traitant cette question de manière explicite sont apparues au début des années 2010. Truong (2010) souligne l'importance de la maîtrise des types de textes, non seulement pour la compréhension, mais aussi pour la production, tant à l'oral qu'à l'écrit (Truong, 2010 : 625). Dans son article, nous pouvons également découvrir une séquence didactique qui a l'avantage d'être très détaillée et surtout adressée à tous types de niveaux dès A1. Par contre, les exercices se concentrent sur la compréhension des textes, et l'aspect production est mis de côté. Notons également que plusieurs auteurs, notamment Hamez (2013, 2014) ainsi que Claudel et Laurens (2016), ont travaillé sur la question des genres textuels en FLE en proposant un panorama de ce qui se trouve dans les manuels. Les genres de la lettre et du courriel sont exploités de manière récurrente à des niveaux élémentaires ou intermédiaires (Claudel, Laurens, 2016 : 8), et l'essai argumenté et l'article de presse à des niveaux plus avancés (Hamez, 2013, 2014).

Les articles de presse ont été grande source d'exploitation dans les manuels de FLE. Par exemple, dans *Production écrite : niveaux C1/C2 du Cadre européen commun de référence* (Causa, Mègre, 2008 : 59), plusieurs procédés linguistiques typiques de ce genre discursif sont évoqués, notamment la phrase nominale et la forme passive, les pronoms relatifs, ainsi que l'emploi fréquent de chiffres et pourcentages, et l'interdiction de l'utilisation du pronom personnel « je », tout cela sous prétexte qu'un article de presse se veut « crédible » et « objectif ». Les connecteurs de cause et conséquence - cf. *Alter ego + 3 : méthode de français B1* (Dollez, Pons, 2013 : 74) - ainsi que les marqueurs concessifs - cf. *Cosmopolite 3 : méthode de français B1* (Hirschsprung et al., 2018 : 106) - sont également présentés dans le cadre des sections dédiées aux textes journalistiques. Dans les manuels, nous recensons également des remarques plus vagues, comme c'est le cas dans *Préparer le DALF C1 et C2* (Debeuckelaere, Hulin, 2018 : 126), qui présente des « tons » à adopter (enthousiaste, défensif, volontaire, virulent, accusateur, etc.)

pour exprimer et défendre un point de vue, dans le cadre de l'écriture d'un essai argumenté. Cette liste nous semble problématique car aucune application concrète n'est proposée. En effet, aucun procédé linguistique n'est explicité, ne serait-ce que le type de mots ou tournures à utiliser pour appréhender un ton défensif plutôt qu'un ton accusateur.

Weber (2021), quant à elle, propose une réflexion plus générale autour de la question du genre et centre son analyse autour des indices de typicalité qui « permettent de classifier et distinguer les faisceaux de caractéristiques des discours » (Weber, 2021 : 213-214). Ce sera grâce à ces indices d'ordre linguistique que l'apprenant pourra reconnaître et produire différents types de textes. Nous pensons que ces indices peuvent prendre plusieurs formes et que les marques de subjectivités sur lesquelles nous nous focalisons dans cet article sont un exemple de ceux-ci.

3. Méthodologie

3.1. Présentation des données et du logiciel

Comme susmentionné, nous avons utilisé quatre types de texte pour cette recherche afin de voir si les champs génériques se différenciaient par rapport aux indices de subjectivité que nous avons sélectionnés. Nous avons ainsi formé un corpus, divisible en quatre parties :

- *Wiki* (1'201'373 tokens) : sous-corpus composé d'articles pris au hasard sur l'encyclopédie collaborative en ligne Wikipédia lors de l'année 2019. Il représente ici le discours encyclopédique.
- *Présidents* (1'273'180 tokens) : sous-corpus composé de discours de Présidents de la République, de Pompidou à Macron, de 1958 à 2020. Il représente ici le discours politique.
- *Journal* (933'246 tokens) : sous-corpus composé d'articles de l'année 2010 du journal national *Le Monde*. Il représente ici le discours journalistique.
- *Etudiant·e·s* (1'084'371 tokens) : sous-corpus composé d'écrits d'étudiants et construit par le Laboratoire de Linguistique et Didactique des Langues Etrangères et Maternelles de Grenoble⁷. Les disciplines représentées sont pour l'essentiel les sciences du langage, la didactique du français et les sciences de l'éducation. Il représente ici le discours scientifique.

Au total, notre corpus est composé de 4'492'170 tokens répartis de manière plus ou moins équivalente selon les sous-parties. Ce corpus est exploité sur la plateforme TXM, développée à l'ENS de Lyon (Heiden et al. 2010).

3.2. Méthodes de calculs

Afin de faire émerger les différences entre nos champs génériques, nous avons pris le parti de représenter nos résultats avec l'indice de spécificité (Lafon, 1980) qui permet

de dégager les écarts d'emploi d'une forme lexicale entre les différentes parties d'un corpus [et qui] rend compte de l'emploi d'une forme dans les parties en réduisant l'incidence de la différence de taille entre les parties. Ces écarts, mesurés à partir d'un indice de probabilité, permettent de mettre au jour les parties suremployant ou sous-employant telle ou telle unité. (Montrichard, 2020 : 221).

Cet outil nous permettra ainsi de distinguer quel sous-corpus emploie de manière significative des indices de subjectivité ou non. Dans TXM, l'indice de spécificité est banal entre -2 et 2. Au-dessus de 2, les paramètres étudiés sont surreprésentés, en dessous, sous-représentés.

4. Analyse des résultats

4.1. Les verbes modaux et les verbes d'attitude propositionnelle

Nous allons, dans un premier temps, nous concentrer sur les verbes modaux et les verbes d'attitude propositionnelle. Il semble pertinent de réunir ces deux types de verbes dans un même paradigme car tous deux modalisent l'énoncé (Gómez-Jordana Ferary, Anscombe, 2015 : 7) et « [...] introduisent une représentation du monde effectuée par la parole, avec spécification du type de parole utilisé, de sa provenance et de la position du locuteur. » (Anscombe, 2015 : 103). Dans notre paradigme, nous avons pris en considération les verbes les plus fréquents du sous-corpus *Présidents*, susmentionnés dans notre introduction, et qui entrent dans cette catégorie. Nous obtenons ainsi le graphique⁸ correspondant à la Figure 2 (Voir note 8, Figure 2 : *Distribution par champs génériques des verbes modaux et verbes d'attitude propositionnelle en français*).

Dans la figure 2, nous pouvons observer une opposition très marquée entre les discours présidentiels et l'encyclopédie collaborative, ce que nous laissait déjà présager l'étude préliminaire que nous avons montrée dans l'introduction. En effet, tous ces verbes apparaissent en surreprésentation dans *Présidents* et en sous-représentation dans *Wiki*. Par contre, si on oppose le sous-corpus *Présidents* à celui du *Journal*, les différences sont moins marquées. Certes, tous les verbes y sont sous-représentés, mais les indices sont moins marqués que pour *Wiki*. Concernant les écrits des étudiants universitaires, il est intéressant de constater

que ce sous-corpus ne semble pas se situer de manière aussi tranchée du côté de la subjectivité ou de l'objectivité. En effet, *devoir*, *pouvoir* et *savoir* y sont surreprésentés, ce qui pourrait ainsi le rapprocher d'un discours plus argumentatif, et donc subjectif, mais *penser* présente un indice banal et tous les autres verbes y sont sous-représentés.

Avec ce premier paramètre, nous pouvons ainsi constater que les champs génériques se démarquent d'une manière plus ou moins forte les uns des autres. Nous avons d'un côté le corpus composé de discours présidentiels qui semble hautement subjectif, et de l'autre, l'encyclopédie et la presse qui n'utilisent pas ces verbes marquant la trace des locutrices. Entre les deux extrêmes, nous trouvons le corpus *Etudiant·e·s* qui possède quelques-uns de ces verbes en surreprésentation, et qui signifierait donc un aspect subjectif de ce type de texte. Cependant, la majorité des marqueurs sont ici tout de même sous-représentés.

4.2. Les pronoms personnels⁹

Pour faire ressortir la subjectivité des champs génériques, nous avons également décidé d'étudier les indices de spécificité liés aux pronoms personnels (désormais PP). Comme le rappelle Reboul (1991 : 131), *je* fait toujours référence aux locutrices et en ceci, il peut ainsi être une marque de subjectivité. Elle rejoint ainsi Kerbrat-Orecchioni (1980 : 41) qui situait déjà les PP *je* et *nous* dans les indices de subjectivité, le *nous* regroupant en tout cas *je* + une autre instance. Ainsi, il y aurait certains PP dotés d'une plus ou moins grande subjectivité (Voir note 9, Figure 3 : *Distribution par champs génériques des pronoms personnels en français*).

Pour ce paramètre, à nouveau, la plus grande opposition se situe entre *Présidents* et *Wikipédia*. Pour les discours présidentiels, tous les PP sont surreprésentés sauf *tu*. Ceci paraît assez logique étant donné qu'un président ne s'adressera que rarement qu'à une seule personne lors de discours officiels et même lorsque c'est le cas, ce serait plutôt le vouvoiement qui serait de rigueur. Dans le corpus encyclopédique, tous les pronoms sont sous-représentés à des mesures différentes. Il peut être étonnant de constater que même *on* y est sous-représenté alors qu'il peut avoir un effet généralisant (Fløttum et al., 2007 : 13) et que cet effet est typique d'un discours informatif, comme dans :

En français, *on* emploie le terme de négationnisme pour qualifier la contestation de crimes contre l'humanité [...]. (*Wiki*)

Concernant le sous-corpus *Etudiant·e·s*, celui-ci ne montre pas une différenciation aussi marquée que la presse écrite ou l'encyclopédie collaborative par rapport au discours politique. En effet, celui-ci possède quelques PP en sous-représentation et d'autres en surreprésentation.

En somme, les PP permettent ici de souligner, eux aussi, des différences concernant la subjectivité. Encore une fois, notre texte prototypique de l'argumentation semble plus ponctué de ces indices que les trois autres types de textes que nous étudions.

4.3. Les catégories morphosyntaxiques

Concernant ce dernier point étudié, nous voulions voir si certaines catégories morphosyntaxiques apparaissent de manière plus marquée dans des sous-corpus. Nous pensons ainsi qu'une surreprésentation de certaines catégories morphosyntaxiques comme les PP ou encore les adverbes pourrait signaler une trace des locutrices. Nous avons décidé de prendre en compte l'ensemble des adverbes afin de dresser un premier tableau des textes plus ou moins subjectifs. Nous sommes conscientes que ce sont usuellement les adverbes de phrase (Molinier, Levrier, 2000) qui sont considérés comme des marques de subjectivité¹⁰ : un tri des valeurs serait donc à envisager dans un deuxième temps.

De plus, ce focus sur les catégories morphosyntaxiques¹¹ nous permet également de voir si les marqueurs propres à certains genres et relevés dans les manuels (cf. la partie 3 de ce travail) apparaissent de manière plus ou moins marquée selon le type de discours (Voir note 11, Figure 4 : *Distribution par champs génériques des catégories morphosyntaxiques en français*).

Il est intéressant de constater que les champs génériques de la presse écrite et de l'encyclopédie collaborative sont ici très proches quant aux catégories morphosyntaxiques mobilisées dans leurs textes. Ils se démarquent des deux autres genres par l'utilisation des noms propres (NAM) mais aussi des nombres (NUM). Ils présentent également une surreprésentation des noms en général (NOM). Les adjectifs (ADJ) sont également en surreprésentation dans ces deux parties, mais cela est sans doute dû au fait qu'ils vont de pair avec les noms. Les noms présentent d'ailleurs aussi un grand indice de spécificité dans les écrits scientifiques. Par contre, ce n'est pas le cas des noms propres. Cela peut être dû au fait que contrairement à des textes à visée informative, les étudiants sont moins dans la diffusion de faits ou de dires relatifs à des instances en particulier. Il peut aussi être intéressant de noter que les verbes sont en surreprésentation dans ce sous-corpus, ce qui n'est pas le cas dans les textes journalistiques ou encyclopédiques. Ceci pourrait être un signe des phrases nominales relevées dans les manuels de FLE que nous avons soulignées en 3.

On peut aussi souligner qu'avec ce paramètre, le corpus des discours présidentiels se démarque à nouveau des trois autres types de texte. Ainsi, il présente une surreprésentation des connecteurs (KON) de manière très marquée, ce qui pourrait être un signe d'une construction plus hiérarchisée de ce type de texte par rapport aux autres. Les connecteurs sont également en surreprésentation dans le corpus *Etudiant·e·s* mais avec un indice bien moins grand (11,7) alors qu'il se situe à 1000 pour les discours présidentiels. Ceci pourrait toutefois montrer que les étudiants organisent davantage de manière logique leurs textes ou alors qu'elles et ils ressentent le besoin de surmarquer l'organisation textuelle, plus que les journalistes ou les auteurs des articles encyclopédiques.

Les paramètres de subjectivité, qui sont ici les PP (PRO:PER) et les adverbes (ADV), sont en surreprésentation uniquement dans les discours politiques, ce qui montre que ce type de texte semble à nouveau répondre à des critères bien particuliers. Relevons également que les pronoms relatifs (PRO:REL) y sont également surreprésentés alors que nous avons relevés dans les manuels de FLE (cf. partie 3 de ce travail) que les phrases subordonnées étaient un trait relevant plutôt de la presse.

Pour résumer, il est très intéressant d'étudier le paramètre des catégories morphosyntaxiques pour faire ressortir les indices de subjectivité selon les types de textes mais aussi pour constater si les remarques des manuels sont pertinentes. Par contre, nous pourrions affiner ces résultats en étiquetant certains marqueurs en particulier, comme certains verbes ou certains adverbes.

4.4. Comparaison avec l'italien

Pour pouvoir effectuer les mêmes recherches pour l'italien, nous avons construit un corpus de 4'855'747 tokens divisé en quatre sous-parties, à savoir :

- *Wiki_it* (1'272'174 tokens) : sous-corpus composé d'articles choisis au hasard sur l'encyclopédie collaborative en ligne Wikipédia pour l'année 2019.
- *Présidents_it* (1'248'782 tokens) : sous-corpus composé de discours de présidents de la République italienne, de Gronchi à Mattarella (1955-2021).
- *Journal_it* (1'325'703 tokens) : sous-corpus composé d'articles choisis aléatoirement entre 2010 et 2015 dans le journal *La Repubblica*.
- *Etudiant·e·s_it* (1'009'088 tokens) : sous-corpus composé de thèses universitaires (niveaux Bachelor et Master) d'étudiantes et étudiants de l'Université de Padoue. Les disciplines représentées sont les sciences du langage et de la communication.

Afin de vérifier les sous- et surreprésentations des verbes modaux et des verbes d'attitude propositionnelle en italien, nous avons choisi d'interroger les correspondants des formes françaises, à savoir : *credere* (croire), *dovere* (devoir), *dire* (dire), *occorrere* (falloir), *pensare* (penser), *potere* (pouvoir), *sapere* (savoir), *volere* (vouloir). Nous remarquons que les tendances sont presque pareilles que celles relevées pour le français¹² (Voir note 12, Figure 5 : *Distribution par champs génériques des verbes modaux et des verbes d'attitude propositionnelle en italien*).

Nous observons une forte dissymétrie entre le discours présidentiel et le discours encyclopédique, avec une surreprésentation de tous les verbes dans *Présidents_it* et une sous-représentation des mêmes verbes dans *Wiki_it*. Le corpus *Journal_it* ne se démarque pas de *Wiki_it* de façon aussi nette : seulement *occorrere* et *potere* se trouvent sous-représentés et on observe une surreprésentation des verbes *dire*, *pensare* et *volere*. Le sous-corpus *Etudiant·e·s_it* montre lui aussi des comportements assez départagés, avec une surreprésentation des verbes modaux *potere* et *dovere* et une sous-représentation des verbes *credere*, *dire*, *sapere*, *volere*. Ces résultats confirment ce que nous avons étudié pour le français, avec une exception pour *Journal_it* qui semble avoir un comportement moins proche de *Wiki_it* par rapport aux deux sous-corpus correspondants en français.

Le comportement des pronoms personnels dans le corpus italien¹³ confirme nos intuitions par rapport aux différentes sensibilités qu'éprouvent les champs génériques par rapport aux marqueurs de subjectivité (Voir note 13, Figure 6 : *Distribution par champs génériques des pronoms personnels en italien*).

Les pronoms personnels de première personne (*io*, *noi*) sont surreprésentés dans *Présidents_it* (*io* : 138,37 / *noi* : 234,41) et sous-représentés dans *Wiki_it* (*io* : -171,83 / *noi* : -219,71) et *Etudiant·e·s_it* (*io* : -22,34 / *noi* : -50,03). Ce résultat n'est pas étonnant du moment que dans un discours à visée informative la locutrice aura tendance à effacer sa propre voix pour donner la priorité aux sources (Marnette, 2004 : 51) alors que dans un discours argumentatif l'image de soi est primordiale et l'ego omniprésent. Le discours de presse italien montre encore une fois un comportement intermédiaire, avec une légère surreprésentation du *io* (5,27) et du *noi* (2,70).

Une analyse plus générale des catégories morphosyntaxiques¹⁴ permet de mettre en lumière des différences de fonctionnement entre le sous-corpus *Présidents_it* d'un côté et les sous-corpus *Journal_it* et *Wiki_it* de l'autre (Voir note 14, Figure 7 : *Distribution par champs génériques des catégories morphosyntaxiques en italien*).

Dans *Présidents_it*, on observe une sous-représentation des noms communs (NOM), des noms propres (NPR) et des chiffres (NUM), catégories qui se trouvent en

surreprésentation dans *Wiki_it* et *Journal_it* (à l'exception des noms pour *Journal_it*). Ceci nous donne déjà une indication importante quant aux propriétés de nominalisation que les discours à visée informative possèdent : il s'agit de reporter des événements de façon neutre, sans engagement des scriptrices. Au contraire, le sous-corpus *Présidents_it* privilégie l'utilisation d'adverbes et de conjonctions, car les indices de ces deux catégories atteignent le plafond de surreprésentation, fixé à 1000. Ces catégories aident à structurer les propos des locutrices tout en amenant un point de vue très construit rhétoriquement. Les verbes sont également surreprésentés dans *Présidents_it* (indice : 9,76), ce qui vient confirmer qu'il s'agit bien d'un discours moins nominal. Inversement, les catégories ADV, CON et VER atteignent des scores très hauts de sous-représentation dans le corpus *Wiki_it*. Si les propriétés génériques du sous-corpus encyclopédique sont transférables, pour le français, au sous-corpus journalistique, ceci n'est pas entièrement le cas pour l'italien. En effet, dans le corpus *Journal_it*, bien que les catégories NPR et NUM soient surreprésentées, les verbes le sont également (indice de 140,30), alors que les noms et les adjectifs sont sous-représentés. Ce n'est pas exactement le profil qu'on attendrait d'un discours à visée informative.

Les tendances présentées pour l'italien montrent, tout comme pour le français, un fort contraste dans la subjectivité entre le discours politique et le discours encyclopédique. Si le discours journalistique français semble se placer de façon assez nette du côté du discours encyclopédique, il est plus difficile de reporter cette observation au discours journalistique italien. Celui-ci semble, certes, partager des propriétés avec l'autre sous-corpus à visée informative, mais de façon beaucoup moins cohérente et départagée que pour le français. On peut alors se demander si les caractéristiques formelles d'un discours ne seraient pas également sujettes aux contraintes imposées par une langue donnée.

5. Didactisation

Nous proposons dans cette section deux exercices exploitables en classe de FLE, exercices construits grâce aux analyses présentées à l'instant. Le but est de montrer comment nos observations et notre méthodologie peuvent servir autant les enseignants dans la préparation des cours que les conceptrices de manuels et donc, par conséquent, les apprenants.

Un premier exercice de compréhension demande aux apprenants de lire et analyser deux paragraphes issus de deux corpus différents. Il s'agit de repérer des procédés récurrents tels que la nominalisation ou le recours à des verbes plus ou moins « subjectifs ». Cette analyse a pour but d'arriver à étiqueter le premier texte

comme « discours à visée argumentative » et le deuxième comme « discours à visée informative ». Un deuxième exercice propose une production écrite exploitant une liste de mots apparaissant de façon fréquente dans notre corpus. Ce type d'activité permet aux apprenants de s'exercer dans l'écriture tout en apprenant de nouveaux mots qui peuvent être plus ou moins porteurs de subjectivité. On attendra que des mots comme *savoir* ou *je* soient plutôt choisis pour l'écriture argumentative, alors que des mots plus neutres comme *exemple* ou *deuxièmement* seront privilégiés pour la rédaction d'un texte informatif.

Exercice 1. *Lisez les textes et essayez de comprendre de quel type de discours il s'agit en expliquant quel type de procédé vous fait pencher pour un genre plutôt qu'un autre.*

Texte 1 :

Alors, j'ajoute que, si je fais le bilan des, par exemple, cinquante années passées, je me dis au total que, malgré tous les problèmes auxquels on est confronté, les progrès l'ont emporté sans aucun doute sur les reculs, dans tous les domaines. Dans le domaine de la paix, c'est l'essentiel. De la démocratie. Dans le monde entier, elle est en progrès. Dans le domaine économique, c'est évident. Dans le domaine social aussi, pas partout, mais dans une partie importante de nos pays, ce qui suppose que l'on réponde à la question de la solidarité et du partage que j'évoquais tout à l'heure. Dans le domaine de la santé : on vit beaucoup plus longtemps qu'il y a cinquante ans, et cetera. Donc, je pense qu'au total, il faut assumer les difficultés auxquelles on est chaque jour confronté, mais il ne faut pas se laisser aveugler par ces difficultés.

Texte 2 :

La compétition a été entachée par des critiques soutenues concernant l'arbitrage, notamment en raison d'un prétendu manque d'expérience de certains arbitres, choisis pour des raisons de politique d'élargissement aux pays émergents. En particulier, l'Italie s'est vue refuser plusieurs buts au cours de la compétition ; un hors-jeu a été signalé dans leur match perdu contre la Corée du Sud en huitièmes de finale et les journaux italiens y ont vu, avec l'expulsion de Francesco Totti dans les prolongations, une erreur d'arbitrage lourde. L'entraîneur italien estime : « Je ne comprends pas pourquoi nous avons à être victimes de mauvaises décisions. Je pense que le gagnant devrait être l'Italie. ». L'équipe de France, alors championne du monde, est également concernée par ces polémiques : contre l'Uruguay, un but de Trezeguet est refusé pour hors-jeu. Le match se termine alors par 0-0. Le refus d'un but en or et un signalé lors du match Espagne-Corée du Sud en quart de finale ont également été critiqués par la presse espagnole.

Des reproches similaires ont été faits à l'arbitrage lors du match de premier tour entre le Portugal et la Corée. D'autres décisions contestées ont été prises ; on pourra citer la rencontre entre la Belgique et le Brésil en huitièmes de finale, où l'arbitre annule de manière litigieuse un but du joueur belge Marc Wilmots¹⁵.

Exercice 2. *Sur la base de cette liste de mots, créez des phrases qui peuvent s'insérer dans un ou l'autre genre : savoir (verbe), je, on, 17%, entreprise, vouloir, monde, deuxièmement, exemple.*

Texte argumentatif :

Texte informatif :

Conclusion et perspectives

Dans cette contribution, nous avons essayé de montrer à quel point les connaissances sur les genres textuels n'étaient pas encore assez exploitées en FLE et que ce manque devait être comblé tout simplement pour permettre une première décharge cognitive chez l'apprenant. En effet, une connaissance des champs génériques facilite tant les activités de compréhension que de production, orales ou écrites. Cette lacune de didactisation des genres textuels peut être due à la difficulté même de qualifier les différents types de textes ou de définir ce qu'est un genre. Notre méthodologie est innovante, car non seulement elle prend en compte des champs génériques qui ont rarement été comparés entre eux - surtout les textes des étudiants qui sont très peu considérés dans les études sur les genres textuels - mais aussi parce que nous nous sommes focalisées sur les traits de la subjectivité pour différencier ces sous-corpus. Les résultats ont été très saillants, surtout pour différencier les textes à visée argumentative, comme les discours présidentiels, qui se distinguent - plus ou moins nettement selon les langues - des textes à visée informative, que ce soit la presse, l'encyclopédie ou encore les écrits universitaires.

Sur la base de ces résultats, nous avons également pu démontrer qu'il était possible d'élaborer des exercices prenant en compte ces traits subjectifs tant sur le plan de la compréhension que de la production.

Concernant les prolongements possibles, nous pensons qu'il serait intéressant d'aller également vérifier les manuels d'italien langue étrangère afin de voir comment la question des genres y est abordée et si oui, de quelle manière. En effet, nous avons pu remarquer quelques différences entre les pratiques francophones et italophones en termes de traces de subjectivité selon le champ générique investigué, notamment en ce qui concerne les deux corpus de presse. Ceci nous permettrait ainsi de constater si ces différences découlent d'idéaux culturels déjà exposés dans les manuels. Nous considérons également que notre méthodologie pourrait s'étendre à d'autres genres, en intégrant par exemple des textes littéraires comme des romans, des nouvelles voire des pièces de théâtre. En outre, il serait envisageable d'appliquer notre démarche à d'autres marqueurs de subjectivité, ou alors, de regarder de plus près certaines catégories morphosyntaxiques comme les adverbes, les adjectifs ou encore les connecteurs qui peuvent être plus ou moins objectifs. Enfin, il reste d'autres outils statistiques à déployer pour étudier les différences entre les champs génériques. En effet, nous pourrions envisager une étude des genres à travers l'Analyse Factorielle des Correspondances afin de représenter les éloignements ou les rapprochements entre les sous-corpus de manière plus visuelle.

Bibliographie

Adam, J.-M. 2017. *Les textes: types et prototypes: récit, description, argumentation, explication et dialogue*. Paris : Armand Colin.

Anscombe, J.-C. 2015. « Verbes d'activité de parole, verbes de parole et verbes de dire: des catégories linguistiques? ». *Langue française*, n° 186 (2), p. 103-122.

Beacco, J. C. 2013. « L'approche par genres discursifs dans l'enseignement du français langue étrangère et langue de scolarisation », *Pratiques*, n° 157-158, p. 189-200. [En ligne] : <https://doi.org/10.4000/pratiques.3838> [consulté le 20 décembre 2022].

Branca-Rosoff, S. 1999. « Types, modes et genres : entre langue et discours ». *Langage & société*, vol. 87, n° 1, p. 5-24. [En ligne] : https://www.persee.fr/doc/lsoc_0181-4095_1999_num_87_1_2851 [consulté le 20 décembre 2022].

Causa, M. Mègre, B. 2008. *Production écrite : niveaux C1/C2 du Cadre européen commun de référence*. Paris : Didier.

Claudiel, C., Laurens, V. 2016. Le genre discursif comme objet d'enseignement en didactique du français. *SHS Web of Conferences*, vol. 27, p. 07005.

Conseil de l'Europe. 2001. *Cadre européen commun de référence pour les langues : apprendre, enseigner, évaluer*. Strasbourg : Conseil de l'Europe.

Debeuckelaere, A., Hulin, H. 2018. *Préparer le DALF C1 et C2 : méthodologie de l'épreuve de production écrite, entraînements, corrigés*. Fontaine : PUG.

Dollez, C., Pons, S. 2013. *Alter ego + 3 : méthode de français : B1*. Paris : Hachette.

- Fløttum, K. et al. 2007. *On : pronom à facettes*. Bruxelles : De Boeck/Duculot.
- Garric, N. 1996. « Pour une stratégie discursive publicitaire : les adverbess assertifs en -ment ». *Langage & société*, vol.78, n° 1, p. 77-88. [En ligne] : https://www.persee.fr/doc/lscoc_0181-4095_1996_num_78_1_2761 [consulté le 20 décembre 2022].
- Gómez-Jordana Ferary, S., Anscombe, J.-C. 2015. « Introduction : Dire et ses marqueurs ». *Langue française*, vol. 186, n° 2, p. 5-12. [En ligne] : <https://doi.org/10.3917/lf.186.0005> [consulté le 20 décembre 2022].
- Hamez, M.-P. 2013. Enseigner l'écriture avec la presse écrite: tour d'horizon des manuels de FLE de l'approche communicative à la perspective actionnelle. In : *Ecarts et apports des médias francophones : lexique et grammaire*. Berne : Peter Lang, p. 245-283.
- Hamez, M.-P. 2014. « L'essai argumenté, un genre enseigné en classe de FLE, étude de manuels publiés entre 1990 et 2013 ». *Synergies Pays Scandinaves*, n° 9, p. 79-92. [En ligne] : <https://gerflint.fr/Base/Paysscandinaves9/hamez.pdf> [consulté le 20 décembre 2022].
- Heiden, S. et al. 2010. TXM : Une plateforme logicielle open-source pour la textométrie - conception et développement. In *10th International Conference on the Statistical Analysis of Textual Data - JADT 2010*, Rome, Italy. <https://halshs.archives-ouvertes.fr/halshs-00549779> [consulté le 20 décembre 2022].
- Hirschsprung, N. et al. 2018. *Cosmopolite 3 : méthode de français B1*. Vanves: Hachette.
- Jacques, M.-P., Rinck, F. 2017. "Un corpus de littéracie avancée: résultat et point de départ". *Corpus*, n° 16, p. 1-17. [En ligne] : <https://doi.org/https://doi.org/10.4000/corpus.2806> [consulté le 20 décembre 2022].
- Kerbrat-Orecchioni, C. 1980. *L'énonciation : de la subjectivité dans le langage*. Paris : Armand Colin.
- Lafon, P. 1980. « Sur la variabilité de la fréquence des formes dans un corpus ». *Mots. Les langages du politique*, n° 1, p. 127-165.
- Leblanc, J.-M. 2015. « Proposition de protocole pour l'analyse des données textuelles : pour une démarche expérimentale en lexicométrie ». *Nouvelles perspectives en sciences sociales*, vol. 11, n° 1, p. 25-63. [En ligne] : <https://doi.org/https://doi.org/10.7202/1035932ar> [consulté le 20 décembre 2022].
- Malrieu, D., Rastier, F. 2001. Genres et variations morphosyntaxiques. *Traitement Automatique des langues*, vol.42, n° 2, p. 548-577.
- Marnette, S. 2004. « L'effacement énonciatif dans la presse contemporaine ». *Langages*, n° 4, p. 51-64.
- Molinier, C., Levrier, F. 2000. *Grammaire des adverbes: description des formes en-ment*. Genève : Librairie Droz.
- Montrichard, C. 2020. *La presse de tranchées : Un espace discursif de mise en scène d'un contre-discours combattant ?* Thèse de doctorat. Université Bourgogne Franche-Comté : Besançon.
- Prendes, E. Á. 2014. "L'expression de la subjectivité dans le discours: le cas de deux adverbes de phrase (sérieusement, peut-être)". *Cahiers de praxématique*, n° 62.
- Reboul, A. 1991. "Le système des pronoms personnels en français contemporain: le problème de la référence". *Revue Tranel (Travaux neuchâtelois de linguistique)*, n° 17, p. 123-146.
- Ringoot, R. 2014. *Analyser le discours de presse*. Paris : Armand Colin.
- Rossari, C. et al. 2022. Pour une approche sémantique des connecteurs au-delà de leurs propriétés relationnelles : étude sur des variations génériques et diachroniques dans des corpus écrits. *SHS Web Conf.*, n° 138, p. 11016. [En ligne] : <https://doi.org/10.1051/shsconf/202213811016> [consulté le 20 décembre 2022].
- Truong, H. L. 2010. Des perspectives didactiques de la notion de genre de discours à une approche didactique de la lecture-compréhension en FLE. *2° Congrès Mondial de Linguistique Française*, p. 042. [En ligne] : <https://doi.org/10.1051/cmlf/2010028> [consulté le 20 décembre 2022].

Vrydaghs, D., Cordeiro, G. S. 2016. Le statut des genres dans l'enseignement du français: quelles réflexions et propositions didactiques aujourd'hui? In : *Statuts des genres en didactique du français: recherche, formation et pratiques enseignantes*. Namur : Presses universitaires de Namur, p. 11-25.

Weber, C. 2021. Chapitre 18. *Grammaire, textes et discours*. In : *Le FLE et la francophonie dans le monde*. Paris : Armand Colin, p. 211-218. <https://doi.org/10.3917/arco.chiss.2021.01.0211> [consulté le 20 décembre 2022].

Notes

1. Aylin Pamuksaç et Linda Sanvido ont contribué à cette recherche dans une même proportion. La première, s'est occupée de la rédaction des sections 1, 2, 4, 5.1, 5.2, 5.3 ; la seconde, des sections 5.4 et 6, 7. Elles ont rédigé de manière égale la section 3.
2. Nous reviendrons sur la terminologie même de la notion de genres en 2.
3. Nos corpus seront présentés dans la partie 4.1.
4. Nous remercions Cristelle Cavalla de nous avoir mises sur la piste de cet article afin de délimiter de manière plus précise notre définition des genres textuels.
5. Figure 1 : Niveaux de classification de nos corpus, inspirés de Malrieu et Rastier (2001 : 549) : <https://drive.google.com/file/d/1ipGw5Muy7W8Ww6VoUAudN9MUe6WdPCx6/view?usp=sharing>
6. Nous soulignons.
7. Corpus développé par Jacques et Rinck (2017) et disponible sur Ortholang à l'adresse suivante : <https://www.ortolang.fr/market/corpora/litteracieavancee/v1> [consulté le 20 décembre 2022].
8. Figure 2: Distribution par champs génériques des verbes modaux et verbes d'attitude propositionnelle en français : <https://drive.google.com/file/d/1oQrSfIAuAPcH1rM3nTKfmdqckOgBl4wx/view?usp=sharing>
9. Figure 3 : Distribution par champs génériques des pronoms personnels en français : https://drive.google.com/file/d/1SrVF4_Ulg13qWHD11CukBcdu9GBceUn/view?usp=sharing
10. cf. les études de Garric (1996), Prendes (2014) ou encore Rossari et al. (2022) qui choisissent ce type d'adverbes comme indicateur de subjectivité.
11. Figure 4 : Distribution par champs génériques des catégories morphosyntaxiques en français : <https://drive.google.com/file/d/1T2tMVOSZ7RwGLAdtC7HenDjYUyXCkmMn/view?usp=sharing>
12. Figure 5 : Distribution par champs génériques des verbes modaux et des verbes d'attitude propositionnelle en italien : <https://drive.google.com/file/d/1jfx8t1XijXCGoFDZvgD-j9NaFK1t7v9/view?usp=sharing>
13. Figure 6 : Distribution par champs génériques des pronoms personnels en italien : https://drive.google.com/file/d/1JqY_AtMzGBkOSmY4hpHrsXENIE9CHof7/view?usp=sharing
14. Figure 7 : Distribution par champs génériques des catégories morphosyntaxiques en italien : https://drive.google.com/file/d/1K2X6qlX9PtL8S_Fg0umaQGAuPn_Y5n-G/view?usp=sharing
15. Texte issu de la page : https://fr.wikipedia.org/wiki/Coupe_du_monde_de_football_2002 [consulté le 20 décembre 2022].