



HAL
open science

Linguistique de corpus et caractérisation des genres : un exemple d'analyse d'un conte didactique

Denise Malrieu

► **To cite this version:**

Denise Malrieu. Linguistique de corpus et caractérisation des genres : un exemple d'analyse d'un conte didactique. Construire et exploiter des corpus de genres scolaires, 2007, Nanterre, France. pp.10-20. halshs-00161678

HAL Id: halshs-00161678

<https://shs.hal.science/halshs-00161678>

Submitted on 11 Jul 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Linguistique de corpus et caractérisation des genres : un exemple d'analyse d'un conte didactique

Denise MALRIEU
CNRS-MODYCO Paris X
dmalrieu@u-paris10.fr

Introduction

La notion de *genre textuel* est devenue depuis une dizaine d'années une notion centrale de la linguistique textuelle, depuis les travaux de Bronckart, Adam ou Rastier ; on peut dire que son usage s'est répandu aussi dans les domaines de la linguistique appliquée, que ce soit en didactique des langues ou en sémantique lexicale pour l'élaboration des terminologies. Même en syntaxe la démarche d'élaboration de grammaires "corpus-driven" et les travaux de syntaxe de l'oral font progresser la prise en compte des variations de la norme linguistique selon les situations d'énonciation et donc selon les genres.

Catherine Boré m'avait demandé de parler des travaux de caractérisation linguistique des genres à partir de corpus étiquetés. Mais, pour ne pas me répéter (les travaux menés en collaboration avec F. Rastier sont publiés et datent de 2001) et après discussion avec elle sur les travaux actuels de son équipe, il m'a paru plus intéressant de prendre l'exemple d'un des textes étudiés en classe et sur lesquels s'appuie son corpus de productions d'élèves, le conte didactique d'Orsenna : *La grammaire est une chanson douce*.

La notion de genre faisant partie du vocabulaire pédagogique officiel, mais sans définition bien précise, l'objectif était de caractériser les corpus de textes soumis aux élèves et de voir si les catégories de genre officiellement attribuées coïncident avec une objectivation des propriétés formelles des textes.

Je précise tout de suite que cet objectif n'est pas atteint dans l'immédiat, et le projet de comparer le texte analysé avec un corpus du même genre n'a pu être rempli faute de corpus numérisé de référence adéquat. L'absence d'étude contrastive n'a pas empêché cependant d'avancer sur la caractérisation du texte.

1- La démarche de caractérisation des genres par des variables morpho-syntaxiques

1.1. L'historique :

Je ne reviendrai pas ici sur les apports des travaux pionniers de Bronckart dans le domaine francophone. Dans le domaine anglophone, les travaux de Biber ont été à l'origine de la *linguistique de corpus dans son acception forte* : la démarche adoptée, dénommée par Biber "corpus-driven" n'utilise pas les corpus du BNC comme simple réservoir d'exemples illustratifs d'une grammaire de référence qui tente de définir une norme unique d'acceptabilité, mais affirme que les normes linguistiques varient selon les registres, les contextes

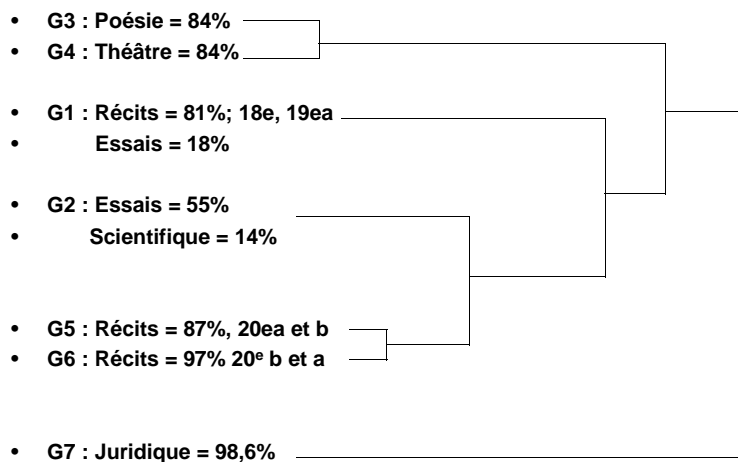
d'énonciation et les genres (cf la discussion de Manning (2002) sur les problèmes de jugement d'acceptabilité et l'intérêt de l'observation de corpus).

Biber observe que les distributions d'un lexème ou d'un morphème sont bien différentes selon les registres et que certains mots sont spécifiques à certains registres; que les ambiguïtés grammaticales de la forme "-ed" en anglais par ex, ou la catégorie grammaticale (verbe/nom/adjectif) pour un même mot se répartissent de façon contrastée selon que l'on est dans des textes expositifs ou fictionnels. Il en vient donc à la définition de grammaires et dictionnaires intégrant les variations de la norme linguistique (*Longman Grammar of Spoken and Written English*, Biber, Johansson, Leech, Conrad, & Finegan, élaborée à partir d'un corpus de 40 millions de mots dans 4 registres : conversation, fiction, articles de presse, textes académiques).

Les analyses de Rastier et Malrieu à partir des variables issues du corpus de Synapse

Une première démarche déductive sur un corpus de 2541 ouvrages étiquetés par Synapse avait permis de montrer (Malrieu et Rastier, 2001) en mettant en œuvre l'analyse univariée, l'existence de variations morphosyntaxiques systématiques selon les domaines, champs génériques et genres textuels; l'analyse en composantes principales montrait que le facteur principal dans l'analyse factorielle est en rapport avec les variables liées à l'oral représenté. La démarche inductive par la classification hiérarchique ascendante sur les 250 variables morphosyntaxiques obtenues par CORDIAL redonne de façon nette les découpages en domaines et champs génériques (juridique vs théâtre vs poésie vs essais); cependant, la classification des genres narratifs s'avérait moins évidente.

Classification hiérarchique ascendante
de 2600 ouvrages de différents genres



1.2 – Les limites de l'approche globale des textes pour leur caractérisation

M'intéressant particulièrement aux variations des récits fictionnels et à la caractérisation des narrateurs, mes essais pour décrire de façon plus fine des familles narratologiques au sein du roman "sérieux" se sont avérés peu concluants en utilisant les mêmes variables ou des sous-ensembles de celles-ci (le premier facteur n'expliquant pas plus de 30% de la variance). La démarche inductive de classification des romans rencontrait plusieurs obstacles :

- non disponibilité de corpus raisonnés de romans et suffisamment étoffés.
- nécessité de repenser les variables morphosyntaxiques à prendre en compte dans la classification (insuffisance des variables disponibles dans Cordial : nous en avons redéfini environ 300).

- les calculs sur le texte dans sa globalité effectuent un lissage qui dilue les différences : le roman est par essence un texte composite : par ex le fort taux de 1PS ne peut différencier le roman homodiégétique et le roman hétérodiégétique fortement dialogué. Le constat est donc que les contraintes interprétatives portées par le genre impliquent la prise en compte des séquences inférieures au texte, séquences informées par le genre mais qu'il faut caractériser en tant que telles.

D'où l'abandon temporaire de la démarche inductive sur textes entiers pour passer à une représentation qui prenne en compte le genre comme dispositif énonciatif.

1.3. – Les dimensions de description des textes narratifs fictionnels

Dans une optique qui viserait à caractériser les familles narratologiques et la place accordée au lecteur par la forme du récit, on peut a priori distinguer différents aspects du dispositif énonciatif :

- le rapport narrateur/lecteur se dessine d'une part dans une déclaration explicite ou implicite du genre qui va définir un horizon d'attentes; l'adresse explicite du narrateur au lecteur et l'entrée dans l'univers fictionnel peuvent être verbalisées de façon plus ou moins riche (allant de la simple étiquette "roman" aux incipits développés des contes du XVIIIème siècle); cette adresse au lecteur peut être très localisée (en introduction et conclusion dans la fable et le conte) ou répartie sur l'ensemble du texte.

Conte et roman n'autorisent pas les mêmes configurations :

- dans le **conte traditionnel**, on a l'expression de vérités universelles empruntant les voies mythologiques en s'appuyant sur des symboliques collectives (d'où possibilité de l'effacement du narrateur et sollicitation médiée du lecteur); le narrateur est effacé, le récit est à la 3PS, le maintien dans l'univers fictionnel merveilleux n'autorise pas l'immixtion du rapport énonciatif primaire,

- alors que le **roman classique** qui ne dessine pas de frontières nettes avec la réalité quotidienne l'autorise. De plus le roman autorise des modalités de dialogue indirect avec le lecteur à travers les faces du narrateur (crédibilité affichée des sources de savoir, scénographie), omniscience affichée ou déniée, guidage du lecteur (ou fausses pistes ludiques).

- en ce qui concerne le rapport du narrateur aux personnages, il semblerait (mais cela reste à vérifier dans le matériau linguistique) que les niveaux d'empathie pour les personnages soient plus prononcés dans le roman. Le narrateur adoptant souvent et de façon successive les points de vue internes des personnages (accès à la vie mentale et à la parole intérieure), sollicitant la participation affective du lecteur. Le point de vue surplombant du narrateur induit activité critique et distanciation du lecteur. Cette activité critique de mise en évidence de l'aveuglement des personnages peut s'appuyer ou pas sur un discours narratorial dissonant à leur propos, elle peut n'être mise en œuvre qu'à travers l'agencement travaillé des séquences de DR et le récit des comportements (cf. Flaubert).

- dans le conte, l'effacement du narrateur, récit à la 3S, psycho-récit rudimentaire ou carrément absent, absence de monologue intérieur vont de pair.

- dans le conte pour enfants, les processus d'identification sont favorisés par l'appariement de l'âge du héros et de celui du public cible.

Comment situer le récit d'Orsenna entre conte et roman?

On peut a priori caractériser les propriétés partagées avec le conte :

Concernant la composante dialectique : le motif de départ = le manque (la séparation des parents, la tempête qui ôte les mots de la tête), à la fin les parents, comme les mots, reviennent, les mots retrouvent un ordre.

On trouve des éléments de merveilleux : les mots sont des êtres vivants animés, ils ont un pouvoir magique, ils ont des attributs humains. Mais il s'agit d'un faux merveilleux (les mêmes actants sont alternativement animaux ou humains sans que ces changements d'état correspondent à des motifs ou à une logique du récit).

Le temps du récit : On a un récit au passé avec usage important du PS.

La caractérisation thématique *par Cordial* décrit des thématiques peu conformes à celles du conte traditionnel:

- Indice de lisibilité : 74 du niveau du Cours Moyen
- Statistiques sur la sémantique :
 - Les deux domaines «linguistique» et «grammaire» sont les domaines les plus saillants de ce texte.
 - Une analyse plus fine encore de la thématique du texte fait apparaître comme thèmes centraux :
 - * catégorie grammaticale (catégorie : grammaire, correction)
 - * gestion du temps (catégorie : durée)
 - * forme grammaticale (catégorie : grammaire, correction)
 - * synchronisation (catégorie : retard, ponctualité)
 - * cause (catégorie : cause)
 - * mot (catégorie : unités de texte)
 - * lexicologie (catégorie : unités de texte)

1.4. - Comparaison des contes traditionnels et du récit d'Orsenna sur quelques variables morphosyntaxiques issues de Cordial.

Remarque préalable : il ne m'est pas possible de comparer le texte d'Orsenna avec un corpus de référence qui serait constitué de contes contemporains. Les seules comparaisons dont je dispose sont celles avec un corpus de contes établi par Synapse et qui comporte essentiellement des contes du 19^{ème} siècle. On compare donc deux états de langue différents. Nonobstant cette limite importante, on peut cerner (à partir des variables de CORDIAL) quelques propriétés distinctives du conte d'Orsenna par rapport au corpus de contes du 19^{ème}.

1.4. 1 - la complexité syntaxique de la phrase :

La phrase est plus courte chez Orsenna :

VARIABLES CORDIAL	Contes	Orsenna
moyenne nb mots entre ponctuations	6,29	4,73
moyenne nb mots/proposition	7,71	6,41
moyenne nb mots/phrased	15,84	9,79
moyenne nb mots/paragraphe	100	25,87
moyenne nb substantifs/proposition	1,67	1,3
% des phrases comportant au moins une prop. subordonnée / l'ensemble des phrases	17,31	6,47

Cela se traduit aussi dans la répartition des ponctuations :

% de points par rapport à l'ensemble des ponctuations	25,5	35,45
% de points virgules par rapport à l'ensemble des ponctuations	4,8	0,6
% de virgules par rapport à l'ensemble des ponctuations	54,71	45,73
% de parenthèses par rapport à l'ensemble des ponctuations	0,49	2,53

Comme dans le poids des conjonctions, pronoms relatifs et prépositions :

% de prépositions par rapport à l'ensemble des mots	10,96	8,87
---	-------	------

% de conjonctions de coordination par rapport à l'ensemble des mots	3,63	2,57
% de pronoms relatifs par rapport à l'ensemble des mots	1,98	1,39

Et dans la répartition des types de propositions :

% propositions indépendantes/ l'ensemble des propositions	80,53	90,42
% de propositions relatives parmi l'ensemble des propositions	2,53	0,92
% de propositions subordonnées parmi l'ensemble des propositions	14,19	7,62

1.4. 2 - Le dispositif énonciatif du récit :

Le récit d'Orsenna est un récit à la 1^{ère} personne, l'héroïne narratrice étant une enfant de l'âge du public destinataire, ce qui n'est pas la norme dans les contes traditionnels.

Cela se traduit évidemment dans le poids de la 1PS et de la 1PP :

% de pron. poss. à la 1PS / pron. poss.	22,4	34,51
% de pron. poss. à la 2PS / pron. poss.	3,7	6,37
% de pron. poss. à la 3PS / pron. poss.	52,19	30,33
% de pron. poss. à la 1PP / pron. poss.	6,13	13,63
% d'adj. poss. à la 1PS / adj. poss.	7,37	50
% d'adj. poss. à la 1PP / adj. poss.	73,89	33,33
% d'adj. poss. à la 2PP / adj. poss.	0,24	16,67
% d'adj. poss. à la 3PP / adj. poss.	27,06	0
% de pron pers à la 1PS / pron pers pers sujets	21,54	30,46
% de pron pers à la 1PS / pron pers pers sujets	3,88	8,3
% de pron pers à la 1PP / pron pers pers sujets	3,61	7,16

Dans la répartition des temps verbaux :

% de verbes au présent / verbes conjugués (temps simples et composés)	35,46	42,01
% de verbes au passé simple / verbes conjugués	22,44	12,69
% de verbes au passé composé / verbes conjugués	4,78	6,89
% de verbes au conditionnel passé / verbes conjugués	0,65	1,55

Et dans le poids relatif de certaines parties du discours comme : plus faible proportion de noms propres, liée au récit à la 1PS :

% de noms communs par rapport aux noms (communs ou propres)	85,98	91,93
% de noms propres par rapport aux noms (communs ou propres)	14,09	8,07
% de pronoms personnels par rapport à l'ensemble des pronoms	11,63	79,78
% de pronoms parmi l'ensemble des sujets	49,97	57,03

Le fait que le récit est à la 1PS se traduit aussi dans le poids de certaines variables liées au monologue intérieur du narrateur : par ex, la fréquence des phrases interrogatives, des adverbes de doute, ou l'usage spécifique des parenthèses qui marquent les *a parte* de parole intérieure:

% de phrases interrogatives par rapport à l'ensemble des phrases	5,19	8,16
% de points d'interrogation par rapport à l'ensemble des ponctuations	2,71	4,49
% d'adverbes de doute parmi l'ensemble des adverbes	3,15	6,89
% de parenthèses par rapport à l'ensemble des ponctuations	0,49	2,53

Le poids relatif des DR et séquences dialogiques ne peut malheureusement pas être comparé avec le corpus de contes, car sa reconnaissance par Cordial est problématique. (On y reviendra dans la seconde partie).

Enfin, les particularités de la thématique du récit d'Orsenna dont l'objectif didactique est une présentation de la grammaire se reflètent indirectement dans les types de substantifs :

% de noms abstraits par rapport à l'ensemble des noms	69,27	75,43
% de noms communs de type uniquement animal parmi l'ensemble des substantifs	2,94	4
% de noms communs de type uniquement abstrait parmi l'ensemble des substantifs	35,14	41,88
% d'adjectifs toujours utilisés avec un substantif abstrait parmi l'ensemble des adjectifs	9,47	21,75
% des verbes à sujet uniquement humain ou humanoïde parmi l'ensemble des verbes	0?	31,68
% verbes à sujet uniquement abstrait/ l'ensemble des verbes	0?	5,73

2. - Les propriétés du conte didactique :

La stratégie pour éviter le didactisme du conte s'exprime dans la contradiction impliquée par le choix d'un narrateur enfant donc sans savoir : le discours didactique doit être assumé par d'autres personnages, adultes. Le narrateur enfant jouant le rôle d'heureux apprenant. Les opinions, jugements évaluatifs sur les institutions étant, eux, assumés par le narrateur enfant, qui jouit du regard naïf et innocent sur ce qui l'entoure.

Les aspects didactiques : ils se traduisent dans le marquage d'activités métalinguistiques fréquentes et dans le lexique.

2.1. – Les activités métalinguistiques :

La méthodologie d'analyse : le texte a été balisé selon la méthodologie XML et de la TEI (voir l'exemplier). Ce balisage porte sur la structure textuelle (chapitres) et sur les séquences qui peuvent être ou non graphiquement marquées. Un premier niveau de balisage retraduit les marques graphiques (italiques, guillemets, gras, etc.) en balises qui encadrent la séquence marquée; on attribue ensuite une valeur sémantique à ces balises tout en les conservant, selon la syntaxe suivante : <quote rend="guillemet">, <mentio rend="RO">.

On peut distinguer les activités le plus souvent graphiquement marquées qui sont liées soit à la mention soit à la citation, et le discours plus général sur la langue.

Définition des différentes balises d'activités métalinguistiques graphiquement marquées:

- la balise "cit" est utilisée pour les citations d'extraits d'oeuvres, de chansons, les écriveaux.
- la balise "quote" est utilisée pour les mentions d'expressions.

- la balise "mentio" est utilisée pour l'activité métalinguistique de mention d'un mot.
- les discours rapportés sont balisés par "q type="XX".
- la balise "term" est utilisée dans l'extrait de dictionnaire pour la définition d'un mot.
- la balise <hi rend="XX"> est utilisée pour les mentions de lettres.
- la balise "gloss" est utilisée pour baliser les gloses.
- la balise "title" est utilisée pour les citations de titres d'oeuvres.

L'ensemble des activités métalinguistiques marquées (non compris les Discours Rapportés) occupe 6% du texte total (en nombre de mots) et représentent 9% de la surface du discours du narrateur, ce qui est énorme. (Le discours sur la langue et ses catégories n'est pas décompté ici). De plus, ces activités étant assumées partiellement par d'autres personnages que la narratrice, apparaissent pour 23% à l'intérieur d'un discours direct (les propos didactiques du gentil Mr Henri). Les discours rapportés (Discours Direct, Discours Direct Rapporté et Monologue Intérieur) représentent 33,4% de la surface.

Le système de marquage des activités métalinguistiques est multiple : dans 59,4% des cas, les guillemets, 12,74% les italiques, 2,83% les majuscules (uniquement pour les citations et écriteaux), et dans 22% des cas, ces activités n'ont pas de marque graphique.

Les DD sont marqués par le tiret, la parole intérieure le plus souvent par des parenthèses.

Texte entier	tiret	Guillemet	italique	majuscule	rien	Total	
		Nb	Nb	Nb	Nb	Nb	Nb mots
Mentio		91	12		30	133	253
quote		19	3		6	28	160
citation		8	12	6	1	27	460
titre					2	2	3
glose					8	8	153
terme						6	10
souligné		8				8	8
Total		126	27	6	47	212	1047
% / Total		59,43	12,74	2,83	22,17		

Marquage graphique des activités métalinguistiques

Il sera intéressant de voir (je renvoie ici aux analyses de C. Boré) comment les élèves utilisent ces marques typographiques dans le marquage des activités métalinguistiques.

2.2. – Le lexique

2.2.1. *La richesse lexicale* : Le texte comporte 17404 mots et 3161 mots (lemmes) différents, soit un taux de 18,16%. Les 147 paragraphes (sur le total de 592, soit 24,8%) contenant au moins une occurrence de l'un des mots suivants : *mot, phrase, nom, verbe, pronom, adjectif*, représentent 42% du nombre total de mots et 58% du nombre de mots différents, ce qui montre bien le poids de la thématique linguistique dans le texte.

	contenant un des 6 mots ling	Texte entier	% § ling/ texte entier
Nb de Paragraphes	142	592	24,8
Nb mots	7311	17404	42
Nb lemmes diff	1833	3161	58
% lemmes/mots	25	18,16	

2.2.2. La fréquence lexicale :

Toujours à l'intérieur des paragraphes contenant un des 6 mots précités, les 20 mots de grammaire les plus fréquents représentent 1% des mots différents et 3,8% du total des mots. Sur l'ensemble du texte, "mot" est le substantif de fréquence la plus forte (131), puis viennent en ordre décroissant les mots "phrase"(34), "nom" (23), "adjectif"(18), "langue"(14), "grammaire"(13), "verbe"(11), "syllabe", "lettre", "dictionnaire", "article" (8), "pronom" (6). Cela nous amène à regarder de plus près la représentation de la langue dans le texte d'Orsenna. Pour cela il faut sortir de l'analyse lexicale brute et examiner les activités prédicatives liées aux mots pôles. *Remarque* : L'analyse de ce texte unique et bref ne justifiait pas une analyse statistique de co-occurrences.

2.2.3. Activités prédicatives liées aux mots pôles :

Il est tout à fait significatif que le lexème le plus fréquent du texte soit le substantif "mot" : le modèle de la langue sous-jacent est le modèle combinatoire qu'on peut dire de la doxa grammairienne dominante : l'unité de base est le mot, lui-même composé de syllabes, elles-mêmes composées de lettres. L'extraction des paragraphes contenant "mot" (pas moins de 53 paragraphes attestent cette idée du mot comme unité de base de la langue, cf. les extraits en Annexe A1) met bien en évidence ce présupposé :

- la capacité de parole est avant tout liée à la possession du lexique (dans la tête comme dans le dictionnaire, la tempête fait perdre les mots et non la syntaxe); l'important est la richesse lexicale (pauvreté vs trésor).
- Le dictionnaire est le pourvoyeur de cette richesse et un objet d'amour.
- Les métaphores récurrentes qui identifient le mot soit à une marchandise (le marché des mots, l'usine à mots, les distributeurs de mots, de verbes, d'adjectifs, de prépositions) soit à un élément d'une réserve d'animaux (la volière, la fourmilière, la ville des noms) présentent l'activité langagière comme une activité de choix de mots dans un sac de mots selon des règles qui mettent en ordre hiérarchique et linéaire les différentes "tribus" de mots.
- L'objectif de cette activité de mise en ordre des mots est de produire une phrase, présentée comme l'arbre décoré selon les bonnes traditions.
- La motivation de cette activité combinatoire est ludique. (avec un bémol qui rassurera les sémanticiens logiciens) : - *Les vrais amis des phrases sont comme les fabricants de colliers. Ils enfilent des perles et de l'or. Mais les mots ne sont pas seulement beaux. Ils disent la vérité.* <pb n="130"/> ; - *Un écrivain a pour métier la vérité.*; - *Qu'est-ce qu'un grand écrivain ?* <pb n="145"/> - *Quelqu'un qui construit des phrases, sans se soucier des modes, seulement pour aller explorer la vérité.*)
- La doxa s'exprime aussi dans la représentation adamique de l'activité de dénomination : la fonction première de la langue est d'étiqueter en nommant (voir le personnage de la nommeuse qui "redonne vie aux mots" en les prononçant). Le nom est le personnage dominant de la phrase, pour construire une phrase, on choisit d'abord deux noms et ensuite le verbe qui va les relier.
- Liée au dernier point, la représentation de la hiérarchie des parties du discours est loin d'être neutre: " *Le premier métier, c'est de désigner les choses. Vous avez déjà visité un jardin botanique? Devant toutes les plantes rares, on a piqué un petit carton, une étiquette. Tel est le premier métier des mots : poser sur toutes les choses du monde une étiquette, pour s'y reconnaître. C'est le métier le plus difficile. [...] Les mots chargés de ce métier terrible s'appellent les noms. La tribu des noms est la tribu principale, la plus nombreuse. Vous comprenez pourquoi dans la ville, au pied de notre colline, les noms pullulaient. Les autres tribus de mots devaient lutter pour se faire une place.*"

- Quant aux pronoms, il leur est prêté des motivations compétitives (remplacer les noms) sur lesquelles on peut s'interroger : ont-elles une valeur normative : il ne faut pas en abuser? " *Les pronoms toisaient tous les autres mots avec un de ces mépris... - On leur a donné un rôle très important : tenir, dans certains cas, la place des noms.*"

Ce texte introduit d'autres confusions massives porteuses de préjugés annexes : l'espèce d'assimilation entre genre et sexe par exemple (voir l'annexe A3, GENRE): on parle de mots-hommes et mots-femmes; l'accord entre nom et adjectif est assimilé au mariage, avec la contradiction qu'il s'agit de mariage homo, puisque " *Rien de plus docile et souple que le sexe d'un adjectif. Il change à volonté, il s'adapte au client.*" Avec un jeu métaphorique qui n'est pas entièrement contrôlé par son auteur puisque c'est un adjectif féminin qui se voit délaissé par les noms, évoquant irrésistiblement la rangée des jeunes filles qui font tapisserie au bal, attendant qu'un nom leur prête attention: " *Hélas pour « hantée », les noms passaient sans lui jeter un regard. C'était à serrer le coeur, tous ces adjectifs abandonnés.*" (Si bien que "nom" (de genre masculin) se voit implicitement attribué la molécule sémique /sexe masculin/), ce qui s'exprime explicitement dans la phrase qui suit : " *l'infidélité congénitale des noms, de vrais garçons, ceux-là, ils changent de qualificatifs comme de chaussettes. À peine accordés, ils jettent l'adjectif, retournent au magasin pour en chercher un autre et, sans la moindre gêne, reviennent à la mairie pour un nouveau mariage.*"

Voir aussi les images garçon/fille : " *douceur, doux et sœur [...] Et je suis sûre qu'il se serait moqué du mien, douceur, un vrai mot de fille*". " *Faire plaisir à son frère, à toute heure du jour et de la nuit, vous connaissez d'autres moyens pour rendre potable la vie de famille ?*"

On peut donc se poser des questions sur le bien fondé de telles représentations de l'activité langagière, bien conforme au modèle logico-grammatical, définissant la phrase comme objet de la linguistique, comme combinatoire de mots, représentations contredites depuis longtemps aussi bien par l'analyse des productions orales que par celle des acquisitionnistes ou les études de pathologie du langage. La vie des mots est loin d'être abordée avec la richesse d'approches qu'autorise l'analyse des productions discursives (*le méchant veut m'ouvrir la tête pour y déposer des mots desséchés* : on pourrait attendre une autre approche du fonctionnement de la langue de bois; la nommeuse est celle qui redonne vie aux mots en les prononçant).

Certes, le genre du conte incite à l'usage des métaphores. Mais on voit bien que leur mise en œuvre dans ce genre de discours didactique opère des glissements mal contrôlés où s'exprime le jeu de doxas et idéologies dépassant les objectifs explicitement poursuivis par leur auteur.

Conclusion :

Une première comparaison du conte didactique contemporain avec le corpus de contes du 19^{ème} siècle a mis en évidence des différences de langue dans la longueur de la phrase, dans le type de narrateur (1PS vs 3PS) qui se répercute sur un grand nombre de variables morphosyntaxiques liées au dispositif énonciatif. Les aspects didactiques du texte s'expriment dans le poids des différentes activités métalinguistiques (mention, citation, gloses), dans le poids du lexique abstrait. Les propriétés du genre conte sont largement présentes (mais je dirai qu'il s'agit plutôt de caricature des propriétés narratives du conte) dans l'attribution de traits animés animaux et humains aux unités linguistiques, dans l'usage de métaphores filées, dans la polarisation binaire des actants, les méchants (Jargonos et Nécrole, représentants du pouvoir dans l'institution Education Nationale et les gentils (la jeune enseignante blonde et sensible et Mr Henri qui rime). Nous avons vu que le conte didactique n'évite pas les écueils d'une présentation très doxique de la langue. L'analyse des productions d'élèves à partir de ce texte confirme d'ailleurs les confusions conceptuelles que ce discours métaphorisant peut entraîner.

Pour la suite des travaux, il serait intéressant de mettre en commun les efforts de constitution de corpus numérisés de textes intégraux faisant partie des textes didactiques de référence pour

poursuivre un travail d'analyse contrastive plus systématique des genres textuels, en particulier des textes fictionnels didactiques. D'un point de vue technique, nous avons pu tester que les méthodes de balisage identiques du texte didactique et des productions d'élèves dans leur version définitive permet, si ces dernières sont suffisamment nombreuses, la comparaison des deux et dans ce cas précis, d'analyser, entre autres, la gestion des activités métalinguistiques à l'écrit.

ANNEXE :

A.1. Extrait du texte balisé :

<p>Justement, une femme en larmes suppliait : </p>
<p><q type="DD" rend="tiret"> - Mon mari m'a sauvagement quittée. Je voudrais un mot pour qu'il comprenne ma douleur, un mot terrible, qui lui fasse honte. </q><pb n="48"/></p>
<p>Le vendeur, un jeunot, sans doute un débutant, commença par rougir, <q type="DDR" rend="guillemet"> tout de suite, tout de suite</q>, plongea dans un vieux volume et se mit à feuilleter comme un forcené <q type="DDR" rend="guillemet"> «j'ai ce qu'il vous faut, une petite seconde. Voilà, vous avez le choix : <mentio>affliction</mentio>... </q></p>
<p><q type="DD" rend="tiret"> - Ça sonne mal. </q> </p>
<p><q type="DD" rend="tiret"> - <mentio>Neurasthénie. </mentio> </q></p>
<p><q type="DD" rend="tiret"> - On dirait un médicament. </q></p>
<p><q type="DD" rend="tiret"> - <mentio>Désespérade. </mentio> </q></p>
<p><q type="DD" rend="tiret"> - Je préfère, celui-là, il me plaît. <mentio>Désespérade</mentio>, je suis en pleine désespérade ! </q></p>
<p>Elle glissa une pièce dans la main du vendeur et s'en alla ragailardie. Elle emportait dans ses bras son mot nouveau, <mentio>désespérade</mentio>, <mentio>désespérade</mentio>... Elle n'était plus seule, elle avait retrouvé quelqu'un à qui parler. </p>
<p>Le client suivant était un vieux, d'au moins quarante ans; à cet âge, je ne croyais pas qu'on s'occupait toujours d'amour. </p>
<p><q type="DD" rend="tiret"> - Voilà. Ma femme ne supporte plus mes <quote rend="RO">je t'aime</quote>. <q type="DDR" rend="guillemet">« Depuis vingt ans, tu pourrais varier; invente autre chose, <seg ana="incise_di">me dit-elle,</seg> ou je m'en vais. »</q> </q> </p>
<p><q type="DD" rend="tiret"> - Facile, vous pourriez lui dire : <q type="DDR" rend="guillemet"> «J'ai la puce à l'oreille. »</q> </q></p>
<p><q type="DD" rend="tiret"> - Pour qu'elle me croie malpropre ? </q></p>
<p><q type="DD" rend="tiret"> - <quote rend="guillemet">«Je suis coiffé de toi. »</quote> </q><pb n="49"/></p>
<p><q type="DD" rend="tiret"> - Ce qui veut dire ? </q> </p>
<p><q type="DD" rend="tiret"> - <gloss rend="RO">L'obsession que j'ai de toi s'est enfoncée sur ma tête comme un chapeau trop grand. Je suis coiffé de toi. Je ne vois plus que toi</gloss>...</q> </p>
<p><q type="DD" rend="tiret"> - Je vais essayer. Si ça ne marche pas, je vous le rapporte. </q></p>
<p>Nous aurions pu rester jusqu'à la nuit. La file des clients s'allongeait. Thomas, comme moi, tendait l'oreille, <quote rend="guillemet"><je vais lui faire une langue fourrée</quote>, on jouera à <quote rend="guillemet"><la bête à deux dos ></quote>. Ses yeux brillaient, il avait l'air de comprendre des choses. Il faisait provision. Il saurait leur parler, aux filles, dès son retour, elles n'en reviendraient pas. Depuis le temps qu'il cherchait une recette pour draguer les grandes, les bien trop grandes pour lui. </p>
<p>Devant les autres boutiques aussi, une foule se pressait. J'aurais volontiers passé du temps avec <cit rend="maj">DIEUDONNÉ
APPELEUR DIPLÔMÉ
DES PLANTES ET DES POISSONS</cit>
ou chez la mystérieuse <pb n="50"/>
<cit rend="maj">MARIE-LOUISE
ÉTYMOLOGISTE EN QUATRE LANGUES</cit></p>

A2. Le mot comme unité de base de la combinatoire (quelques unes des 53 occurrences)

La nuit, je rêvai qu'avec des pincettes quelqu'un se préparait à m'ouvrir la tête pour y installer un tas de mots qu'il avait près de lui, des mots aussi desséchés que des squelettes. Heureusement, un lion, un moucheron et une tortue prenaient ma défense, mettaient en fuite le méchant et ses pincettes. tout ce que je savais, le vent me l'arrachait, ça ressortait par l'autre oreille, mes leçons d'histoire, les dates que j'avais eu tant de mal à apprendre, les verbes irréguliers anglais... Bientôt je serais tout à fait creuse. Et vide.

Un petit mot m'est apparu, le dernier qui me restait, blotti dans un coin, deux syllabes minuscules, tout aussi terrorisées que moi

Muets. La tempête nous avait arraché tous nos mots.

C'était à moi de parler mais je ne pouvais pas. Je n'avais pas encore retrouvé mes mots. Je n'étais pas encore guérie de la tempête.

Un mot qui flottait sur l'eau verte, un mot plat comme une méduse ou une limande. Inutile d'être bien maligne pour deviner ce qui s'était passé. La tempête avait tant secoué, comme nous, les dictionnaires que les mots s'en étaient détachés. Et maintenant les dictionnaires, vidés de leur contenu, devaient reposer sur le fond de la mer, à côté de leurs amis, les champions de Scrabble.

La mer nous rendait ce que le vent nous avait volé. Des milliers de mots, un banc immense clapotait tranquillement devant nous. Il suffisait de tendre les bras pour les pêcher. Je me souviens des premiers que j'ai pris dans ma main.

-Je vous présente notre marché aux mots, dit Monsieur Henri. C'est ici que je fais mes courses. Vous y trouverez ou retrouverez tout ce dont vous avez besoin.

Les habitants s'étaient fait, comme vous, nettoyer de tous leurs mots.

Comment, privés des mots et encore plus des chiffres, aurions-nous pu lui répondre ? Je vous rappelle qu'après les cahots de la tempête et les agressions du vent, nos pauvres têtes ne pouvaient plus fabriquer la moindre phrase ! Nous parvenions tout juste à comprendre ce qu'on nous disait.

Mais le mot demeura, ses cinq syllabes brillantes voletant dans l'air tel un papillon. Il disparut bientôt en agitant les ailes, pour dire merci, merci de m'avoir prononcé.

Et, de sa même voix douce, attendrie, d'amoureuse, la nommeuse, sans plus s'occuper de nous, lut lentement les quatre syllabes d'un autre mot

Elle redonne vie aux mots rares. Sans elle, ils disparaîtraient à jamais dans l'oubli.

Oh, la joie de ces mots sortis de l'oubli. Ils s'étiraient, ils s'ébrouaient, certains n'avaient pas dû voir le grand air depuis des siècles.

De temps en temps, il nous envoie des hélicoptères équipés de lance-flammes, et fait brûler une bibliothèque... chaque dimanche soir, avant de m'endormir, je flâne quelques minutes au fond d'un dictionnaire, je choisis un mot inconnu de moi (j'ai le choix : quand je pense à tous ceux que j'ignore, j'ai honte)

Pour être franche, moi qui déteste ranger ma chambre, j'aurais volontiers mis un peu d'ordre dans ma tête. Les mots s'étaient entassés partout, sous mes cheveux, derrière mon front, derrière mes yeux. Je les sentais amoncelés au petit bonheur la chance dans les moindres recoins de mon crâne. Je sentais revenir à grands pas la migraine...

La ville des noms est à neuf kilomètres.

Les rues n'étaient parcourues que de mots. Des mots innombrables, radieux sous le soleil. Ils se promenaient comme chez eux, ils étiraient dans l'air tranquillement leurs syllabes,

Ils passent leurs journées à se déguiser, à se maquiller et à se marier.

Du haut de ma colline, je n'ai d'abord rien compris. Les mots étaient si nombreux. Je ne voyais qu'un grand désordre. J'étais perdue dans cette foule.

Vous comprenez pourquoi dans la ville, au pied de notre colline, les noms pullulaient.

L'hôpital ? Un hôpital pour les mots ?

Tous les noms, mes amis de la ville des mots, étaient là. Ils m'avaient reconnue, ils se pressaient contre le grillage, ils me faisaient fête.

Nous nous avançâmes de quelques pas, vers une grande vitre derrière laquelle, sur plusieurs étages, s'activaient d'autres mots. Par leur manière de s'agiter perpétuellement et en tout sens, on aurait dit des fourmis.

Tu vas construire ta première phrase.

- Commence par le plus simple. Va là-bas, dans la volière, choisis deux noms. Après, pour le verbe, tu viendras choisir dans la fourmilière.

À peine la porte poussée, je fus assaillie, étouffée, aveuglée, les noms se battaient, ils m'entraient dans les yeux, les narines, les oreilles, j'éternuai, je toussai, je faillis mourir, ils voulaient tous que je les retienne, ils devaient tellement s'ennuyer dans leur prison. Au moment de m'évanouir, j'en saisis deux par les ailes, au hasard, « fleur » et « diplodocus », et je refermai la porte, pâle, tremblante, à demi morte.

- Allez, maintenant, tu pêches un verbe.

puis je me retirai avec l'un d'entre eux, pris au hasard, « grignoter ».

- Parfait, maintenant tu t'assieds là, à ce bureau, tu déposes tes mots sur la feuille de papier et tu formes ta phrase. Mes mots, si péniblement attrapés, je les retenais toujours par les ailes, je ne voulais pas les laisser, je craignais qu'ils ne s'échappent. Après tout, une phrase, pour un mot, c'est une prison.

Tu en connais de plus utiles, des usines ? Que peut-on fabriquer au monde de plus nécessaire pour les êtres humains que des phrases ? Tu as compris le principe. Tu trouveras le magasin des adjectifs derrière la volière des noms. Et aussi un distributeur de prépositions pour les compléments indirects : aller à Paris, revenir de New York. Bonnes phrases !

J'ai joué toute la journée. J'avais l'impression de retrouver les cubes de mon enfance. Je combinais, j'accumulais, je développais. J'avais découvert, en founant dans l'usine, d'autres distributeurs. Celui des interjections (Ah ! Bon

! Hélas!), celui des conjonctions (mais, ou, et, donc, or, ni, car...), petits mots bien utiles pour relier les morceaux de phrase.

L'usine est à toi. Comme à tous les habitants de l'île qui veulent s'amuser avec les phrases.

Les vrais amis des phrases sont comme les fabricants de colliers. Ils enfilent des perles et de l'or.

A.3.Extraits du texte portant sur le genre : 9

douceur, doux et sœur

Et je suis sûre qu'il se serait moqué du mien, douceur, un vrai mot de fille.

Faire plaisir à son frère, à toute heure du jour et de la nuit, vous connaissez d'autres moyens pour rendre potable la vie de famille ?

Il y a des noms-hommes, ce sont les masculins, et des noms-femmes, les féminins.

Les articles marchent devant les noms, en agitant une clochette : attention, le nom qui me suit est un masculin, attention, c'est un féminin! Le tigre, la vache.

Plutôt des amitiés. Comme dans les écoles d'autrefois, quand elles n'étaient pas mixtes. Au royaume des mots, les garçons restent avec les garçons et les filles avec les filles.

L'article entraînait par une porte, l'adjectif par une autre. Le nom arrivait le dernier. Ils disparaissaient tous les trois. Le toit de la mairie me les cachait. J'aurais tout donné pour assister à la cérémonie. J'imagine que le maire devait leur rappeler leurs droits et leurs devoirs, qu'ils étaient désormais unis pour le meilleur et pour le pire. Ils ressortaient ensemble se tenant par la main, accordés, tout masculin ou tout féminin : le château enchanté, la maison hantée... Peut-être qu'à l'intérieur le maire avait installé un distributeur automatique, les adjectifs s'y ravitaillaient en e final pour se marier avec un nom féminin. Rien de plus docile et souple que le sexe d'un adjectif. Il change à volonté, il s'adapte au client.

Je vais vous dire un secret : les adjectifs ont l'âme sentimentale. Ils croient que leur mariage durera toujours...

C'est mal connaître l'infidélité congénitale des noms, de vrais garçons, ceux-là, ils changent de qualificatifs comme de chaussettes. À peine accordés, ils jettent l'adjectif, retournent au magasin pour en chercher un autre et, sans la moindre gêne, reviennent à la mairie pour un nouveau mariage.

- Ah, ces adverbes! De vrais invariables, ceux-là! Pas moyen de les accorder. Les femmes auront beau faire avec eux, elles n'arriveront à rien.

Bibliographie

Adam, J.-M., Lugin, G., Revaz, F., 1998, "Pour en finir avec le couple récit/discours", *Pratiques*, 100, p 81-98.

Bakhtine, M., 1984 [1952-53], *Esthétique de la création verbale*, Paris, Gallimard.

Benveniste, E., 1966, *Problèmes de linguistique générale*, Paris, Gallimard.

Biber, D., 1998, *Corpus Linguistics: Investigating Language structure and Use*, Cambridge University Press.

Boré, C. 2004, "L'écriture scolaire : langue, norme, "style", quelques exemples dans le discours rapporté", *LINX*, n° 51, 91-106

Bouquet, S. (éd.), 2004, Les genres de la parole, *Langages*, n° 153.

Bronckart, J.P., 1996, *Activité langagière, textes et discours*, Lausanne, Delachaux et Niestlé.

Cohn, D., 1981, *La transparence intérieure*. Paris, Le Seuil.

Declerck, R., 2003, "How to manipulate tenses to express a character's point of view", *Journal of Literary Linguistics*, 85-112.

De Mattia, M., Joly, A., 2001, *De la syntaxe à la narratologie*, Paris, Ophrys.

Fillietaz, L. et Grobet, A., 1999, "L'hétérogénéité compositionnelle du discours : quelques remarques préliminaires", *Cahiers de linguistique française*, n° 21, 213-260.

Genette, G., 1972, *Figures III*, Paris, Le Seuil.

Genette, G., 1983, *Nouveau discours du récit*, Paris, Le Seuil.

Genette, G., (éd.), 1986, *Théorie des genres*, Paris, Le Seuil.

Hamburger, K., 1986, (trad.), [1957], *Logique des genres littéraires*, Paris, Le Seuil.

Henrot, G., 2000, *L'usage de la forme. Essai sur les Fruits d'or de Nathalie Sarraute*, Biblioteca Francese, Unipress, Padova.

Kuyumcuyan, A. 2002, Diction et mention. Pour une pragmatique du discours narratif, Berne, Peter Lang.

Langue française, 2001, "La Parole intérieure." n° 132.

Lintvelt, J., 1989, *Essai de typologie narrative. Le "point de vue". Théorie et analyse*. Corti. Paris.

Lips, M., 1926, *Le Style indirect libre*. Payot. Paris.

Mainueneau, D., 1986, *Eléments de linguistique pour le texte littéraire*, Paris : Bordas.

Mainueneau, D., 2000, Instances frontières et angélisme narratif, *Langue française*, 128; pp. 74-95.

Malrieu, D., Rastier F. 2001, "Genres et variations morpho-syntaxiques ", *T.A.L.*, 42, 2, 547-577.

- Malrieu D., 2002, "Stylistique et Statistique textuelle: À partir de l'article de C. Muller sur les pronoms de dialogue", JADT 2002, 6èmes Journées internationales d'analyse des données textuelles, St-Malô, 13-15 mars 2002.
- Malrieu, D., 2004, Linguistique de corpus, genres textuels, temps et personnes, *Langages*, 153, 73-85.
- Malrieu, D., "Discours rapportés et typologie des narrateurs dans le genre romanesque", Actes du colloque Ci-Dit de Cadix "Dans la jungle des discours", 11-14 mars 2004, (à paraître chez l'Harmattan).
- Malrieu, D., 2006, Type de narrateur et place du lecteur dans *Le ravissement de Lol V. Stein*, Janvier 2006, <http://www.modyco.fr/>
- Malrieu, D., 2006, Quel balisage du roman contemporain ? Actes des Journées d'étude internationales "Corpora et questionnements du littéraire", Si@t et Paris X, 15-16 novembre 2005, (à paraître dans LINX, 2007).
- Malrieu, D., 2006, Familles narratologiques et balisage du roman contemporain, Actes de l'ALLC (Association for Linguistic & Literary Computing), Paris, 5-9 juillet 2006.
- Marnette, S., 2002, "Étudier les pensées rapportées en français parlé: Mission impossible?". *Faits de Langues*, 19, p 211-20.
- Marnette, S., 2002, "Aux frontières du discours rapporté". *Revue Romane*, 37.1, p 3-30.
- Marnette, S., 2001, "Du discours insolite: Le discours indirect sans que". *French Studies*, 55.3. 297-313.
- Marnette, S., 2001, "The French Théories de l'Énonciation and the Study of Speech and Thought Presentation". *Language and Literature*, 10.3, p 261-80.
- Mellet, S. & Vuillaume, M. (éds.), 2000, *Le style indirect libre et ses contextes*. Rodopi. Amsterdam-Atlanta.
- Newman A.S., 1976, *Une poésie des discours. Essai sur les romans de Nathalie Sarraute*, Genève : Droz.
- Philippe, G. (ed.) (2000) : L'ancrage énonciatif des récits de fiction, *Langue française*, 128.
- Rabatel, A., 1998, *La construction textuelle du point de vue*, Lausanne, Paris, Delachaux & Niestlé
- Rastier, F., 2001, *Arts et sciences du texte*. Paris, PUF.
- Rivara, R., 2000, *La langue du récit*, Paris, L'Harmattan.
- Rosier, L., 1999, *Le discours rapporté. Histoire, théories, pratiques*. Paris, Bruxelles, Duculot.