

Chapitre 10 : Conclusion

10.1. Contributions

La présente thèse a proposé une description quantifiée, mais qualifiée du genre de l'article scientifique de revue linguistique, en corpus et en exploitant les méthodologies du traitement automatique des langues et des statistiques textuelles.

L'originalité de l'étude est d'avoir pris le genre, non comme cadre, mais comme point de départ descriptif : la plupart des travaux partent en effet d'un marqueur, ou d'une hypothèse descriptive qui sera évaluée en genre ; au final, les descriptions proposées sont localement plus fines que les nôtres, mais demeurent insatisfaisantes au niveau global, dans la mesure où les observations sont autonomisées du genre.

Le genre ne pouvant être observé qu'en corpus génériquement homogène, et de taille suffisamment importante pour être représentatif de l'objet considéré, c'est à partir d'une collection représentative de 224 textes que l'article a été saisi.

Etant donné son intérêt descriptif et son caractère discriminant, c'est le niveau morphosyntaxique que nous avons adopté comme point de départ exploratoire ; outre leur fiabilité mitigée, les étiqueteurs disponibles se sont rapidement avérés peu adaptés aux spécificités des textes scientifiques. Nous avons donc choisi d'adapter un étiqueteur dédié au genre de l'article de revue linguistique, en exploitant la solution d'entraînement ; un système de 145 descripteurs adapté aux caractéristiques de nos textes a été construit dans cette perspective.

Cette adaptation des descripteurs et des outils à l'entreprise linguistique engagée et aux objectifs souhaités nous a semblé cruciale, et nous paraît même devoir présider à toute entreprise descriptive de corpus systématique comme la nôtre ; les outils sont en effet largement généralistes, ce qui garantit leur appropriation par le plus grand nombre. Si cette orientation convient parfaitement aux entreprises appliquées – et notamment aux travaux de catégorisation textuelle, qui visent davantage à classer les textes plutôt qu'à les caractériser, elle dessert généralement les études descriptives comme la nôtre, qui cherchent d'abord à *interpréter* les phénomènes et les regroupements textuels obtenus. La linguistique étant vouée à fonder ses descriptions sur des corpus de plus en plus importants qui ne sauraient être appréhendés sans recourir aux outils du TAL, cette possibilité d'entraînement devrait être de plus en plus exploitée.

Au final, quatre plans descriptifs principaux ont été considérés : la morphosyntaxe du genre, la structuration des textes, leurs composantes et la thématique générique. La division est naturellement méthodologique plutôt que linguistique, dans la mesure où les quatre niveaux de description sont fédérés par le texte et se répondent constamment :

- ✓ Le chapitre 3 dresse un profil morphosyntaxique du genre de l'article, appréhendé par sous-systèmes descriptifs et en tant que structure organisée et stabilisée ; le genre de l'article apparaît articulé autour de quatre pôles opposant les temps narratifs du passé au présent et deux types d'application linguistique : formalisation et exemplification (dimension interlocutive et plus orale) ;

- ✓ Le chapitre 4 s'intéresse à la structuration des textes et précise les résultats précédemment obtenus sur l'ensemble des corps d'articles, dans la mesure où les jeux de corrélations et les sous-systèmes descriptifs s'affinent lorsqu'on examine les sections de l'article (introduction, conclusion et développement) ;
- ✓ Le chapitre 5 explore la thématique du genre ; nous avons cherché à évaluer et à élaborer en corpus un ensemble de critères typologiques discriminants. Bien que préliminaires, les critères définis permettent de définir et d'opposer certaines catégories de concepts (*e.g.* problématisés/instrumentaux).

Une fois la collection initiale appréhendée, le travail d'exploration s'est poursuivi de manière contrastive, en confrontant ASLF à d'autres corpus recoupant d'autres typologies ayant par hypothèse une incidence sur les modes de détermination et de formation du genre :

- ✓ L'examen des styles d'auteur nous a ainsi permis d'observer les modes de constitution et d'appropriation d'un genre et d'une rhétorique scientifique : les contraintes imposées par le genre de l'article semblent intériorisées et assimilées dans les représentations, dans la mesure où les auteurs les plus hardis ne s'en écartent que relativement, et dans les limites des pôles génériques mis au jour. On voit donc ici comment un espace représentationnel et rhétorique se construit et se stabilise ;
- ✓ Les axes d'organisation générique mis au jour semblent bien caractéristiques du genre de l'article : les genres scientifiques linguistiques du *compte rendu* et de la *présentation de revue* ont une structure morphosyntaxique bien distincte. Par exemple, nous n'avons pas pu observer une opposition *présent/passé* qui organisait l'article, et que l'on retrouve pourtant dans les textes de *mécanique* : malgré les différences importantes qui séparent les deux disciplines, on observe une proximité structurelle des deux collections sur le plan morphosyntaxiques, qui serait à approfondir et à préciser sur d'autres domaines, et d'autres genres ;
- ✓ ASLF a enfin été contrasté à ASLA : le genre de l'article a été mis à l'épreuve des langues anglaise et française. Bien que préliminaire, dans la mesure où l'analyse contrastive des genres est un domaine encore inexploré, notre étude fait déjà apparaître des lieux de similarité entre les deux structures génériques. On retrouve ainsi trois des quatre pôles mis au jour : orientation narrative passé/présent et opposition des textes exemplifiés contenant des marqueurs de l'oral et des textes formalisés.

L'intérêt de notre étude aura donc été double : nous avons d'une part caractérisé les différents aspects d'un objet normatif complexe et multidimensionnel en couvrant différents niveaux de phénomènes et de régulations linguistiques, et d'autre part élaboré une méthodologie d'observation des genres impliquant des outils et des techniques de traitement automatique des langues, de statistique multidimensionnelle et de traitement de corpus.

La description proposée étant systématique et objectivée, elle pourra servir de support de comparaison aux entreprises comparatistes futures, qui pourront valider, ou invalider les résultats obtenus ; de surcroît, elle pourrait éventuellement intéresser certains domaines d'application : la recherche d'information, en quête de critères discriminants, pourrait

exploiter certains aspects de notre étude dans des perspectives de classification, d'indexation ou de fouille de texte, tandis que la traduction automatique gagnerait à identifier les genres en amont du processus de traduction, afin de réduire les problèmes de polysémie et d'optimiser les traductions. Notons enfin que l'ensemble des résultats obtenus pourraient donner lieu à la publication d'un manuel de rédaction scientifique destiné aux jeunes linguistes francophones.

10.2. Directions de recherche

Notre travail étant exploratoire, il comporte naturellement ses limites, et ouvre différentes voies d'analyse.

S'il serait pertinent de valider et de préciser les résultats obtenus sur ASLF en augmentant la taille et la représentativité du corpus – et ceci vaut pour l'ensemble des corpus pris en compte, les relations entre les quatre plans descriptifs abordés (morphosyntaxe, structure, configurations et thématique) doivent être approfondies et appréhendées de manière plus systématique ; par exemple, la structuration des textes n'est pas corrélée avec le plan thématique, tandis que les relations entre concepts et morphosyntaxe doivent encore être explorées.

De manière générale, les méthodes de statistique multidimensionnelle employées n'ont été appliquées qu'à la morphosyntaxe des textes (bien que nous ayons inclus des variables supplémentaires illustratives) ; il serait probablement pertinent d'inclure l'ensemble des données et des résultats dont nous disposons au final au sein d'une même table, afin d'obtenir une représentation plus unifiée et plus exhaustive du genre.

Il conviendrait par ailleurs d'enrichir, et d'améliorer le système d'étiquetage morphosyntaxique de départ : les passifs ne sont pas exemple pas pris en compte, tandis que certaines catégories sont peut-être trop générales. Les résultats obtenus devraient permettre d'affiner le jeu d'étiquettes en fonction des nouvelles hypothèses élaborées.

D'autres systèmes de descripteurs pourraient d'ailleurs être exploités ; comme nous l'avons déjà évoqué, la dimension syntaxique des textes n'est pas prise en compte ici. Le lexique épistémique de F. Rinck, qui a démontré son intérêt au cours du chapitre 7, pourrait également être appliqué à la description du corpus ASLF.

L'analyse des styles d'auteur, qui nous semble devoir être étendue à un panel plus large de linguistes, pourrait être précisée au niveau sociologique, à partir de variables additionnelles relatives à l'école d'origine de l'auteur, sa date de naissance, la date d'obtention de sa thèse, son laboratoire, etc.

Soulignons également que nous n'avons observé que la configuration optative de l'exemple, la citation n'ayant été abordée que relativement à cette composante. Nous envisageons d'analyser les notes de bas de page et de fin d'article, de même que les références bibliographiques des textes, qui pourraient être corrélées de manière plus claire aux noms propres, aux dates et aux citations des corps d'articles.

Les genres héritant comme les discours de la tradition comparatiste linguistique, leurs caractéristiques ne se déclinent que par champs de contraste et d'opposition, que nous avons tenté d'appréhender à différents niveaux : l'étude contrastive des genres au niveau interlangue nous a ainsi semblé particulièrement stimulante, et le travail préliminaire que nous avons accompli propose déjà des résultats encourageants, qui ne demandent qu'à être creusés. Les structures génériques des deux genres font d'ailleurs l'objet d'une étude menée en partenariat avec N. Serpollet, qui sera présentée à *Digital Humanities* en juillet 2006.

Enfin et si nous avons pu considérer, ou plutôt approcher les niveaux des genres, des domaines et des langues dans le cadre de la présente thèse, celui des *discours* nous paraît également pertinent. Une étude portant sur les lieux de contraste et de similarité entre discours philosophique, critique et linguistique est en cours, en collaboration avec S. Loiseau et D. Ablali (v. Loiseau, Poudat et Ablali, 2006).