

CHAPITRE 7 : Genre et styles d'auteur

7.1. *Pour une approche des styles en tant qu'usages singuliers du genre*

La question des styles en sciences peut paraître paradoxale : communément entendu comme propre à l'écrivain, le style serait une dimension neutralisée de textes censément objectifs, qui honorent le seul *logos*, au point d'être conçus comme des textes « sans locuteurs »¹, ou dont l'auteur serait à envisager comme « non-auteur »². L'écrit de recherche impose en effet une forme d'effacement énonciatif, comme l'ont bien mis en évidence les typologies énonciatives des textes établies par Benveniste et, à sa suite, par Simonin-Grumbach et Bronckart ; mais la dimension collective caractéristique de l'activité scientifique renforce aussi cette idée d'absence d'une dimension personnelle dans l'écrit de recherche, encore décrit comme fortement contraint, voire « routinisé » (De Nuchèze). Finalement, quand il n'est pas taxé de fioriture, voire de déviance journalistique, le style ne semble être admis que dans la mesure où il est au service de qualités scientifiques : soyez clair, soyez concis, et soyez pertinent³.

C'est sans doute pour ces raisons que la stylométrie, ou analyse quantitative et statistique de traits stylistiques, a largement délaissé l'étude des styles scientifiques, alors qu'elle s'est beaucoup intéressée au discours littéraire et dans une moindre mesure au discours politique également – qu'il s'agisse de dater les textes, d'en déterminer la paternité, de débusquer les apocryphes, ou encore plus simplement, de comparer les textes d'un ou plusieurs auteurs.

La question des styles en science prend pourtant sens pour peu que l'on intègre le niveau du genre à l'analyse des styles d'auteur. Comme le soulignait déjà Bakhtine, une étude des styles ne saurait faire l'économie, et les genres semblent en effet l'emporter sur les styles en matière de classification⁴. Une analyse réalisée sans référence au genre permet au mieux de dégager des types linguistiques de textes ; elle n'autorise pas comme le montre Brunet à conclure que les ressemblances et différences entre textes sont le seul fait de leur paternité : « pourquoi chercher à confondre Molière et Corneille, quand le partage du même genre comique et de la même expression versifiée suffit à expliquer la ressemblance ? »⁵. Inversement, dans quelle mesure un auteur conserve-t-il ses spécificités quand il change de genre ?⁶

C'est donc une définition du style relative au genre qui intéresse notre étude : nous considérons avec Rastier (2001) les styles en tant qu'*usages singuliers du genre* et faisons l'hypothèse que le genre de l'article de revue, aussi normé soit-il, ne va pas sans une relative

¹ Philippe, 2002.

² Cf. D. Pontille

³ On remarquera cette même négation de toute dimension stylistique dans les représentations que les étudiants et apprentis-chercheurs se font de l'écriture de recherche (cf. Reuter, 1998).

⁴ Muller et Brunet, 1988.

⁵ Cité par Mézaille, 2003.

⁶ Brunet cité par Mézaille, 2003.

variation stylistique – relative dans la mesure où les fondations même du genre ne sauraient être ébranlées.

Le présent chapitre s'articulera ainsi en deux temps : après avoir exploré le corpus « Auteurs » (axes de contraste interne et spécificités des 15 styles d'auteur observés), on s'intéressera plus particulièrement aux relations entre styles d'auteur scientifique et genre en évaluant la distinctivité des styles au sein du genre.

7.2. Corpus d'étude et systèmes de descripteurs adoptés

L'analyse est fondée sur deux corpus : le corpus générique de 224 textes (précisément décrit chapitre 2) et le corpus « Auteurs » spécifiquement dédié à l'étude des styles, qui rassemble un ensemble de 122 textes⁷ de 15 linguistes français accrédités dans le champ. Hormis cette spécificité liée à l'auteur, les textes sont sélectionnés selon les mêmes critères que ceux du corpus « genre » (date de parution, rattachement disciplinaire, etc.).

Etant donné leur grande incidence sur les estimations statistiques, les exemples du corpus « Auteurs » ont également été extraits, selon la procédure adoptée dans les chapitres précédents : afin de ne pas briser la continuité syntaxiques, seuls les exemples délimités dans les textes ont été isolés ; les textes portent donc la trace de leur objet linguistique, mais de manière toutefois moindre.

Bien qu'il eût été parfaitement envisageable de recourir au même système de descripteurs que celui que nous avons adopté jusqu'à présent, nous avons tenu, dans un souci d'objectivation, à adjoindre aux variables morphosyntaxiques deux ensembles supplémentaires de descripteurs : un ensemble de 97 variables morphosyntaxiques et syntaxiques extraites du jeu d'étiquettes du système Cordial (Synapse développement). Les variables sélectionnées sont complémentaires plutôt que concurrentes de notre jeu d'étiquettes, qui ne contenait ainsi aucune variable syntaxique (e.g. % des proposition relatives, subordonnées ou coordonnées par rapport à l'ensemble des propositions, % des propositions ayant un complément d'objet direct ou indirect, etc.) ; et un lexique épistémique de 100 descripteurs lemmatisés de type « hypothèse », « méthode »⁸ qui sert la mise en scène de la recherche dans ses composantes et ses procédures, et qui, par hypothèse, est transversal aux sous-domaines considérés.

Ces trois ensembles de descripteurs ont servi de fondement à la détermination des spécificités des 15 styles d'auteur observés, tandis que les relations entre styles et genre ont essentiellement été analysées dans leurs aspects morphosyntaxiques (étiquetage TnT).

⁷ Nous ne disposons malheureusement pas de textes de genres différents ; si le discours scientifique comprend bien des genres distincts, il est plus délicat de constituer un corpus textuel équilibré en termes de diversité de genres et d'auteurs quand il s'agit de textes scientifiques que de textes littéraires.

⁸ Rinck, 2005.

7.3. Exploration du corpus « Auteurs »

7.3.1. Première exploration du corpus « Auteurs » : spécificités des 15 styles observés

Un *Tri Systématique de Signification* (TSS) a été effectué au moyen du logiciel DTM en confrontant le corpus « Auteurs » avec une variable supplémentaire stylistique positionnée *a posteriori* sur les axes principaux au moyen d'une *valeur-test* exprimant « la signification statistique de la coordonnée de la variable sur l'axe » (Lebart, 2004, p. 712).

Cette méthode nous a permis de déterminer les caractéristiques morphosyntaxiques et épistémiques de chaque auteur ; à ces variables ont été ajoutés les spécificités mots des textes, qui ont autorisé la mise au jour de certaines des caractéristiques lexicales des styles observés. Nous avons en effet choisi de sélectionner des variables *a priori* indépendantes de la thématique des textes pour les classer : les variables lexico-thématiques nous semblent susceptibles d'entraver l'émergence des styles en privilégiant celle des sous-domaines scientifiques dont relèvent les articles et leurs auteurs.

L'ensemble des résultats statistiques obtenus est présenté en annexe 11. Nous présentons ici les caractéristiques des 15 styles d'auteur observées, obtenues à partir de l'analyse des spécificités.

Grille de lecture des caractéristiques lexicales, épistémiques et morphosyntaxiques des 15 styles d'auteur

En positif : du plus significatif au moins significatif ;

En négatif : (ordre inverse) du moins significatif au plus significatif ;

En gras dans l'analyse des variables morphosyntaxiques : caractéristiques partagées par Cordial et TnT.

7.3.1.1. Authier

A. Lexique

En positif : dire, méta, mots, autonymique, formes

En négatif : politique, français, les, nous

B. Lexique épistémique

En positif : appeler, relever, terme, commentaire, choix, exemple, viser, signifier, question

En négatif : expérimental, hypothèse, schéma, présumé

C. Variables morphosyntaxiques :

En positif :

Avec Cordial : **tirets, verbes conjugués avec *je, je*, points de suspension, nombre moyen de pronoms par proposition**, pronoms indéfinis, **nombre total de ponctuations, crochets et accolades**, pronoms possessifs, propositions relatives, verbes à COI obligatoire, verbes transitifs indirects

Avec TnT : **tirets, % ponctuations, crochets**, participes présent, ***je***, préfixes, **points de suspension**, disjoints *lui/elle*, adjectifs réflexifs, renvois dans le texte, déterminants possessifs *son, sa, ses*, subordonnants, amalgames singulier, clitiques *me*, connecteurs d'exemplification, guillemets, longueur du texte, pronoms possessifs *sien(ne)*, points-virgules, **% pronoms**

En négatif :

Avec Cordial : verbes conjugués avec *ils/elles*, **deux points**, nombre moyen de lettres par mot, **points**

Avec TnT : dates, adjectifs pluriel, amalgames pluriel, % adverbes et connecteurs, numéraux ordinaux, % noms, % adjectifs, **deux points**, impersonnels *il*, parenthèses, connecteurs de conséquence, **points**

Parmi les **variables morpho-syntaxiques** qui caractérisent les textes d'Authier au sein du corpus auteur, les **ponctuations** ont une place importante : la forte présence de ponctuation de manière générale, et de points de suspension en particulier sont les deux caractéristiques les plus fortes des textes d'Authier comparativement aux autres textes du corpus. Ses textes présentent par ailleurs significativement plus de guillemets simples, de slashes, de tirets, de points d'exclamation, de crochets et de parenthèses, et significativement moins de deux points d'interrogation et de points.

Si la ponctuation suggère que ces textes sont très exemplifiés, voire énonciativement marqués, on en trouve un autre indice dans l'usage particulièrement abondant de **pronoms personnels** de la première personne du singulier, et dans une moindre mesure, celui des clitiques pour cette même personne. Cela se traduit avec le jeu de variables Cordial par une sur-représentation des verbes à la première personne du singulier.

Outre une présence importante d'exemples qui contiennent *je*, Authier s'exprime souvent par le biais de la première personne du singulier :

Je ne ferai que noter que c'est dans le cadre de cette opposition, discours autre approprié/associé, qu'il convient, à mon sens, de traiter, comme un cas particulier d'émergence de mots d'ailleurs (le fragment "X") *appropriés à l'objet du discours...* (001)

Cette sur-représentation de la première personne (singulier) va de pair avec une sous-représentation du pronom de 3^{ème} personne du singulier – et, avec Cordial, de verbes à la 3^{ème} personne du pluriel. Mais inversement, les pronoms possessifs et les pronoms disjoints de troisième personne du singulier sont bien représentés avec le jeu de variables TnT. Plus largement, les pronoms sont une classe discriminant les textes d'Authier par leur sur-représentation, les adjectifs, noms, et les adverbes et connecteurs dans une moindre mesure les caractérisant à l'inverse par leur sous-représentation.

La pratique de l'incise (cf les tirets et parenthèses, préférés aux crochets, sous-représentés) s'accompagne d'un usage fort de verbes au participe présent. On peut également rattacher ce dernier à un usage important de subordonnants qui peut être l'indice d'une syntaxe faisant

large place aux subordonnées. Cependant, les subordonnants *comme, lorsque, que, quand, quoique, si + sinon* sont sous-représentés, et, dans une plus large mesure encore, les connecteurs de conséquence. Les connecteurs d'exemplifications sont par contre particulièrement présents.

La caractérisation avec TnT des numéraux ordinaux et de datation sont significativement moins présents que chez les autres auteurs. Par contre, on observe chez Authier un usage plus abondant de numéraux de renvoi. On peut penser que cette tendance est liée à un usage abondant d'exemples numérotés dans ses textes. La forte présence de tirets serait alors à mettre au compte des indices de structuration qui servent à introduire les exemples.

Du côté des déterminants, les textes d'Authier se caractérisent positivement par l'usage de déterminant positifs de troisième personne du singulier et de déterminants contractés au singulier également, négativement par les déterminants contractés pluriels. Les adjectifs au pluriel sont aussi une variable discriminant les textes d'Authier par leur faible présence.

La forte présence d'adjectifs réfléchis (« même » dans « lui-même » par exemple) peut marquer une tendance à l'emphase, qu'on pourrait rattacher à l'idée qu'il s'agit de textes énonciativement marqués.

Le jeu de variables Cordial montre que les textes d'Authier se distinguent également en termes syntaxiques, en particulier au niveau du SV, fait inédit chez les autres auteurs : les textes d'Authier se caractérisent par l'usage de verbes transitifs directs, de verbes à COI obligatoires et de propositions relatives parmi l'ensemble des subordonnées.

A noter, les exemples auxquels recourt Authier sont susceptibles de parasiter l'analyse, voire la parasitent complètement si l'on considère les résultats obtenus avec Cordial (pas d'extraction des exemples) ; c'est ainsi que l'on peut expliquer sans doute l'usage des points d'exclamation.

En ce qui concerne les **variables lexicales**, les variables discriminant positivement les textes d'Authier semblent liées aux thématiques abordées : « dire », qui chez Authier peut avoir le statut de verbe ou de substantif : *le(s) dire(s), méta* - dont l'usage fort peut expliquer la forte présence de préfixes, *mots, autonymique* et enfin *formes*. Par rapport au corpus, l'usage de *politique, français, les* et *nous* est faible, ce que l'on peut expliquer par les thématiques, un usage faible des formes plurielles en général (cf les déterminants), et une préférence du *je* sur le *nous* (cf. le taux fort de pronoms personnels de 1^{ère} personne, clitiques ou non).

En ce qui concerne les **termes épistémiques**, les termes mis en avant comme sur-représentés chez Authier sont *appeler, relever, terme* et *commentaire*.

En résumé, beaucoup de variables discriminent, positivement et négativement les textes d'Authier, que ce soit dans les résultats fournis par le jeu Cordial ou par le jeu TnT. Différentes, les deux caractérisations montrent pourtant des tendances complémentaires, qui concernent en particulier l'usage des ponctuations et des pronoms personnels de 1^{ère} du singulier. Ces tendances observées dans le corpus auteur restent constantes quand on compare les textes d'Authier au corpus « genre » (variables Cordial et épistémiques) : un grand nombre de variables, notamment les mêmes variables relatives aux personnes et à la ponctuation caractérisent le style d'Authier.

7.3.1.2. Barberis

A. Lexique

En positif : rue, i, actualisation, tu, là

En négatif : contexte, b, sujet, sémiotique, discours

Les textes de Barbéris sont d'abord caractérisés par le mot *rue*, employé 209 fois dans l'ensemble du corpus, et 147 fois par Barbéris, ce qui est remarquable. Si l'un des articles porte effectivement sur le substantif *rue* (" Rue X " : La grammémisation à l'oeuvre dans la parole), on notera que 8 textes sur 9 contiennent plusieurs occurrences de *rue* (dans la même acception) ; le texte 015 aborde en effet un thème différent, celui de la mise en scène de la présence et de l'absence dans *Così fan tutte* – cela dit, l'article 013, qui porte sur la *Légende des siècles* contient plusieurs occurrences de *rue*.

On observe également la présence positive du *i*, *tu* et *là*, marqueurs caractéristiques de l'oral (« bon *i* suffit que tu te souviennes ... »), et du substantif *actualisation*, concept très employé par l'auteur (plus de la moitié des occ. du corpus entier).

En revanche, on observe un emploi plus faible, voire inexistant, de *sujet* (aucune occ. relevée sur un total de 540), discours (2 occ. sur 1026) et *sémiotique* (4 occ. sur 963). Barbéris est en effet plus intéressée par le concept de *parole*, souvent employé comme *synonyme* de *discours*. On note également un usage modéré de *contexte* (179 occ. sur 4962).

B. Lexique épistémique

En positif : description, explication, proposer, montrer, savoir, phénomène, comprendre

En négatif : questionnement, expérimental, présupposé

Les textes de Barbéris semblent particulièrement intéressés par le descriptif et l'explicatif plus que par le questionnement.

C. Variables morphosyntaxiques

En positif :

Avec Cordial : **deux points**, *tu*, **futur**, **verbes conjugués avec tu**, propositions indépendantes, **vous**, **verbes conjugués avec vous**, **points d'interrogation**

Avec TnT : interjections, slashes, disjoints *toi*, connecteurs de spatialité, **tu**, **points d'interrogation**, **vous**, points d'exclamation, **deux points**, auxiliaires futur et **futur**, *moi*, clitique *vous* et *te*, *tien*, *je*, symboles, disjoints *vous*, structuration, adverbes et connecteurs de conséquence

En négatif :

Avec Cordial :nb moyen de mots par phrase (= phrases plus courtes), **virgules**, **pronoms relatifs**, **négations**, **phrases avec une subordonnée**, **subordonnées**

Avec TnT : connecteurs de conclusion, de disjonction, **subordonnants**, connecteurs de reformulation, de causalité, d'exemplification, **pronoms relatifs**, **virgules** et **négations**.

Globalement dédiés à l'oral, les textes de Barbéris sont d'abord marqués par un usage important de marques de seconde personne du singulier et du pluriel (disjoints *toi* et *vous*, pronoms *tu* et *vous*, clitiques *vous* et *te*). On y relève de nombreux points d'interrogation, également caractéristiques de l'oral.

On relève avec l'étiquetage TnT une représentation importante d'interjections (0,35% de l'ensemble des tokens dans les textes de Barbéris vs. 0,04 dans les autres textes) et de slashes, liée à la présence de segments oraux dans les corps de texte, et à l'objet même de la recherche de l'auteur (l'un des articles y est même dédié « L'interjection : de l'affect à la parade, et retour »).

Les textes de Barbéris sont sur le plan stylistique caractérisés par une présence importante de verbes conjugués au futur ; on observe en effet dans les articles de nombreuses annonces de ce qui va suivre, en début d'article mais également en son corps, e.g. :

Je me limiterai cependant à l'observation des exemples fournis par le corpus, dans la mesure où ils permettent de mieux cerner les conditions d'usage de ces expressions circonstancielles. (...) Le type *rue X* sera plus particulièrement étudié (011)

Je prendrai pour exemple les énoncés (de type descriptif) par lesquels les enquêtés répondent à la question des enquêteurs sur les délimitations du quartier Saint Roch (Montpellier) (017)

On y relève en outre de nombreuses ponctuations deux points (quasi deux fois plus que dans les autres textes), très employés par l'auteur pour justifier ou exemplifier son propos :

D'autre part, les noms de lieux se référant à la voirie urbaine (*rue*, mais aussi *place*, *boulevard* etc.) ne sont pas les seuls à introduire un complément illatif/locatif sans préposition : d'autres sortes de lieux servent de support au même type de construction. (011)

Tout texte exprime un *ethos* : il possède une manière de se dire et de s'énoncer qui fait sens (Maingueneau 1993). (013)

Ce recours fort aux deux points explique les sous-représentations des connecteurs de conclusion, disjonction, reformulation, causalité et exemplification observées. Le style de Barbéris est ainsi caractérisé par une non-explicitation de certaines relations logiques, implicitement marquées par des ponctuations deux points.

La syntaxe employée semble de surcroît moins complexe : on remarque tant avec l'étiquetage Cordial que celui de TnT une représentation plus modérée des pronoms relatifs et des subordinants. Les articles de Barbéris contiendraient a fortiori 1,5 fois moins de négations que les autres textes du corpus, ce qui est remarquable.

7.3.1.3. *Bergounioux*

A. Lexique

En positif : psychologie, linguistique, française, parole, la

En négatif : nous, c, sémiotique, politique, peut

Les mots employés par Bergounioux reflètent globalement les thèmes abordés dans ses textes. Fait intéressant, on remarque une sous-représentation du pronom personnel *nous*, qui n'a d'ailleurs été relevée par aucun des deux systèmes d'annotation, contrairement au modal *peut*, corréléable aux quelques modaux présent relevés.

B. Lexique épistémique

En positif: démontrer, méthodologie, étudier, expérimentation, matériau, impliquer, typologie, présupposé, réflexion

En négatif: -

On voit ainsi un intérêt marqué pour l'ensemble des étapes de la démarche scientifique : élaboration des hypothèses (*présupposé*), construction des observables (*expérimentation, matériau*), analyse et conclusions (*démontrer, impliquer*), sans oublier le temps de l'analyse (*étudier, réflexion*).

C. Variables morphosyntaxiques :

En positif :

Avec Cordial : **conditionnel, plus-que-parfait, pronoms relatifs**, futur antérieur, **imparfait, nombre moyen de noms par proposition**, nombre total de mots signifiants, propositions avec un COI, subjonctif présent, nombre moyen de mots par proposition, verbes intransitifs, nombre moyen de mots par paragraphe, **subjonctif passé, passé simple**, nombre moyen de mots par phrase, verbes pronominaux

Avec TnT : auxiliaires conditionnel, connecteurs d'addition, **auxiliaires imparfait**, symboles linguistiques (ERREUR, assurément), modaux participe passé, **pronoms relatifs, % de noms**, participes passé, dates, **imparfait**, noms propres, modaux passé simple, auxiliaires participes passé, % de prépositions, **passé simple**, auxiliaires passé simple, **conditionnel, il** anaphorique, **auxiliaires subjonctif**, clitiques, déterminants *son, sa, ses*, % déterminants

En négatif :

Avec Cordial : **adverbes, présent, déterminants démonstratifs**

Avec TnT : pronoms démonstratifs, connecteurs de spatialité, **adverbes**, noms communs singulier, PUL, futur, infinitifs, % d'adverbes et de connecteurs, pronom *on*, connecteurs de conséquence, modaux au présent, **présent**, connecteurs de justification, **déterminants démonstratifs**

Bergounioux étant spécialiste d'histoire de la linguistique, il n'est pas surprenant que l'on observe la présence d'une dimension historico-narrative dans ses textes, particulièrement visible au regard des temps verbaux employés : on relève en effet un usage important de l'imparfait et du plus-que-parfait, du passé simple et de ses auxiliaires et modaux. Le présent et ses modaux sont en revanche sous-représentés (un tiers de moins en moyenne pour le présent – v. test de $-4,11$ – et plus de deux fois moins de modaux au présent que dans les autres textes). A fortiori les textes contiennent-ils beaucoup de noms propres et de dates.

On relève également une proportion plus importante de marques de troisième personne (*il* anaphoriques et déterminants possessifs), qui nous semble particulièrement caractéristique des textes historiques. Plus narratifs et descriptifs, les textes historiques semblent en effet davantage focalisés sur l'objet observé, qui constitue un thème plus constant de l'article. Par exemple :

Il y a un cas Albert Dauzat. Soixante années durant, **il** fut un des acteurs majeurs du champ de la linguistique et jamais **il** n'y occupa une position centrale. Les témoignages d'admiration et de gratitude ne **lui** ont pas été ménagés et pourtant **il** reste en marge, par **ses** origines et son cursus, par **ses** objets de prédilection qui n'étaient pas de ceux que l'Université reconnaît volontiers. C'est que, si Dauzat est un universitaire, **il** l'est en dehors des lieux de reconnaissance académique, hors des facultés de lettres et d'abord de la première d'entre elles, la Sorbonne. **Il** n'est parvenu ni au Collège de France, ni à l'Institut... (074)

Cette proportion importante de marques de troisième personne suggère également des textes dont la prise en charge fonctionne principalement sur le mode du **délocuté**.

On remarque également que les propositions et les paragraphes sont de longueur plus importante, ce qui dénote l'utilisation d'une syntaxe plus complexe, a fortiori marquée par de nombreuses subordonnées relatives.

Outre le passé simple et l'imparfait (et leurs homologues), on remarque un emploi remarquable du conditionnel, spécifique aux textes plus spéculatifs :

Au rebours, les idées sans mots sont illustrées par la relation d'un rêve où, se voyant dans une ville italienne dont il ne retrouve pas le nom et qui est Pérouse, Charcot l'identifie à son réveil en se rappelant auparavant le nom du Pérugin. Bref, une idée sans mot **serait** une *image*, au sens visuel du terme, sans que rien ne soit dit sur ce qui **pourrait** en demeurer hors de cette apparence sensible. (072)

On note enfin un usage significatif du mode subjonctif (présent et passé), ce qui associerait peut-être le style de Bergounioux avec un registre plus soutenu.

7.3.1.4. Calvet

A. Lexique

En positif : langues, langue, français, anglais, pays

En négatif : contexte, b, sujet, sémiotique, discours

Les mots caractérisant le style de Calvet reflètent bien les thèmes de recherche abordés dans ses articles.

B. Lexique épistémique

En positif : tendance, montrer, principe, constituer, travail, proposer, problème, choix, recherche, terme, exemple, système

En négatif : présupposé

Les termes épistémiques semblent révéler une dimension plus pragmatique des articles de Calvet, particulièrement visible avec les lexèmes *problème*, *choix*, *proposer*.

C. Variables morphosyntaxiques :

En positif :

Avec Cordial : verbes auxiliaires *avoir*, nombre moyen de mots par phrase, **imparfait**, subjonctif passé, *nous*, nombre moyen de mots par paragraphe, *ils/elles*

Avec TnT : connecteurs de concession, **imparfait**, modaux imparfait, noms propres, clitiques *nous*, cardinaux, pronoms possessifs *tien* (étonnant, erreur !), auxiliaires *été*, disjoints *eux/elles*, sigles, connecteurs de temporalité, modaux conditionnel, connecteurs de présupposition

En négatif :

Avec Cordial : **points-virgules**

Avec TnT : structuration, **points virgules**, noms communs singulier

On relève de manière générale un nombre restreint de variables caractéristiques : les textes de Calvet semblent ainsi proches de la norme générique à l'œuvre dans le corpus.

Sociolinguiste, Calvet semble pourtant avoir un style bien singulier, qui se manifeste d'abord au niveau des temps verbaux employés : contrairement à de nombreux textes d'histoire de la linguistique, qui traitent des faits historiques sur un mode narratif articulé autour de l'imparfait et du passé simple, Calvet emploie à l'instar des textes journalistiques de très nombreux imparfaits narratifs, non corrélés cette fois au passé simple :

Deux ans plus tard, lors de la cinquante-quatrième session (19 octobre 2000), le Secrétaire général de l'ONU **présentait** un nouveau rapport sur le multilinguisme, exprimant les mêmes principes que dans le texte précédent, avec des précisions supplémentaires. Puis le 27 novembre 2001 il **présentait** un nouveau rapport sur le multilinguisme (022)

Si les textes historiques contiennent quasi systématiquement dates et noms propres, Calvet emploie une proportion de noms propres qui singulariseraient son style au sein du corpus, contrairement aux proportions de dates recensées. A l'instar de l'ensemble des textes plus historiques, les textes sont peu structurés (peu de niveaux de structuration et de listes), deux fois moins en moyenne que l'ensemble du corpus.

On note en outre un emploi important de connecteurs de temporalité reliant les événements décrits, de même que de nombreux sigles, essentiellement présents dans les deux textes de l'auteur qui se rapportent aux politiques linguistiques : Calvet y énumère par exemple les (nombreux) organismes liés à la francophonie.

Le style de Calvet semble, eu égard aux connecteurs employés, plus concessif ; l'auteur emploie en outre de nombreux connecteurs de présupposition. On remarque une sous-représentation de points-virgules (0,27% de l'ensemble des ponctuations vs. 1,41% corpus), qui dénote un style peut-être plus tranché.

Enfin, l'auteur s'exprime essentiellement par le biais de *nous*, dont on observe un emploi important et particulièrement dialogique avec le lecteur (de très nombreux *nous verrons* par exemple) : c'est ainsi un *nous*-guide, tant dans ses emplois exclusifs :

Nous allons ci-dessous en présenter trois autres, reposant sur trois principes organisateurs différents : (024)

qu'inclusifs :

bilinguisme que l'on retrouve, comme **nous** allons le voir, sur les plaques des rues et des voitures, les enseignes des boutiques, etc. On y trouve tout un discours qui **nous** dit que cette ville a quelque chose de particulier. (021)

7.3.1.5. Combettes

A. Lexique

En positif : syntagme, proposition, grammaticalisation, verbe, topique

En négatif : a, politique, sémiotique, discours, et

B. Lexique épistémique

En positif : considérer, examiner, exemple, constater, unité, système, phénomène, constituer, commentaire, tendance, relever, fonction, montrer, valeur, étude, présenter

En négatif : définir, constat, présupposé, questionnement, théorie, expérimental

C. Variables morphosyntaxiques :

En positif :

Avec Cordial : **conditionnel**, **passé antérieur**, **passé simple**, **points-virgules**, verbes intransitifs, phrases contenant au moins une subordonnée, **pronoms relatifs**, propositions relatives, **déterminants indéfinis**, subjonctif plus-que-parfait, verbes à COI obligatoire, **déterminants démonstratifs**, **propositions subordonnées**, adverbes, subjonctif imparfait, **points de suspension**, déterminants possessifs *son, sa, ses*

Avec TnT : éléments de langue étrangère, connecteurs de justification, auxiliaires infinitifs, modaux conditionnel, **passé simple**, **conditionnel**, modaux présent, **points de suspension**, connecteurs de doute, *il* impersonnel, **points virgules**, structuration, **subordonnants**, modaux infinitifs, **auxiliaires passé simple**, **déterminants indéfinis**, **déterminants démonstratifs**, subjonctif présent, pronoms possessifs *leur*, auxiliaires participes présent, **pronoms relatifs**, pronoms possessifs *mien(ne)*, futur, déterminants possessifs *votre, vos*, auxiliaires conditionnel

En négatif :

Avec Cordial : nombre moyen d'adjectifs par proposition, **nombre moyen de mots par proposition**, déterminants possessifs, nombre moyen de noms par proposition, **points d'interrogation**, **points**, cardinaux, propositions avec un COD, verbes à COD obligatoire, propositions indépendantes, auxiliaires *être*, verbes transitifs indirects

Avec TnT : déterminants définis, **points d'interrogation**, connecteurs de reformulation, cardinaux, présent, déterminants possessifs *leur(s)*, clitiques, disjoints *lui, elle*, **points**, connecteurs de causalité, déterminants possessifs *son, sa, ses*, préfixes, % déterminants, % **noms**

La caractéristique **morpho-syntaxique** la plus forte des textes de Combettes est qu'il recourt à des **termes** identifiés comme **étrangers** : il s'agit d'un effet du corpus, ses articles portant souvent sur des textes en ancien français.

Autres caractéristiques fortes, l'usage de modaux, significativement plus important que dans le corpus, en particulier pour les **modaux au conditionnel** (mais aussi au présent, et à l'infinitif). On peut y voir deux tendances : une tendance à des assertions fortes, de type

déontiques (*falloir, devoir*), et une tendance à la nuance, dont on trouve d'autres indices : l'usage fort du conditionnel (presque le double en moyenne que dans le reste du corpus) et de connecteurs de doute.

Les **connecteurs de justification** sont également très représentés (c'est la seconde plus forte caractéristique des textes de Combettes via TnT).

Inversement les connecteurs de causalité et les connecteurs de reformulation font partie des variables caractérisant négativement les textes de Combettes.

Les verbes au **passé simple** sont une autre caractéristique, mise en évidence par les deux jeux d'étiquettes et l'on retrouve comme autre variable discriminante les auxiliaires au passé simple (TnT). On peut penser que c'est un effet du corpus, ou que des séquences historiques ou narratives interviennent dans les textes de l'auteur. Cela est à relier sans doute à un usage faible du présent. On notera également l'usage fort de passé antérieur mis en évidence au moyen du jeu Cordial / auxiliaires passé simple chez TnT.

La présence de verbes au subjonctif, d'auxiliaires à l'infinitif et au participe présent signale une forme de **complexité syntaxique**, à laquelle on peut indexer l'usage des pronoms relatifs et des connecteurs envisagés supra. Les textes de Combettes se distinguent ainsi par une sur-représentation des subordinants (tous types confondus).

Cette tendance syntaxique se manifeste également avec le jeu Cordial, qui montre une sur-représentation des phrases comportant une subordonnée, en particulier les relatives puisque le nombre de relatives par rapport au nombre de propositions est une autre variable discriminante, de même que le nombre de pronoms relatifs par rapport à l'ensemble des pronoms.

Le subjonctif peut se rattacher à ce qu'on a dit du doute et du conditionnel ; la présence importante de futur suggère à nouveau un mode narratif, mais permet de penser aussi que l'auteur fournit des propositions d'ordre programmatique, ou encore qu'il recourt à un abondant métalangage guidant le lecteur dans les parties du texte.

Concernant les ponctuations, les slashes et les parenthèses sont bien représentés, tandis que les crochets et les guillemets le sont peu.

Les indices de liste et de titrage se manifestent comme une caractéristique supplémentaire, alors que les numéraux sont quant à eux sous-représentés.

A noter encore, un usage fort de *il* impersonnel, mais qui rivalise avec un usage fort également de pronoms de première personne du singulier. On aurait donc un double effet relativement fort d'effacement/ présence de l'auteur, qui rejoint ce qu'on a dit des assertions.

D'autres variables restent plus difficilement interprétables (un goût pour les déterminants démonstratifs et indéfinis, une tendance à ne pas employer les déterminants possessifs singulier et pluriel, la faible représentation des déterminants définis, des pronoms clitiques et des préfixes).

Le **lexique** privilégié des textes de Combettes est nettement rattaché à leur univers conceptuel : *syntagme, proposition, grammaticalisation, verbe et topique* pourraient quasiment être envisagés comme mots-clés. *Politique, sémiotique* et *discours* sont au contraire presque totalement exclus, et l'on voit se profiler ici des paradigmes structurant le champ de la linguistique : analyse des propositions et syntagmes *versus* analyse des discours, grammaticalisation dans le sens d'une perspective historique, etc.

Concernant le **lexique épistémique**, *examiner* et *constater* ressortent comme des variables discriminant positivement les textes de Combettes. S'y ajoutent d'autres termes plus spéculatifs : *considérer*.

En résumé, un grand nombre de variables discriminantes à nouveau, qui concernent ici principalement les termes étrangers, les modaux et les temps verbaux (conditionnel d'une part, passé simple et passé antérieur d'autre part) et enfin la syntaxe des phrases, qui tend à la complexité en raison d'un usage abondant de subordonnées. On ajoutera également la préférence pour des termes épistémiques relativement spéculatifs. La comparaison des textes de Combettes par rapport aux textes du corpus genre corrobore la prégnance de certains de ces éléments dans le style de l'auteur : le conditionnel et le passé simple notamment. Par contre, l'usage des subordonnées ne se distinguerait pas tant chez Combettes par leur taux plus important, mais par une plus grande variabilité entre ses textes qu'au sein du corpus « genre ».

7.3.1.6. François

A. Lexique

En positif : 0, x, prédications, ex, prédication

En négatif : formes, qu, discours, politique, qui

B. Lexique épistémique

En positif : propriété, donnée, méthode, interprétation, viser, présenter, voir, valeur, recherche, désigner, comprendre, conceptuel, fonction, condition, observation

En négatif : présupposé

C. Variables morphosyntaxiques :

En positif :

Avec Cordial : nombre total de paragraphes, **crochets et accolades**, **passé composé**, mots signifiants, nombre total de mots signifiants, nombre total de phrases, points d'interrogation, cardinaux, propositions indépendantes, **parenthèses**, **pronoms indéfinis**, verbes conjugués à la troisième personne du singulier

Avec TnT : % symboles, sigles et abréviations, **accolades**, **auxiliaires présent**, numéraux, **crochets**, **parenthèses**, pronoms réflexifs, slashes, symboles, **pronoms indéfinis**, infinitifs, adjectifs singulier, connecteurs de temporalité, auxiliaires participes présent, prépositions

En négatif :

Avec Cordial : déterminants démonstratifs, nombre moyen d'adjectifs par proposition, propositions relatives, **conditionnel**, propositions avec un attribut du sujet, % **adverbes**, % verbes, % propositions subordonnées, nombre moyen de mots par proposition, nombre moyen de proposition par phrase, % phrases avec une subordonnée, propositions avec un COD, nombre moyen de mots par paragraphe, % mots-outils, nombre moyen d'adverbes par proposition, nombre moyen de mots par phrase, nombre moyen de pronoms par proposition, virgules, propositions avec un COI

Avec TnT : adverbess interrogatifs, dates, % déterminants, subjonctif présent, adjectifs pluriel, modaux présent, abréviations, connecteurs de concession, pronoms relatifs, **conditionnel**, % **adverbess et connecteurs**, connecteurs de justification, présent, % pronoms, virgules

Moins de variables discriminantes ici que chez Combettes : les textes de François se rapprocheraient davantage d'une norme du genre – du moins telle qu'elle se profile dans le corpus. Aucune caractéristique globale ne semble par ailleurs se dégager véritablement des quelques **variables morpho-syntaxiques** propres au style de l'auteur.

La **punctuation** cependant est l'un des axes où se manifeste le plus la singularité du style, via TnT comme via Cordial : les deux points sont la variable la plus délaissée (et les virgules, avec le jeu Cordial), les accolades sont au contraire une des plus forte caractéristique, de même que les guillemets simples, les points-virgules, et les points d'interrogation, qui vont de pair avec un usage fort d'adverbess interrogatifs.

Par ailleurs, la **modalisation** se caractérise négativement : les verbes modaux au présent sont délaissés, de même que l'usage du conditionnel, ce que confirme le jeu Cordial.

Les verbes au présent le sont également, et ce sont les auxiliaires au présent et au participe présent qui apparaissent comme des caractéristiques positives.

Les connecteurs de justification et de concession sont également sous-représentés. Avec la tendance à une modalisation peu marquée, on peut penser que ce sont des textes faiblement argumentatifs ; l'utilisation de connecteurs temporels et de numéraux de date pourrait alors suggérer que les textes comportent une dimension historique.

On peut y voir aussi une tendance générale signalée par Cordial à préférer **des phrases et des paragraphes courts** (cf. le nombre moyen de mots par P et par §), et à éviter l'usage des subordonnées. Cette tendance irait de pair avec un taux fort de numéraux et indices de structuration signalé par TnT.

Les **variables lexicales** ne semblent pas davantage faire émerger de caractéristique forte du style de François, hormis la thématique forte de la *prédication* (le terme est significativement plus présent que dans le corpus, au singulier comme au pluriel), et l'usage de *ex*, et plus largement des abréviations, mais aussi des symboles (cf. supra, et l'usage de *x*). *Sémiotique*, *discours* et *formes* se distinguent par leur absence.

Le **lexique épistémique** indique une tendance à faire un usage abondant de termes dédiés au corpus : *donnée*, *présenter*, *voir* avec Cordial. D'autres termes cependant concernent plutôt le positionnement doctrinal : *méthode*.

Les quelques caractéristiques observées restent identiques lorsque l'on compare les textes de François au corpus « genre » : en particulier, le taux faible de verbes au conditionnel, l'usage important des points d'interrogation et adjectifs interrogatifs et enfin l'usage de phrases et de paragraphes courts (les phrases comportent peu de mots et de propositions, les paragraphes peu de mots) caractérisent le style de l'auteur au regard des tendances du genre.

7.3.1.7. Kerbrat

A. Lexique

En positif : politesse, vous, femmes, interaction

En négatif : nous, du, politique, sémiotique, la

B. Lexique épistémique

En positif : questionnement, expérimental, problématique, schéma, indiquer, présupposer

En négatif : -

C. Variables morphosyntaxiques :

En positif

Avec Cordial : *vous*, **deux points**, verbes conjugués avec *vous*, déterminants possessifs *notre*, *nos*, adverbess, déterminants indéfinis, impératif, **déterminants démonstratifs**, verbes conjugués avec *je*, **points-virgules**, parenthèses, phrases exclamatives, points d'exclamation, *je*, auxiliaires *être*, propositions coordonnées

Avec TnT : guillemets, **deux points**, déterminants possessifs *vo*tre *vos*, % adverbess et connecteurs, **déterminants démonstratifs**, structuration, **points virgules**, impératif, connecteurs de reformulation, préposition, % subordonnants, modaux présent, adverbess

En négatif

Avec Cordial : nombre moyen de mots par proposition, nombre total de mots signifiants, nombre moyen de noms par proposition, **points**

Avec TnT : % prépositions, virgules, participes passés, **points**

Moins de variables discriminantes encore, que ce soit via le jeu Tnt ou celui de Cordial. Les deux modes de caractérisation présentent ici nombre de points communs qui permettent d'avancer quelques tendances fortes :

On retrouve la part importante jouée par les **punctuations** dans la caractérisation, déjà perçue pour les textes d'Authier et de François. Comme chez Authier se manifeste chez Kerbrat une tendance à faire un usage important de punctuations énonciativement marquées – qu'il s'agisse des points d'exclamation, variable discriminante la plus forte dans le jeu TnT, ou de marques d'incise, les parenthèses étant sur-représentées. A ces mêmes variables s'ajoutent dans le jeu Cordial les deux points et points-virgules. La forte présence de points d'exclamation va également de pair dans Cordial avec une sur-représentation des phrases de type exclamatif.

Le second axe fort qui distingue les textes de Kerbrat des autres textes du corpus concerne les **personnes** : on observe via le jeu Cordial un usage fort de la 1^{ère} et de la 2^{ème} personne du pluriel, qui va de pair avec une proportion forte de verbes à ces personnes. TnT met en avant l'importance des déterminants possessifs de la 2^{ème} personne du pluriel. Il faut sans doute voir dans la présence de cette dimension interlocutive 1^{ère}/2^{ème} personne un effet du corpus : avec les points d'exclamation, on peut penser que l'on a affaire à un corpus oral. C'est ce qui explique peut-être la sur-représentation des tirets et indices de structuration.

Les variables lexicales corroborent la présence forte de la 1^{ère} personne : *nous* est une des variables discriminant positivement les textes de Kerbrat. Les autres termes concernent

davantage le domaine d'investigation (*politesse, interaction*), ce qui confirme l'hypothèse de corpus oraux, de type conversationnalistes.

La comparaison des textes de Kerbrat avec les textes du corpus « genre » fournit comme pour les autres auteurs des résultats proches de ceux obtenus au sein du corpus « auteur ». Ce sont ainsi les ponctuations, en particulier les points d'exclamation, les deux points et les tirets qui caractérisent le style de Kerbrat au regard des tendances centrales du genre. Par ailleurs, les verbes aux 1^{ère} et 2^{nde} personne du pluriel représentent une autre caractéristique forte, dont on a dit, comme de la première, qu'elle relève vraisemblablement du corpus conversationniste employé par l'auteur.

Différence notable entre l'analyse au sein du corpus « auteur » et celle menée par comparaison des textes de Kerbrat avec le corpus « genre » : les pronoms personnels de 1^{ère} et 2^{nde} personne du pluriel ; ils sont suffisamment fréquents pour influencer significativement sur les verbes, mais restent en moyenne plus faiblement représentés que dans le corpus « genre ». Cette première différence importante observée entre les deux analyses (corpus auteur/ corpus genre) permet d'avancer que le corpus « auteur » manque peut-être d'individus pour être représentatif en lui-même du genre.

7.3.1.8. Kleiber

A. Lexique

En positif : noms, ça, sens, dénomination, nom

En négatif : politique, discours, dans, de, la

B. Lexique épistémique

En positif : question, notion, idée, discussion, caractérisation, observable, choisir, approche, conséquence

En négatif : -

C. Variables morphosyntaxiques :

En positif

Avec Cordial : auxiliaires *avoir*, % **adverbes**, **négations**, phrases avec une subordonnée, propositions subordonnées, adjectifs interrogatifs, % **verbes**, **articles indéfinis**, auxiliaires *être*, nombre moyen d'adverbes par proposition, propositions avec attribut du sujet, propositions principales, verbes conjugués avec *nous*, pronoms démonstratifs, **déterminants indéfinis**

Avec TnT : % **adverbes et connecteurs**, % **verbes**, % subordonnants, connecteurs de causalité, guillemets simples, déterminants possessifs *nos*, *notre*, % pronoms, connecteurs de disjonction, connecteurs de présupposition, connecteurs de certitude, **négations**, subjonctif présent, **déterminants indéfinis**, clitiques *nous*, modaux infinitif, clitiques, *il* impersonnel, futur

En négatif

Avec Cordial : nombre moyen de mots par proposition, **points-virgules**, articles définis/articles, articles définis/déterminants, **nombre moyen de noms par proposition**, propositions indépendantes, nombre total de mots signifiants

Avec TnT : **points-virgules**, participes présent, pronoms indéfinis, % ponctuations, adverbes, connecteurs d'addition, % prépositions, **% noms**

La lecture de ces textes ne fait pas de doute : Kleiber a un style personnel, une façon bien à lui d'écrire le genre, qui va jusqu'à se jouer de ses normes, tout en les respectant. La caractérisation et la classification automatiques font-elles émerger ces spécificités ? Ou mettent-elles en avant au contraire l'inscription de ces textes dans les régularités du genre ?

Le faible nombre de **variables morpho-syntaxiques** discriminantes dans la caractérisation indiquent que c'est plutôt la seconde piste qui l'emporte. Notons cependant que les subordinants d'une part, les adverbes et connecteurs d'autre part sont très représentés, on le voit via TnT et Cordial ; TnT permet de préciser que les connecteurs de causalité, de disjonction, de présupposition, de certitude sont significativement plus nombreux que chez les autres auteurs, les connecteurs d'addition étant quant à eux significativement sous-représentés.

Ils signalent une **syntaxe complexe**, bien mise en évidence au moyen du jeu Cordial qui signale un taux fort de phrases comportant une subordonnée, et de manière redondante, de subordonnées parmi l'ensemble des propositions (indépendantes, principales, subordonnées).

Par ailleurs, Cordial et Tnt indiquent que les **adverbes de négation** sont particulièrement abondants, ce qui suggère, avec les connecteurs, des textes comportant une importante dimension argumentative. L'usage fort des clitiques pourrait ainsi s'expliquer par leur rôle double de structuration et d'emphase.

L'usage abondant de subjonctif présent peut s'expliquer en termes syntaxiques, en lien avec l'abondance de connecteurs, et en termes énonciatifs, en lien avec l'usage de modaux (à l'infinitif).

Le futur peut signaler des textes prospectifs, programmatiques, ou témoigner d'un usage abondant de métalangage, hypothèse que suggère également l'emploi de déterminants possessifs de 1^{ère} personne du pluriel (« notre étude »).

D'autres variables restent difficiles à interpréter : la forte présence de verbes et de pronoms, contre celle de noms et de prépositions.

Lexicalement, les variables discriminantes rejoignent le constat fait pour les autres auteurs : *politique* et *discours* apparaissent comme des caractéristiques négatives, tandis que les variables positives signalent les spécificités des domaines thématiques abordés dans les textes, et leur dimension conceptuelle ; aussi trouve-t-on ici, avec *sens* et *dénomination*, la forme singulier et plurielle de *nom*. La sur-représentation de *ça* est plus problématique, d'autant que c'est là un terme quasi absent du genre de l'article : a-t-on affaire à un élément autonome étudié par l'auteur ?

Les termes épistémiques discriminants corroborent l'hypothèse de textes prospectifs et argumentatifs, dédiés à la discussion théorique : *question*, *notion*, *idée*, *discussion* caractérisent positivement les textes de Kleiber.

La comparaison des textes de Kleiber avec ceux du corpus « genre » permet d'asseoir certaines tendances observées dans le corpus « auteur » : les indices de complexité syntaxique (subordonnées notamment) discriminent le style de l'auteur au regard des tendances centrales

du genre, de même que la négation, variable discriminante dans le style de cet auteur seulement.

On notera par ailleurs que l'analyse menée avec le corpus « genre » fournit davantage de termes épistémiques discriminants ; certains sont ambigus, car ils entrent dans le champ conceptuel de l'auteur et jouent ainsi le rôle de lexique de spécialité (par ex., désigner). D'autres renforcent l'idée avancée à partir du corpus « auteur » qu'il s'agit de textes théoriques, faisant large place à l'argumentation (*argument, concept, conceptuel, présupposé* sont significativement plus fréquents dans les textes de Kleiber que dans le genre)

7.3.1.9. Lamizet

A. Lexique

En positif : sémiotique, politique, symbolique, espace, public

En négatif : plus, x, 1, pas, 2

On retrouve l'ensemble des thèmes abordés par l'auteur, qui semblent très spécifiques dans la mesure où la plupart des autres auteurs sont négativement caractérisés par les thèmes de Lamizet.

On remarque en négatif un emploi faible des numéraux, également observé au niveau morphosyntaxique.

B. Lexique épistémique

En positif : résultat, critère, investigation, définition, systématique, conception, classification, concept, approche, matériau, questionnement, taxinomie, conséquence, déterminer

En négatif : travail, présenter

On observe un emploi important des termes épistémiques relatifs au travail de classification/comparatif (*classification, taxinomie* et éventuellement *critère*).

C. Variables morphosyntaxiques :

En positif :

Avec Cordial : **virgules**, mots-outils, nombre moyen de mots par paragraphe, **présent**, nombre moyen de mots par phrase, propositions avec un COI, nombre moyen de mots par proposition, **articles définis**, verbes à COI obligatoire, nombre moyen de pronoms par proposition, **points, déterminants possessifs PP3PL**, nombre moyen de propositions par phrase, propositions indépendantes, propositions avec un COD, **pronoms relatifs**

Avec TnT : numéraux ordinaux, % déterminants, **virgules**, connecteurs de conclusion, PUL, connecteurs de doute, **déterminants possessifs PP3PL, présent**, % prépositions, connecteurs d'addition, symboles grammaticaux, **points**, abréviations, **déterminants définis, pronoms relatifs**, % noms, connecteurs de conséquence

En négatif :

Avec Cordial : **déterminants démonstratifs, verbes conjugués à la première personne du singulier, points de suspension**, nombre moyen de phrases par paragraphe, **propositions subordonnées, points virgules, je, adjectifs cardinaux**, impératif, pronoms personnels sujets, **deux points**, négations, nombre moyen d'adverbes par proposition, adverbes, **futur**, verbes intransitifs, mots signifiants, **parenthèses**

Avec TnT : guillemets, **points virgules, je, points de suspension**, symboles, noms propres, **déterminants démonstratifs**, structuration, connecteurs de concession, % punctuations, préfixes, déterminants indéfinis, **deux points**, connecteurs de disjonction, % **subordonnants**, connecteurs d'opposition, % **numéraux**, participes passés, **futur, parenthèses**

Le style de Lamizet se caractérise d'abord par un usage important de virgules (1,5 fois plus que la moyenne des autres textes), ce qui confère à ses textes un rythme, manifeste dans ses textes :

C'est par la main que le corps signifie, car c'est la main qui, à la fois, établit la relation entre le corps et les choses ou entre mon corps et celui de l'autre, et permet d'orienter *le corps*, de lui donner son sens, puisque c'est par rapport à elle que mon corps distingue la gauche et la droite. La main, en ce sens, *institue le code du corps*, c'est elle qui fait du corps une forme symbolique, que l'on va pouvoir lire, interpréter, à laquelle on va pouvoir reconnaître une signification en en reconnaissant l'orientation. (091)

De surcroît observe-t-on un emploi plus important de points, qui nous semble aller de pair avec les proportions plus faibles de points-virgules, de deux points et de points de suspension relevés. Le style de Lamizet serait peut-être ainsi plus arrêté.

Les paragraphes, phrases et propositions de l'auteur sont de manière générale de taille plus importante, tandis que les textes sont en moyenne trois fois moins structurés.

Le style de l'auteur semble globalement moins polémique, comme en témoignent les faibles proportions observées de connecteurs d'opposition (quasi trois fois moins, alors que ce sont les troisièmes connecteurs les plus représentés du genre) et de concession ; Lamizet privilégie ainsi les connecteurs de conclusion, de doute, d'addition et de conséquence.

On remarque également une proportion plus importante de déterminants définis, et un emploi faible de verbes intransitifs : Lamizet privilégierait ainsi les verbes à deux, voire trois arguments.

Si les articles sont fortement rédigés au présent, ils contiennent enfin très peu de futurs (7 fois moins, ce qui est notable). On relève également peu de noms propres, de numéraux, et de pronoms *je*.

7.3.1.10. Moirand

A. Lexique

En positif : folle, vache, scientifiques, libération, science

En négatif : x, nous, dire, sémiotique, est

B. Lexique épistémique

En positif : expliquer, observer, observation, expérimentation, notion, réflexion, typologie, démonstration, fonction, paradigme, examen, étude, étudier, schéma, analyser

En négatif : -

Neveu paraît privilégier un lexique épistémique peut-être plus conceptuel.

C. Variables morphosyntaxiques :

En positif :

Avec Cordial : **points de suspension**, verbes conjugués à la 3^e personne du pluriel, articles définis/articles, nombre moyen de noms par proposition, nombre moyen de lettres par mot, articles définis/déterminants, nombre moyen de mots par proposition, nombre moyen d'adjectifs par proposition, verbes intransitifs, propositions relatives, verbes à COI obligatoire

Avec TnT : noms communs pluriel, adjectifs pluriel, amalgames pluriel, sigles, dates, modaux participes passés, participes passés, connecteurs de disjonction, *on*, guillemets simples, % prépositions, disjoints *eux/elles*, **points de suspension**

En négatif :

Avec Cordial : déterminants possessifs, verbes conjugués avec *nous*, articles indéfinis, auxiliaires *être*, propositions avec attribut du sujet, verbes, **points**

Avec TnT : auxiliaires participes passés, impersonnels *il*, structuration, négations, % verbes, *on*, **points**, noms communs singulier, prépositions, adjectifs singulier

Les textes de Moirand semblent se situer dans un entre deux entre caractéristiques centrales du genre et spécificités stylistiques - **morpho-syntaxiques** d'abord. Les variables discriminantes ne sont pas aussi nombreuses que chez Authier ou Lamizet. Des tendances stylistiques émergent cependant ; ainsi, l'usage du **pluriel des syntagmes nominaux** : les noms, adjectifs et déterminants contractés pluriels sont significativement plus abondants que dans le corpus, et les noms et adjectifs singuliers significativement sous-représentés.

Les textes de Moirand semblent se situer dans un entre deux entre caractéristiques centrales du genre et spécificités stylistiques - **morpho-syntaxiques** d'abord. Les variables discriminantes ne sont pas aussi nombreuses que chez Authier ou Lamizet. Des tendances stylistiques émergent cependant ; ainsi, l'usage du **pluriel des syntagmes nominaux** : les noms, adjectifs et déterminants contractés pluriels sont significativement plus abondants que dans le corpus, et les noms et adjectifs singuliers significativement sous-représentés.

Les aspects énonciatifs ne paraissent pas particulièrement discriminants ; notons toutefois que les verbes modaux au participe passé sont bien représentés, mais c'est le cas aussi des verbes en général.

D'autres variables sont plus difficilement interprétables, telles la sur-représentation de locutions prépositionnelles, contrastant avec la sous-représentation de prépositions. Ou encore, la forte présence de connecteurs de disjonction, qui ne s'accompagne pas toutefois d'une sous-représentation significative d'autres types de connecteurs.

D'un point de vue **lexical**, ce n'est pas tant avec cet auteur le domaine de spécialité, en termes conceptuels, qui semble l'emporter que son corpus de prédilection dans les articles ici étudiés : *folle* et *vache* constituent ainsi les premières places des variables positives discriminantes. Viennent ensuite *scientifiques*, *Libération* et *science*, où l'on retrouve le poids du corpus (Libération), conjoint à celui de l'objet étudié (discours médiatiques et scientifiques).

On notera qu'outre *dire* et *sémiotique*, *nous* est une des variables caractérisant négativement les textes de Moirand, qui semblent finalement peu recourir aux marques personnelles (cf. supra, aucune marque personnelle ne s'avère discriminante), sans nécessairement préférer des tours impersonnels (cf. supra : il impersonnel = caractéristique négative).

Le **lexique épistémique** est tout particulièrement important dans la caractérisation des textes de Moirand : ce sont eux qui émergent d'abord de l'analyse (surtout avec le jeu d'étiquettes Cordial) : *expliquer*, *observer*, *observation* sont les plus fréquents, suivis par *expérimentation*. En raison de leur taux important, on peut se demander dans quelle mesure ces termes peuvent être ambigus, et faire sens par rapport au corpus plutôt qu'en qualité de termes épistémiques, en particulier pour les termes *expérimentation* et *expliquer*.

L'analyse menée avec le corpus « genre » met en évidence la forte présence de substantifs et d'adjectifs dans les textes de Moirand. Ces indices d'un style descriptif s'accompagnent également d'indices énonciatifs, en particulier les types assertifs (phrases interrogatives) et les points de suspension, virgules et parenthèses.

7.3.1.11. Neveu

A. Lexique

En positif : apposition, support, segment, détaché, référentielle

En négatif : x, les, politique, sémiotique, nous

B. Lexique épistémique

En positif : objectif, hypothèse, argument, paradigme, présupposé, analyser, définir, tableau, concevoir, désigner, théorie

En négatif : -

Neveu paraît privilégier un lexique épistémique peut-être plus conceptuel.

C. Variables morphosyntaxiques :

En positif :

Avec Cordial : **nombre moyen d'adjectifs par proposition**, nombre moyen de lettres par mot, pronoms relatifs, **adjectifs**, nombre moyen de mots par proposition, verbes conjugués à la 3^e personne du sg., nombre moyen de noms par proposition, nombre moyen d'adverbes par proposition

Avec TnT : pronoms relatifs, % **adjectifs**, connecteurs de conséquence, participes passés, adjectifs singulier, connecteurs de reformulation et de causalité, noms communs singulier, connecteurs de concession, virgules, subordonnants

En négatif :

Avec Cordial : **nous sujet, verbes conjugués avec nous**, pronoms personnels sujets, verbes, **deux points**

Avec TnT : points de suspension, **nous, clitiques nous**, noms communs et adjectifs pluriel, **deux points**

Le style de Neveu semble plus conforme aux canons du genre, dans la mesure où peu de variables lui sont caractéristiques. C'est apparemment au niveau de son emploi des adjectifs qu'il se distinguerait : il en emploierait de 1/5^e à 1/3 de plus.

Les mots des textes semblent en outre de longueur plus importante (5,2 lettres vs. 4,9 dans l'ensemble du corpus, v. *test* de 4).

On note une sous-représentation des marques de première personne du pluriel et des ponctuations deux points.

7.3.1.12. Rabatel

A. Lexique

En positif : pdv, narrateur, personnage, perceptions, mi

En négatif : langues, politique, français, sémiotique, langue

B. Lexique épistémique

En positif : vérifier, comparaison, interprétation, condition, désigner, classification

En négatif : -

C. Variables morphosyntaxiques :

En positif :

Avec Cordial : imparfait, **subjonctif plus-que-parfait**, verbes conjugués à la 3^e personne du singulier, passé simple, **déterminants démonstratifs**, plus-que-parfait, verbes transitifs directs, verbes, auxiliaires *être*

Avec TnT : sigles linguistiques, renvois dans le texte, **déterminants démonstratifs**, amalgames singulier, pronoms démonstratifs, connecteurs d'opposition, % subordonnants, présent, **subjonctif plus-que-parfait**, % adverbes et connecteurs

En négatif :

Avec Cordial : présent, passé composé, verbes à COD obligatoire, futur, verbes transitifs directs

Avec TnT : modaux infinitif, modaux présent, auxiliaires participes passés, déterminants possessifs *leur*, abréviations

Les textes de Rabatel signalent une tendance relativement centripète par rapport aux textes du corpus. Au niveau **morpho-syntaxique**, les **acronymes** (de type *SN*, *SV*), et les **abréviations** linguistiques (*prep.*) caractérisent le style de l'auteur. Par contre, les abréviations en général le caractérisent négativement.

Les **numéraux à fonction de renvoi** sont particulièrement abondants ; ils peuvent en particulier être liés à la numérotation des exemples.

Concernant les verbes, on notera que **l'imparfait et le passé simple** sont mis en avant comme variables discriminantes par Cordial. La dimension historique qui se dessine potentiellement ici peut cependant résulter du corpus employé, le subjonctif imparfait ressortant des deux jeux utilisés comme une caractéristique forte également.

Les verbes modaux au présent et à l'infinitif sont sous-représentés. Dans une perspective énonciative, ajoutons que les connecteurs d'opposition caractérisent les textes de Rabatel, ainsi indicieux d'une importante dimension argumentative, à moins qu'encore une fois, ce ne soit là que l'effet de l'objet traité par l'auteur et/ou de son corpus.

Concernant les personnes, on observe une sur-représentation des verbes à la troisième personne du singulier (étiquetage Cordial), ce qui suggère des textes dont la prise en charge fonctionne principalement sur le mode du **délocuté**.

Enfin, on remarquera que l'auteur a tendance à employer beaucoup de pronoms et déterminants démonstratifs, et qu'il préfère ces derniers aux déterminants possessifs - une des unités des plus délaissées par rapport aux autres textes du corpus.

En termes **lexicaux**, ce qui distingue les textes de Rabatel de ceux des autres auteurs tient aux concepts qu'il mobilise, et leur singularité en linguistique : on trouve en effet *point de vue*, sous la forme abrégée *pdv* - qui peut expliquer en partie l'abondance d'abréviations linguistiques, *narrateur*, *personnage* et *perceptions*.

Le **lexique épistémique** dans les textes de Rabatel suggère à la fois une dimension empirique et une démarche herméneutique : on relève ainsi, tous jeux confondus, les termes *vérifier*, *interprétation*, *comparaison*, *classification*.

Les textes de Rabatel se distinguent des textes du corpus « genre » sur des variables confirmant les tendances observées avec le corpus « auteur » : l'imparfait est particulièrement sur-représenté dans les textes de Rabatel par rapport aux tendances centrales du genre, et dans une moindre mesure le passé simple. A contrario, le futur et le passé composé caractérisent négativement le style de l'auteur.

A la différence de l'analyse au sein du corpus « auteur », la comparaison avec le corpus genre montre que le subjonctif imparfait n'est pas sur-représenté chez Rabatel. Par contre, elle rejoint l'hypothèse avancée supra de textes fonctionnant beaucoup sur le mode du délocuté : les textes de Rabatel contiennent significativement plus de verbes aux 3^{ème} personnes du singulier et du pluriel que les caractéristiques du genre ne le laissent attendre.

7.3.1.13. Rastier

A. Lexique

En positif : ontologie, textes, corpus, genres, texte

En négatif : dans, politique, ce, c, x

B. Lexique épistémique

En positif : théorique, méthodologique, examen, définir, constat, expérimental, problématique, problème, procéder, définition, catégorisation, représenter, méthodologie, réflexion, tableau, conclusion, conceptuel, décrire, étudier, démarche, conception, traiter, investigation, discussion

En négatif : indiquer

Le nombre de termes épistémiques caractérisant positivement le style de Rastier est particulièrement frappant.

C. Variables morphosyntaxiques :

En positif :

Avec Cordial : nombre moyen de lettres par mot, articles définis, **points**, **verbes conjugués à la 3^e personne du pluriel**, verbes intransitifs, nombre moyen de propositions par phrase, nombre total de phrases verbales, **négations**

Avec TnT : connecteurs d'exemplification, anaphoriques singulier *il/elle*, abréviations, **points**, amalgames pluriel, noms communs pluriel, **anaphoriques pluriel *ils/elles***, modaux futur, connecteurs de conclusion, % déterminants, connecteurs de causalité, **négations**, déterminants possessifs de 3^e personne du pluriel, clitiques, adjectifs pluriel, modaux présent

En négatif :

Avec Cordial : **pronoms démonstratifs**, nombre moyen de phrases par paragraphe, **points de suspension**, *je*, nombre moyen de **pronoms par proposition**, **déterminants démonstratifs**

Avec TnT : **pronoms démonstratifs**, subordonnants, adjectifs singulier, **points de suspension**, % **pronoms**, symboles, prépositions, participes présents, **déterminants démonstratifs**

Le style de Rastier se singularise surtout au niveau de son emploi massif d'un nombre important de termes épistémiques. On relève finalement peu de variables morphosyntaxiques qui le caractérisent.

Rastier emploie d'abord un nombre moyen de lettres par mot plus important (5,15 vs. 4,9), lié à l'érudition des textes. On note en outre une proportion plus importante de négations et de points, qui va de pair avec un usage très faible (et d'ailleurs constant) des points de suspension (moins de 0,1% de l'ensemble des ponctuations vs. 1,3% en moyenne).

On relève dans les textes de Rastier de nombreux éléments au pluriel (amalgames⁹, noms communs et adjectifs), et un emploi important des anaphoriques pluriel *ils/elles* et singulier *il/elle*, de même qu'une proportion importante de déterminants possessifs de 3^e personne du pluriel.

L'auteur emploie comparativement peu de démonstratifs (déterminants et pronoms).

⁹ Prépositions + déterminants.

Au niveau des connecteurs employés, on note un emploi important de connecteurs d'exemplification (plus de deux fois plus), de conclusion et de causalité.

Enfin, on relève peu de symboles dans les textes, et un emploi privilégié des verbes intransitifs.

7.3.1.14. *Rosier*

A. Lexique

En positif : p, nous, discours, espèce, mon

En négatif : deux, sciences, sens, espace, sémiotique

B. Lexique épistémique

En positif : expérimental, théorie, présumé

En négatif : présenter

C. Variables morphosyntaxiques :

En positif :

Avec Cordial : verbes conjugués avec *nous*, *nous*, phrases exclamatives, points d'exclamation, déterminants possessifs, impératif, phrases interrogatives, pronoms personnels sujets, déterminants possessifs *mon*, *ma*, *mes*, parenthèses, points d'interrogation

Avec TnT : *nous*, déterminants possessifs *notre*, *nos*, points d'interrogation, impératif, disjoints *nous*, deux points, déterminants possessifs *mon*, *ma*, *mes*, déterminants possessifs *ton*, *ta*, *tes*, parenthèses, clitiques *nous*

En négatif :

Avec Cordial : verbes conjugués à la 3^e personne du singulier, articles définis, pronoms relatifs

Avec TnT : virgules, pronoms démonstratifs

Les variables discriminant positivement les textes de Rosier évoquent les caractéristiques **morpho-syntaxiques** de ceux d'Authier et de Kerbrat : ils semblent se distinguer en termes de prise en charge énonciative, de ponctuation et de types assertifs.

L'usage du pronom personnel *nous* est la variable qui caractérise le mieux ces textes au sein du corpus, avec l'usage de déterminants possessifs de cette même personne. Le jeu Cordial montre des résultats tout à fait convergents de ce point de vue : les verbes et les pronoms personnels à la 1^{ère} personne du pluriel caractérisent les textes de Kerbrat par leur taux important. Avec TnT, les pronoms disjoints de cette même personne représentent une autre variable discriminant les textes de Rosier. On remarque que les déterminants possessifs de 1^{ère} et deuxième personne du pluriel sont aussi plus abondants que dans le corpus, ainsi que les adjectifs possessifs (toutes personnes confondues selon Cordial).

Ces variables sont généralement très marginales dans le genre, si bien que leur taux important peut être envisagé ici comme le résultat de l'objet d'étude et/ou du corpus de Rosier. Les déterminants possessifs de 1^{ère} du singulier pourraient s'expliquer de la même façon, sauf à signaler une alternance entre l'emploi du *je* et *nous* de l'auteur – qui au regard des pratiques d'écriture du genre reste cependant assez peu probable (*nous* inclusif ? Emploi en tout cas universalisant de *je* :

La construction *Vdire + que* superpose cette modalisation (je prends position) et la mention des marques typographiques (je cite), produisant un effet d'hypperréalisme : c'est bien comme cela que s'opère le travail de citation dans la presse. (084)

qui explique également les marques possessives de première personne du singulier :

Si j'occupe une position dominante dans l'interaction ou une position sociale qui *autorise* mon discours, quelles sont les conséquences en termes de construction de ma position d'énonciateur dans mon discours et, ce faisant, de l'ethos que je construis également ? (086)

L'usage des **verbes à l'impératif** corrobore l'hypothèse que ces spécificités énonciatives relèvent du corpus utilisé par l'auteur et de son objet d'étude, de même que les points d'interrogation, qui peuvent signaler une modalité d'énonciation particulière (la question rhétorique et l'attitude d'implication du lecteur qu'elle suggère) ou relever d'un corpus conversationnel, journalistique, etc. Le jeu Cordial confirme les résultats obtenus via TnT, soulignant le taux important de phrases interrogatives, et y ajoutant également le taux de points d'exclamation et de phrases exclamatives.

Lexicalement, on retrouve la prégnance de la première personne du pluriel, *nous* étant le terme qui discrimine le plus les textes de Rosier au sein du corpus. Mais *mon* est également discriminant, ce qui soulève à nouveau la question de savoir si ces indices de personne sont le fait du corpus, et dans quelle mesure. Un article porte d'ailleurs sur *nous autres*, ce qui explique l'abondance de *nous* disjoints remarquée par TnT.

Le terme de *discours* identifie quant à lui un domaine de spécialité (cf. article « La circulation des discours à la lumière de 'l'effacement énonciatif' : l'exemple du discours puriste sur la langue. »), tandis que le terme *espèce* est davantage ambigu : s'agit-il d'une forme de « tic » de langage de l'auteur, ou d'une des unités soumises à l'analyse ? *Sciences, sens, espace* et *sémiotique* sont a contrario des indices des concepts délaissés par l'auteur au regard de leur présence dans le corpus.

Les **termes épistémiques** récurrents dans les textes de Rosier sont *expérimental, théorie, présumé*.

Ce sont les aspects énonciatifs que la comparaison entre les textes de Rosier et le corpus « genre » montre comme discriminants : l'usage de *nous*, de verbes à la 1^{ère}, mais aussi à la 2^{nde} personne du pluriel, de phrases interrogatives et exclamatives, et de ponctuations déterminant ces types assertifs (! et ?), de points de suspension, et de parenthèses. A cela s'ajoute encore l'usage de l'impératif. On le voit, l'analyse menée sur le corpus « genre » aboutit à des conclusions identiques à celles menées sur le corpus « auteur ». Elles sont sans doute pour beaucoup le fait de la spécificité du domaine d'investigation de l'auteur.

7.3.1.15. *Siblot*

A. Lexique

En positif : Congo, prédicat, thème, nom, dénomination

En négatif : exemple, ou, sémiotique, espace

B. Lexique épistémique

En positif: démonstration, démarche, analyser, conséquence, choisir, catégorisation, procéder, problème, constater, indiquer, systématiser, conclusion

En négatif: -

C. Variables morphosyntaxiques :

En positif :

Avec Cordial : nombre moyen de propositions par phrase, **points**

Avec TnT : **points**, pronoms possessifs de 1^e personne du pluriel, guillemets, amalgames singulier, connecteurs de présupposition, négations

En négatif :

Avec Cordial : -

Avec TnT : adverbes, connecteurs de reformulation

Peu de variables morphosyntaxiques caractérisent le style de Siblot : Cordial ne décèle qu'une longueur plus élevée des propositions employées et un nombre plus élevé de points, confirmé par TnT, qui note en outre un usage singulier des pronoms possessifs de 1^e personne du pluriel, des guillemets, connecteurs de présupposition et négations.

Cordial ne décèle aucun emploi négatif qui distinguerait le style de Siblot ; TnT note une présence plus faible d'adverbes et de connecteurs de reformulation (trois fois moins).

7.3.1.16. Synthèse

A. Style et lexique

1. Style et épistémologie

Le lexique joue un rôle relativement important dans la caractérisation des styles personnels. Ce sont principalement les spécificités-mots qui présentent ce potentiel discriminant. Ainsi le style de Lamizet manifeste-t-il sa spécificité par les concepts qu'il mobilise dans ses textes : *sémiotique*, *politique*, *symbolique*, *espace*, *public* sont des caractéristiques qui distinguent positivement ses textes des autres textes du corpus, par l'usage abondant que l'auteur en fait.

Ces variables lexicales positives rendent nettement compte des principaux thèmes abordés par l'auteur, qui semblent très spécifiques dans la mesure où la plupart des autres auteurs sont caractérisés négativement par ces mêmes thèmes. Exemples les plus parlants, *politique* et *sémiotique* caractérisent positivement les seuls textes de Lamizet et représentent respectivement les caractéristiques négatives de 8 et 11 auteurs.

Tous les auteurs sont loin d'être concernés par l'importance du lexique en tant que caractéristique positive de leur style. Outre Lamizet, Calvet se distingue également par un usage important de *langues*, *langue*, *français*, *anglais*, *pays*, termes qui reflètent encore une fois les thématiques privilégiées de l'auteur.

On peut faire l'hypothèse que les auteurs dont les styles se caractérisent par des termes essentiellement rattachés à la dimension thématique et conceptuelle des textes ont des domaines d'investigation relativement à la marge dans le corpus ; une telle approche, menée sur des corpus de taille plus importante, permettrait ainsi de dessiner une cartographie du champ, les concepts discriminants ébauchant des paradigmes structurant le corpus et la discipline.

Le graphique qui suit, qui représente les mots non-vides des textes sur les deux premiers axes factoriels, permet ainsi de voir où se positionnent les auteurs par rapport aux termes qu'ils emploient : la singularité de Lamizet est particulièrement frappante, tandis que Combettes se démarquerait également en positif sur le facteur 2 :

Le poids du corpus ne se manifeste pas seulement au niveau lexical : il joue un rôle important au niveau morphosyntaxique, comme on a pu le voir *infra* (textes avec/sans exemples).

B. Style et morphosyntaxe

On insistera ici, à titre illustratif, sur le style d'auteur de Combettes, qui se distingue particulièrement au niveau morphosyntaxique.

Parmi les variables morphosyntaxiques qui caractérisent l'usage singulier que cet auteur fait du genre, on observe un emploi plus important de modaux, et de modaux au conditionnel en particulier (bien que les modaux au présent, et à l'infinitif soient également sur-représentés dans les textes). On peut y voir deux tendances opposées mais non contradictoires : (i) un *penchant pour les assertions fortes*, de type déontique (*falloir, devoir*), corrélées à un emploi important de connecteurs de justification (seconde caractéristique observée des textes de Combettes via TnT) et (ii) une *tendance à la nuance*, dont on relève d'autres indices : un usage important du conditionnel (deux fois plus en moyenne que dans le reste du corpus) et des connecteurs de doute. Soulignons que ces caractéristiques rappellent celles de la dimension plus spéculative observée au sein du corpus générique (chapitre 3), auquel étaient déjà rattachés les textes de Combettes.

Associés à des modalités de narration, les temps verbaux semblent particulièrement discriminants en matière de styles : on retrouve ainsi la présence d'une dimension historico-narrative particularisée par un usage important de l'imparfait et du plus-que-parfait, du passé simple et de ses auxiliaires et modaux, des dates et des noms propres dans les textes de Bergounioux, spécialiste d'histoire de la linguistique. Le présent et ses modaux sont en revanche sous-représentés (un tiers de moins en moyenne pour le présent – v. test de -4,11 – et plus de deux fois moins de modaux au présent que dans les autres textes), et on relève une proportion plus importante de marques de troisième personne (*il* anaphoriques et déterminants possessifs), qui nous avait semblé particulièrement caractéristique des textes historiques.

Ces spécificités font encore une fois écho aux dimensions de l'article mises au jour dans le chapitre 3, d'autant que le numéro thématique de Bergounioux sur la parole intérieure se trouvait bien démarqué du corpus.

Signalons enfin que les ponctuations et les indices de complexité syntaxique (nombre moyen de mots par propositions, subordonnées et subordonnants) se sont également révélés particulièrement discriminants dans notre entreprise.

C. Styles et épistémiques

Le nombre de termes épistémiques caractérisant positivement le style de Rastier est particulièrement frappant. C'est même à ce niveau que son style se singularise le plus, davantage qu'au niveau morphosyntaxique, qui s'avère globalement peu distinctif :

En positif : théorique, méthodologique, examen, définir, constat, expérimental, problématique, problème, procéder, définition, catégorisation, représenter, méthodologie, réflexion, tableau, conclusion, conceptuel, décrire, étudier, démarche, conception, traiter, investigation, discussion

En négatif : indiquer

On notera plus largement que ce lexique épistémique est particulièrement productif pour la caractérisation des styles (cf. *supra*), essentiellement en termes de caractéristiques *positives* : 60 termes contribuent par exemple à la caractérisation de quatre auteurs et plus. C'est donc en termes de sur-représentation, plutôt que de sous-représentation, que les termes épistémiques jouent un rôle discriminatoire ; chez Rastier, l'abondance de ces termes peut signaler des textes fortement orientés dans une perspective théorique, et offrant des éléments de discussion épistémologique.

7.3.2. Seconde exploration du corpus « Auteurs » : axes d'organisation interne et distinctivité des styles

Une fois les spécificités des 15 auteurs mises au jour, nous avons observé les pôles d'opposition morphosyntaxiques du corpus « Auteurs », qui ont été contrastés aux axes d'organisation du corpus « Genre » (cf. chapitre 3) ; les styles qui se démarquaient le plus au sein du corpus « Auteurs » ont ensuite été examinés au regard de ces pôles de contraste.

Dans cette perspective, une ACP a été menée sur le corpus « Auteurs ». On examinera les premiers axes factoriels mis au jour, avant d'apprécier la distinctivité des 15 styles observés à l'aide d'ellipses de confiance, et leur positionnement sur les axes obtenus.

7.3.2.1. Les styles d'auteur et leurs axes de contraste interne

Comme on avait pu l'observer pour le corpus « Genre », on relève un palier dans la décroissance des premières valeurs propres. Soulignons que les 2 premiers facteurs extraient cette fois 16.38% du nuage de points (*vs.* 12.41%), ce qui est lié à la taille plus restreinte du corpus observé. On s'intéressera particulièrement à ces deux premiers axes factoriels, qui se démarquent encore une fois significativement des autres.

Nb	Valeur propre	% d'inertie	% cumulé	
1	12.89	9,28	9,28	*****
2	9.86	7,10	16,38	*****
3	7.86	5,66	22,03	*****
4	7.11	5,12	27,15	*****
5	6.66	4,79	31,94	*****
6	5.63	4,05	35,99	*****
7	5.32	3,83	39,82	*****

Figure : Diagramme des 7 premières valeurs propres (corpus « Auteurs »)

A titre de rappel, reconsidérons – sans le réinterpréter – le premier plan factoriel obtenu sur le corpus « Genre » :

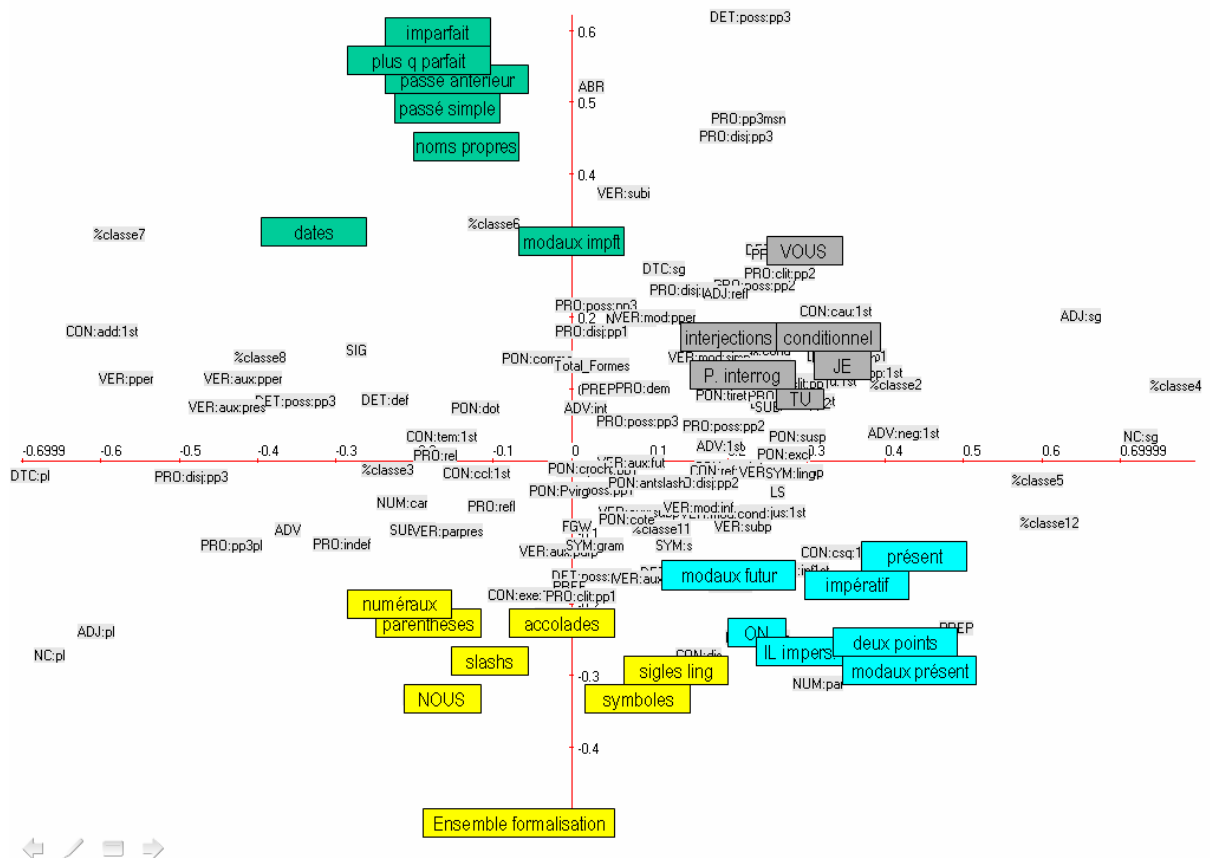


Figure : Positionnement des variables sur les deux premiers axes factoriels – corpus « Genre »

Afin de faciliter la comparaison, les groupements de descripteurs précédemment observés ont été repositionnés sur le nouveau plan factoriel, obtenu sur le corpus « Auteurs » :

renvoyant à une distinction entre textes historiques et appliqués et d'un pôle contenant les caractéristiques énonciatives du discours scientifique les plus observées dans la littérature (*il* impersonnel, *on*, présent et modaux présent, etc.).

Les deux corpus « Genre » et « Auteurs » présentent ainsi des similarités frappantes : le corpus « Auteurs » porte bien les traces d'une organisation morphosyntaxique que l'on peut tenir pour *générique*, et l'on peut *a fortiori* considérer les pôles mis au jour dans les deux corpus comme *caractéristiques* du genre de l'article de linguistique. En d'autres termes, une représentation forte des styles d'auteur ne bouleverse pas outre mesure les pôles d'organisation générique, ce qui attesterait bien de la présence d'un noyau dur générique au sein du corpus « Auteurs ».

7.3.2.2. *Distinctivité des styles d'auteur*

Nous avons ensuite cherché à déterminer si les styles d'auteur étaient significativement distincts au sein du corpus « Auteurs ». Dans cette perspective, nous avons utilisé la méthode du bootstrap (Lebart, 2004) via DTM afin de tracer des zones de confiance autour des 15 modalités de la variable supplémentaire « Auteurs » prise en compte, ce qui nous donne la représentation graphique suivante :

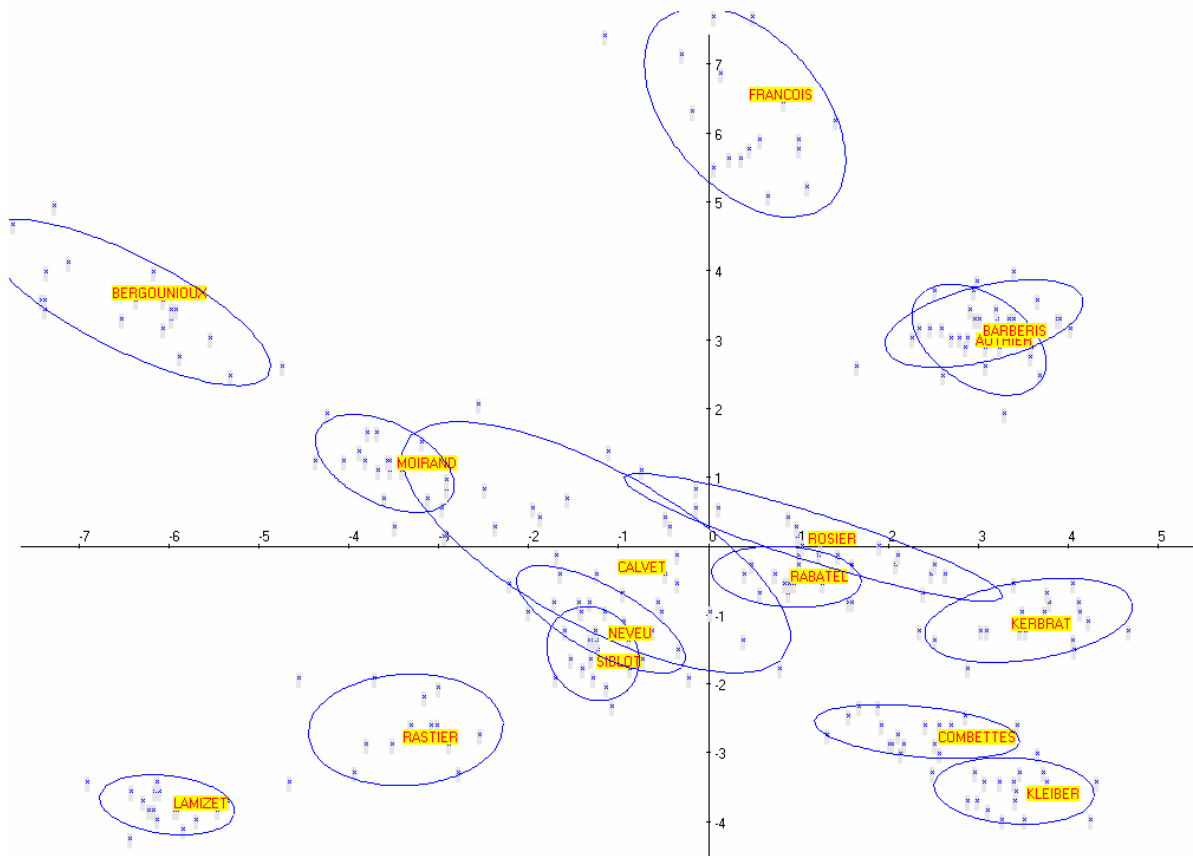


Figure : Ellipses de confiance autour des 15 auteurs observés sur le premier plan factoriel – corpus « Auteurs »

Notons d'emblée que les ellipses sont relativement petites, ce qui atteste bien d'une significativité de la variable stylistique.

On remarque d'abord que six styles sur quinze sont significativement distincts (les ellipses ne se recouvrent pas) : les styles de Combettes, François, Kleiber, Bergounioux, Lamizet et Rastier.

Si l'on se réfère à l'examen des axes factoriels de la section précédente, les styles isolés renvoient aux différents pôles mis au jour : le style de Bergounioux est effectivement marqué par un emploi intensif du passé simple, de l'imparfait (et de leurs homologues composés), des dates et des noms propres – Bergounioux étant précisément spécialiste d'histoire de la linguistique –, tandis que le style François est marqué par de nombreuses formalisations.

On notera également le recouvrement entre Authier et Barbéris, spécialistes du discours travaillant sur matériel oral.

Le style d'auteur le moins distinct sur le plan factoriel est le style de Calvet : on note en effet 5 recouvrements stylistiques et l'ellipse est plus grande.

7.4. Genre de l'article et styles d'auteurs scientifiques : exploration des relations entre style et genre

On s'intéresse ici plus particulièrement aux relations entre styles et genre : si le corpus « Auteurs » semble bien porter les traces d'une organisation morphosyntaxique générique, il convient d'apprécier la distinctivité des styles au sein d'un corpus homogène en genre, mais stylistiquement moins marqué. On réitérera donc le test du bootstrap sur les deux corpus fusionné (7.4.1.). Le caractère significativement distinct des styles sera enfin apprécié par le biais d'une CAH, en termes de bonne / mauvaise classification (7.4.2.).

7.4.1. Distinctivité des styles d'auteurs au sein du genre

Si les styles d'auteur se distinguent relativement les uns des autres au sein de leur corpus, il convient maintenant d'apprécier leur significativité au sein du corpus « Genre », dans lequel la dimension stylistique est neutralisée. Nous avons ajouté la modalité « Autre », qui réfère à l'ensemble des auteurs du corpus genre, aux 15 modalités précédentes de la variable supplémentaires « Auteurs ».

L'expérience, qui se veut comparable à celle menée en 7.4.2.2., est fondée sur le même système de description (étiquetage par TnT).

Le caractère distinct des styles au sein de la combinaison des deux corpus démontre bien la présence de singularités stylistiques au sein du genre.

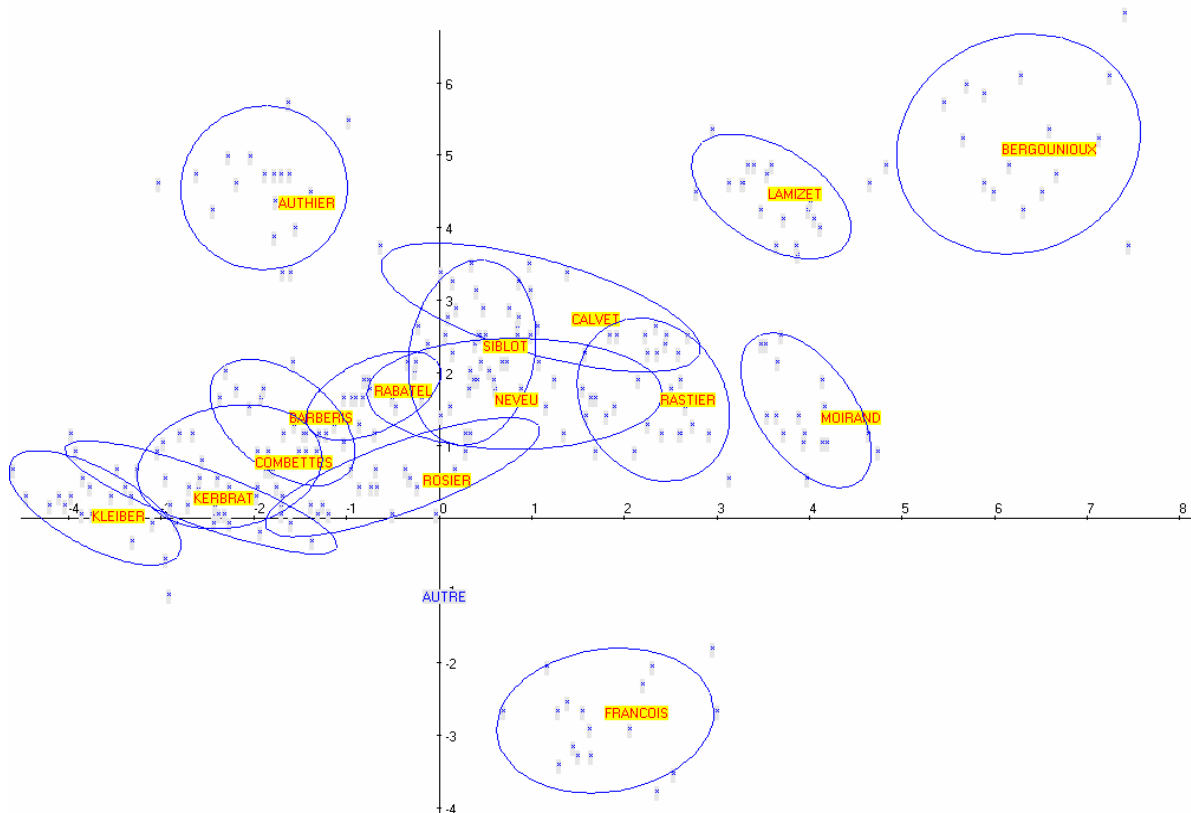


Figure : Ellipses de confiance autour des 15 auteurs observés sur le premier plan factoriel – corpus « Genre » + « Auteurs »

Cinq styles sur quinze sont cette fois significativement distincts : on retrouve les styles de François, Bergounioux et Lamizet déjà isolés dans le corpus « Auteurs ». En revanche, si les styles de Authier et Barbéris étaient recouverts dans le premier corpus, on note qu'ils sont cette fois bien distincts sur le plan factoriel.

Les styles qui se démarquent significativement sont naturellement positionnés aux extrémités des pôles morphosyntaxiques mis au jour (mode de narration plus historique, formalisation et travaux sur matériel oral).

On observe que si la plupart des styles d'auteur restants se recouvrent, ils s'opposent globalement à l'ensemble des autres auteurs (variable « Autre »), ce qui démontre que la variable stylistique est bien neutralisée dans le corpus.

7.4.2. Analyse complémentaire par classification ascendante hiérarchique avec prise en compte de trois systèmes de descripteurs

Une entreprise de classification ascendante hiérarchique a été menée parallèlement au bootstrap, mais à partir des trois systèmes de descripteurs considérés.

Un style a été considéré comme *homogène* si la (quasi) totalité des textes le constituant étaient regroupés dans une même partition, et *singulier* si cette même classe s'avérait isolée. En assignant des scores de 1 en cas d'*homogénéité*, et de 2 en cas d'*homogénéité ET de singularité*, on obtient le tableau suivant :

Légende

Style homogène et singulier : 2

Style homogène (et non singulier) : 1

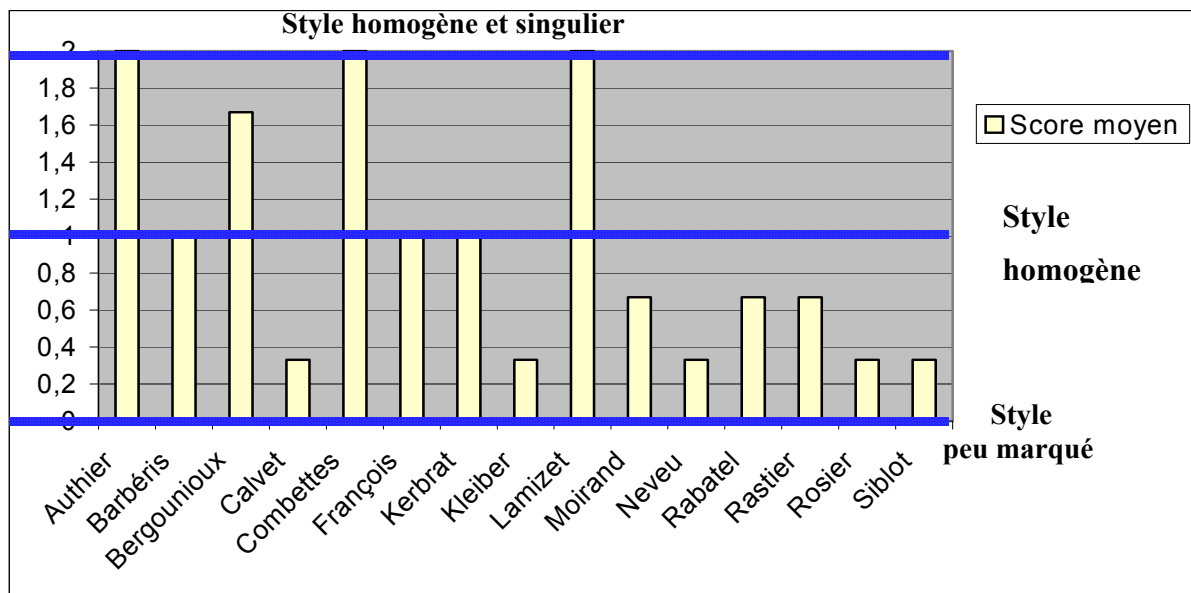
Style non marqué : 0

	TnT	TnT + épisté	Cordial + épisté	Moyenne	Total
Authier	2	2	2	2	6
Barbérís	2	1	0	1	3
Bergounioux	2	2	1	1,67	5
Calvet	0	1	0	0,33	1
Combettes	2	2	2	2	6
François	2	1	0	1	3
Kerbrat	1	1	1	1	3
Kleiber	0	0	1	0,33	1
Lamizet	2	2	2	2	6
Moirand	1	1	0	0,67	2
Neveu	0	1	0	0,33	1
Rabatel	1	0	1	0,67	2
Rastier	0	1	1	0,67	2
Rosier	1	0	0	0,33	1
Siblot	1	0	0	0,33	1

Tableau : Scores de singularité et d'homogénéité des styles d'auteur

Les résultats obtenus diffèrent sensiblement lorsqu'on considère les deux types d'annotation morphosyntaxique, d'ailleurs complémentaires comme nous l'avons déjà souligné dans la section 3.1. Ainsi, les styles de Kleiber et de Rastier sont moins marqués avec TnT qu'avec Cordial. A l'inverse, les textes de Barbérís, François et Moirand qui semblaient homogènes, voire singuliers avec l'étiquetage TnT pour Barbérís et François (cas 1), sont non marqués avec Cordial.

On remarque que certains auteurs, comme Combettes ou Lamizet ressortent fortement, indépendamment du système de description pris en compte. Si l'on calcule la moyenne des scores obtenus par les auteurs avec les trois systèmes de variables, on obtient la représentation suivante, qui distingue les styles selon leur degré de singularité :



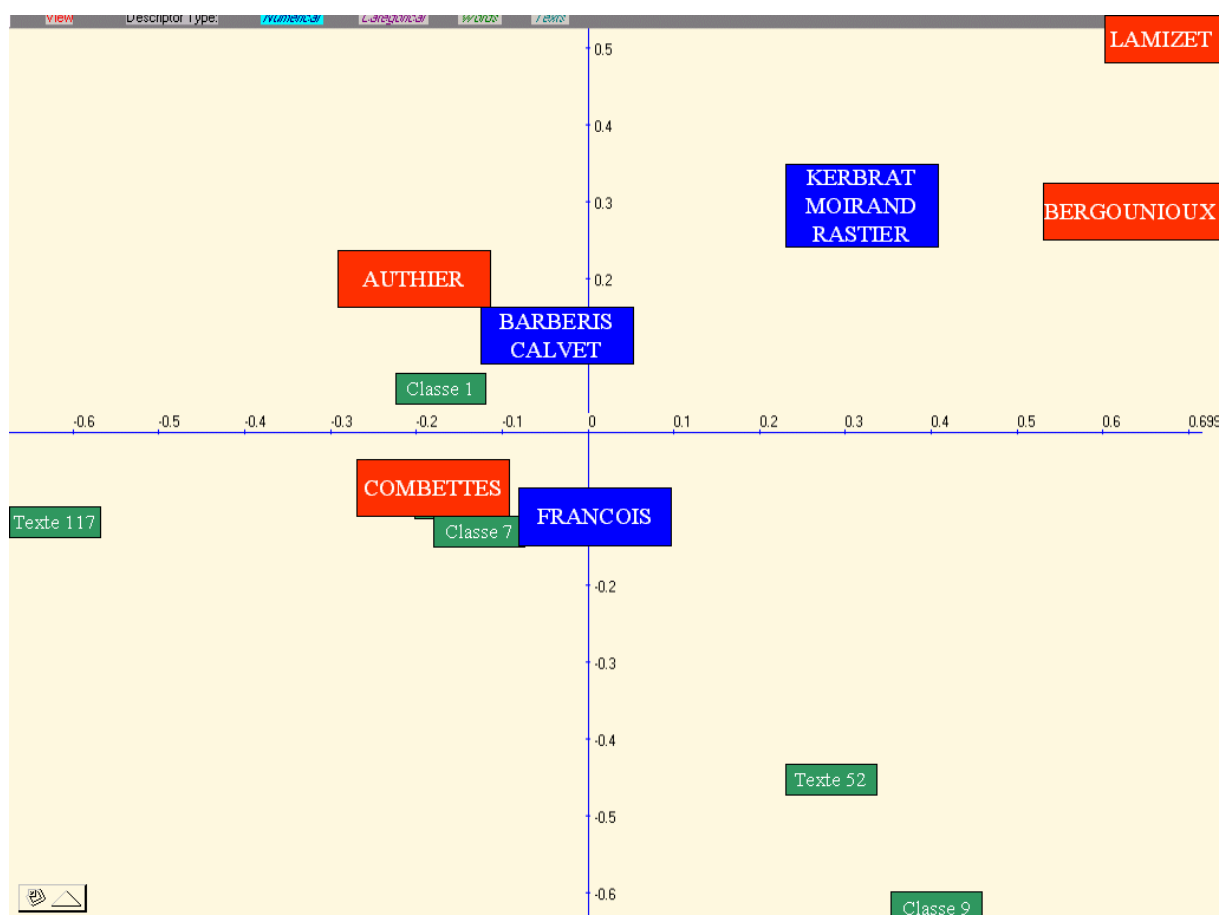
Graphique : Singularité et homogénéité des styles d'auteur

Authier, Lamizet, Combettes et Bergounioux sont ainsi les quatre styles qui se distinguent le plus : les textes des trois premiers auteurs sont en effet systématiquement regroupés, ce qui atteste de la présence de caractéristiques stylistiques stabilisées, et ils sont isolés des autres au sein d'une classe qui leur est quasi dédiée, ce qui dénote une singularité stylistique.

On observera que le style de Bergounioux se distingue également très honorablement, mais de manière moindre si l'on considère la classification fondée sur l'étiquetage Cordial, qui ne prend en compte que les temps verbaux, et non les noms propres et les dates, fortement caractéristiques du style (et plus exactement du domaine scientifique de rattachement) de l'auteur. Les textes de Bergounioux sont quasi systématiquement rassemblés, ce qui témoigne d'une grande homogénéité stylistique, et isolés dans 2 cas sur 3, ce qui démontre une singularité de l'auteur.

On observe le même phénomène avec les textes de Kerbrat, qui sont également systématiquement réunis, mais au sein de classes plus importantes : le style de Kerbrat serait ainsi identifiable, mais les caractéristiques de l'auteur seraient finalement moins saillantes, dans la mesure où elles seraient partagées par de nombreux textes. On notera que Barbéris, François, Calvet, Rastier et Moirand se trouvent dans un cas similaire.

La sortie de DTM qui suit illustre bien la singularité de ces styles par rapport aux autres, en termes de partitions stylistiques :



Graphique : Annotation TnT + épistémique

Eu égard aux résultats obtenus à partir de notre entreprise de caractérisation, les descripteurs « responsables » de la distinction des auteurs sont relatifs aux spécificités des textes se rapportant à l'oral (chez Authier nombre important de *je*, de points de suspension ou encore de crochets), aux temps verbaux relevant du narratif comme on l'a vu chez Bergounioux, aux indices d'un style plus spéculatif comme celui de Combettes et enfin, pour Lamizet, à un usage important de virgules au sein de paragraphes, de phrases et de propositions de taille plus importantes dans des textes moins structurés.

En d'autres termes, ces différences sont naturellement liées à une sur- ou sous-représentation de descripteurs, et on note que ces descripteurs semblent avoir d'abord trait aux ponctuations (et aux temps verbaux en ce qui concerne deux des quatre premiers auteurs), ce qui restera à approfondir de manière plus claire.

Les analyses que l'on a menées pour l'heure à ce sujet montrent que dans l'ensemble les termes épistémiques ont un potentiel discriminant particulièrement élevé, et que les deux termes les plus discriminants, en tant que caractéristique positive ou négative des styles, sont *experimental* et *présupposé*. D'un point de vue morpho-syntaxique, ce sont les variables en lien avec la complexité syntaxique et les ponctuations qui seraient les plus discriminantes.

Il convient bien sûr d'interroger la valeur, et la nature discriminante des systèmes de description pris en compte : que ce soit à partir des variables TnT, Cordial, ou du lexique épistémique, la classification met en évidence certains styles d'auteur au détriment d'autres. Authier, Combettes ou encore Lamizet semblent ainsi singuliers, d'une part parce que leurs textes ont des caractéristiques morphosyntaxiques et épistémiques stables, et d'autre part

parce que ces spécificités ne semblent pas partagées par les autres textes de l'ensemble du corpus « genre ».

Cela ne signifie naturellement pas que les auteurs que notre analyse a distingué ont un style qu'on tiendra pour particulièrement original ou interpellant à la lecture : ainsi, Kleiber, dont les linguistes connaissent bien les spécificités stylistiques et l'originalité des articles – notamment sa tendance à se jouer des normes du genre, en les explicitant, en les commentant, ou même en usant d'images et d'expressions idiomatiques inattendues et complètement marginales dans le genre de l'article, e.g. :

Conclusion

C'est fini. Du moins, pour le moment. (05)

n'aurait pas un style singulier – ni même homogène – si l'on s'en fie à la présente étude. Plus justement, on peut dire que les particularités du style de Kleiber ne se situent pas aux niveaux morphosyntaxiques et épistémiques, dont il intègre parfaitement les normes.

Les styles qui se singularisent par des biais non pris en compte ici sont ainsi difficilement repérables et peuvent facilement être tenus pour parfaitement « conformes » aux caractéristiques morphosyntaxiques et épistémiques du genre. Les éventuelles marques humoristiques des textes n'ont ainsi pas été balisées...

Etant donné qu'il n'est pas envisageable, et encore moins réalisable, d'être exhaustif en termes de descripteurs, nous nous en sommes tenue au niveau morphosyntaxique, bien connu pour son impact sur les styles, et au niveau épistémique, visiblement discriminant.