

Contrastes internes et variations stylistiques du genre de l'article scientifique de linguistique

Céline Poudat¹, Fanny Rinck²

¹CORAL, UFR Lettres, Langues et Sciences Humaines – 10, rue de Tours – 45072 Orléans
Cedex 02

²Lidilem - B.P. 25 - 38040 Grenoble Cedex

Abstract

The present study focuses on the relations between the genre of scientific articles, and author styles in the linguistic field. Our objective is to describe the internal variations of scientific articles and their singular use within two corpora “Authors” and “Genre”. The study is based on the morphosyntactic level, and we use exploratory techniques to assess the stability of the two corpora – and the possible presence of a generic kernel – as well as the significance of the stylistic variable. The presence of similar places of stability in both corpora, and the visibility of the stylistic variables in the “Genre” corpus help in validating the two notions, although the results call for further analysis taking into account other genres.

Résumé

L'étude porte sur les relations entre genre et style dans l'article de recherche en linguistique. Il s'agit de décrire la variabilité interne du genre en s'intéressant à l'usage singulier que chaque auteur en fait, i.e. le style, à partir de deux corpus, un corpus « auteur » et un corpus « genre ». À partir de méthodes d'analyse exploratoire de données textuelles, on se fonde sur la dimension morphosyntaxique des textes pour apprécier la stabilité des axes de contraste interne des deux corpus – et la présence d'un noyau générique – et évaluer la significativité de la variable stylistique. La mise en évidence de lieux de stabilité dans la variabilité de l'article et de la visibilité des styles d'auteur dans le genre valide en partie les deux notions, bien qu'elle appelle à analyser les relations genres/styles dans un corpus intégrant des genres de textes variés.

Mots-clés : analyse exploratoire de données textuelles, statistique textuelle, linguistique de corpus, analyse des genres, styles d'auteur.

1. Le genre et son usage singulier, le style

En tant que principe de structuration et instance de normalisation des pratiques langagières, le genre se caractérise par des propriétés linguistiques régulières et conventionnelles, qui ne doivent cependant pas occulter la diversité des textes. Aussi, si l'article de recherche est généralement envisagé comme un genre contraint, fortement normé voire routinisé, il se définit comme tout genre du double point de vue de sa stabilité et de sa variation intrinsèque : y compris au sein d'un même domaine disciplinaire, les textes produits et publiés en tant qu'articles font état de cette diversité des pratiques, dont on propose de rendre compte ici en s'intéressant aux axes de contraste internes au genre, d'une part en termes de spécificités morphosyntaxiques des textes, d'autre part en examinant les relations entre le genre et l'usage singulier que chaque auteur en fait, le *style*.

En considérant avec Rastier (2001) les styles en tant qu'usages singuliers du genre, nous avons dans une précédente étude (Poudat & Rinck, 2005, à paraître) cherché à caractériser les styles d'auteur en linguistique et montré que certains styles se démarquaient plus que

d'autres. Entreprise sur la base d'un corpus d'articles de 15 linguistes, cette exploration des spécificités stylistiques appelle à approfondir les relations entre le genre de l'article et les styles d'auteur scientifiques de manière plus précise.

L'analyse de styles d'auteur doit se coupler d'un retour sur le genre : elle met le genre à l'épreuve du style, en montrant si les caractéristiques des textes restent stables quels que soient les auteurs entrant dans le corpus. Autrement dit, il s'agit de tester l'hypothèse de la permanence du genre en observant si l'on retrouve les mêmes axes de contraste dans un corpus stylistiquement marqué.

D'autre part, les relations entre le genre et l'usage singulier que chaque auteur en fait, le style, appellent une seconde question : les styles personnels sont-ils significativement distincts une fois leurs textes inclus dans un corpus d'articles produits par des auteurs variés, à même donc de neutraliser cette variable « style » ?

La question de la relation entre genre et style participe de la description du genre dans ses lieux de variation et de stabilité ; en confrontant les usages du genre de l'article, elle vise à s'interroger sur la validité des deux notions par leur mise en relation en corpus.

2. Aspects méthodologiques

L'analyse est fondée sur deux corpus « auteur » et « genre » décrits ci-après. Elle s'effectuera en deux temps : il s'agit dans un premier temps de déterminer si les deux corpus sont organisés selon les mêmes axes morphosyntaxiques. On s'intéressera dans un second temps aux styles, pour apprécier leurs (non-)recouvrements au sein des deux corpus.

2.1. Corpus d'étude

Deux corpus sont mobilisés : un corpus « Genre » constitué de 224 articles extraits de 32 numéros de revues (soit 11 revues) francophones de sciences du langage, dans lequel la variation stylistique est neutralisée (226 auteurs y sont représentés) et un corpus « Auteurs », qui rassemble 130 articles de 15 linguistes français accrédités dans le champ. L'ensemble des articles rassemblés a été publié entre 1998 et 2003.

Étant donné les particularités de l'étude, nous avons choisi d'écarter les exemples des textes, fortement susceptibles de corrompre leurs caractéristiques linguistiques. Afin de ne pas briser la continuité syntaxique des textes, nous avons choisi cependant de n'isoler que les exemples délimités dans les textes : les textes portent donc la trace de leur objet linguistique, mais de manière toutefois moindre.

En effet, bien que les corpus soient globalement homogènes en termes de domaine, dans la mesure où seuls des articles de linguistique y sont représentés, ils sont inévitablement partitionnés – de manière d'ailleurs souvent peu transparente¹ – en différentes sous-disciplines linguistiques (sémantique, syntaxe, phonologie, etc.), induisant d'autres types de variation malheureusement inévitables. On notera que le corpus « Auteurs » contient un nombre beaucoup plus restreint de domaines, étant donné la spécialisation des chercheurs dans leur champ.

¹ En effet, il n'existe pas, à notre connaissance, de cartographie précise et délimitée des branches de la linguistique française, ce qui pose des problèmes de représentativité et d'homogénéité certains. On note en outre une plus grande polyvalence des linguistes français par rapport aux chercheurs anglo-saxons, plus spécialisés.

2.2. Système de descripteurs adopté

Les niveaux de description possibles d'un genre et d'un style étant multiples, nous nous sommes essentiellement attachés au niveau morphosyntaxique, d'abord parce qu'il a montré son efficacité en matière de typologies textuelles² (validation et exploration) et ensuite parce qu'il représente un niveau intéressant particulièrement la description des langues. En outre, c'est encore le niveau le plus développé en terme d'annotation de corpus et de nombreux outils d'étiquetage automatique sont disponibles.

Nous avons donc adopté un jeu de 140 descripteurs dédiés à l'étude du discours scientifique³, étiquetés après entraînement par TnT (Trigrams'n'Tags, Univ. Saarlandes, Brants) – et manuellement vérifiés. Les variables sont morphosyntaxiques, voire sémantiques, et incluent des caractéristiques supposées du discours scientifique (types de connecteurs, *il* impersonnels vs. anaphoriques, modaux, symboles, indices de structuration et de renvois dans le texte, etc.) et linguistique (* et ? d'acceptabilité linguistique, abréviations linguistiques comme SN, SV ou GN, etc.).

C'est à partir de ce système de description que seront observés les deux corpus « Genre » et « Auteurs ».

2.3. Analyses statistiques

Nous avons mené sur les deux corpus une ACP associée à une CAH, à partir du jeu de catégories adopté. Cette opération nous a permis d'obtenir les champs de contraste et les axes d'organisation principaux des deux corpus.

Le caractère significativement distinct des styles d'auteur a été apprécié avec la technique du bootstrap.

L'ensemble des tests statistiques a été réalisé avec le logiciel DTM développé par Lebart⁴.

3. Comparaison des axes d'organisation des deux corpus

L'étude des relations entre genre et styles d'auteur nécessite une description précise des deux corpus observés. On tentera d'abord de dégager les axes d'organisation du corpus « Genre » (3.2.), avant d'observer le corpus « Auteurs » selon la même procédure.

3.1. Le genre de l'article et ses axes de contraste interne

On observe un pallier dans la décroissance des quatre premières valeurs propres de l'ACP, qui fournissent 19.99% de l'inertie. Dans la mesure où ils se dégagent significativement, on s'intéressera particulièrement à la description des deux premiers axes factoriels.

Nb	Valeur propre	% d' inertie	% cumulé	
1	10.90	7.79	7.79	*****
2	6.46	4.62	12.41	*****
3	5.40	3.86	16.27	*****
4	5.20	3.72	19.99	*****
5	4.09	2.93	22.92	*****

Figure 1 : Diagramme des 5 premières valeurs propres

² Pour les entreprises spécifiquement attachées à la description des genres, voir par exemple Malrieu et Rastier (2001), Poudat (2003).

³ Poudat (2004).

⁴ DTM étant la version universitaire de SPAD-T <http://www.enst.fr/egsh/lebart/>

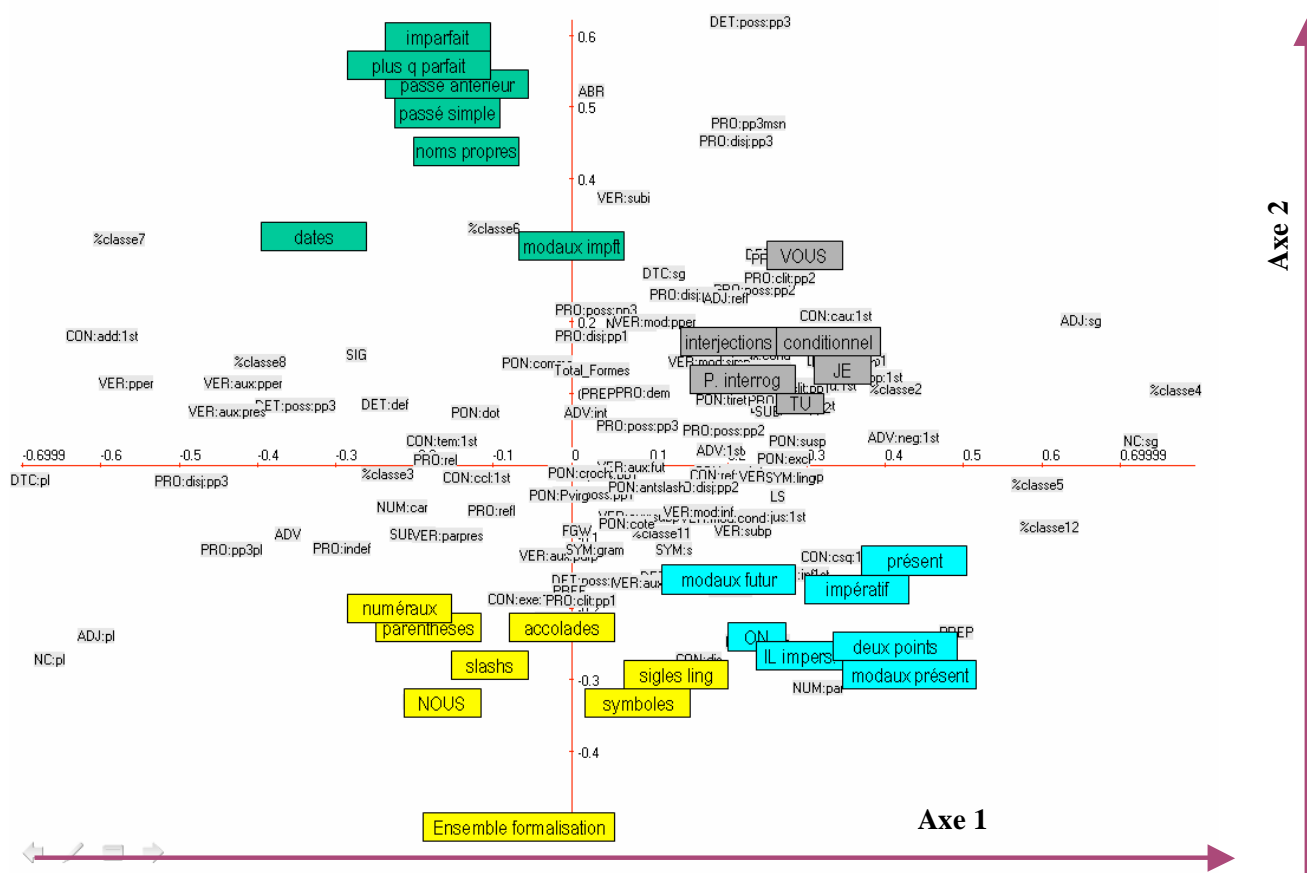


Figure 2 : Positionnement des variables sur les deux premiers axes factoriels – corpus « Genre »⁵

Deux grands modes narratifs s'opposent sur les facteurs 1 et 2 (voir Figure 2) : un *mode narratif à visée atemporelle* typique des textes scientifiques, et d'ailleurs caractérisé par l'usage intercorrélé des temps les plus représentés du corpus et des pronoms impersonnel et indéfini *il* et *on* (*en bleu* : présent, impératif, futur...), et un *mode de narration plus historique* et plus proche du discours romanesque, défini par l'emploi intercorrélé du passé simple et de l'imparfait, et de leurs homologues composés et modaux (*en vert*).

Associés aux dates et aux noms propres, qui confirment bien la présence d'une composante historique, ce dernier groupement verbal s'oppose sur le premier axe à un ensemble de descripteurs caractéristiques de l'oral (*en gris* : marques de première et de seconde personne, points d'interrogation et d'exclamation, interjections, etc.), spécifiques des articles plus exemplifiés travaillant sur corpus oraux, et sur le deuxième axe aux marqueurs de formalisation (*en jaune* : symboles, formalisations linguistiques, slashes, parenthèses et accolades).

On observe ainsi quatre pôles morphosyntaxiques du genre, qui reflètent un contraste entre les textes plus appliqués (plus formalisés et exemplifiés) et les articles de linguistique historique.

⁵ Pour des raisons de lisibilité du graphique, nous avons fait ressortir les groupes de descripteurs les plus notables.

Cette tension des textes plus exemplifiés et historico-narratifs se confirme lorsqu'on examine le positionnement des textes sur les deux premiers facteurs : on retrouve en effet l'ensemble des textes appartenant aux numéros thématiques et aux revues plus historiques en positif sur l'axe 2. Trois grandes tendances – de nature toutefois distincte – régulent ainsi la morphosyntaxe du genre de l'article : son mode narratif (historique ou atemporel), son degré de formalisation et l'usage d'un corpus d'exemples oraux.

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

Si l'on observe également un pallier 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 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 3 : Diagramme des 7 premières valeurs propres

Nous avons repositionné sur le premier plan factoriel les groupements de descripteurs observés précédemment :

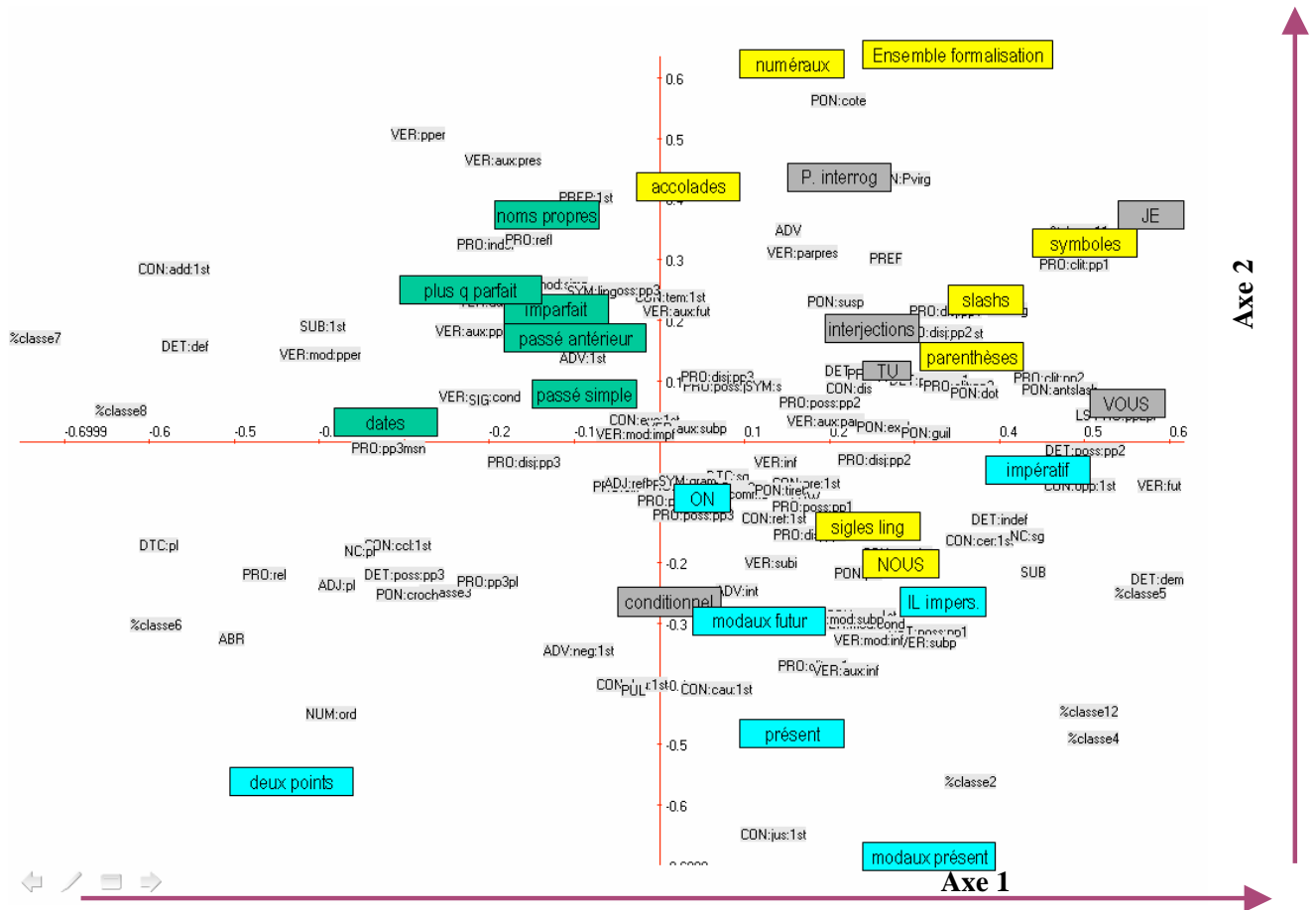


Figure 4 : Positionnement des variables sur les deux premiers axes factoriels – corpus « Auteurs »

S'il est clair que l'on relève des différences d'organisation du corpus « Auteurs », on observe la présence persistante du groupement *mode de narration historique* mis à jour dans la section précédente, qui s'oppose encore une fois à l'ensemble des caractéristiques les plus représentées du discours scientifique – nettement moins intercorrélées que dans le graphique précédent.

Les différences d'organisation les plus manifestes se situent au niveau des marques de formalisation et d'exemplification, qui ne s'opposent plus sur les deux axes : on a ainsi un pôle clairement plus appliqué qui rassemble les deux types de caractéristiques.

Notons également que le pronom NOUS n'est plus associé au pôle *formalisation* mais aux pronoms ON et IL impersonnel, tandis que le conditionnel n'est plus corrélés aux descripteurs de l'oral, et rejoint l'ensemble des éléments traditionnellement considérés comme caractéristiques des textes scientifiques (présent, modaux au présent, deux points, etc.).

Malgré les changements observés, on relève dans le corpus « Auteurs », pourtant plus restreint en termes de taille, et moins représentatif du domaine scientifique linguistique⁶ et du discours qui lui est associé, les traces des oppositions mises à jour dans le premier corpus, stylistiquement non marqué : opposition de deux grands modes narratifs, présence de pôles

⁶ Eu égard à la spécialisation des auteurs dans le champ scientifique, un nombre nettement plus restreint de domaines de la linguistique y est en effet représenté.

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 » sont ainsi globalement comparables : en dépit de sa plus petite taille et de sa représentativité générique relative, 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 à 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 atteste bien la présence d'un noyau dur générique au sein du corpus « Auteurs ».

4. Singularisation des usages du genre

Si les caractéristiques du genre sont visibles dans le corpus « Auteurs », il convient d'interroger si l'inverse est vrai, et de déterminer dans quelle mesure les styles d'auteur se démarquent et transparaissent dans le corpus « Genre ». On cherchera d'abord à évaluer l'homogénéité et la singularité des 15 styles au sein du corpus « Auteurs » (4.1.) avant d'examiner leur position et leur visibilité dans le corpus « Genre » (4.2.).

4.1. Distinctivité des 15 auteurs entre eux

Nous avons d'abord 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 :

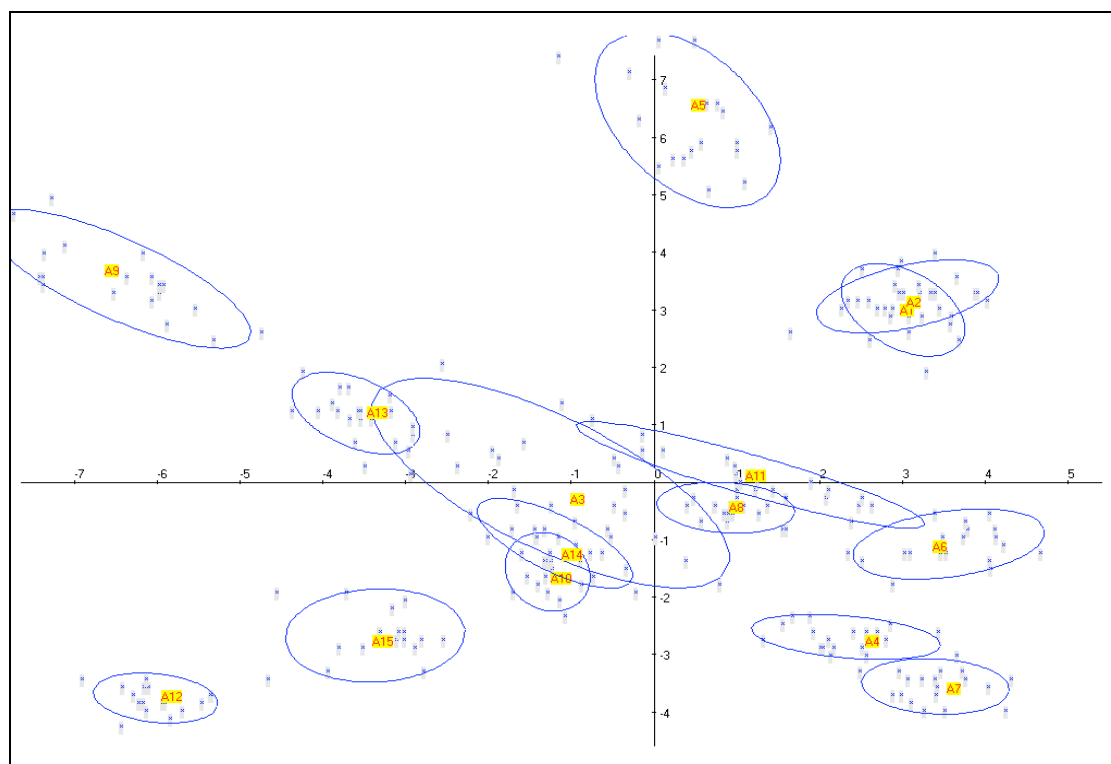


Figure 5 : Ellipses de confiance autour des 15 auteurs observés sur le premier plan factoriel – corpus « Auteurs »

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

Afin d'interpréter les données de manière plus pertinente, nous avons effectué un *Tri Systématique de Signification* (TSS) en confrontant l'ensemble du corpus avec la variable supplémentaire « Auteurs », 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 : 712). Cette méthode nous a permis d'obtenir les spécificités des 15 styles d'auteur, et d'interpréter les positions des styles sur le plan factoriel de manière plus satisfaisante.

On remarque d'abord que six styles sur quinze se présentent ici comme particulièrement singuliers, dans le sens où les ellipses ne se recouvrent pas : les styles A4, A5, A7, A9, A12 et A15. Si l'on se réfère à l'examen des axes factoriels de la section précédente, les styles isolés renvoient aux différents pôles mis à jour : le style de A9 est effectivement marqué par un emploi intensif du passé simple, de l'imparfait (et de leurs homologues composés), des dates et des noms propres – A9 étant précisément spécialiste de linguistique historique -, tandis que le style A5 est marqué par de nombreuses formalisations.

On notera également le recouvrement entre A1 et A2, spécialistes du discours travaillant sur matériel oral.

Le style d'auteur le moins distinct sur ce plan factoriel serait le style de A3 : on note en effet 5 recouvrements stylistiques et l'ellipse est plus grande.

4.2. Distinctivité des 15 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 ».

Le caractère distinct des styles au sein de la combinaison des deux corpus démontrerait bien la présence de singularités stylistiques au sein du genre : les ellipses demeurent suffisamment resserrées pour attester de la présence de spécificités stylistiques relativement homogènes dans les textes d'un même auteur. Cependant, le recouvrement de la plupart d'entre elles suggère que ces caractéristiques sont souvent proches les unes des autres, la position attendue des textes d'un auteur pouvant aller jusqu'à s'inscrire dans l'ensemble caractéristique des textes d'un autre.

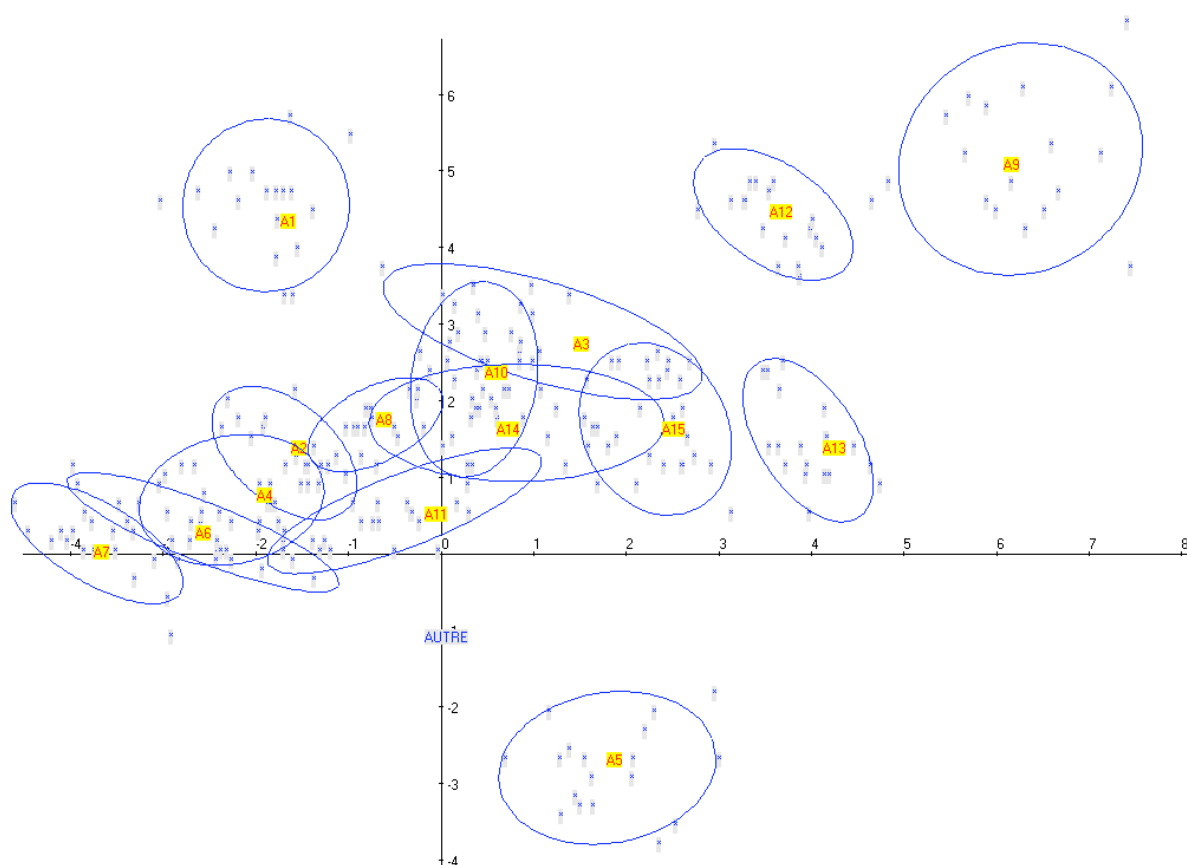


Figure 6 : Ellipses de confiance autour des 15 auteurs observés sur le premier plan factoriel – corpus « Genre » + « Auteurs »

Cinq styles sur quinze sont cette fois significativement distincts : on retrouve les styles A5, A9 et A12 déjà isolés dans le corpus « Auteurs ». En revanche, si les styles A1 et A12 étaient recouverts dans le premier corpus, on note qu'ils sont bien distincts sur le plan factoriel.

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

On observe que les styles d'auteur restant se recouvrent largement, et s'opposent globalement à l'ensemble des autres auteurs (variable « Autre »).

5. Conclusion et perspectives

L'analyse des relations entre genre et style mise en œuvre au moyen des deux corpus « auteur » et « genre » suggère, pour le dire caricaturalement, que l'on retrouve le genre dans les styles et les styles dans le genre.

D'une part en effet, la caractérisation différentielle des textes au sein du corpus « auteurs » fait émerger des axes d'opposition relativement proches des axes de contraste interne au genre de l'article en linguistique : opposition de deux grands modes narratifs, présence de pôles 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.). Cette double description montre d'abord que le genre est loin de se limiter aux caractéristiques attendues du discours scientifique, puisque celles-ci vont jusqu'à constituer un pôle spécifique qui signale leur pertinence tout autant que la nécessité de ne pas restreindre la diversité des textes à ces seuls traits canoniques. Par ailleurs, la description du corpus « auteur », malgré les changements qu'elle introduit, en particulier le fait que le pôle appliqué y soit plus marqué que dans le corpus « genre », valide la permanence du genre du point de vue de sa variabilité : elle suggère que la diversité des articles se structure de manière relativement stable, quel que soit le nombre d'auteurs et le nombre de textes par auteurs représentés dans le corpus.

D'autre part, l'analyse des styles d'auteur dans le corpus « auteur » et dans le corpus « genre » corrobore la pertinence de cette variable stylistique : tous les styles ne se caractérisent pas par le même degré d'homogénéité et de singularité des textes de leur auteur, mais tendent à se distinguer s'ils se situent aux extrémités des pôles ; ils conservent d'ailleurs cette visibilité dans le corpus « Genre ». Le style ainsi défini comme singularisation de l'usage du genre chez un même auteur intègre pour beaucoup des caractéristiques domaniales, relatives au domaine d'investigation d'un auteur : le sous-champ concerné (e.g. la linguistique historique), de même que l'objet étudié (e.g. l'oral) induisent des spécificités nettes dans les usages du genre de l'article. Aussi peut-on faire l'hypothèse que la distinctivité des styles tient pour partie à l'homogénéité et à la singularité du domaine dont relèvent les textes d'un auteur en regard des pratiques dans l'ensemble du champ. Reste que cette question est difficile à traiter : comment décider de l'homogénéité et de la singularité domaniale des textes d'un auteur, en l'absence de cartographie précise et délimitée du champ de la linguistique ?

On soulignera pour finir que cette question des relations entre genre et style demande à être étudiée au sein d'un corpus plus vaste, incluant des genres de textes différents – corpus toutefois plus délicats à rassembler en sciences qu'en littérature. Dans une perspective de validation des deux notions en corpus, il importe en effet d'évaluer dans quelle mesure le genre de l'article et les styles d'auteur émergent d'une entreprise de classification basée sur une multiplication des genres et des styles pris en compte. Comme on s'appuie sur une définition des styles relative à celle des genres, retrouve-t-on, dans le cas de l'article scientifique, les résultats mis en évidence pour les textes littéraires, qui indiquent que les genres l'emportent sur les styles en matière de classification ?

Références

- Bakhtine, M. (1986). *Speech Genres and Other Late Essays*. Trad, Vern. W. McGee. University of Texas Press, Austin.
- Lebart, L. (2004). Validité des visualisations de données textuelles. In *Actes des 7èmes Journées internationales d'Analyse statistique des Données Textuelles*, Louvain-la-Neuve, 10-12 mars 2004 : 708-715.
- Malrieu, D. and Rastier, F. (2001). Genres et variations morpho-syntaxiques. In Daille, Romary (dir.), *Linguistique de corpus*, Traitement automatique des langues, vol. 42 n°2, Paris, Atala/Hermès : 547-577.
- Mézaille, T. (2003). Thématiques littéraires : enseignement des textes numériques. In *Texte !* [en ligne]. Disponible sur : http://www.revue-texto.net/Inedits/Mezaille_Thematiques.html. (Consultée le 25/04/2005).
- Molinié, G. and Cahné, P. (1994). *Qu'est-ce que le style ?* Presses Universitaires de France, Paris.

- Philippe, G. (2002). L'appareil formel de l'effacement énonciatif et la pragmatique des textes sans locuteur. In Amossy, R., dir., *Pragmatique et analyse des textes*, Université de Tel-Aviv : 17-34.
- Pontille, D. (2004). La signature scientifique. Une sociologie pragmatique de l'attribution. CNRS Éditions, Paris.
- Poudat, C. (2003). Characterization of French linguistic research papers with morphosyntactic variables. In. Fløttum K. & Rastier F., éd. (2003), *Academic discourses — Multidisciplinary Approaches*, Novus, Oslo.
- Poudat, C. and Rinck, F. (2005). Genres scientifiques et style d'auteur : des variations stylistiques de l'article de revue linguistique. In *4e Journées Internationales de la Linguistique de Corpus*, Lorient, 15-17 septembre 2005.
- Rastier, F. (2001). Arts et Sciences du texte. Collection Formes sémiotiques. Presses Universitaires de France, Paris.
- Rinck, F. (2005). Images of scientific activity through the research article : a comparison between linguistics and literary studies. In K. Fløttum and O. Korsnes (eds), *Academisk Prosa*, n° 3, Department of Romance Studies, University of Bergen.

