

Un modèle exploratoire d'analyse de données textuelles au service de la didactique de l'écrit dans les collèges québécois

Libersan, Lucie

Cégep Ahuntsic, Université du Québec à Montréal
lucie.libersan@collegeahuntsic.qc.ca

Foucambert, Denis

Université du Québec à Montréal
foucambert.denis@uqam.ca

1 Introduction

Nombreux sont les manuels québécois de rédaction technique, scientifique ou administrative qui comportent une section spécialement consacrée à la qualité de la langue écrite. On y retrouve généralement des guides théoriques sur les catégories et fonctions syntaxiques, des listes de connecteurs logiques, des recommandations générales sur les difficultés orthographiques (participes passés, homophones, etc.). Il n'est pas rare d'y lire également des « conseils pratiques » sur la syntaxe (*faites des phrases complètes, évitez les phrases trop longues*), la cohérence du texte (*accordez une attention particulière aux mots de liaison*) et le vocabulaire (*évitez les termes familiers*). Ces rappels ne sont certainement pas nuisibles en eux-mêmes : nul ne contestera l'utilité de bien construire ses phrases, d'être cohérent, de châtier son style. Bien souvent, ils sont malheureusement trop généraux pour être vraiment utiles dans des contextes où l'enjeu est de produire des documents bien écrits et efficaces – sans nécessairement avoir à réviser l'ensemble des règles grammaticales. Par exemple, les policiers transcrivent des propos souvent très grossiers dans leurs rapports quand ils rendent compte de leurs interactions avec les suspects : ils doivent pour cela maîtriser les règles du discours rapporté, dans le but précis de respecter les écarts de registre. Les graphistes emploient peu de connecteurs logiques dans leurs argumentaires de projets, car ils doivent « séduire » leurs clients sans leur donner l'impression d'argumenter; leur défi est justement de trouver des tournures permettant d'éviter l'emploi des adverbes, subordinées et groupes prépositionnels à valeur argumentative. Dans les devis, les technologues en architecture emploient plus souvent des phrases infinitives que des phrases conformes au modèle canonique; pour eux, la difficulté consiste à s'assurer de la cohérence syntaxique de ces énumérations de groupes verbaux « sans sujets ». Ces personnes et tant d'autres rencontrent de réels défis d'écriture, et pourtant ont disposé à ce jour de peu d'outils pour développer des compétences rédactionnelles particulières aux différents programmes d'études (*rédiger des rapports* en Techniques policières, *écrire et interpréter des devis* en Technologie de l'architecture, *justifier une proposition visuelle* en Techniques de graphisme, etc.).

Le projet *Stratégies d'écriture dans la formation spécifique*¹ vise à fournir aux élèves qui poursuivent des études postsecondaires en français des outils didactiques leur permettant de réinvestir leurs connaissances linguistiques dans des situations d'écriture propres à leurs cursus de formation. L'objectif du projet est de produire des fascicules² sur des genres de textes en usage au cégep (*cf. infra, 2.1*) – certains typiquement scolaires, d'autres directement liés aux secteurs professionnels que les élèves se préparent à intégrer. S'adressant en premier lieu aux élèves et aux enseignants des cégeps, ces outils visent à cibler un nombre restreint de structures linguistiques significativement sollicitées dans des genres textuels et à mettre ces structures en parallèle avec des habiletés communicationnelles mises en œuvre dans les programmes. Le présent article passe en revue les différentes étapes de la démarche heuristique suivie pour déterminer les genres en usage dans le cursus du cégep et pour identifier les structures linguistiques caractéristiques de

chacun d'eux. Une attention toute particulière sera accordée à la méthode d'analyse de corpus utilisée pour obtenir ces données.

2 Problématique de la langue écrite dans les cégeps francophones

2.1 Un rapide tour d'horizon

Au Québec, l'enseignement supérieur comprend, en plus de l'ordre universitaire, un niveau collégial accessible aux élèves titulaires d'un diplôme d'études secondaires. Les *collèges d'enseignement général et professionnel (cégeps)*, majoritairement francophones, offrent à ces élèves le choix entre une formation préuniversitaire d'une durée prévue de deux ans, et une formation technique donnant accès au marché du travail ou à l'université, d'une durée de trois ans. Le secteur préuniversitaire, qui accueille environ la moitié de l'effectif³, propose neuf programmes parmi lesquels figurent *Sciences de la nature*, *Arts et lettres* et *Sciences humaines*. L'autre moitié de l'effectif se répartit dans près de 170 programmes techniques de divers domaines, par exemple la santé (*Radio-oncologie*), les bâtiments et travaux publics (*Génie civil*), l'administration (*Archives médicales*), les services juridiques (*Techniques policières*) et les arts (*Graphisme*). En plus des cours de la *formation spécifique*, tous les candidats au diplôme d'études collégiales doivent réussir les cours obligatoires de la formation générale (français, philosophie, langue seconde et éducation physique), de même que l'épreuve uniforme de français (ÉUF), une dissertation littéraire de 900 mots écrite en temps restreint. Le cursus collégial s'adresse aux finissants du cours secondaire, mais également aux adultes de plus de 20 ans effectuant un retour aux études; ces derniers, qui constituent environ 40 % de l'effectif⁴, s'inscrivent surtout dans les programmes techniques.

2.2 Le problème de l'orientation pédagogique des mesures de valorisation

Pour un grand nombre d'élèves, l'adaptation aux exigences du cégep représente un défi considérable. En témoignent un taux d'échecs élevé dans les cours du premier trimestre⁵ et, corollairement, une diplomation relativement faible dans les délais prévus selon le régime d'études⁶; le nombre important d'échecs à l'ÉUF (15,8% en 2010-2011; cf. DGAUC-DEC, 2011) vient appuyer ces constats. L'évaluation de cette épreuve certificative révélant surtout des faiblesses en français écrit⁷, il peut sembler légitime, de prime abord, que les mesures d'aide à la réussite – cours d'appoint, tutorat par les pairs, etc. – mettent l'accent sur la révision du code écrit ou sur la préparation à l'examen proprement dit (CCDMD, 2006; Table Éducation Montréal, 2002). Or, il est douteux que ces mesures à elles seules suffisent à améliorer les habiletés rédactionnelles des élèves dans un contexte d'enseignement supérieur. Plusieurs études constatent en effet l'efficacité limitée des interventions ponctuelles comme le tutorat par les pairs (Barbeau, 2007; Lefrançois et al., 2005); parallèlement, des critiques s'élèvent à l'égard d'une didactique de l'écrit lacunaire, axée principalement sur des exercices décontextualisés de grammaire et d'orthographe, et largement répandue au collégial (cf. par exemple Lecavalier, 2003).

Dans cette conjoncture où les interventions visant à améliorer et à valoriser la langue écrite s'attachent principalement au code orthographique, les habiletés rédactionnelles en contexte de communication authentique passent souvent au second plan. Pourtant, les exigences à cet égard sont élevées, et ce, dans la formation spécifique autant que dans les cours de littérature et de philosophie. Tous les programmes incluent dans leur cursus des critères de performance concernant les compétences rédactionnelles; ces dernières s'exercent dans des *genres textuels* (cf. *infra*, 3.1) typiquement scolaires comme le rapport de laboratoire et le compte rendu critique, mais aussi dans des genres orientés vers des contextes professionnels comme le devis technique (dans les programmes de génie) ou l'argumentaire de projet (dans les arts appliqués). Faisant face à des situations d'écriture plus exigeantes que celles qu'ils ont connues précédemment, les élèves sont appelés à maîtriser simultanément plusieurs genres différents, dont certains leur sont inconnus. Or, le laps de temps dont ils disposent pour assimiler les conventions particulières à chacun d'eux est bref compte tenu de l'ampleur des apprentissages à réaliser.

2.3 Des pistes de solutions

Une fois admis que les mesures d'appoint misant sur la révision systématique du code orthographique ne contribuent que de manière limitée à améliorer les habiletés rédactionnelles des élèves dans des situations de communication complexes, il devient évident que la valorisation de la langue écrite incombe à l'ensemble de la communauté collégiale, et qu'elle demande à être recadrée. Parce que les enseignants de la formation spécifique ne sont pas des didacticiens du français écrit, les interventions, idéalement intégrées au parcours pédagogique régulier, doivent être associées aussi concrètement que possible à des compétences requises dans les programmes. Il est donc souhaitable que ces interventions soient liées aux aspects communicatifs des *genres textuels* (cf. *infra*, 3.1) en usage dans le cursus collégial et, par conséquent, aux *configurations* de structures linguistiques (Bronckart et al., 1985) – structures syntaxiques, énonciatives, pragmatiques, textuelles – les plus sollicitées dans chacun d'eux; autrement dit, l'effort de valorisation de la langue écrite dans la formation spécifique devrait tenir compte de la *compétence générale* (Maingueneau, 2007). Les outils didactiques proposés dans le cadre du projet *Stratégies d'écriture dans la formation spécifique* visent à soutenir les élèves et les enseignants dans cette démarche. L'approche privilégiée s'inspire d'une conception « spiralaire » des apprentissages, qui définit le développement linguistique comme une série d'approfondissements plutôt que comme une accumulation de connaissances (Chartrand, 2008). On fait ainsi l'hypothèse que les jeunes scripteurs deviennent compétents en réinvestissant leurs connaissances et habiletés dans des situations de communication de plus en plus spécialisées, et ce, tout au long de leur formation. Pour qu'ils puissent prendre conscience de cette évolution, il semble logique d'orienter la révision linguistique vers des configurations restreintes de structures linguistiques, présentées dans des *activités sociales* (Bronckart, 1996, 2004) où elles sont significativement représentées. Nous avons déterminé pour ce faire de planifier les outils didactiques à partir des données d'une analyse de corpus cherchant à découvrir ces configurations (Bronckart et al., 1985) dans les différents écrits de la formation spécifique, avec la conviction qu'au final, une telle approche devrait aider les élèves à parfaire leurs habiletés dans un cadre concret et réaliste. Des modèles d'outils pédagogiques prenant pour appui l'étude des propriétés linguistiques des genres (Chartrand, 2011; Dolz et al., 2001) offraient d'ailleurs des pistes intéressantes en ce sens.

Cette étude préalable était selon nous nécessaire pour que la planification des fascicules soit fondée sur des données objectives. Le cadre méthodologique dont nous nous dotions évitait d'une part de fonder la description des propriétés linguistiques des genres uniquement sur des connaissances pratiques ou intuitives, parfois empreintes de subjectivité, et permettait d'autre part d'obtenir des données précieuses sur certains écrits spécialisés dont les caractéristiques nous étaient peu familières.

3 Genres et structures langagières

3.1 Quelques éléments théoriques

La proposition de planifier des outils pédagogiques en français écrit sur la base d'une étude de corpus repose sur une « linguistique de la parole » selon laquelle la langue n'est qu'une virtualité hors de l'acte d'énonciation qui la réalise dans le discours (Benveniste, 1974). Une telle orientation pose en principe que les unités de la langue prennent leur sens lorsqu'un sujet parlant ou un scripteur se les approprie dans un contexte social précis; elle conduit également à postuler que l'interaction entre un locuteur et un ou plusieurs interlocuteurs, réels ou virtuels, est une condition nécessaire à la réalisation de l'acte énonciatif (principes défendus par exemple dans Bronckart, 1996 ou Maingueneau, 2007). Une fois affirmée la primauté de l'interaction dans l'acte de parole, le *genre de discours* (cf. *infra*) peut se concevoir, suivant Bakhtine (1984), comme une interface nécessaire entre la langue – au sens saussurien – et sa réalisation dans le discours; fournissant le cadre conventionnel à l'intérieur duquel les membres d'une communauté reconnaissent la structure et le contenu des énoncés, le *genre* apparaît ainsi comme « l'outil » essentiel pour structurer la parole et rendre cette dernière efficace et compréhensible dans l'interaction. Dans un contexte pédagogique, adopter une description linguistique *descendante*, qui va du discours effectif

considéré dans sa dimension générique (ou typologique) jusqu'à la morphosyntaxe (Bronckart, 1996), c'est possiblement donner aux élèves les moyens de comprendre la portée sociale et communicative des faits de langue et d'ancrer leur produit dans cette réalité. Une telle approche semble convenir tout particulièrement à de jeunes adultes qui préparent leur entrée à l'université ou sur le marché du travail.

Dans la conception adoptée ici, le *genre* désigne «une forme langagière [...] relativement conventionnelle et stable dans le temps et dans une culture donnée, qui présente des caractéristiques de divers ordres» (Chartrand, 2008 : 11). Ces caractéristiques ont trait à des *conditions de réussite* (Maingueneau, 2007); la pratique légitime d'un genre – telle que l'entend Bourdieu (1982), par exemple – nécessite en effet de se conformer aux paramètres d'un certain cadre énonciatif (rôle des partenaires, temps et lieu de l'énonciation), à une finalité spécifique (décrire, convaincre, séduire, raconter, inciter, etc.), à une structure que l'on peut reconnaître (plan d'ensemble, construction des paragraphes), ainsi qu'à des formes linguistiques conventionnelles (construction des phrases, vocabulaire, marqueurs textuels et énonciatifs). Ainsi définie, la notion de *genre* s'éloigne sensiblement des notions de *type de texte* ou de *type de discours* (Bronckart et al., 1985). Un texte se classe dans un *genre* en référence à des repères culturels et sociaux (Bronckart, 2004; Chartrand, 2008) et à une compétence acquise (Maingueneau, 2007) : on reconnaît un argumentaire de projet, un rapport d'intervention, un devis d'architecte pour avoir déjà fréquenté ces genres d'écrits; les conventions des mêmes textes seront difficiles à décoder pour quiconque ne possède pas les repères idoines. Par comparaison, le type textuel apparaît soit comme une unité discursive uniforme, identifiable au niveau de la séquence (Adam, 1999), soit comme un archétype (ex. : dialogal, narratif, théorique) réunissant un certain nombre de productions ayant des caractéristiques communes (Bronckart et al., 1985). Qu'elle soit considérée d'un point de vue séquentiel ou archétypal, la notion de *type* demeure une abstraction théorique; dans sa forme concrète, un texte est rarement composé uniformément d'un seul type, mais résulte au contraire d'un «montage» complexe de plusieurs séquences typologiques (Adam, 1999; Chartrand, 2008).

Pour rendre compte de la richesse des pratiques d'écriture dans les cours de la formation spécifique, le *genre* est ainsi apparu comme l'unité d'analyse la plus prometteuse. Or, même en admettant que la compétence générique permette de reconnaître et d'utiliser des genres dans des contextes particuliers, les critères de catégorisation, nombreux et hétérogènes (Bronckart, 2004; Maingueneau, 2007), rendaient difficile l'établissement d'une taxonomie cohérente des genres en usage dans le cursus collégial. La grande diversité des programmes entraîne forcément une grande diversité dans les pratiques d'écriture; s'il est aisé de reconnaître et de regrouper les variantes d'un genre scolaire aussi codé que le rapport de laboratoire, il est plus ardu par exemple de faire cet exercice avec les divers écrits regroupés sous le terme de «document d'information technique», étroitement liés à la pratique professionnelle et usuels dans plusieurs programmes de génie (civil, industriel, mécanique, etc.). En l'absence d'un critère unique permettant de hiérarchiser les variantes des genres et de les classer sous des «hyperonymes», nous avons résolu d'établir une classification sur la base de renseignements fournis par des enseignants du réseau collégial (cf. *infra*, 4.2), puis de vérifier empiriquement la validité de la classification et de dégager les particularités langagières des différents genres de cette classification au moyen de statistiques exploratoires. En adoptant cette démarche heuristique, nous espérons au final établir des bases «objectives» pour caractériser les genres.

3.2 La caractérisation des genres : cadre méthodologique

La méthode observée pour obtenir les données destinées à guider la planification du matériel pédagogique s'inspire d'une étude effectuée sous la direction de J.-P. Bronckart (1985). L'hypothèse de travail, exploratoire, était «qu'à des conditions de production différentes, suffisamment contrastées, correspondraient des types de textes différents» et que «les conditions de production devraient permettre de prévoir les caractéristiques morphosyntaxiques du texte» (1985 : 67). L'analyse portait principalement sur un corpus de textes classés a priori dans trois types, à savoir le *discours en situation*, le *discours théorique* et la *narration*. L'appartenance à l'une des trois catégories était la variable de classement a priori; 27 unités linguistiques sélectionnées au terme d'essais successifs constituaient la grille des variables discriminantes (1985 : 62). Les données recueillies à la suite du relevé des variables ont été

compilées sous forme d'indices calculés en fonction d'un référentiel (1985 : 74-75), puis soumises à une analyse factorielle discriminante. En établissant des corrélations entre l'appartenance à l'un des trois types et les 27 unités linguistiques, le traitement statistique devait révéler l'existence d'axes discriminants correspondant aux oppositions entre les types, vérifier la validité de la classification a priori des textes dans l'un ou l'autre des trois types et faire apparaître, aux extrémités des axes, des systèmes multidimensionnels de variables en interaction, discriminatives par leur fréquence ou leur rareté, opposant les types textuels les uns aux autres.

La méthode d'analyse statistique utilisée a effectivement fait ressortir des configurations d'unités permettant de distinguer les trois types entre eux (1985 : 85-88) et a permis de vérifier la validité du classement initial. L'hypothèse exploratoire étant confirmée, ce cadre méthodologique se révélait prometteur pour notre propre étude, moyennant quelques nécessaires réaménagements dus aux différences opposant les unités d'analyse respectives. L'étude de J.-P. Bronckart et *al.* portait en effet sur des *types textuels* très généraux, fortement contrastés, alors que notre projet s'intéressait à des *genres de textes*, certains assez proches les uns des autres, comme le rapport de recherche, le rapport de stage et le compte rendu critique (*cf. infra*, 4.2). Afin que le traitement statistique puisse révéler des contrastes entre des genres présentant plusieurs points communs, il a semblé pertinent de sélectionner un plus grand nombre de variables de diverses natures; cette décision présentait par ailleurs l'avantage de préserver le caractère exploratoire de l'analyse, ce qui n'aurait pas été le cas si nous avions utilisé la grille des « 27 variables », dont le caractère explicatif avait été démontré dans un autre contexte.

4 Méthodologie

4.1 Rappel des objectifs

L'objectif terminal du projet était d'offrir aux élèves et aux enseignants du collégial des outils offrant des stratégies de rédaction adaptées à la réalité de la formation spécifique. Le défi à relever dans le cadre de la recherche préalable à la planification du matériel pédagogique était de cerner les structures linguistiques distinctives des genres en usage dans les différents programmes d'étude, plus précisément des indices du cadre énonciatif, ainsi que certains indices de construction syntaxique et de structuration textuelle (*cf. infra*, 4.3). L'hypothèse de recherche, exploratoire tout comme celle de notre modèle méthodologique, était qu'à des genres différents correspondraient des configurations particulières d'unités discriminatives par leur fréquence ou leur rareté. Nous postulions en corollaire que les configurations obtenues au terme de l'analyse seraient interprétables sur le plan discursif, en d'autres termes, qu'il serait possible d'établir un parallèle entre les données de l'analyse et les compétences mises en œuvre dans les différents genres. L'intention était de fournir aux élèves et aux enseignants des stratégies concernant l'énonciation, la morphosyntaxe et la cohérence du texte, mais surtout, à donner un sens aux structures de la langue ainsi observées dans leur contexte communicatif.

4.2 Sélection des genres textuels et collecte données

Nous avons d'abord réalisé une enquête auprès d'enseignants du réseau collégial afin de déterminer les genres les plus représentatifs des pratiques d'écriture dans la formation spécifique (*cf. Libersan et al.*, 2010). Près de 45 % des cégeps y ont participé. Après un travail de sélection, nous avons établi une liste de 14 genres touchant 31 programmes d'études :

- argumentaire de projet⁸ (ARG)
- compte rendu critique (CR)
- correspondance administrative (ADM)
- correspondance d'affaires⁹ (AFF)
- curriculum vitæ (CV)
- dépliant publicitaire¹⁰ (PUB)

- document d'information technique¹¹ (TECH)
- dissertation argumentative (DIS)
- procès-verbal (PV)
- rapport d'intervention¹² (INT)
- rapport de laboratoire (LAB)
- rapport de recherche (RECH)
- rapport de stage (STA)
- résumé (RÉS)

La liste établie, nous avons constitué un corpus d'analyse de 20 textes par genre, soit un total de 280 textes, en assurant une représentation équitable des programmes d'études ayant participé à l'enquête. Comme il s'agissait de mettre en évidence les caractéristiques des genres à travers des exemples positifs, seuls des textes rédigés par des scripteurs reconnus comme compétents ont été retenus. La plupart proviennent d'ouvrages pédagogiques québécois ou européens ; certains, nettement moins nombreux, sont l'œuvre d'élèves réputés forts. Une préparation minimale des textes s'est révélée nécessaire afin de rendre les données « lisibles » pour les logiciels utilisés en vue du relevé automatique des variables. Nous avons ainsi abrégé les textes de plus de dix pages, en nous efforçant d'assurer une représentation équilibrée des grandes articulations textuelles (ex. : introduction, développement, conclusion). Dans le cadre de cette étude, nous avons par ailleurs pris la décision de supprimer le paratexte (tableaux, graphiques, figures), les formules scientifiques et les passages en langue étrangère, illisibles pour les logiciels d'analyse; ces unités, qui ne correspondaient pas directement à notre objectif de découvrir des configurations d'unités linguistiques dans les genres, seront toutefois réintroduites dans la phase de planification pédagogique et, le cas échéant, commentées dans les fascicules. Dans les textes d'élèves, nous avons corrigé les erreurs liées au code écrit (orthographe, syntaxe), de même que certaines erreurs de cohérence lorsqu'elles nuisaient à la clarté du propos; nous nous sommes bien sûr abstenus d'intervenir sur le contenu et sur la structure des textes.

4.3 Sélection des variables et relevé « automatique » des données

La grille utilisée pour l'analyse discriminante comportait au départ 72 variables (*tableau 1*). Notre intention au moment de constituer cette grille était de conserver le plus possible d'unités susceptibles de servir d'indices du cadre énonciatif, de la « complexité syntaxique » et de la cohérence, et ce, afin d'accorder toute la latitude souhaitée à la démarche exploratoire. Les variables sont regroupées ici dans des « catégories d'indices » à titre indicatif; dans les faits, l'interprétation discursive des variables discriminantes a tenu compte des dynamiques propres à chacune des configurations et des données contextuelles. Le tableau 1 présente comme des indices du cadre énonciatif les unités qui « accentuent la relation discursive au partenaire » (Benveniste, 1966 et 1974). Les marques de personne grammaticale ainsi que les marqueurs de temps et de lieu sont de ce nombre, mais également les marqueurs de modalité (Benveniste, 1974; Charaudeau, 1992; Genevay, 1994), qu'ils se situent au niveau de l'énonciation (interrogation, intimation, exclamation) ou de l'énoncé (adverbe modal, négation, etc.). Le cadre énonciatif comprend aussi les verbes performatifs (Austin, 1970; Ducrot, 1984) et les unités axiologiques comme les adjectifs et les adverbes (Charaudeau, 1992). Les subordinées et les compléments verbaux sont classés comme des indices de la complexité syntaxique (Riegel et al., 1994); ont également été incluses dans cette catégorie d'indices des unités susceptibles de servir de référentiels comme les verbes conjugués et les substantifs. Enfin, les différents types de connecteurs, ainsi que les procédés de reprise thématique (déterminants, pronoms) sont présentés à titre d'indices de la cohérence du texte (Chartrand, 2008; Genevay, 1994; Riegel et al., 1994).

La saisie des données a été effectuée à l'aide des logiciels *Antidote RX* (©Druide Informatique inc., 2007) et *Analyse de textes* (©AFL, 1994) et a fait l'objet d'une vérification systématique. Relevées en nombre d'occurrences dans un premier temps, les données ont été rapportées ensuite à un référentiel défini en fonction de la nature des variables, c'est-à-dire, dans la plupart des cas, au nombre de verbes ou au

nombre de mots. Font exception l'adjectif complément du nom, rapporté au nombre de noms communs, et la complexité, qui établit le rapport entre le nombre de subordonnées et le nombre de verbes.

| Pistes d'interprétation | Variables | |
|---|---|--|
| Indices de présence ou d'absence des interlocuteurs | <ul style="list-style-type: none"> • 1^{re} personne du singulier • 2^e personne du singulier • 1^{re} personne du pluriel | <ul style="list-style-type: none"> • 2^e personne du pluriel • pronom on |
| Indices de temps et de lieu (cadre énonciatif) | <ul style="list-style-type: none"> • Auxiliaire d'aspect • Infinitif • Indicatif présent • Imparfait • Futur simple • Passé composé | <ul style="list-style-type: none"> • Passé simple • Plus-que-parfait • Futur antérieur • Marqueurs de temps • Marqueurs de lieu |
| Indices de modalité | <ul style="list-style-type: none"> • Phrase interrogative • Phrase interrogative négative • Pronom interrogatif • Déterminant interrogatif • Adverbe interrogatif • Phrase exclamative • Phrase passive • Phrase impersonnelle • Phrase emphatique • Négation | <ul style="list-style-type: none"> • Adverbe exclamatif • Auxiliaire pouvoir • Auxiliaire modal (devoir/falloir/vouloir) • Subjonctif • Impératif • Conditionnel • Discours direct • Adverbe modal |
| Verbes performatifs ¹³ | <ul style="list-style-type: none"> • Verdictif • Exercitif • Promissif | <ul style="list-style-type: none"> • Comportatif • Expositif |
| Indices d'un point de vue axiologique | <ul style="list-style-type: none"> • Adjectif verbal • Gentilé¹⁴ • Adverbe de manière (-ment) • Adverbe comparatif | <ul style="list-style-type: none"> • Adverbe de degré • Attribut • Adjectif complément du nom |
| Indices et référentiels de complexité syntaxique | <ul style="list-style-type: none"> • Verbe conjugué • Nom propre • Nom commun • Complément transitif • Subordonnée complétive • Sub. interrogative indirecte • Sub. corrélatrice | <ul style="list-style-type: none"> • Sub. complément de phrase • Sub. adjointe • Sub. participiale • Sub. relative • Nombre de mots • Complexité syntagmatique (% de phrases avec subordonnées) |
| Indices de connexion | <ul style="list-style-type: none"> • Conjonction de coordination • Organisateur logique (adverbe, groupe prépositionnel, subordonnée, conj. de coord.) | <ul style="list-style-type: none"> • Préposition • Nombre romain (org. textuel) |
| Indices de progression et de reprise thématique | <ul style="list-style-type: none"> • Pronom possessif/démonstratif • Pronom indéfini • Déterminant défini • Déterminant démonstratif • Déterminant indéfini | <ul style="list-style-type: none"> • Déterminant possessif • Quantifieur • Nombre cardinal • Nombre ordinal |

Tableau 1 – Variables sélectionnées pour l'analyse discriminante (n=72)

4.4 Traitement statistique

Pour comparer les genres textuels et en extraire les particularités les plus remarquables, nous avons utilisé la technique statistique de l'Analyse Factorielle Discriminante (AFD), à l'aide du module « Modèles généraux d'analyse discriminante » intégré au logiciel *Statistica v7* (©Statsoft inc., 1984-2009). Les deux objectifs principaux de l'AFD sont de séparer du mieux possible l'ensemble des groupes (ici 14 genres textuels) à l'aide des 36 variables finalement calculées (*cf. infra, tableau 2*) et de mettre à jour les combinaisons de variables spécifiques à chacun des groupes/genres initiaux. Cette technique regroupe certaines des potentialités de l'analyse de variance – analyse inférentielle, classement a priori – et de l'analyse en composantes principales – analyse descriptive, regroupement a posteriori. Dans notre travail, la discrimination de populations textuelles est rendue possible par le classement, a priori, des productions à leur genre respectif. Le propos de l'analyse est de trouver les combinaisons particulières qui permettent – si l'AFD fonctionne – de reclasser automatiquement les textes produits dans leur genre initial. Une analyse reclassant plus de 90% des individus (les textes) permet d'observer les combinaisons de variables qui sont discriminantes des genres auxquels appartiennent ces textes.

Comme le rappelle Cibois (1987), l'usage des modélisations statistiques exploratoires à l'aide des différents types d'analyse factorielle nécessite un long travail d'ajustement des données, de création de sens à partir des premiers résultats et de leur validation conceptuelle. Nous ne présenterons ici que la partie finale de ce travail.

5 Résultats

Afin d'éviter les problèmes de multicolinéarité qui grèvent sérieusement les résultats d'une AFD, nous avons calculé les corrélations entre les variables et éliminé 13 variables présentant un coefficient de corrélation supérieur à 0,7 avec une ou plusieurs variables. Les résultats généraux de l'AFD sont très satisfaisants dans la mesure où ils permettent de reclasser 100 % des textes de départ. Par ailleurs, le modèle final présente une très forte puissance discriminatoire, puisque toutes les « distances de Mahalanobis au carré » entre les centroïdes des genres sont non seulement importantes, mais également statistiquement significatives. Les figures 1 et 2 présentent les projections des textes produits sur les 4 premiers axes de l'AFD. Au total, ces 4 facteurs représentent près de 80% de la variance (respectivement pour les axes de 1 à 4 : 41 %, 19%, 10% et 8% – valeurs propres respectivement de 14,05, 6,81, 3,36 et 2,99). On voit très clairement se dessiner des communautés de textes très regroupés autour de leur centre de gravité respectif.

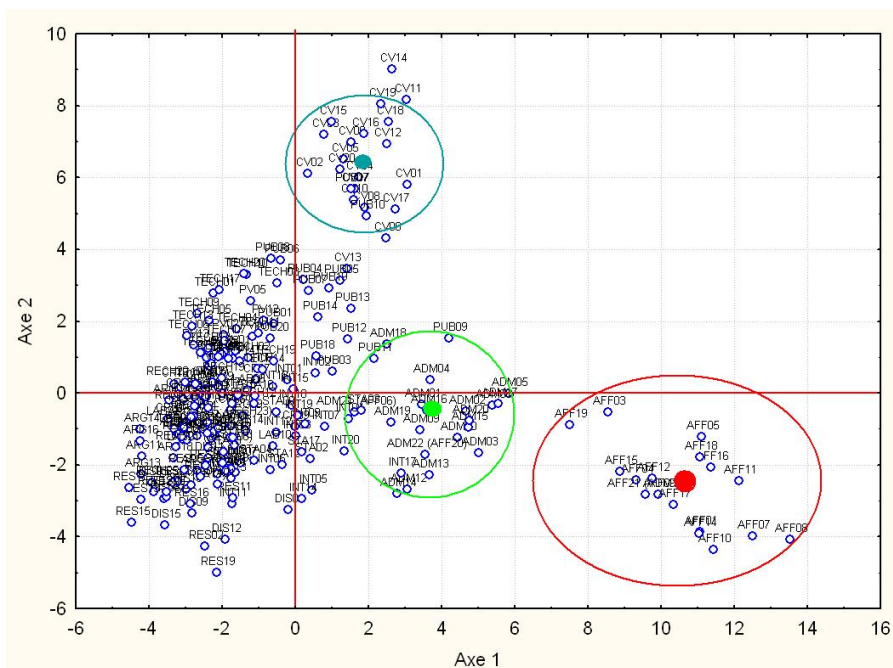


Figure 1 – Premier plan factoriel de l'AFD. Les points pleins représentent les centres de gravité des genres textuels, ici le rouge pour AFF (correspondances d'affaires), le vert pour ADM (correspondance administrative) et le bleu pour CV (curriculum vitæ).

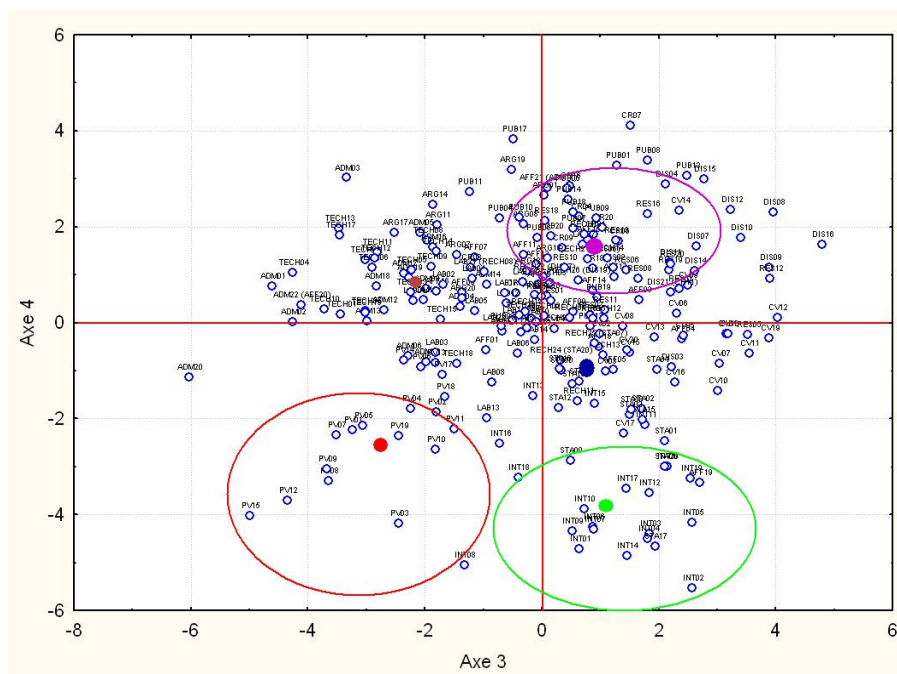


Figure 2 – Deuxième plan factoriel de l'AFD. Les points pleins représentent les centres de gravité des genres textuels ici le rouge pour PV (procès-verbal), le vert pour INT (rapport d'intervention) et le mauve pour RES (résumé)

On pourrait également projeter sur les mêmes axes les variables qui ont servi à les construire de manière à examiner leurs contributions respectives, en se méfiant toutefois des problèmes de parallaxe dus à la projection des points ayant des coordonnées dans un espace à 14 dimensions sur une succession de plans factoriels chacun en deux dimensions. Des moyens géométriques sont bien évidemment disponibles pour mesurer la véritable proximité entre variables et facteurs. La place nous manque dans cette contribution pour exposer l'ensemble du travail de cette analyse exploratoire des données, mais une autre manière d'observer les contributions des variables à chacun des groupes est de nous servir du modèle linéaire général qui est appliqué par l'AFD pour résoudre la classification (après avoir recodifié – de façon interne – les membres de groupes enregistrés dans la variable dépendante catégorielle en plusieurs variables dépendantes codées et factices). Les résultats issus de l'analyse de régression¹⁵ sont résumés dans le tableau 2 qui présentent la significativité des variables dans le modèle statistique utilisé (pour simplifier la présentation, nous ne présentons ici ni les paramètres estimés courants ni les paramètres estimés centrés-réduits – bêta – qui permettent de comparer le poids respectif des variables indépendantes).

Ces résultats permettent non seulement de repérer les variables significativement associées à chacun des genres textuels (par l'indice de significativité), mais également d'en mesurer le sens, donc la rareté ou la fréquence dans un genre particulier (de manière interne à partir du signe des bêtas). Par exemple, si l'on s'arrête un instant sur le rapport de recherche (RECH), on observe qu'il se distingue par un emploi important de subordonnées complétives et d'adjectifs compléments du nom (*cf. infra, tableau 2*). En revanche, on n'y trouve que très peu de marques de la 1^{re} personne du singulier et de noms communs (ou plus précisément, une proportion faible de noms communs par rapport au nombre total de mots), en comparaison avec les autres genres du corpus. Il est important de noter que la variable subordonnée complétive n'est pas discriminative uniquement du rapport de recherche : on la retrouve (en positif) dans le procès-verbal (PV). En revanche, pour ce dernier genre, on remarque que les noms communs sont particulièrement présents, contrairement à ce que l'on observe dans les rapports de recherche – ce qui ne signifie nullement, bien entendu, qu'ils en soient totalement absents.

Un examen détaillé du tableau 2 montre clairement la complexité des combinaisons de variables qui discriminent au mieux les genres étudiés; il s'agit toujours, lorsqu'on décrit ce type de résultats, d'analyser des constellations dynamiques de variables opposant, ici, les genres les uns aux autres. Autrement dit, il faut garder à l'esprit que les données obtenues sont relatives en ce sens qu'elles révèlent des configurations permettant de discriminer les genres à l'étude les uns par rapport aux autres.

| Genre | Unités distinctives fréquentes | | Unités distinctives peu fréquentes | |
|-------|---|---|--|--|
| ADM | Aux. d'aspect* Aux. de mode*** Futur antérieur*** | 2 ^e pers. du pluriel** Nombre cardinal*** Marqueur de temps*** | Indicatif P-q-P* 1 ^{re} pers. du sing.* | Nom propre* |
| AFF | Aux. d'aspect* Aux. de mode*** Sub. relative* Indicatif P-q-P* 1 ^{re} pers. du sing.*** 1 ^{re} pers. du plur.*** | 2 ^e pers. du sing.*** 2 ^e pers. du plur.*** Nom propre*** Dét. possessif** Nombre cardinal* Marqueur de temps*** | Discours direct** Phrase passive* Pronom <i>on</i> * Déterminant défini* Organisateur logique* | |
| ARG | Pronom <i>on</i> * Adjectif verbal** Conj. de coord.** | Dét. démonstratif*** Adj. compl. du nom** Mod. énonciative*** | Préposition*** | |
| CR | Phrase excl.*** Discours direct* | Org. logique* | Auxiliaire d'aspect* 2 ^e pers. du sing.** | Phrase passive** |
| CV | Sub. relative*** Indicatif PC** Nom commun** Marqueur de lieu** | Conj. de coord.* | Phrase excl.** Auxiliaire d'aspect* Auxiliaire de mode* Phrase passive** | Ind. prés.*** Impératif* 2 ^e pers. du plur.* Dét. défini* |
| DIS | Phrase excl.** Phrase interro. nég.*** Discours direct*** Attribut* | 2 ^e personne du singulier** Pronom <i>on</i> * Organisateur logique* | Indicatif PC * | |
| INT | Discours direct* Indicatif PC*** Indicatif P-q-P*** 2 ^e pers. du sing.* | Nom propre* Verbe conjugué** Marqueur de lieu* | Sub. relative*** Pronom <i>on</i> ** Adj. compl. du nom* | Attribut* 1 ^{re} pers. du plu.* Org. logique** |
| LAB | Attribut* 1 ^{re} pers. du plur.*** Pronom <i>on</i> *** | Déterminant défini** Nombre cardinal*** Préposition* | Verbe promissif*** Marqueur de temps*** | Sub. relative*** |
| PUB | Phrase excl.*** Verbe promissif* Impératif* | Conj. de coord.*** Indicatif présent*** Marqueur de temps* | Aux. d'aspect* Aux. de mode* Nom propre* | |
| PV | Phrase passive** Sub. complétive*** Nom propre*** | Préposition* Dét. défini* | Aux. de mode* 1 ^{re} pers. du sing.* Marqueur de lieu*** | Attribut* Indicatif PC* Indicatif P-q-P* |
| RECH | Sub. complétive* | Adj. compl. du nom* | 1 ^{re} pers. du sing.* | Nom commun* |
| RES | Sub. relative* Indicatif présent* Verbe conjugué** Dét. possessif** | Adj. compl. du nom*** Org. logique*** Verbe promissif*** | Discours direct** Passé composé** Mod. énonciative*** Phrase interro.-nég.* | 1 ^{re} pers. du sing.* 1 ^{re} pers. du plur.** Impératif** N. commun*** |
| STA | Phrase impersonnelle*** 1 ^{re} pers. du sing.*** | | Phrase interro.-nég.** Discours direct** Nombre cardinal* Adj. compl. du nom* Dét. défini*** | Nom propre** Nom commun** Dét. possessif*** Adj. verbal* |
| TECH | Aux. de mode*** Phrase passive*** | Adjectif verbal** Impératif*** | Marqueur de temps*** Adj. compl. du nom* | 2 ^e pers. du plur.* Nom propre* |

Tableau 2 – Contribution des variables à chacun des genres, issue des équations de régressions. Les unités distinctives fréquentes indiquent une présence plus importante de la variable dans le genre considéré que dans la moyenne générale des textes étudiés, tandis que les unités distinctives peu fréquentes dénotent l'inverse. N.B. : * = <0,05 ; ** = <0,01 ; *** = <0,001

6 Discussion

6.1 La généralisabilité des résultats

Le recours à un modèle exploratoire d'analyse textuelle a donné des résultats satisfaisants dans l'ensemble; quelques remarques sur nos choix méthodologiques s'imposent toutefois. Les analyses effectuées en vue de caractériser les genres sélectionnés présentent l'avantage de fonder l'observation sur une méthode reconnue; les configurations d'unités fréquentes et rares obtenues au terme des analyses factorielles discriminantes constituent donc une base qui peut être justifiée théoriquement. Ayant effectué la recherche avec le soutien du ministère de l'Éducation dans le but de produire des outils pédagogiques pour le réseau collégial, nous tenions à la transparence qu'apportait cette démarche, d'autant que l'approche pédagogique consistant à porter une attention soutenue à un nombre restreint de structures linguistiques en fonction des paramètres de la communication et des propriétés des genres bouscule certaines pratiques solidement ancrées dans la tradition scolaire. La méthode présentait également l'avantage non négligeable de guider l'observation des genres qui nous étaient moins familiers, notamment dans le cas du document d'informations techniques et de ses nombreuses variantes.

Malgré la confiance que nous accordons à cette démarche et à ses bases théoriques, nous avons conscience des nuances à apporter quant à la généralisabilité des résultats obtenus. Ainsi, tout en militant en faveur d'une description linguistique *descendante* (cf. *supra*, 3.1), nous reconnaissons que les observations effectuées portent sur un corpus relativement restreint et, surtout, sur un univers conceptuel dont la construction résulte d'une série de choix éventuellement discutables. Rappelons en premier lieu qu'en l'absence d'un critère taxonomique unique, la sélection et la classification des genres a été effectuée de manière empirique, sur la base des témoignages d'enseignants ayant accepté de répondre à l'enquête. Cette façon de procéder présentait l'avantage de sensibiliser le milieu, c'est-à-dire de faire en sorte que les enseignants de la formation spécifique se sentent directement concernés par le projet, et ce, dès l'étape de la recherche préalable. Cependant, comme les données de l'enquête ne couvrent pas la totalité des programmes du cégep selon les informations dont nous disposions, la représentativité réelle de certains genres sélectionnés peut sembler incertaine. Ce questionnement a d'ailleurs entraîné l'élimination de quatre d'entre eux à l'étape de la planification des fascicules quand, pour des raisons ayant trait aux contingences de l'échéancier, nous avons dû nous résoudre à restreindre le plan de travail : il s'agit du dépliant publicitaire, du procès-verbal, de la correspondance administrative et de la dissertation critique (la production d'outils sur ce dernier genre ayant été reportée à une étape ultérieure en raison de sa plus grande représentativité dans les cours de la formation générale). Par ailleurs, la sélection des variables discriminantes a été réalisée dans le souci que le plus grand nombre possible de structures de la langue soient représentées dans la grille (cf. *supra*, 4.3). Il faut noter toutefois que le choix des variables a été déterminé en partie par les possibilités des logiciels dont nous disposions pour la saisie des données. Cette contrainte a eu deux conséquences : d'une part, nous avons dû adopter la classification grammaticale proposée par ces logiciels et accepter d'introduire des redondances et des recoupements entre les catégories; d'autre part, nous avons composé avec les limites du relevé automatique, et ce, même en contrôlant au mieux la qualité de la saisie. En particulier, nous avons dû nous rendre à l'évidence que ce mode de relevé donnait des résultats décevants pour les verbes performatifs, qui, de fait, sont peu représentés dans le modèle final, ce qui nous incite à envisager de futures recherches visant à améliorer les outils d'analyse automatique. En définitive, nous comptons sur le nombre élevé de variables pour pallier certaines lacunes des classifications, et sur les analyses discriminantes pour réduire le modèle jusqu'à ce qu'il devienne explicatif. Il reste que le modèle tel qu'il apparaît au terme des analyses statistiques reflète le corpus que nous avons étudié, et non pas, de manière absolue, les genres étudiés. D'où la pertinence d'avoir consulté des enseignants au moment de la planification et de la rédaction du matériel, afin de compléter les données obtenues au terme de l'analyse (cf. *infra*).

6.2 L'interprétation des résultats et leur utilisation pédagogique

D'autres remarques peuvent être formulées, cette fois en lien avec les particularités de l'analyse factorielle discriminante et l'interprétation des résultats dans un cadre pédagogique. Les unités linguistiques désignées comme contributives à la distance entre les genres ne sont pas *discriminatives* dans l'absolu, mais dépendent au contraire d'une dynamique interactive. En principe, ce sont donc 1) les configurations de variables, et non les variables prises individuellement et 2) les configurations considérées les unes par rapport aux autres, et non chacune des configurations isolée des autres, qui révèlent les caractéristiques des genres. Lorsque les résultats d'une analyse factorielle discriminante sont considérés comme la finalité d'une démarche scientifique, la prise en compte de leur caractère interactif s'impose comme une évidence. Il en va tout autrement lorsqu'il est question d'utiliser ces résultats dans un cadre pédagogique; surgissent alors des objections qui induisent un questionnement sur l'interface entre l'analyse textuelle et la didactique de la langue. Ainsi, les résultats de l'analyse discriminante ont guidé la planification du matériel didactique; les fascicules ne présentent toutefois pas les caractéristiques des genres en les resituant dans la dynamique du corpus, sinon à travers des remarques visant à mettre en relief certaines finalités communicatives. Une présentation « interactive » des caractéristiques génériques tenant compte de l'ensemble du corpus aurait certes été intéressante en elle-même, mais se serait révélée contre-productive dans des fascicules destinés à des élèves du collégial. Dans les fascicules, chaque configuration d'unités a donc été prise en grande partie pour elle-même, et ce, afin de satisfaire à des nécessités pédagogiques et éditoriales.

Quant aux unités proprement dites, elles ont dû faire l'objet d'une interprétation discursive avant d'être intégrées à la planification du matériel. Nous avons suivi dans les grandes lignes des classifications générales (*cf. supra*, 4.3) désignant les unités comme des indices du cadre énonciatif, de la cohérence du texte, ou alors comme des indices d'une certaine complexité syntaxique. Conscients qu'une unité linguistique ne prend son sens qu'en contexte (*cf.* Charaudeau, 1992), nous avons recueilli en parallèle le plus d'informations possible sur les genres, dans des manuels de rédaction spécialisés lorsqu'ils étaient disponibles, mais surtout auprès d'enseignants du réseau collégial ayant une bonne expertise des matières concernées. Ces derniers nous ont fourni des textes d'élèves et des textes de leur cru, avec l'autorisation de les utiliser dans le matériel; ils ont également transmis des renseignements essentiels sur le contexte de communication propre aux genres et, le cas échéant, sur les particularités du milieu professionnel dans lesquels ils sont en usage; enfin, ils ont validé les manuscrits à différentes étapes de la rédaction. C'est ainsi qu'il a été possible de donner un sens aux unités distinctives des genres et de les utiliser comme repères pour la planification des fascicules.

Le cas du rapport d'intervention, en usage notamment dans le programme de Techniques policières, servira d'exemple¹⁶. L'analyse statistique (*cf. supra*, section 5) donne, comme unités discriminatives, 1) le discours direct, le passé composé, le plus-que-parfait, le verbe conjugué à un mode personnel, la deuxième personne du singulier, le marqueur de lieu, le nom propre (unités fréquentes), et 2) la subordonnée relative, l'attribut, l'adjectif complément du nom, l'organisateur logique, la première personne du pluriel, le pronom *on* (unités rares). Dans ce genre textuel dont la finalité est de décrire aussi objectivement que possible un conflit, un accident ou un crime, le discours direct a pour fonction de rendre le verbatim des interactions entre les personnes impliquées et les intervenants – précaution particulièrement importante si les propos sont incriminants ou irrespectueux (École nationale de police, 2007); le passé composé, temps principal de la description, situe l'événement dans un passé proche sans installer le compte rendu dans une dynamique narrative, ce que feraient possiblement le passé simple et le présent (Charaudeau, 1992). Dans ce genre essentiellement descriptif où les organisateurs logiques (ou connecteurs argumentatifs) sont rares, la cohérence du texte est assurée par des marqueurs de lieu qui localisent l'événement dans l'espace (Charaudeau, 1992) et par la cohésion temporelle qui reconstitue la chronologie des faits, ce que suggère la fréquence du passé composé et du plus-que-parfait qui, implicitement, suppose aussi la présence de l'imparfait. Par ailleurs, la rareté des unités qualifiantes (adjectif complément du nom, attribut, subordonnée relative) peut être mise en parallèle avec la fréquence des verbes conjugués à un mode personnel : la personne qui rédige un rapport d'intervention ne donne pas de point de vue axiologique, mais rapporte une série de faits dont la description très détaillée est

susceptible d'induire un jugement. En d'autres termes, un policier écrira avec force détails que *le suspect titubait, parlait avec difficulté et sentait l'alcool*, plutôt que de mentionner de manière plus économique – mais aussi plus subjective – qu'*il était ivre*. Par ailleurs, la relative fréquence des noms propres s'explique en partie par l'importance des renseignements nominatifs que le rédacteur doit fournir sur les personnes impliquées : dans ce genre qui consiste à décrire une situation avec exactitude et objectivité, les « protagonistes » autres que les rédacteurs (*cf. infra*) sont en effet désignés de la manière la plus neutre possible, par leur nom propre plus souvent que par des pronoms de reprise qui peuvent, dans certains contextes, être porteurs d'ambiguïté. Parmi les résultats de l'analyse discriminante, cette particularité offre une piste éclairante pour décrire le cadre énonciatif du rapport d'intervention : elle suggère qu'une importante partie du compte rendu, qui consiste à rapporter des faits avec objectivité et exactitude, se réalise dans une dynamique « non personnelle » (Benveniste, 1966). À côté de cette donnée apparemment aisée à interpréter, les indices des personnes grammaticales se sont révélés plus difficiles à intégrer dans le matériel didactique. Ainsi, de la deuxième personne du singulier est apparue comme une information d'importance secondaire sur le plan de la position énonciative du rédacteur dans la mesure où elle est directement associée à l'emploi fréquent du discours direct¹⁷ (vérification faite dans les textes du corpus); ainsi, dans une logique pédagogique, il a semblé moins intéressant d'insister sur la relative fréquence de cette unité dans les passages au discours direct que de rappeler l'importance de rapporter intégralement les échanges verbaux, écarts pragmatiques inclus – le fait pour un suspect de tutoyer un policier n'étant qu'un exemple parmi d'autres. Par ailleurs, il a paru hasardeux d'insister sur la rareté de la première personne du pluriel : cette unité est effectivement peu fréquente dans les vingt rapports d'intervention analysés dans le corpus, mais n'est pas pour autant « proscrite » du genre selon les témoignages que nous avons recueillis. En effet, c'est généralement la personne en charge de l'intervention qui rédige le rapport et assume l'entière responsabilité des informations qu'elle y consigne (École nationale de police, 2007). Même si cette personne doit faire un compte rendu « distancié » de l'événement, il arrive qu'elle se désigne elle-même comme un protagoniste de l'intervention – en incluant ses coéquipiers, le cas échéant – dans des tournures du type : *L'agente Ducharme et moi-même, constable Arsenault, nous sommes rendus au domicile de M. Tremblay pour y effectuer une perquisition*¹⁸. Dans ces passages, le *je* et le *nous* ne sont pas des indices d'implication subjective dans l'énonciation (Benveniste, 1974), puisqu'en principe, ces unités ont pour seule fonction de rendre compte objectivement de la contribution du rédacteur aux actions rapportées. Par conséquent, la première personne (du singulier ou du pluriel) n'a pas été traitée dans le matériel comme une unité « à proscrire », puisqu'elle apparaît plus ou moins fréquemment selon les exemples; il est toutefois question de l'importance pour le rédacteur de ne pas transmettre son point de vue, qu'il soit ou non impliqué dans l'événement. Quant au pronom *on*, sa rareté dans le rapport d'intervention aurait pu être mise en relation avec sa fréquence dans des textes à visée explicative ou argumentative comme le rapport de laboratoire (*cf. supra*, tableau 2), où il permet au locuteur de s'exprimer en tant que membre d'une « communauté » (Charaudeau, 1992) dans des formulations du type : *on constate que...*, *on en déduit que...*, etc. Cette donnée n'a pas été retenue telle quelle, notamment parce qu'il est apparu au moment de l'édition qu'elle entraînait des développements trop lourds dans un fascicule de dimension relativement modeste; elle a toutefois inspiré des remarques générales sur l'importance de décrire l'événement « sans argumenter ».

Ces dernières remarques suscitent des questions sur le statut pédagogique que l'on devrait accorder aux unités que l'analyse factorielle discriminante donne pour « peu fréquentes ». Rappelons en premier lieu que l'intention première de la présente étude était de fournir des données pertinentes, d'une part pour attirer l'attention des jeunes scripteurs sur les conventions énonciatives liées à des situations de communication particulières, d'autre part pour orienter stratégiquement la rédaction et la révision linguistique vers des configurations restreintes d'unités selon les genres. Les résultats d'une recherche comme celle que nous avons menée doivent, on l'a vu, faire l'objet d'une interprétation tenant compte des données contextuelles afin d'être utilisables dans un but « normatif ». En ce sens, les unités fréquentes posent moins de problèmes : d'un point de vue énonciatif, il paraît logique de considérer ces structures comme des « modèles positifs à imiter », dans la stricte mesure où les textes du corpus sont l'œuvre de scripteurs reconnus pour leur compétence; par ailleurs, on attire plus facilement l'attention sur les embûches morphosyntaxiques que posent des unités concrètement identifiables dans les exemples que sur

des unités peu fréquentes. En fait, c'est le sens que l'on donne à la rareté relative de certaines unités qui pose problème dans une perspective pédagogique. Dans certains cas, elle peut être interprétée d'un point de vue normatif, à condition d'être mise en relation avec la fréquence d'autres unités : c'est en suivant cette logique que l'attribut (unité potentiellement axiologique) et l'organisateur logique (unité potentiellement argumentative) peuvent être présentés comme des structures « à éviter » dans le rapport d'intervention. Ce statut ne peut toutefois être reconnu à toutes les unités peu fréquentes. Il est en effet possible qu'une variable soit donnée pour rare dans les résultats sans que sa présence dans un genre constitue un écart normatif : c'est le cas, on l'a vu, de la première personne du pluriel dans le rapport d'intervention (*cf. supra*); c'est le cas également du nom commun dans le rapport de recherche (*cf. supra, section 5*), qui apparaît en faible proportion par rapport au nombre total de mots (en comparaison avec les autres genres du corpus), mais qui n'en est certainement pas absent. Pour être en mesure d'interpréter sans ambiguïté les unités peu fréquentes et juger de la pertinence de les utiliser, dans un cadre pédagogique, afin d'encadrer la rédaction et la révision linguistique, il serait certainement utile de mener davantage d'analyses de même nature avec des corpus différents, incluant possiblement une comparaison entre des textes d'experts et des rédactions jugées faibles. Les limites du présent projet n'ayant pas permis de poursuivre dans cette direction, nous nous sommes fiés aux données contextuelles et aux avis des professeurs que nous avons consultés à toutes les étapes du projet pour compléter les données de l'étude.

7 Conclusion

Réalisée afin de servir de base à la production d'outils didactiques, une analyse textuelle comme celle que nous avons conduite n'est pas une fin en soi, mais une étape : ses résultats, tout intéressants soient-ils, doivent être mis en perspective et complétés par des témoignages issus de la pratique pédagogique. Malgré le fait que notre étude repose sur un nombre relativement restreint de textes et qu'elle nécessiterait probablement d'être complétée par des analyses plus approfondies, nous demeurons convaincus de la pertinence de cette étape d'observation empirique préalable à la planification des outils. La méthode adoptée n'a pas seulement ouvert des pistes à suivre pour la planification; elle nous a également conduits à transmettre un message qui a été entendu – et favorablement reçu – du milieu collégial, à savoir que la compétence rédactionnelle est étroitement liée à une connaissance des paramètres de la communication. Les fascicules recommandent par exemple aux policiers – et autres utilisateurs du rapport d'intervention – de *relater les actions avec objectivité* en décrivant, au moyen de verbes d'action, ce que leurs sens ont pu percevoir; aux rédacteurs d'argumentaires, il est conseillé notamment de justifier implicitement leur concept en employant diverses formes de complément du nom (comme les relatives appositives) plutôt que des subordinées ou groupes prépositionnels à valeur argumentative; les personnes qui rédigent des documents techniques sont invitées à porter une attention particulière aux enchaînements de groupes du verbe à l'infinitif, de manière à éviter les confusions avec les phrases impératives. Les diverses stratégies, associées à des défis que posent les situations de communication, sont illustrées par des textes authentiques. Nous verrons à l'usage si les outils didactiques produits à la suite de la démarche que nous proposons seront effectivement profitables notamment aux futurs policiers, graphistes et architectes.

Références bibliographiques

- Adam, J.-M. (1999). *Linguistique textuelle. Des genres de discours aux textes*, Paris : Nathan.
- Austin, J.L. (1970 [1962]). *Quand dire, c'est faire*, traduit par Gilles Lane, Paris : Seuil.
- Bakhtine, M. (1984 [1979]). Les genres de discours. *Esthétique de la création verbale*, Paris : Gallimard, p. 263-308.
- Barbeau, D. (2007). *Interventions pédagogiques et réussite au cégep. Méta-analyse*, Québec : Les Presses de l'Université Laval.
- Benveniste, E. (1966). Structure des relations de personne dans le verbe. *Problèmes de linguistique générale 1*, Paris : Gallimard, p. 225-236.
- Benveniste, E. (1974). L'appareil formel de l'énonciation. *Problèmes de linguistique générale 2*, Paris : Gallimard, p. 79-88.

- Bolduc, J., Parent, S. et Perron, J. (2004). *Rapport synthèse. L'évaluation des plans d'aide à la réussite dans les collèges*, Québec : Commission d'évaluation de l'enseignement collégial (CEEC).
- Bourdieu, P. (1982). *Ce que parler veut dire : L'économie des échanges linguistiques*, Paris : Fayard.
- Bronckart, J.-P., Bain, D., Schnewly, B., Davaud, C. et Pasquier, A. (1985). *Le Fonctionnement des discours. Un modèle psychologique et une méthode d'analyse*, Lausanne : Delachaux et Niestlé.
- Bronckart, J.-P. (1996). *Activité langagière, textes et discours. Pour un interactionnisme socio-discursif*, Lausanne : Delachaux et Niestlé.
- Bronckart, J.-P. (2004). Les genres de textes et leur contribution au développement psychologique. *Langages*, 153, p. 98-108.
- Centre collégial de développement de matériel didactique [CCDMD] (2006). Portrait des mesures d'aide à la réussite de l'épreuve uniforme de français offertes par les établissements. *Correspondance*. [En ligne] : www.ccdmd.qc.ca/correspo/Corr12-2/Sondage.html
- Charaudeau, P. (1992). *Grammaire du sens et de l'expression*, Paris : Hachette Éducation.
- Chartrand, S.-G. (2008). *La progression dans l'enseignement du français langue première au secondaire québécois*, Québec : Les Publications Québec français.
- Chartrand, S.-G. [dir.] (2011). *Portail pour l'enseignement du français*, Québec : Centre de services et de ressources en technopédagogie de la FSÉ (Université Laval). [En ligne], www.enseignementdufrancais.fse.ulaval.ca/
- Cibois, P. (1987). *L'analyse factorielle*. Paris: PUF.
- Direction générale des affaires universitaires et collégiales, Direction de l'enseignement collégial [DGAUC-DEC] (2011). *Résultats à l'épreuve de français 2010-2011. Épreuves uniformes, langue d'enseignement et de littérature*, Québec : Gouvernement du Québec, ministère de l'Éducation, du Loisir et du Sport.
- Dolz, J., Noverraz, M. et Schnewly, B., dir. publ. (2001). *S'exprimer en français. Séquences didactiques pour l'oral et pour l'écrit. Notes méthodologiques. Volume IV. 7^e/8^e/9^e*, Bruxelles : De Boeck & Larcier.
- Ducrot, O. (1984). *Le dire et le dit*, Paris : Les Éditions de Minuit.
- École nationale de police (2007). *Guide de présentation des rapports et des formulaires*, 4e éd., Québec : Programme de formation nationale en patrouille-gendarmerie.
- Genevay, E. (1994). *Ouvrir la grammaire*, Lausanne : Loisir et Pédagogie.
- Hill, T., & Lewicki, P. (2007). *Statistics Methods and Applications*, Tulsa, OK: StatSoft.
- Lecavalier, J. (2003). *La didactique de l'écriture : les marqueurs de relation dans les cours de français au collégial*. Thèse de doctorat en Sciences de l'éducation, Montréal : Université de Montréal.
- Lefrançois, P., Laurier, M.D., Lazure, R. et Claing, R. (2005). *Évaluation de l'efficacité des mesures visant l'amélioration du français écrit du primaire à l'université*. Rapport de recherche, Montréal : Université de Montréal.
- Libersan, L. (2012). *Stratégies d'écriture dans la formation spécifique*.
En ligne : www.ccdmd.qc.ca/fr/strategies_écriture
- Libersan, L., Claing, R. et Foucambert, D. (2010). *Stratégies d'écriture dans la formation spécifique. Rapport 2009-2010*, Montréal : CCDMD / Collège Ahuntsic.
En ligne : www.ccdmd.qc.ca/media/doc_theo_div_Rapport_Formation_specifique.pdf.
- Mangueneau, D. (2007). *Analyser les textes de communication. 2^e édition*, Paris : Armand Colin.
- Ouellet, R. (2008). *Statistiques de l'éducation. Enseignement primaire, secondaire, collégial et universitaire*, Québec : ministère de l'Éducation, du Loisir et du Sport (MELS).
- Paradis, J. (2003). Les élèves en échec après une première session au collégial : les causes, les facteurs d'adaptation et des moyens d'intervention. *Pédagogie collégiale*, 14-1, p. 18-23.
- Riegel, M., Pellat, J.-C. et Rioul, R. (1994). *Grammaire méthodique du français*, Paris : PUF.
- Table Éducation Montréal (2002), *100 façons d'améliorer le français écrit des élèves*, Table de concertation interordres montréalaise. En ligne : www.ccdmd.qc.ca/100facons.

Logiciels :

- Association Française pour la Lecture [AFL] (1994). *Analyse de textes*.
- Druide Informatique inc. (2007). *Antidote RX*.
- Statsoft inc. (1984-2009). *Statistica v7*.

¹ Le projet, entrepris à l'initiative de Robert Claing (professeur retraité du cégep Ahuntsic), a été rendu possible grâce au soutien du MELS (ministère de l'Éducation, du Loisir et du Sport, Direction des affaires universitaires et collégiales), de la Direction des études du cégep Ahuntsic et du CCDMD (Centre collégial de développement de matériel didactique). La recherche décrite dans le présent article a été réalisée par Lucie Libersan avec la collaboration de Robert Claing, selon un cadre méthodologique élaboré sous la supervision de Denis Foucambert (voir Libersan et al., 2010).

² En tout, dix fascicules seront mis en ligne, en format PDF, sur le site *Amélioration du français* du CCDMD (voir Libersan, 2012).

³ Selon les *Statistiques de l'enseignement 2008* (tableau 2.3.3).

⁴ Selon les *Statistiques de l'enseignement 2008* (tableaux 2.3.14 et 2.3.15).

⁵ De 1998 à 2002, 57 % des élèves ont réussi tous leurs cours en première session (Bolduc et al., 2004 : 27); il y a autant d'échecs en formation spécifique qu'en formation générale (Paradis, 2000).

⁶ De 1996 à 2000, environ 35 % des élèves ont obtenu leur diplôme au terme de la durée prévue (Bolduc et al., 2004 : 35).

⁷ Pendant l'année 2010-2011 (DGAUC-DEC, 2011), 14,1 % des candidats ont échoué en langue écrite (orthographe, syntaxe et vocabulaire), contre 0,4 % en structure du texte et 3,3 % en argumentation.

⁸ Dans les arts appliqués, l'argumentaire de projet a pour finalité de justifier auprès d'un client les choix esthétiques et conceptuels ayant présidé à la création d'un design.

⁹ La correspondance d'affaires désigne les communications professionnelles qu'un individu ou un organisme adresse à l'externe, tandis que la correspondance administrative désigne les communications internes (notes de service, avis, etc.) dans une entreprise.

¹⁰ Texte publicitaire utilisant différents médiums (papier, page Internet).

¹¹ Le terme général de *document d'information technique* regroupe plusieurs sous-genres dont l'objectif consiste, entre autres : 1) à énoncer les normes et procédures en vigueur dans la construction immobilière, l'élaboration d'infrastructures routières, la production industrielle, l'entretien de systèmes, etc.; 2) à fournir des renseignements sur le fonctionnement d'équipements mécaniques; 3) à décrire des parcelles de terrain en vue d'une certification foncière. Exemples de documents techniques : *devis, cahier des charges, procédure, manuel d'assurance de la qualité, norme*.

¹² Le terme *rapport d'intervention* désigne un certain nombre de documents officiels qui décrivent les actions d'une ou plusieurs personnes en situation d'autorité, à la suite d'événements comme un accident ayant entraîné des blessures ou des dommages, une infraction à une loi ou à un règlement, un acte criminel ou un conflit entre des personnes.

¹³ Ce sont les catégories de verbes performatifs définies par Austin (1970 [1962]). En principe, chacune de ces catégories de verbes correspond à des actes de langage. Verdictifs : formuler un jugement (*apprécier, estimer*); exercitifs : exercer une influence ou un pouvoir (*désigner, commander, supplier*); promissifs : s'engager à faire quelque chose (*promettre, consentir*); comportatifs : réagir à la conduite ou au sort d'autrui (*remercier, féliciter*); expositifs : asserter, exposer quelque chose (*affirmer, remarquer, distinguer*).

¹⁴ Les gentilés (noms et adjectifs désignant les habitants d'un lieu) peuvent constituer des unités axiologiques dans certains contextes; par exemple, dans un cas comme le rapport d'intervention où la situation oblige à décrire des personnes de manière objective, l'usage de gentilés pourrait traduire des jugements de valeur inappropriés.

¹⁵ Une autre approche était encore possible, celle de s'en tenir aux résultats univariés issus d'une succession d'ANOVA pour chacun des groupes. Les résultats entre les deux procédures peuvent parfois légèrement différer, dans la mesure où les résultats que nous présentons ici s'appliquent plus spécifiquement à un modèle par essence multivarié (Hill & Lewicki, 2007).

¹⁶ Cette section de la discussion réfère au fascicule sur le rapport d'intervention (Libersan, 2012), qui peut être consulté à l'adresse suivante : http://www.ccdmd.qc.ca/fr/strategies_ecriture/.

¹⁷ L'analyse statistique donne la deuxième personne du singulier comme une unité distinctive par sa fréquence dans le rapport d'intervention en raison de sa quasi-absence dans le reste du corpus; il faut noter toutefois que cette variable demeure très rare dans les rapports d'intervention étudiés et que son indice de significativité est faible (p légèrement inférieur à 0,05).

¹⁸ Situation fictive.