

Styles d'auteur et singularité des textes. Approche stylométrique du genre de l'article en linguistique

Fanny Rinck

E.A. 609, Laboratoire Lidilem, Université Stendhal Grenoble 3

L'étude que nous présentons s'intéresse à l'article de recherche en linguistique et à la question de savoir en quoi le genre varie en fonction de l'auteur. Elle s'inspire d'une approche stylométrique pour mettre au jour les spécificités des articles de 15 linguistes français au niveau morpho-syntaxique et lexical ⁽¹⁾. Ce ne sont donc ni les effets de style, ni le travail stylistique, ni même l'idée que tel ou tel auteur a ou n'a pas de style qui font l'objet du présent article. L'objectif est de rendre compte des caractéristiques linguistiques qui distinguent les textes des 15 auteurs pour montrer que seules certaines s'assimilent à un style au sens d'un usage idiomatique du genre, et discuter alors la singularité des articles d'un auteur à l'aune de la diversité des textes et des recherches au sein du champ considéré.

1. Le style en sciences, une question paradoxale

Qu'on y voie un effet expressif et esthétique, ou une réalité individuelle, le style est supposé être une dimension neutralisée du discours scientifique, que l'on dit fondé sur une logique du vrai et sur le seul logos. Les historiens des sciences s'emparent pourtant de la notion de styles, pour évoquer des styles de recherche, des styles de pensée ou de raisonnement (Crombie, 1994, Gayon, 1996). Ils remettent ainsi en question l'idée que la production de connaissances universelles ne peut reposer que sur une méthode unitaire, conçue comme un ensemble de procédures exclusivement guidées par leur logique interne. C'est ce tournant aussi qui caractérise les études de la vie de laboratoire (Latour et Woolgar, 1979) et la prise en compte des dispositifs langagiers à l'œuvre dans la construction du savoir scientifique.

Secondaire ou absente des réflexions sur les styles de recherche en histoire des sciences, la dimension textuelle de la science est traitée dans des études rhétori-

(1) Ce travail a été mené avec C. Poudat (Laboratoire ERTIM, INALCO), que nous tenons à remercier pour cette collaboration. Il a fait l'objet de deux textes, dont l'un reste à paraître (Poudat et Rinck, 2005, 2006).

ques et linguistiques, qui décrivent les standards en usage dans des genres variés, et mettent en évidence des variations culturelles, liées aux disciplines ou aux nationalités⁽²⁾. Dans ce cadre intervient l'hypothèse de variations individuelles et de spécificités propres dans l'usage qu'un auteur fait d'un genre. Cette hypothèse a partie liée avec une théorie de l'action qui envisage la science non plus seulement dans sa dimension collective, mais du point de vue aussi des acteurs et de leur singularité. Elle se présente en outre comme un éclairage complémentaire à la question de savoir en quoi l'auteur scientifique imprime sa marque au texte, à côté des études de la subjectivité langagière et de l'ethos.

L'extension de la notion de styles à un champ comme la science destitue les écrivains du privilège d'une auctorialité réservée aux pratiques d'excellence. Elle doit permettre de mieux cerner le singulier des pratiques langagières et leur fonction individuante, dans des genres réputés fortement contraints voire routinisés. Elle peut contribuer à terme aussi à une réflexion sur la science, et sur la dimension contextuelle et contingente d'une production de connaissances qui n'interdit pas leur accréditation et leur universalité.

2. Une définition du style relative aux genres

De manière très générale, le style s'associe à des manières d'être ou de faire, et pointe leur singularité et leur diversité. D'un point de vue linguistique, le style est d'abord conçu comme un écart, qui suppose, et voudrait informer, un « fond de langue » ou un niveau « non marqué » de ses usages (Gueunier, 1969). Il introduit par ailleurs un sujet à l'origine de choix linguistiques et s'interprète comme le résultat d'une intention stylistique ou comme une pratique idiomatique non nécessairement conscientisée, qui le rattache alors à la problématique de l'idiolecte (Détrie et Neveu, 2005). Les différences inter-individuelles justifient l'intérêt de ces notions, mais la question du singulier dans la langue n'échappe pas au problème du référent que l'on se donne, et aux difficultés posées par la comparaison.

Celle-ci est au cœur des études stylométriques, qui explorent la fonction individuante du style à travers l'unité et la spécificité des pratiques textuelles d'un auteur. Se donnant pour objet l'analyse quantitative et statistique de traits stylistiques, les études stylométriques envisagent ces derniers comme des caractéristiques textuelles – souvent restreintes à la dimension lexicale – qui sont propres à un auteur et distinguent son écriture entre toutes. Cette définition du style comme « manière d'être distinctive »⁽³⁾ trouve en la statistique un outil adapté : le fait de style n'est plus tant abordé en termes d'écart qu'en termes de variation ; c'est l'évaluation statistique qui vérifie une différence significative, par la caractérisation différentielle des textes de divers auteurs ou d'un même auteur.

En s'intéressant prioritairement au discours littéraire et dans une moindre mesure au discours politique, de telles études ont ainsi cherché à identifier ce qui fait la spécificité d'un auteur en comparant des textes à seule fin descriptive, ou dans une perspective d'aide à la décision, qu'il s'agisse de déterminer la paternité du texte ou de débusquer les apocryphes. Elles reposent donc sur ce postulat que les textes d'un même auteur présentent des caractéristiques qui le distinguent de ceux d'un autre au point de pouvoir être conçues comme indicielles de l'identité du locuteur qui en est à l'origine.

(2) Cf. notamment les travaux menés dans le cadre du projet KIAP (Cultural identity in academic prose) <http://www.uib.uib.no/kiap> et K. Fløttum *et al.* (2006) pour une synthèse.

(3) M. Cressot (1974) cité par D. Monière et D. Labbé (2002).

Les fréquentes controverses qu'elles suscitent révèlent la tentation de les voir cerner les fondements de la créativité, l'essence de l'auteur et/ou de l'œuvre, le talent du premier, l'esprit et la portée de la seconde et, en un mot, leur « génie ». Elles montrent pour le moins les limites qu'il y a à envisager le style de manière isolée, sans l'interroger dans un cadre d'analyse des pratiques langagières, au-delà de l'évidence d'un sujet qui s'exprime. Or, comme le souligne E. Brunet (cité par Mézaille, 2003), les ressemblances et les différences entre textes mises au jour par les entreprises stylométriques ne suffisent pas à conclure à leur paternité : « pour quoi chercher à confondre Molière et Corneille, quand le partage du même genre comique et de la même expression versifiée suffit à expliquer la ressemblance ? ». Inversement, dans quelle mesure un auteur conserve-t-il ses spécificités quand il change de genre ⁽⁴⁾ ?

C'est une définition du style relative au genre qui intéresse notre étude ; nous considérons avec F. Rastier (2001) les styles en tant qu'usages du genre propres aux textes d'un même auteur. Les deux notions peuvent *a priori* paraître diamétralement opposées : l'une renvoie au sujet et à la singularité de ses textes, l'autre repose sur l'intersubjectivité et montre que les pratiques langagières respectent des modèles textuels stables. La singularité des textes n'est cependant pas compromise par le « déjà-là » des genres et ceux-ci se définissent du point de vue de ce qui est commun aux textes mais aussi par les variations qui existent entre eux.

La notion de style au sens d'usages du genre propres à un auteur représente ainsi une entrée pour rendre compte de la diversité des textes : elle est un des facteurs potentiels de variation du genre et renvoie à l'idée que les textes d'un même auteur au sein d'un même genre présentent des propriétés distinctives qui font leur singularité.

C'est l'hypothèse que nous avons testée sur l'article de recherche, en recourant à des méthodes de caractérisation différentielle des textes. Deux objectifs complémentaires ont présidé à l'analyse :

- mettre en évidence les spécificités propres aux textes de 15 linguistes et évaluer des degrés de distinctivité des styles (Poudat et Rinck, 2005) ;
- s'interroger plus largement sur la diversité des usages du genre et sur ce qui singularise les textes en fonction de leur auteur et indépendamment de l'auteur, au sein du champ considéré (Poudat et Rinck, 2006).

Une fois la méthodologie exposée, nous rendrons compte des traits linguistiques ou des faisceaux de traits qui distinguent les articles des différents auteurs par rapport aux autres articles, pour discuter la spécificité des textes d'un auteur en regard de la diversité des usages du genre.

3. Aspects méthodologiques : linguistique de corpus et statistique multidimensionnelle

3.1. Corpus

Pour décrire en quoi le genre varie avec l'auteur (et indépendamment de lui), deux corpus ont été étudiés séparément puis comparés, le corpus « auteurs » et le corpus « genre ». Le corpus « auteurs » rassemble 130 articles de 15 linguistes

(4) L'étude de E. Brunet (1991), consistant à comparer le vocabulaire des différents textes de Victor Hugo, montre que les différences de vocabulaire se superposent à celles des genres.

français⁽⁵⁾, soit entre 5 et 8 textes par auteur. Le corpus « genre » réunit 330 articles de linguistique⁽⁶⁾, avec une diversité beaucoup plus importante au niveau des auteurs puisqu'il en compte 332. Les articles des deux corpus sont récents, et issus de revues et de numéros variés.

L'analyse automatique requiert un nettoyage des textes. Les exemples que contiennent les articles de linguistique sont ainsi écartés, car ils sont fortement susceptibles de corrompre leurs caractéristiques morphosyntaxiques. Mais pour éviter de briser la continuité syntaxique, ce qui nuirait à l'analyse, nous avons choisi de n'isoler que les exemples délimités dans le texte sous la forme de paragraphes spécifiques. Il reste que, même de manière moindre une fois les exemples ôtés, les textes portent nécessairement la trace de leur matériau.

3.2. Traits textuels retenus

Conséquence directe des possibilités offertes par l'outil informatique, les analyses lexicométriques du style, dominantes en stylométrie, se trouvent aujourd'hui complétées par la linguistique de corpus, dans des études portant sur d'autres niveaux de la textualité⁽⁷⁾.

Notre étude des textes des 15 linguistes porte sur le lexique et sur des traits morpho-syntaxiques, parmi lesquels la ponctuation. Le lexique est analysé à l'aide du logiciel Cordial (version 8 de Cordial Analyseur, Synapse développement). Les traits morpho-syntaxiques sont décrits à partir de deux systèmes d'annotation, ce qui permet d'objectiver la description en palliant les limites de ces systèmes qui varient d'un étiqueteur à un autre. L'ensemble se compose de 97 descripteurs issus du jeu d'étiquettes de Cordial et de 140 descripteurs étiquetés par TnT (Trigrams'n'Tags, Univ. Saarlandes, DE, Brants) et vérifiés manuellement.

À côté de descripteurs généraux, communs aux deux jeux (nombre de mots par phrase, nature syntaxique des unités, temps verbaux, personnes linguistiques, etc.), les descripteurs de TnT incluent des caractéristiques propres au discours scientifique (types de connecteurs, *il* impersonnels vs. anaphoriques, modaux, symboles, indices de structuration et de renvois dans le texte, etc.) et à la recherche en linguistique (signes d'acceptabilité linguistique* et ?, abréviations linguistiques comme *SN*, *SV* ou *GN*, etc.).

3.3. Analyses statistiques multidimensionnelles

Notre approche se base sur les méthodes d'analyses statistiques multidimensionnelles et vise la caractérisation différentielle des textes. Nous recourons au logiciel DTM (*Data and Text Mining*), logiciel académique développé par L. Lebart⁽⁸⁾. Deux grandes familles de méthodes sont employées de manière complémentaire par le logiciel DTM, comme dans toute entreprise de *Text Mining* (Benzécri *et al.*, 1981, 1992) : les méthodes factorielles et les méthodes de classification, qui servent l'analyse descriptive et la prédiction sur l'appartenance d'un texte ou de ses propriétés à une classe.

(5) L'auteur est compris à la fois comme rédacteur et signataire du texte, ce qui ne serait pas envisageable dans le cas des articles à signatures multiples, comme ceux décrits par D. Pontille (2004).

(6) Il résulte de la réunion du corpus de C. Poudat (2006) et du nôtre (Rinck, 2006).

(7) On peut citer la tentative originale de V. Beaudoin et F. Yvon (2004) d'identifier genres et auteurs à partir du décompte métrique des vers qui les caractérise.

(8) <http://www.enst.fr/egsh/lebart/>

Ces analyses nous permettent de tester la variation du genre en fonction de l'auteur et indépendamment de l'auteur à partir des deux aspects suivants :

- les spécificités des textes de chaque auteur par rapport aux autres textes du corpus « auteur » et par rapport au corpus « genre », le degré de spécificité des textes d'un même auteur et le caractère plus ou moins discriminant des traits morpho-syntaxiques et lexicaux retenus ;
- les axes de contraste internes au corpus « auteur » et au corpus « genre », axes factoriels constitués de la corrélation de traits textuels et rendant compte de la diversité interne des corpus. On cherche à voir si les facteurs qui structurent les différences entre les textes restent stables quel que soit le corpus considéré et dans quelle mesure la variation liée à l'auteur recouvre la variation du genre.

Nous proposons ici de montrer la productivité et les limites d'une approche stylo-métrique dans l'article de recherche ; nous isolons, d'une part, les caractéristiques linguistiques qui peuvent être mises au compte de styles d'auteur et d'usages idiomatiques du genre de l'article. Nous abordons ensuite celles qui distinguent les articles d'un linguiste sans s'assimiler à son style, et qui permettent de discuter la singularité des usages du genre et la diversité des profils de texte.

4. Des styles d'auteur au sens d'usages idiomatiques de la langue et du genre

Parmi les traits linguistiques qui caractérisent les textes d'un auteur par comparaison avec ceux des autres, certains suggèrent que l'on peut parler, dans le genre de l'article comme en littérature, de style au sens de spécificités idiolectales. Ce sont principalement les traits morpho-syntaxiques qui sont concernés, et nous en traiterons en abordant les ponctuations, les aspects énonciatifs, et ceux qui se rattachent à l'argumentation.

4.1. Ponctuation, complexité syntaxique et relations logiques

Les ponctuations représentent un des traits qui varie beaucoup avec l'auteur et notamment les marques de pause. L'usage de la virgule chez A12⁽⁹⁾ en est un exemple : ses textes se caractérisent par 1,5 fois plus de virgules que la moyenne des autres textes.

- (1) 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, *instiue 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. [A12-091]

Les virgules ne représentent pas seulement une alternative au découpage du texte en unités phrastiques séparées par le point, et les textes de A12 se distinguent en effet par un usage important des points également. Elles se substituent à un ensemble plus complexe de signes au niveau paradigmatique, et sont considérées comme un des signes de ponctuation les plus stylistiques (Riegel *et al.*, 1994). El-

(9) Les 15 auteurs sont anonymés et désignés de A1 à A15.

les montrent une structuration et un rythme spécifiques de la phrase, en isolant des éléments, notamment ici des locutions adverbiales en incise, des termes juxtaposés et de multiples subordonnées.

Cet usage des virgules montre que ce qui fait la singularité des textes d'un auteur dans le discours scientifique rejoint ce qu'il est convenu d'appeler le style dans le cas des auteurs littéraires. On a affaire à des « choix » scripturaux, au sens d'un usage idiomatique, conscient ou non, des possibilités de la langue et du genre.

Le style de A12 se caractérise en outre par des phrases plus longues que ne le veut le genre, et l'on voit l'intérêt de retenir l'auteur comme entrée pour mieux cerner les facteurs de variation de l'article, à côté d'autres différences observées en fonction des disciplines et du statut de l'auteur : les articles de lettres, par contraste avec ceux de linguistique, et les articles de non-doctorants par contraste avec les articles de doctorants affichent cette même prédilection pour des phrases plus longues (Rinck, 2006).

Comme les textes de l'auteur ici concerné, A12, se distinguent également par des concepts spécifiques (*politique* et *sémiotique*, nous y revenons *infra*), il est permis aussi de voir dans son usage de la phrase et des virgules un effet disciplinaire, à l'instar de la différence que nous venons d'évoquer entre les articles de lettres et de linguistique. Il est possible alors d'avancer dans la définition des usages idiomatiques de certains traits du genre et de se demander s'ils ne pourraient pas être compris comme un effet de la formation d'un auteur : l'ancrage disciplinaire de celle-ci pourrait déterminer des pratiques d'écriture spécifiques.

Le cas de A12 est révélateur avec d'autres de l'importance des indices de complexité syntaxique dans la caractérisation d'usages du genre propres à un auteur. On peut regrouper dans cette catégorie les indices de longueur de phrases et de proposition (en nombre de mots), et tout ce qui concerne les subordonnées et les subordonnants. A nouveau, on voit ici en quoi le style en sciences reste proche des spécificités stylistiques observables dans les textes littéraires : certains auteurs affichent une préférence pour des phrases et des paragraphes courts (comme A5) tandis que d'autres (A7 notamment ou A12 comme on l'a vu) multiplient les phrases complexes (taux fort de phrases comportant une subordonnée, de subordonnants, d'adverbes et connecteurs toutes catégories confondues).

A l'inverse de A7, les textes de A2 se distinguent par la présence significativement plus faible de phrases avec une subordonnée, de subordonnées, de subordonnants, et de divers types de connecteurs (de conclusion, disjonction, reformulation, causalité, exemplification). La sous-représentation de ces traits semble pouvoir s'expliquer en lien avec une sur-représentation des deux-points (quasi deux fois plus que dans les autres textes), employés par l'auteur pour justifier ou exemplifier son propos :

- (2) 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. [A2-011]
- (3) Tout texte exprime un *ethos* : il possède une manière de se dire et de s'énoncer qui fait sens (Maingueneau, 1993). [A2-013]

Ce recours fort aux deux-points et la faible présence de connecteurs montre une tendance des textes de A2 à ne pas expliciter certaines relations logiques. Ce constat montre l'intérêt des ponctuations pour l'analyse des usages singuliers du genre

au niveau de la syntaxe mais aussi en termes d'argumentation, comme les différences évoquées à propos des connecteurs. C'est vrai également des ponctuations à valeur énonciative, comme les points d'interrogation, particulièrement fréquents dans les textes du même auteur, A2, et que l'on peut rattacher au caractère particulièrement discriminant des aspects énonciatifs de manière générale.

4.2. Les aspects énonciatifs : modes d'inscription du scripteur et implication du lecteur

L'importance de la dimension énonciative se vérifie au niveau des possibilités offertes dans l'usage des marques de personne et de non-personne, et dans le mode d'inscription de l'auteur dans son texte qui en résulte. Il s'agit d'un trait particulièrement discriminant, que ce soit par l'usage abondant de la troisième personne (A5, A8, A12, A13, A14, A15), du *on* (A13) des formes de *je*⁽¹⁰⁾ (A1 et A2), ou de *nous* (A3, A7, A10, A11).

Dans certains cas, les textes d'un auteur se distinguent par l'usage moindre d'une marque de personne ou de non-personne, comme ceux de A9 où *on* est sous-représenté. Ils peuvent alors indiquer un choix net dans l'alternative entre le mode de prise en charge personnel et le mode délocuté comme chez A12, A13, A14 et A15 qui font un usage abondant de la troisième personne et délaissent *je* ou *nous*, ou au contraire chez A1 dont les textes se caractérisent positivement par l'usage de *je* et négativement par une sous-représentation de la troisième personne.

A côté de l'usage abondant d'une marque, et du choix tendanciellement exclusif d'un mode au détriment de l'autre, chez A4, la sur-représentation des formes de *je* rivalise avec un usage fort également du *il* impersonnel. Un double effet d'effacement et de marquage de la personne dans les textes se manifeste ainsi dans l'usage singulier que cet auteur fait du genre.

L'examen des textes qui se caractérisent par un usage fort du *je* ou du *nous* et de ceux où les marques de personne sont sous-représentées suggèrent l'affinité de ces marques avec les balises méta-textuelles assurant le guidage du lecteur dans le texte, et qui prennent la forme d'expressions lexicalement figées, du type *nous verrons* :

- (4) [...] 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 [A3-021]

La forte présence de *je* ou *nous* dans le style d'un auteur pourrait ainsi être liée à ces balises. Mais elles restent indépendantes du choix pour *je*, *nous* ou un mode mixte d'usage de *je* et de *nous*, et semblent donc pouvoir être considérées en elles-mêmes comme un trait idiomatique dans l'usage du genre. Le cas de A2 va dans ce sens avec une forte présence de *je* corrélée à celle du futur, ce que l'on peut expliquer d'un point de vue syntagmatique par les nombreuses annonces que fait l'auteur de ce qui va suivre, en début d'article mais également en son corps (« *je me limiterai...* », « *je prendrai pour exemple...* »). Les textes de cet auteur ont pour autre spécificité comme on l'a dit l'usage de points d'interrogation dans des questions rhétoriques et ses préférences idiomatiques concourent ainsi à un mode spécifique d'implication du lecteur dans le texte.

(10) Ici comme dans ce qui suit, nous citons *je* et *nous* mais il peut s'agir aussi des autres formes du pronom personnel (*me* par exemple), des pronoms et déterminants possessifs correspondants, et enfin des marques flexionnelles de personnes au niveau des verbes.

Dans les textes de A7, le *nous* renvoie sans ambiguïté à la personne énonciative du locuteur et intervient dans des balises méta-textuelles, ou, sous la forme du déterminant possessif, en s'associant à un lexique transdisciplinaire (Rinck, 2006, Tutin, 2007), dans des mentions récurrentes comme « *notre travail* », « *notre propos* », « *notre conception* », « *notre hypothèse* ». Les variations dans l'usage des marques de personne, particulièrement discriminantes d'un auteur à l'autre, permettent d'affiner la description du genre dans la diversité de ses usages : elles révèlent la complexité des profils d'articles au-delà de l'alternative entre personne et non-personne, et rendent tangible, en fonction de leurs associations syntagmatiques, une énonciation diversement représentée dans les textes ; ils se caractérisent à des degrés variables par la mise en scène de leur auteur, et/ou l'implication du lecteur, avec un rapport plus ou moins didactique à ce dernier.

4.3. La dimension argumentative des textes

L'argumentation représente une autre entrée intéressante pour l'analyse des styles et de la diversité des usages du genre. Nous l'avons indiqué à propos de la syntaxe de la phrase, et l'usage d'adjectifs réfléchis le suggère aussi, en signalant une tendance à l'emphase (« *condition même de sa production* », « *l'image qu'elle en dessine en elle-même par ses non-coïncidences* », « *non seulement à repérer exactement mais même à localiser* »).

Cette tendance caractérise les textes de A7, avec d'autres traits comme l'usage d'un *nous* auto-centré que l'on vient d'évoquer (« *nous rappellerons* », « *notre hypothèse* », etc.), et de négations, indicelles dans le cas des négations polyphoniques d'une argumentation polémique. Les connecteurs argumentatifs abondent également dans ses textes (notamment des connecteurs de présupposition et de certitude), comme chez A8 où ce sont les connecteurs d'opposition qui sont concernés. Ils sont au contraire délaissés dans les textes d'A12 (quasi trois fois moins, alors que ce sont les troisièmes connecteurs les plus représentés du genre), de même que les connecteurs de concession, ce qui peut indiquer une tendance à des textes globalement peu polémiques, les connecteurs de conclusion, de doute, d'addition et de conséquence l'emportant dans la visée explicative et argumentative des textes de l'auteur.

De manière *a priori* paradoxale, le conditionnel s'avère particulièrement prisé chez des auteurs dont d'autres spécificités (négation, connecteurs) suggèrent des textes comportant une dimension argumentative forte, voire polémique. L'effet de nuance introduit par le conditionnel peut à lui seul être considéré comme un trait discriminant, car il représente une des diverses formes de modalisations possibles, à côté du lexique notamment. Mais le niveau paradigmatique n'est pas le seul concerné, car l'examen des textes montre que le conditionnel est souvent utilisé au sein d'assertions fortes,

- (5) On voit que le traitement de la polysémie de l'adj. sec en termes d'homonymie **serait** particulièrement desséchant ! Car on y **perdrait** tout un échantillonnage de glissements de sens dont nous sommes largement conscients dans l'usage langagier. [A5-047]

marquées, comme ici, par la modalité exclamative, ou encore par le mode déontique : de manière générale, parmi les formes au conditionnel dans les textes des auteurs concernés (A3, A4, A5, A9), on trouve de nombreuses fois le verbe *devoir*.

Le cas de A4 est sans doute celui qui illustre le mieux l'importance de la dimen-

sion argumentative dans l'usage du genre que fait cet auteur. Ses textes se caractérisent par un emploi plus important de modaux, et de modaux au conditionnel en particulier. Ils indiquent deux tendances dans ses textes en termes d'argumentation : d'une part, un penchant pour les assertions fortes, sur le mode déontique (*falloir, devoir*), et auquel on peut aussi relier l'abondance de connecteurs de justification (seconde caractéristique de ses textes) ; d'autre part une tendance à la nuance, avec l'importance du conditionnel (deux fois plus en moyenne que dans le reste du corpus) et des connecteurs de doute.

En termes énonciatifs, cette double tendance marquée par des corrélations de traits est particulièrement intéressante : elle montre qu'on n'a pas d'un côté une modalisation au sens de nuance dans l'assertion et de l'autre une modalisation qui favoriserait des assertions plus péremptoires, mais qu'au contraire le recours abondant à l'un va de pair avec le recours abondant à l'autre. Or, nous avons déjà évoqué le style de cet auteur à propos de la prise en charge énonciative, avec deux tendances opposées, mais là encore non contradictoires : ses textes se caractérisent à la fois par un usage fort du *je* et du *il* impersonnel. Un double effet d'effacement et de marquage de la personne dans les textes se manifeste ainsi. Le cas de A4 révèle des pistes intéressantes pour l'analyse énonciative des textes, et celle du rôle de la dimension énonciative dans l'argumentation. L'exploitation idiomatique de l'effacement et du marquage de la personne d'une part, et de la nuance et du renforcement de l'assertion d'autre part sont caractéristiques de coups de force dans l'assertion (Grossmann et Rinck, 2004) et de profils d'articles marqués, plus encore que d'autres, par la « surénonciation » (Rabatel, 2004).

L'approche stylométrique montre sa productivité pour la mise en évidence de traits textuels spécifiques dans l'usage qu'un même auteur fait du genre de l'article. Certains de ces traits peuvent être compris comme des spécificités d'ordre idiomatique : ainsi en est-il au niveau de la syntaxe de la phrase, en termes de complexité et de relations logiques, au niveau énonciatif par rapport à l'inscription de l'auteur dans son texte et l'implication de son lecteur, et au niveau argumentatif avec des textes plus ou moins polémiques.

On voit que la tension entre normes singulières et collectives est résolue de diverses manières ⁽¹¹⁾. Un auteur qui fait usage de phrases courtes dans ses textes se conforme à une tendance caractéristique du genre et ne verra pas ce trait le distinguer. Les usages singuliers du genre ne sauraient pourtant se restreindre à la transgression de règles communes ⁽¹²⁾. S'ils permettent de dépasser l'idée de standards immuables dans l'article de recherche, ils montrent aussi que le genre ne se restreint pas à ce qui est commun aux textes mais se présente comme un espace de possibilités diversement exploitées. Les choix scripturaux offerts au scripteur comme l'usage des *je, nous, on*, du *il* impersonnel, et l'exploitation plus ou moins massive voire exclusive du délocuté en attestent. Les normes individuelles doivent être questionnées par rapport à la pluralité des normes en usage dans des textes publiés et jugés en cela conformes au genre. Le genre varie avec l'auteur, mais indépendamment aussi.

Nous avons isolé ici les traits qui pouvaient renvoyer à l'individualité de l'auteur dans ses choix scripturaux et être envisagés en prise avec l'idiolecte. Nous

(11) C'est ainsi que l'on peut expliquer les degrés divers de distinctivité des styles, décrits dans (Poudat et Rinck, 2005).

(12) Ni *a fortiori* à l'idée d'une innovation, car dans le cas de la complexité syntaxique, il semble qu'on ait affaire à un trait qui tend à se marginaliser.

abordons dans ce qui suit des traits qui distinguent les textes d'un même auteur mais sont étroitement liés au contenu des textes ; ils rendent délicate une approche en termes de styles d'auteur, et montrent en quoi les usages singuliers du genre renvoient à l'individualité de l'auteur autant qu'à la diversité des textes et à la structure du champ.

5. La singularité des usages du genre

A côté des caractéristiques propres aux textes d'un même auteur que l'on peut mettre au compte de spécificités idiomatiques, d'autres sont induites par le corpus qu'il étudie, par la thématique de ses recherches et par le type d'étude menée ; le lexique est prioritairement concerné, mais également certains traits morpho-syntaxiques.

L'approche stylométrique telle que nous l'avons mise en œuvre produit ce que l'on peut considérer comme un biais. L'exemple le plus frappant est sans doute celui de A13 : *folle* et *vache* viennent en tête des traits qui distinguent ses textes par leur fréquence importante, suivis de *scientifiques*, *Libération* et *science*. On retrouve le poids du corpus (*Libération*), conjoint à celui de l'objet étudié (discours médiatiques et scientifiques), l'auteur analysant dans ses différents articles le traitement journalistique d'objets comme la *vache folle*. Le même type de biais se manifeste au niveau morpho-syntaxique, dans les textes de A2 et A6 notamment, avec des marqueurs d'oralité, tels que les exclamatifs, les interrogatifs, et les marques de deuxième personne, qui tiennent à la présence de segments relevant des corpus oraux étudiés par les auteurs. L'analyse fait ressortir ces traits dès lors qu'ils sont marqués par rapport aux caractéristiques du genre : des articles qui prendraient comme matériau des extraits d'articles scientifiques ne poseraient pas ce problème.

Nous avons précisé que la méthodologie adoptée nous a fait supprimer des textes les exemples décrochés dans un paragraphe spécifique ; le poids du corpus dans la caractérisation des usages singuliers du genre justifie cette extraction mais en montre aussi les limites : il n'est pas possible d'éliminer complètement ce biais, sauf à modifier profondément les caractéristiques des textes, dont le matériau reste inextricable.

Les résultats de la caractérisation automatique des textes et de leur classification demandent donc à être évalués *a posteriori* : il faut opérer un tri et convenir que les traits mis en évidence comme spécifiques des textes d'un auteur par comparaison avec les autres ne peuvent pas tous être conçus comme des traits propres à un style d'auteur, ni comme des traits renvoyant à un idiolecte. Plus même, d'un auteur à un autre, les singularités qui se dégagent de la confrontation des textes relèvent de dimensions foncièrement hétérogènes et ce sont ces dimensions qu'il s'agit de mieux cerner.

L'intérêt d'une étude de la variation stylistique définie en référence au genre reste donc de discuter ce qui fait la diversité de l'article en linguistique : ce qui distingue les textes d'un auteur ne permet pas tant d'approcher des réalités individuelles que la tension constitutive du genre entre les productions singulières et collectives d'un champ. Ainsi en est-il des caractéristiques lexicales liées à la thématique des textes, mais c'est ce que nous aborderons aussi à partir de certains traits morpho-syntaxiques et des types d'étude auxquels ils sont susceptibles de renvoyer.

5.1. Style et lexique : une entrée pour l'étude des concepts et des épistémologies

Parmi les spécificités propres aux articles d'un même auteur, par comparaison avec les autres textes du même genre, le lexique joue un rôle relativement important. Hormis certains termes qui renvoient au corpus étudié par l'auteur (*vache, folle, Congo, femme, Libération*), comme dans le cas de A3 cité ci-dessus, les termes les plus discriminants relèvent en grande majorité d'un lexique spécialisé, à valeur conceptuelle, et montrent le poids de la thématique sur les caractéristiques linguistiques des textes. Encore faut-il que le candidat concept se retrouve dans plusieurs, voire dans l'ensemble des textes d'un même auteur, et qu'il soit suffisamment spécifique dans le champ pour distinguer ses textes des autres.

Ainsi les textes de A12 doivent-ils leur singularité à l'usage abondant des termes *sémiotique, politique, symbolique, espace, public*. Ces termes rendent clairement compte des principaux thèmes abordés par l'auteur, qui présentent une singularité très forte, en particulier *politique* et *sémiotique*.

Les textes de A12 sont les plus clairement concernés, mais A4 se distingue également par l'usage important de *syntagme, proposition, grammaticalisation* et *topique* notamment, et A3 par celui de *langues, langue, français, anglais, pays*, autant de termes qui reflètent là encore les thématiques privilégiées des auteurs dans leurs articles.

On peut alors faire l'hypothèse que les auteurs dont les articles se caractérisent par ces termes liés à la dimension conceptuelle et thématique des textes ont des domaines d'investigation relativement à la marge dans le corpus. Du moins une telle approche doit-elle permettre de le montrer, à condition de la mener sur un corpus de taille plus importante. La caractérisation différentielle des articles se présente comme un moyen intéressant pour fonder en corpus une cartographie du champ disciplinaire : les concepts discriminants donnent en effet à voir des paradigmes qui structurent le corpus, et, partant, la discipline. Pour revenir à l'exemple le plus parlant, le fait que les termes *politique* et *sémiotique* soient sur-représentés chez A12 et sous-représentés chez les autres auteurs signale la place marginale des problématiques politique et sémiotique dans la recherche linguistique.

Le graphique qui suit (*voir figure 1 page suivante*) rend compte des mots signifiants des textes (nom, verbes, adverbess et adjectifs) sur les deux premiers axes factoriels. Il révèle que les termes les moins centraux sont liés aux spécificités du corpus ou de la thématique plutôt qu'à des préférences idiomatiques au sein du lexique général, et permet en outre de voir où se positionnent les auteurs par rapport aux termes qu'ils emploient : la singularité de A12 (*politique, sémiotique, etc.*) est particulièrement frappante (facteur 1), et A4 (*grammaticalisation, topique, etc.*) se démarque également (facteur 2).

La distribution des mots signifiants sur le graphique suggère trois tendances opposées qui pourraient renvoyer à des pôles de recherche en linguistique – sans doute de manière encore trop caricaturale ici, car il faudrait bien entendu augmenter le corpus. On voit en effet se distinguer :

- les approches grammaticales (en haut, à droite), avec des termes tels que *syntagme, syntaxique, structures, constituant, déterminant, etc.*, et le rôle d'attracteur joué par *grammaticalisation*, tout en haut de ce pôle,
- l'analyse de discours (en bas, à droite), avec les termes *énonciation, modalisation, science, écriture*, ou encore *dialogisme, politesse, point de vue, etc.*,

Nous avons vu plus haut que la morpho-syntaxe spécifique aux textes d'un auteur par rapport à un autre manifestait des usages idiomatiques de la langue et du genre, par exemple avec la longueur des phrases, la préférence pour le marquage de la personne ou pour le mode délocuté, ou encore la tendance à une argumentation polémique. Ces usages ne sont cependant pas l'apanage d'un seul et même auteur : rien ne s'oppose à ce qu'ils soient partagés par plusieurs d'entre eux, car ils correspondent à des profils de textes et non à des profils d'individus.

Dans ce qui suit, nous montrons dans le même sens en quoi les traits morpho-syntaxiques qui discriminent les articles de tel ou tel auteur mettent en évidence des différences au niveau du type d'études présentées. Il sera question plus précisément de la place et la nature des exemples d'une part et de la dimension historique de certains articles d'autre part.

5.2.1. Des textes diversement exemplifiés

Les textes des différents auteurs se distinguent pour certains au niveau morpho-syntaxique par des traits qui suggèrent un recours abondant à l'exemplification. Nous pouvons l'illustrer avec le cas de A1 : on observe dans ses textes significativement moins de numéraux ordinaux et de numéraux de datation mais un usage plus abondant de numéraux de renvoi. Leur présence s'explique par l'abondance d'exemples numérotés dans les articles de cet auteur. La forte présence de tirets serait alors à mettre au compte des indices de structuration qui servent à introduire les exemples, et c'est également un trait caractéristique des textes de A6 qui cite des extraits de ses corpus oraux. Ceux de A8 sont eux aussi concernés par les numéraux à fonction de renvoi qui s'associent à un usage important d'acronymes (de type *SN*, *SV*), et d'abréviations linguistiques (*prep.*), et l'examen de ses textes confirme la place importante des exemples et de leur analyse syntaxique et énonciative.

Chez A1, cette large place faite à l'exemple se manifeste en outre par l'abondance de connecteurs d'exemplification. Ces derniers cependant ne sont pas indicatifs en eux-mêmes d'un taux fort d'exemples dans les textes ; le cas de A15 est à ce titre frappant : ses articles se distinguent par un usage massif de ce type de connecteurs (deux fois plus en moyenne que dans le reste du corpus), mais à y regarder de plus près, on ne trouve dans ses textes que de très rares exemples, si l'on s'en tient du moins à une définition des exemples restreinte à des échantillons de la langue (typiquement, des exemples construits, ou des extraits de corpus oraux ou écrits) cités par l'auteur dans son texte.

L'usage de *par exemple* dans les textes de cet auteur permet de souligner la complexité de ce que l'on entend par « exemple » dans l'article en linguistique et dans l'écrit de recherche de manière générale⁽¹³⁾.

- (6) Or, un genre définit précisément un rapport normé entre signifiant et signifié au palier textuel : par **exemple**, dans le genre de l'article scientifique, le premier paragraphe, sur le plan du signifiant, correspond ordinairement à une introduction, sur le plan du signifié ; dans le genre de la nouvelle, il s'agit le plus souvent d'une description. [A15-088]

Les différents marqueurs d'exemplification suggèrent, en fonction de leur nature et de leurs combinaisons dans les textes, des tendances quant aux types d'étu-

(13) Cf. P. Ouellet (1989) pour une réflexion sur le statut de l'exemple en linguistique.

des privilégiées par un auteur : une association de connecteurs d'exemplification et d'indices de structuration comme les numéraux de renvoi et les tirets indiquent une approche plus descriptive chez A1, qui cite des énoncés tirés de son corpus (de même que A2 et A8 évoqués *supra*) tandis que le recours à des formules du type « par exemple » alimente chez A15 une discussion théorique et un positionnement doctrinal qui ne mobilisent pas de référence aux observables en eux-mêmes, mais à des observations faites sur les usages de la langue.

L'analyse des caractéristiques marquées dans l'usage qu'un auteur fait du genre se présente ainsi comme une entrée intéressante pour dépasser une analyse de l'exemple restreinte aux échantillons de la langue. Elle signale des tendances régissant la diversité interne du genre en fonction de ce qui est pris comme matériau par l'auteur. Il est possible ainsi d'affiner l'opposition entre des textes plus théoriques et des textes plus empiriques. Nous discutons dans (Rinck, 2006) des limites qu'il y a à cerner cette opposition avec le seul trait distinctif des échantillons de langue cités dans le texte, ou avec l'exemple *versus* la citation de travaux du champ ; on voit ici que des critères multiples interviennent :

- certains textes se distinguent par la trace de leur corpus (A1 pour le recours à des exemples d'énoncés, A2 pour l'oral, ou encore A13, évoqué à propos de *vache et folle*) ;
- d'autres par l'analyse d'exemples (recours à des exemples numérotés et formalisation linguistique chez A8) ;
- d'autres enfin, comme c'est le cas chez A15, par des exemples qui ne correspondent pas à des échantillons de la langue mais à des résultats de travaux de recherche, proches en cela des citations de travaux antérieurs.

5.2.2. *Des textes comportant une dimension (un style ?) historique*

Les temps verbaux font l'objet d'usages contrastés dans l'article de recherche, entre les disciplines et en fonction du statut de l'auteur, et au sein même de catégories homogènes à chacun de ces deux niveaux (Rinck, 2006). L'analyse des spécificités des textes de 15 auteurs montre qu'ils ressortent là aussi parmi les traits les plus discriminants (Poudat et Rinck, 2005). C'est le cas notamment des temps du passé, dont l'usage est marqué par rapport à une tendance forte du genre de l'article, celle d'un texte où domine le présent. Les occurrences de verbes au passé se trouvent par ailleurs corrélées à d'autres indices d'un mode que l'on peut qualifier d'historico-narratif. Celui-ci correspond à un profil singulier d'articles de linguistique, comme l'a mis en évidence l'entreprise de classification automatique menée sur ce genre (Poudat et Rinck, 2006). Ce profil s'avère être celui des textes de 2 auteurs dont nous traiterons ici, A3 et A9.

Les textes de A9 doivent leur singularité à un usage important de l'imparfait et du plus-que-parfait, du passé simple, ainsi que des dates et des noms propres. Le mode historico-narratif qui les caractérise s'explique aisément par le fait que l'auteur est spécialiste de linguistique historique. Le présent de l'indicatif est en revanche sous-représenté (un tiers de moins en moyenne pour le présent et plus de deux fois moins de modaux au présent que dans les autres textes), et les spécificités des textes de cet auteur vérifient donc les tendances observées dans l'analyse du genre de l'article en linguistique (Rinck, op. cité) : les corrélations négatives entre l'usage du présent et de l'imparfait, et entre le présent et le passé simple, et la corrélation positive entre l'imparfait et le passé simple.

On relève également dans les textes de A9 une proportion plus importante de

marques de troisième personne (et plus précisément de *il* anaphoriques et de déterminants possessifs), qui semble caractéristique des textes historiques. Elles signalent en effet une progression thématique spécifique, dans des textes qui se focalisent sur un acteur ou un événement et s'organisent à partir d'un thème constant :

- (7) 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... [A9-074].

Le mode historico-narratif semble par ailleurs pouvoir être indépendant du passé simple, comme c'est le cas dans les textes de A3. L'auteur emploie à l'instar des textes journalistiques de très nombreux imparfaits narratifs :

- (8) La séparation et la non communication entre les approches de Saussure et dans une moindre mesure de Meillet, d'une part, et celle de la Grasserie ou de Van Genep d'autre part, nous montrent que la linguistique structurale dominante **s'est construite** en marginalisant les faits qui n'**entraient** pas dans le cadre ou dans le modèle. C'est-à-dire qu'elle se **donnait** la langue qui l'**arrangeait** : ou bien les faits **rentraient** dans le modèle ou bien on les **jetait** (ou on les **rejetait** vers la psycholinguistique, la sociolinguistique, approches considérées comme périphériques).[A3-001]

Dans les deux cas cités, A9 et A3, le mode historico-narratif est étroitement lié à l'objet d'étude et les occurrences de verbes au passé permettent d'en dire davantage sur le domaine de spécialité. Les textes de A9 ont en commun de fournir une analyse historique du champ de la linguistique, et ceux de A3 concernent les politiques linguistiques. Ils intègrent également une analyse historique de la linguistique comme nous l'avons vu avec l'exemple cité *supra*, mais recourent au mode historico-narratif pour traiter par exemple de la situation plurilingue de l'Égypte à la charnière du dix-neuvième et du vingtième siècle (A3-002), ou du statut des langues dans les relations internationales (A3-003). Le mode historico-narratif ne s'explique cependant pas seulement par l'objet, car il requiert en outre son traitement historicisant. Certains articles ou certaines parties d'articles traitent ainsi d'événements passés au présent ou au passé composé. Parler de mode historico-narratif ou de profil de texte ne revient-il pas de fait à tenter de cerner un « style historique » dans l'article en linguistique ?

De toute évidence, l'analyse des usages du genre propres à un auteur montre à quel point il est difficile de rendre compte de ce qui fait la diversité des textes. Les caractéristiques des textes de A3 et A9 correspondent, plus qu'à un style d'auteur, à un « style historique », c'est-à-dire à un profil de texte, lié à la fois à des objets de recherche spécifiques et à une construction linguistique spécifique de ces objets. Il singularise certains textes au sein du genre, et parfois les textes d'un même auteur.

Les usages du genre propres à un auteur intègrent cependant d'autres spécificités ; quant au « style historique », on l'a défini statistiquement par un faisceau de critères, mais il se présente dans les textes sous des formes variées. Ainsi, chez A3, la dimension historique signalée par l'usage de l'imparfait va de pair avec un

usage fort des noms propres comme chez A9, mais contrairement à ce que l'on a noté chez ce dernier, les dates ne sont pas discriminantes. A3 recourt abondamment par ailleurs aux connecteurs de temporalité reliant les événements décrits. Dans le même sens, si A3 et A9 ont en commun l'imparfait, le style du second associe à ce temps le passé simple, que l'on retrouve comme spécificité de A4.

Ce dernier critère est corrélé à l'imparfait au niveau du genre, mais le style historique s'actualise tout aussi bien avec les deux temps qu'avec l'un ou l'autre. Outre la différence aspectuelle entre les deux, on peut interpréter l'usage du passé simple comme une spécificité stylistique du fait de la littérarité qui caractérise ce temps verbal selon R. Barthes (1972 :25). Il peut attester en cela d'un certain « goût de l'écriture » (Sansot, 1994), ce qui se manifeste avec d'autres indices dans les articles de lettres, où il est significativement plus utilisé que dans ceux de linguistique (Rinck, op. cité). Il renvoie en outre à des spécificités idiolectales, dans la mesure où, dans les deux disciplines, les doctorants en font un usage moindre que les autres chercheurs, et l'on peut alors comprendre ces spécificités en référence aussi à la variation diachronique de la langue, avec le recul signalé du passé simple en français.

6. Conclusion

L'analyse des styles au sens d'usages singuliers du genre vérifie l'existence de traits homogènes entre les textes d'un même auteur et suffisamment singuliers pour les distinguer de ceux des autres. Menée au niveau morpho-syntaxique et lexical, sans d'autre *a priori* que l'intérêt des traits retenus pour la description du genre de l'article en linguistique, notre approche pose la question de savoir en quoi les traits qui distinguent les textes d'un même auteur peuvent être compris comme des traits stylistiques.

Nous avons ainsi isolé les caractéristiques qui peuvent s'assimiler à un style d'auteur, au sens de spécificités idiomatiques dans ses textes. Mais peut-on alors parler de styles non marqués, lorsque leurs traits se confondent avec les caractéristiques du genre ? Rien ne permet par ailleurs de dire dans quelle mesure ces traits sont conscients et intentionnels, ni en quoi ils s'identifient ou non à ce qui, pour les lecteurs, fait style, ou fait le style d'un auteur en particulier. Quant à l'ensemble des spécificités propres aux textes d'un même auteur, elles montrent la complexité des usages singuliers du genre et de leur fonction individuante. L'opposition entre celles qui se rattachent à un style d'auteur ou à un idiolecte et les autres ne suffit pas à rendre compte des multiples dimensions qui font l'identité des textes de l'auteur-chercheur dans son champ.

La tension entre normes individuelles et collectives joue de manière variée, selon le degré auquel les caractéristiques des textes d'un auteur s'assimilent aux tendances caractéristiques du genre. Mais le genre ne se restreint pas à ce qui est commun aux textes : leur diversité elle-même se structure de manière relativement stable, en fonction de configurations phrastiques ou énonciatives par exemple, comme en fonction d'objets de recherche et de types d'études menées.

Les usages singuliers du genre correspondent à des profils de textes et peuvent dans nombre de leurs traits être partagés par plusieurs auteurs. Ils renvoient à des spécificités linguistiques, à un possible « goût de l'écriture », à des effets de formation et à des différences diachroniques peut-être, à des objets et à des domaines de référence dans lesquels les productions s'inscrivent et vis-à-vis desquels elles sont accréditées. Les méthodes de la linguistique de corpus favorisent la mise en

évidence de faisceaux de traits et doivent permettre d'affiner la description de ces profils de texte, et avec eux celui du champ considéré. La diversité du genre de l'article est à la fois celle des pratiques scripturales et celle des pratiques de recherche, et la tentative de mieux cerner ce qui fait l'individu dans ses textes aboutit à toujours mieux comprendre à quel point il est social.

Références bibliographiques

- BAILEY, R. W. (1969) : « Statistics and style : a historical survey », in Dolezel, L. et Bailey, R. W. (eds), *Statistics and style*, New York, Elsevier, 48-85.
- BARTHES, R. (1972) : *Le degré zéro de l'écriture*, Paris, Seuil (1^{re} édition 1953).
- BEAUDOIN, V. et YVON, F. (2004) : « Contribution de la métrique à la stylométrie », Journées d'Analyse des Données Textuelles, Louvain, en ligne sur http://www.cavi.univ-paris3.fr/lexicometrica/jadt/jadt2004/pdf/JADT_009.pdf
- BENZÉCRI, J.-P. (1992) : *Correspondence analysis Handbook*, New York, Dekker.
— (1981) : *Pratique de l'analyse des données : linguistique et lexicologie*, Paris, Dunod.
- BRUNET, E. (1991) : « What Do Statistics Tell Us ? », *Research in Humanities Computing*, 1, 70-92.
- BRUNET, E. et MULLER, C. (1988) : « La statistique résout-elle les problèmes d'attribution ? », *Strumenti critici*, Sett., 367-387.
- CROMBIE, A. C. (1994) : *Styles of Scientific Thinking in the European Tradition : The history of argument and explanation especially in the mathematical and biomedical sciences and arts*, London : Duckworth.
- DÉTRIE, C. et NEVEU, F. (2005) : *L'idiolecte. Du singulier dans le langage. Cahiers de Praxématique*, 44, Praxiling, Université de Montpellier 3.
- FLØTTUM, K., DAHL, T. et KINN, T. (2006) : *Academic Voices. Across languages and disciplines*, Amsterdam/Philadelphia, John Benjamins.
- GAYON, J. (1996) : « De la catégorie de style en histoire des sciences ». *Alliage*, 26, en ligne sur <http://www.tribunes.com/tribune/alliage/26/gayo.htm>
- GROSSMANN, F. et RINCK, F. (2004) : « La surénonciation comme norme du genre. L'exemple de l'article de recherche et du dictionnaire en linguistique », *Langages*, 156, Paris, Larousse, 34-50.
- GUEUNIER, N. (1969) : « La pertinence de la notion d'écart en stylistique », *Langue Française*, 3, Paris, Larousse, 34-45.
- HOLMES, D.I. (1994) : « Authorship Attribution », *Computers and the Humanities*, 28-2, Kluwer Academic Publishers, 87-106.
- LATOUR, B. et WOOLGAR, S. (1979), *Laboratory Life : the Social Construction of Scientific Facts*, Los Angeles/Londres, Sage.
- LEBART, L. et SALEM, A. (1994) : *Statistiques textuelles*, Paris, Dunod.

- MÉZAILLE, T. (2003) : « Thématiques littéraires : enseignement des textes numériques », en ligne sur *Texto !* à l'adresse http://www.revue-texto.net/Inedits/Mezaille_Thematiques.html
- MONIERE, D. et LABBÉ, D. (2002) : « Essai de stylistique quantitative. Duplessis, Bourassa et Lévesque », en ligne sur <http://www.upmf-grenoble.fr/cerat/Recherche/PagesPerso/LabbeMoniere.pdf>
- OUELLET, P. (1989) : « “Par exemple...” : statut cognitif et portée argumentative de l'exemplification dans les sciences du langage », in Reicher-Béguelin, M.-J., (éd.), *Perspectives méthodologiques et épistémologiques en sciences du langage, Actes du Colloque de Fribourg* 11-12 mars 1988, Berne, Peter Lang, 95-114.
- PONTILLE, D. (2004) : *La signature scientifique. Une sociologie pragmatique de l'attribution*, Paris, CNRS Editions.
- POUDAT, C. (2006) : *Etude contrastive de l'article de revue linguistique dans une perspective d'analyse des genres*. Thèse de doctorat sous la direction de G. Bergounioux, Université d'Orléans.
- POUDAT, C., et RINCK, F. (à paraître) : « Genres scientifiques et styles d'auteur : des variations stylistiques de l'article de revue linguistique », *Actes des Journées linguistique de corpus*, septembre 2005, Lorient.
- POUDAT, C., et RINCK, F. (2006) : « Contrastes internes et variations stylistiques du genre de l'article en linguistique », *JADT06, Actes des 8^{èmes} Journées d'analyse des données textuelles*, avril 2006, Université de Besançon, 785-796.
- RABATEL, A. (ed.), (2004) : *Effacement énonciatif et discours rapportés, Langages*, 156, Paris, Larousse.
- RASTIER, F. (2001) : *Arts et sciences du texte*, Paris, Presses Universitaires de France.
- RIEGEL, M., PELLAT, J.-C. et RIOUL, R. (1994) : *Grammaire méthodique du français*, Paris, Presses Universitaires de France.
- RINCK, F. (2006) : *L'article de recherche en sciences du langage et en Lettres. Figure de l'auteur et identité disciplinaire du genre*. Thèse de doctorat sous la direction de F. Boch et F. Grossmann, Université Stendhal Grenoble 3.
- SANSOT, P., (1994) : « Le goût de l'écriture : une dérive épistémologique ? », *Communications*, 58, Paris, Seuil, 61-67.
- TUTIN, A. (2007) : « Traitement sémantique par analyse distributionnelle des noms transdisciplinaires des écrits scientifiques », *Conférence TALN*, juin 2007, Toulouse.
- VALETTE, M. (2006) : « La genèse textuelle des concepts scientifiques. Étude sémantique sur l'œuvre du linguiste Gustave Guillaume », *Cahiers de Lexicologie*, 89, Paris, Honoré Champion, 125-142.