



Grammaire textuelle et informatique

Pierre Muller

► To cite this version:

Pierre Muller. Grammaire textuelle et informatique. Revue de l'EPI (Enseignement Public et Informatique), 1995, 78, pp.163-175. edutice-00001144

HAL Id: edutice-00001144

<https://edutice.hal.science/edutice-00001144>

Submitted on 15 Nov 2005

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.

GRAMMAIRE TEXTUELLE ET INFORMATIQUE

Pierre MULLER

1. LA GRAMMAIRE TEXTUELLE DANS L'ENSEIGNEMENT DU FRANÇAIS

La grammaire textuelle est une approche de la langue qui se développe depuis plusieurs années et qui se propose de prendre en compte un certain nombre de phénomènes, jusqu'ici mal ou insuffisamment traités dans le cadre de la phrase : il s'agit en particulier du fonctionnement des temps verbaux, des repères énonciatifs, des organisateurs textuels et des relations anaphoriques. Souvent, en effet, on ne peut expliquer la présence et la valeur d'un élément linguistique qu'en étudiant les relations qu'il établit avec d'autres éléments figurant dans les phrases qui précèdent ou qui suivent.

Outre sa capacité d'explication, cette démarche présente un grand intérêt didactique. En liant l'étude de la langue aux textes, elle facilite, par là-même, la liaison avec les deux autres activités de la classe que sont la compréhension et la production des textes. En même temps, elle s'accompagne du souci de faire découvrir les phénomènes au lieu d'imposer magistralement des notions grammaticales.

On ne s'étonnera donc pas que la grammaire textuelle pénètre peu à peu l'enseignement de la langue à l'école et au collège et apparaisse même dans les instructions officielles actuellement en vigueur, par exemple dans le paragraphe intitulé "De la phrase au texte" : "L'enchaînement des phrases et la cohérence du texte : mots de liaison, notamment les adverbes et les conjonctions ; mots de reprise (pronoms, déterminants, substituts lexicaux) ; emploi des temps du verbe ; éléments repris et éléments nouveaux d'une phrase à l'autre ; thème et développement (on dit aussi «propos» ou «rhème») ; relation entre l'ordre des mots et l'enchaînement des phrases." ¹

1. "Instructions et Programmes pour la classe de seconde", février 1987, in *Français, classes de seconde, première et terminale*, brochure C.N.D.P., réédition 1992, p.51.

Les propositions du Comité national des programmes sont encore plus explicites sur ce point, puisqu'elles recommandent de "relier les apprentissages grammaticaux aux tâches de production : la grammaire ne doit intervenir qu'à l'occasion du travail sur les textes et des activités d'écriture. On privilégiera la grammaire de texte et non celle de la phrase, les entraînements grammaticaux centrés sur les significations et non l'analyse formelle".²

On aboutit ainsi à une généralisation de la notion de grammaire de texte, puisqu'il ne s'agit plus seulement de prendre en compte les phénomènes qui dépassent le cadre de la phrase, mais qu'on envisage aussi d'étudier tous les éléments linguistiques à partir de textes. C'est à cette conception étendue de la grammaire textuelle que nous nous référerons désormais.

Certes, des manuels scolaires ont commencé à intégrer cette dimension³, mais les textes qu'ils proposent pour traiter chaque question grammaticale ne sont le plus souvent que de brefs extraits. On peut aussi avoir recours à des œuvres intégrales, mais la recherche préalable à toute analyse risque d'être longue et fastidieuse. Ici, l'étude de la langue trouve dans l'informatique des outils privilégiés. A condition de disposer d'un logiciel approprié, le recours à des banques de données textuelles, préalablement lemmatisées et catégorisées, permet d'avoir accès à des textes complets et de les parcourir exhaustivement pour traiter une question donnée. Ce sont ces outils, à la fois informatiques et linguistiques, que nous avons conçus dans le cadre d'une recherche menée à l'Institut national de recherche pédagogique.

2. LES OUTILS INFORMATIQUES ET LINGUISTIQUES

Trois aspects ont été pris en compte dans la conception de ces outils : les catégories linguistiques à étudier, les banques de textes, les logiciels capables de les traiter.

2. "Projet pour l'évolution du Collège", *Le français aujourd'hui*, supplément au N° 96, janvier 1992.

3. Citons en particulier le manuel *De la phrase au texte* (classes de 4^{ème} et de 3^{ème}), édité par la librairie Delagrave.

2. 1. Les catégories linguistiques

La définition des catégories linguistiques ne laissait pas toute la liberté souhaitable, car pour une application destinée à l'enseignement, il fallait respecter la nomenclature grammaticale officielle et rester compatible avec les ouvrages utilisés par les élèves (grammaires et dictionnaires). En ce qui concerne la lemmatisation, nous avons généralement suivi les conventions habituelles des dictionnaires, mais nous avons parfois opéré des regroupements qui devraient faciliter les recherches pour l'utilisateur. Pour la catégorisation proprement dite, nous avons retenu les catégories suivantes : la partie du discours, la conjugaison, le genre, le nombre, la personne, la voix, le mode, le temps. Nous y avons ajouté pour le latin la déclinaison et la conjugaison.

2. 2. Les banques de données

Dans chacun des domaines retenus a été constitué un corpus de textes qui tient compte de plusieurs critères : apprentissages visés, niveau des élèves concernés, programmes scolaires.

- un corpus de textes narratifs destiné aux classes de quatrième ou de troisième en français, comprenant trois œuvres du XIX^e siècle : *Le pied de momie*, de Théophile Gautier (1840), *Un cœur simple*, de Gustave Flaubert (1876), *Sur l'eau*, de Guy de Maupassant (1876).
- un corpus de textes argumentatifs, destiné à la classe de seconde pour une étude interdisciplinaire en lettres et en histoire, comprenant trois textes concernant les problèmes de l'enseignement et en particulier celui de la laïcité : *Projet de loi sur l'instruction publique* (18 juin 1849), discours d'Alfred de Falloux ; *La liberté de l'enseignement* (15 janvier 1850), Discours de Victor Hugo, prononcé au cours d'un débat autour du projet précédent ; *Lettre aux instituteurs* de Jules Ferry (17 novembre 1883).
- un corpus latin, comportant pour le premier cycle des textes choisis pour permettre un apprentissage des notions fondamentales et des connaissances de base de la langue latine sans présenter de difficultés majeures pour des débutants ; pour le second cycle les *Catilinaires* de Cicéron.⁴

Chacun de ces corpus a été saisi, lemmatisé et catégorisé.

4. cf. *Analyses des Catilinaires et de textes latins choisis*, C.N.D.P. et I.N.R.P., 1994.

2. 3. Les logiciels

Pour réaliser la codification des textes et automatiser les recherches linguistiques, nous avons conçu deux logiciels spécifiques qui s'inscrivent dans le prolongement de la collection Logitexte.⁵

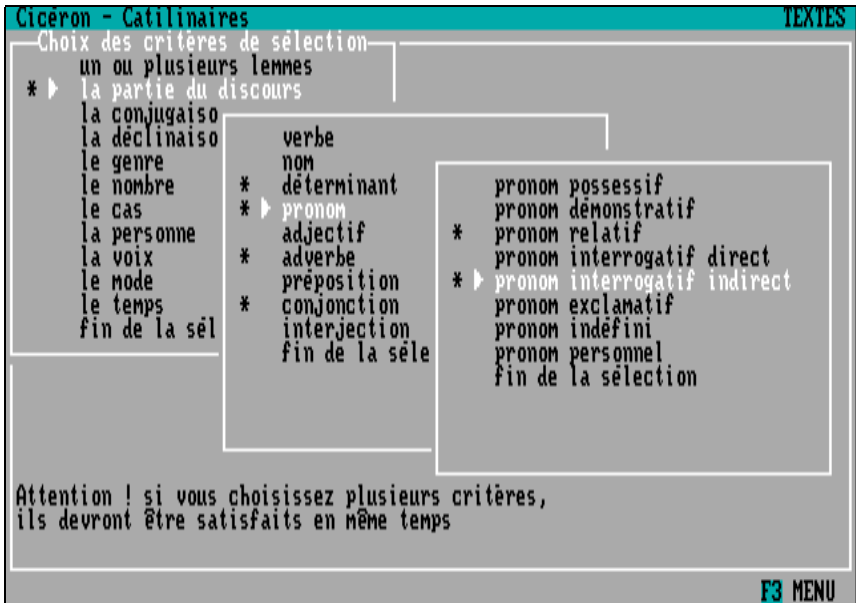
- Le premier, qui effectue la lemmatisation et la catégorisation, affecte à chaque forme du texte un descripteur qui contient les informations linguistiques décrites ci-dessus, c'est-à-dire le lemme et les catégories grammaticales. Cette affectation se fait de manière semi-automatique, une partie des informations étant fournies par l'utilisateur, l'autre étant calculée par l'ordinateur.⁶
- Le deuxième qui est une extension du logiciel PISTES déjà diffusé, ajoute aux fonctions de celui-ci la possibilité d'effectuer des recherches sur toutes les formes qui correspondent à une série de critères d'ordre linguistique et de signaler ces formes, sur écran par l'inversion vidéo, le soulignement ou une couleur prédéterminée, sur imprimante par des caractères gras, italiques ou soulignés. On peut éditer de cette manière soit l'ensemble du texte, soit les phrases correspondant à la demande. Le document 1 ci-dessous montre comment s'effectue le choix des critères qui président à la recherche dans le texte.

5. La collection LOGITEXTE, publiée en coédition avec le C.N.D.P., comprend, outre le logiciel *PISTES*, les *Analyses du Tartuffe, des Fleurs du Mal, des déclarations des droits de l'homme et de textes sur la guerre en 1915*.

6 Ce logiciel, non encore diffusé, a été réalisé en grande partie par Roger BARRAU, professeur au lycée Joffre à Montpellier, et inclut un module de traitement de la conjugaison française qui est l'oeuvre de Frédéric ROBERT (I.N.R.P.)

DOCUMENT 1

Choix des critères



3. L'EXPÉRIMENTATION

Avec les outils ainsi élaborés, cette approche a pu être expérimentée pendant deux ans, de 1991 à 1993, dans une dizaine de classes de collège et de lycée.

3. 1. Les exercices

Les traitements que le logiciel effectue sur les textes ne sont pas exploitables immédiatement par des utilisateurs non formés, qu'il s'agisse des enseignants ou des élèves. Avant qu'ils acquièrent une certaine autonomie dans l'utilisation d'un logiciel ouvert, comme nous l'ont montré déjà des recherches antérieures, il est nécessaire de les former aux méthodes nouvelles, en leur fournissant des plans d'étude qui les guident dans leurs premières investigations. C'est seulement après quelques séances de ce type qu'il est possible de laisser aux élèves la liberté de définir eux-mêmes le sujet de leur recherche.

Les programmes de travail ainsi fournis aux élèves sont généralement constitués de la manière suivante : ils donnent les questions qui doivent être posées à l'ordinateur pour obtenir les données nécessaires ; ils indiquent les informations à noter pendant la lecture, souvent à l'aide d'un tableau qui en facilite le relevé ; ils proposent les analyses susceptibles d'aboutir à des conclusions. Parmi les sujets qui ont pu être étudiés de cette manière, citons en français le système temporel dans la narration, les adjectifs dans la description, les indices de l'énonciation et les connecteurs logiques dans des textes argumentatifs ou encore, en latin, les paradigmes des conjugaisons et des déclinaisons, les emplois des différentes conjonctions et prépositions. Nous en développerons trois exemples.

3. 2. L'emploi des temps dans *Sur l'eau*

Si l'on prend comme instrument d'analyse la description des temps verbaux telle qu'elle a été proposée par Benveniste ⁷, puis Weinrich ⁸, on peut présenter sous la forme du document 2 leur répartition page par page dans la nouvelle de Maupassant.

Cette répartition fait apparaître que l'ensemble de la nouvelle est dominé par les temps de l'histoire. Cependant il y a quelques temps du commentaire, qui sont presque tous concentrés dans les deux premières pages. Pour une analyse plus fine du fonctionnement de ces temps, on peut observer leur répartition phrase par phrase (autre présentation du tableau fournie par le logiciel), qui permet de les envisager dans leur enchaînement, et leur situation précise dans le texte en demandant l'édition du texte avec la mise en relief des trois sortes de temps suivants :

- imparfait et plus-que-parfait de l'indicatif et du subjonctif (arrière-plan de l'histoire),
- passé simple et passé antérieur (premier plan de l'histoire),
- présent, futur, passé composé de l'indicatif, du subjonctif et de l'impératif (temps du commentaire).

Il serait trop long de faire relever tous les emplois, mais il suffit ici de demander aux élèves de s'intéresser aux changements de temps, en leur proposant par exemple de remplir le tableau ci-dessous et d'expliquer les changements de registre et de plan observés.

7. Emile BENVENISTE, *Problèmes de linguistique générale*, Paris, Gallimard, collection TEL, tome 1, 1976, 356 p.

8. Harald WEINRICH, *Le temps*, Paris, Editions du Seuil, 1973, 334 p.

DOCUMENT 2
Répartition des temps dans *sur l'eau*

page	HISTOIRE					COMMENTAIRE		
	rétro		degré 0		prosp.	rétro	0	prosp.
	arr.	P. PL.	arr.	P. PL.				
137	2		6	3			18	2
138			5	1		3	32	
139	1		23	34	1			
140	3		22	24	2			
141	2		20	15	3		3	1
142	7		18	16	1		5	
143			5	9			1	
total	15		99	102	7	3	59	3

On peut ainsi étudier dans quels cas Maupassant utilise les temps du commentaire. La plupart d'entre eux sont situés au début du discours tenu par le deuxième narrateur qui se présente précisément comme un long commentaire sur la rivière, son caractère étrange et inquiétant, donnant ainsi l'ambiance du récit qui va suivre et qui en est en quelque sorte l'illustration. Quelques-uns sont des interventions brèves au milieu du récit, soit remarques générales sur la nature humaine, soit indices de rattachement au hic et nunc de l'énonciateur, les uns et les autres conférant à cette surprenante aventure un effet de réel. Enfin, il faut noter les présents de narration appartenant au récit lui-même, mais dont l'auteur n'abuse pas, puisqu'on ne les trouve que dans la première partie.

A l'intérieur des temps de l'histoire, la répartition entre premier plan et arrière-plan est assez classique. On trouve à l'imparfait le portrait du personnage, ses activités habituelles, et les passages descriptifs. Les actions du personnage sont plutôt au premier plan. Mais on remarque aussi que certains verbes peuvent être alternativement à l'imparfait et au passé simple, en particulier les verbes de sensation ou d'opinion. Il y a donc un choix possible pour l'écrivain qui lui permet de donner plus ou moins de relief à certaines actions.

DOCUMENT 3
Changement de temps *sur l'eau*

page	ligne	verbe	temps	registre	plan
137	1	j'avais loué	p-q-parfait	histoire	arrière-plan
137	3	je fis	passé simple	histoire	premier-plan
137	5	était	imparfait	histoire	arrière-plan

3. 3. L'argumentation dans un discours de Victor Hugo

Ce discours, prononcé au cours du débat sur la loi Falloux, présente un intérêt aussi bien pour le cours d'histoire que pour le cours de français et permet d'étudier le fonctionnement d'un texte argumentatif. Menée dans le cadre d'un travail interdisciplinaire en classe de première, son étude a été centrée sur l'emploi de la catégorie de la personne telle qu'elle se manifeste à travers les pronoms personnels et les pronoms-adjectifs possessifs.

L'argumentation de Victor Hugo se déroule en effet en trois mouvements : il expose d'abord sa conception de l'enseignement opposée à celle que révèle la loi Falloux ; puis il met en cause le parti clérical qu'il considère comme responsable de cette loi et dont il révèle les crimes antérieurs ; enfin il annonce les menaces que le vote de cette loi ferait peser sur l'ordre public.

Or la consultation des index révèle que la répartition des trois personnes utilisées (1^{re} du sing., 1^{re} et 2^{ème} du plur.) est très irrégulière entre ces trois parties (comme le montre le tableau ci-dessous, où les signes + et - révèlent les suremplois et les sous-emplois, calculés en tenant compte de la longueur des parties).

DOCUMENT 4
Locuteur et allocataire dans le discours de Hugo

	I	II	III	TOTAL
nombre total de mots	1737	1879	1038	4654
1ère pers. sing	83 +	34 -	37	154
je	75 +	33 -	33	141
mon	8	1 -	4	13
1ème pers. plur.	11	7	5	23
nous	6	6	4	16
notre	5	1	1	7
2ème pers. plur.	9-	101+	43	153
vous	9 -	76 +	41 +	126
votre	0 -	25 +	2 -	27
Messieurs	11 +	0 -	6	17

Mais il importe aussi de savoir quels sont les référents qui se cachent derrière le *nous* et le *vous*. On peut faire relever ces informations dans un tableau. Le retour au contexte permet ainsi de constater que les articulations du texte correspondent à des changements de référents du *vous*. Dans la première partie, Hugo s'adresse aux membres de l'assemblée, ce que confirme le suremploi de *Messieurs* qui désigne également les députés. Dans la deuxième partie au contraire, il s'adresse directement au parti cléricale. Enfin la troisième partie marque un retour au premier destinataire, mais les protestations d'une partie de l'assemblée qui n'accepte pas ses attaques contre la loi l'oblige à partager les emplois du *vous* entre l'ensemble de l'assemblée et ses adversaires.

Quoique les emplois du *nous* soient beaucoup moins nombreux, il est intéressant de noter que le *nous* inclusif est lié au *vous* désignant l'assemblée alors que le *nous* exclusif se situe plutôt dans les passages polémiques, où Hugo s'en prend directement à l'adversaire.

Chaque partie se caractérise donc par une stratégie argumentative spécifique qui est fortement liée à l'énonciation. Dans la première partie, pour exposer son projet personnel, l'orateur fait une place importante à la

première personne du singulier et s'adresse à l'ensemble de l'assemblée, sans trop la solliciter, en privilégiant plus le *messieurs* que le *vous*, et en recherchant la connivence par le *nous* inclusif. Dans la seconde, il s'en prend directement et violemment à l'adversaire, en faisant une place beaucoup plus importante au *vous* qui désigne maintenant le parti clérical, non comme une partie de l'assemblée, mais comme une réalité extérieure. Enfin, dans la dernière partie, il s'adresse de nouveau à l'assemblée qu'il essaie de séparer du parti clérical, mais devant les réactions d'une partie de ses membres il est obligé d'attaquer un adversaire devenu présent. On peut donc dire que Victor Hugo a choisi une stratégie, mais qu'il ne peut la maintenir jusqu'au bout, parce que l'adversaire n'accepte pas de se laisser réduire au silence.

DOCUMENT 5

Les références du vous dans le discours de Hugo

page, ligne	référent du vous	termes associés
15, l. 17	les députés	Messieurs
18, l. 41	parti clérical	parasites de l'Eglise
		maladie de l'Eglise
		metteurs en scène de la sainteté
p. 22, l. 30	les députés	Messieurs
p. 23, l. 20	la droite	Messieurs
p. 23, l. 27	les députés	Messieurs
p. 23, l. 40	la droite	Messieurs

3. 4. L'apprentissage de la langue latine

En latin, comme en français, il est possible de faire découvrir des règles de grammaire à partir des inventaires systématiques que fournit le logiciel.

a) - En morphologie, on peut ainsi découvrir les règles de formation des déclinaisons et des conjugaisons latines en pratiquant des exercices de classement des formes mises en valeur sur l'écran. Le document 6 fournit un exemple d'exercice pour l'étude du présent de l'indicatif de la deuxième conjugaison. Un exercice de ce type, destiné à des élèves débutants, doit être suivi d'une mise en commun au cours de laquelle le

professeur rectifie les interprétations erronées et fait dégager les notions principales (ici celles de radical et de désinence). Des exercices semblables ont été conçus pour les principaux temps des autres conjugaisons.

DOCUMENT 6

Étude du présent de l'indicatif de la deuxième conjugaison

Choisissez le traitement "édition du texte" et demandez au logiciel de vous fournir, dans le texte étudié, tous les verbes qui sont au présent de l'indicatif de la deuxième conjugaison à la voix active. Vous sélectionnez donc, dans le menu des critères, les lignes suivantes :

- partie du discours,
- conjugaison,
- voix,
- mode,
- temps.

Relevez et classez, dans un tableau du modèle ci-dessous, les différentes formes verbales repérées par le logiciel. Vous reporterez dans une même colonne les formes qui commencent de la même manière (radical identique) et sur une même ligne les formes qui se terminent de la même manière (désinence identique).

	verbe 1	verbe 2	verbe 3	verbe 4
désinence 1				
désinence 2				
désinence 3				
désinence 4				
désinence 5				
désinence 6				

b)- En syntaxe, on peut étudier le fonctionnement des divers types de subordinées en observant les différents mots subordonnants et les verbes qu'ils introduisent. Le document 6 présente un exercice de révision sur les emplois de la conjonction *ut* pour la classe de seconde. Des exercices semblables peuvent être réalisés pour les autres

conjonctions, en latin comme en français. On peut aussi réviser de la même manière les valeurs des différents cas en liaison avec l'emploi des prépositions.

DOCUMENT 7

La conjonction *ut*

Choisissez le traitement "Recherche de contextes" et demandez au logiciel de vous signaler, dans le texte étudié (*Deuxième Catilinaire* de Cicéron), toutes les occurrences de la conjonction *ut*. Vous sélectionnerez donc, dans le menu des critères, les lignes suivantes :

- un ou plusieurs lemmes, en indiquant le mot *ut*,
- partie du discours, en demandant uniquement les conjonctions de subordination.

Demandez ensuite (deuxième sélection) le repérage de tous les verbes.

Pour chaque emploi de *ut*, remplissez une ligne du tableau suivant. Une fois le tableau rempli, comparez les lignes correspondant à une même valeur de *ut* et essayez de formuler les règles qui correspondent à chaque valeur (temps et modes du verbe de la subordonnée, relation avec le verbe de la principale, corrélatifs éventuels).

§	verbe de la subordonnée	temps	mode	corrélatif	verbe de la principale	valeur de <i>ut</i>

4. CONCLUSION

Les observations faites par les enseignants pendant cette expérimentation permettent de formuler quelques conclusions. On a pu vérifier que ce travail met les élèves en situation de recherche et leur

permet de pratiquer une méthode expérimentale grâce à laquelle ils ont pu découvrir les structures élémentaires d'une langue qu'ils ne connaissaient pas comme le latin. En français, ils ont pu étudier systématiquement certains phénomènes linguistiques tout en relevant les caractéristiques de différents types de textes. On a pu aussi utiliser cette démarche dans le cadre d'un travail interdisciplinaire, en lettres et en histoire. L'étude des marques linguistiques que l'enseignant d'histoire pourrait difficilement effectuer seul permet des conclusions sur l'usage d'une argumentation propre à tel personnage ou à tel groupe en même temps qu'elle est susceptible d'interprétations littéraires.

Pierre MULLER

I.N.R.P. et I.U.F.M. de Versailles