



**HAL**  
open science

## Le mot texte dans les textes

Étienne Brunet

► **To cite this version:**

Étienne Brunet. Le mot texte dans les textes. T. Fitch, A. Oliver. Textes, Revue de critique et de théorie littéraire, 12, pp.91-110, 1992, Textes, Revue de critique et de théorie littéraire. hal-01552407

**HAL Id: hal-01552407**

**<https://hal.science/hal-01552407>**

Submitted on 2 Jul 2017

**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.

Étienne Brunet  
Le mot *texte* dans les textes

Au moment où l'image et le son tendent à se substituer à l'écrit et où semble s'éloigner la galaxie Gutenberg, on devrait s'attendre à quelque déclin du texte, qu'il s'agisse de la chose et du mot. Les professionnels de l'imprimé se plaignent de la mévente du livre, de la crise de la presse écrite et de la concurrence triomphante des médias. Pour leur donner tort ou raison, il faudrait s'engager sur le plan économique et procéder à des études de bibliométrie. Or nous ne voulons et ne pouvons employer ici que les méthodes de la lexicométrie et, au risque de lâcher la proie pour l'ombre, nous abandonnerons le référent culturel et social pour fixer notre attention sur la seule étude du mot. Or ce mot apparaît plus vivant que jamais. C'est lui qui orne - à la française - la couverture et le titre de la revue américaine qui nous accueille présentement. C'est lui qui apparaît - sous une livrée anglaise - dans la base de données *Frantext* où nous nous proposons de mener notre enquête<sup>1</sup>. Quand une revue ou un produit affronte le marché le nom qu'on leur donne est leur meilleur soutien et il est judicieux de le choisir parmi les thèmes en vogue et les termes de l'actualité. Nous ne disposons pas de compteur Geiger qui puisse mesurer au jour le jour le rayonnement lexical et le développement du vocabulaire. Mais différents observatoires de la langue ont été installés aux quatre coins de la francophonie<sup>2</sup> qui permettent de conclure à une fécondité soudaine d'une racine jusque-là peu sollicitée.

Jusqu'à notre époque en effet la lignée qui remonte au latin *textus* (=tissu, trame) n'a eu qu'une descendance étroite, partagée au demeurant entre le sens propre et originel qui se maintient dans *textile*, *texture*, *texturer* ou *texturiser*, et le sens figuré qui est celui de *texte*, de *textuel*, de *textuellement* et parfois de *texture*. Aucun verbe dans la famille, sauf dans le composé *prétexter*, qui vient lui aussi de la source latine (la fameuse *robe prétexte*) et qui a emprunté à son préfixe l'essentiel de son contenu sémantique (*prae* = par devant, d'où *prétexter* = mettre en avant). Les composés de *texte* sont rares et dans la base de données *Frantext* on ne trouve guère que *contexte* (195 occurrences) *contextuel* (4 occ.), *contexture* (205 occ.) et *microtexture* (2 occ.). Reste un indice qui mérite

---

<sup>1</sup> Le mot se retrouve aussi dans le titre du CD-ROM (ou DOC) que l'Institut National de la langue française vient de réaliser et qui porte le nom plaisant de DISCOTEXT, le calembour sous-jacent donnant au terme un relief inattendu.

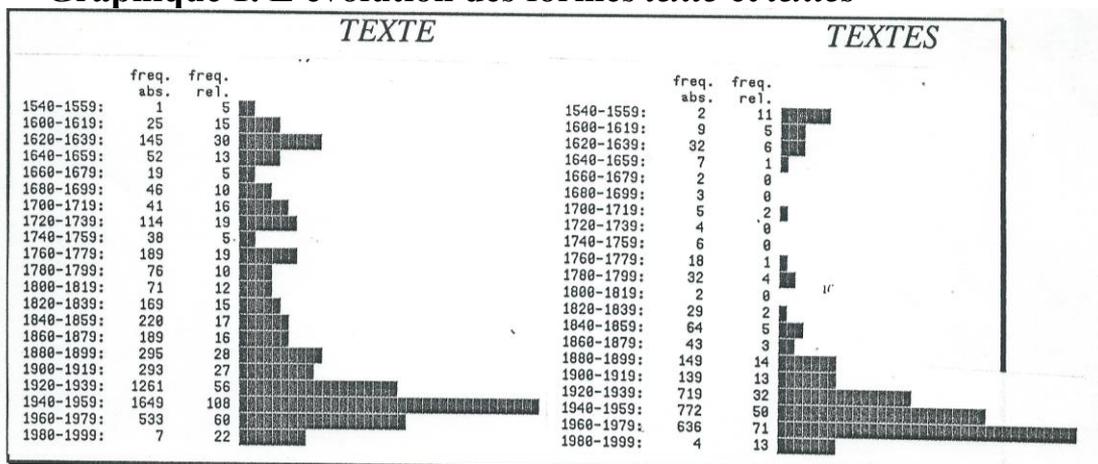
<sup>2</sup> Ce vaste projet - que coordonne Bernard Quemada - couvre l'ensemble des territoires et des disciplines. On a placé des guetteurs sur différents points du globe, pour relever les variations régionales et nationales, mais aussi aux endroits stratégiques de la technologie, là où naît la néologie.

attention et qui témoigne que la sève n'a pas déserté la racine: la présence de la forme *extratexte* (8 occ.). Mais nulle part nous n'avons rencontré dans la base le vocabulaire de la textologie, non plus que de la textique, où divers préfixes viennent croiser la trame du texte: pas le moindre *intertexte*, pas de *paratexte*, ni d'*hypertexte*. Le trousseau habituel des préfixes (*sur*, *sous*, *hypo*, *pseudo*, *trans*, *avant*, *anti*, *contre*, *poly*, *télé*) n'ouvre aucune porte, et ne fournit aucune occurrence<sup>1</sup>. Ainsi mesure-t-on l'écart inévitable entre l'actualité du discours et le reflet qu'en donnent les bases de données, même si les dépouillements qui se poursuivent à l'Institut National de la langue française tentent de combler la distance et de rendre compte de l'usage d'aujourd'hui. Mais le choix des textes contemporains qui sont incorporés à la base s'inspire comme dans le passé de critères surtout littéraires et laisse peu de chances aux jargons critiques où le mot *texte* multiplie sa descendance.

- I -

1 - Mais si la prolifération récente du mot échappe à nos données, peut-on y trouver la trace d'une **évolution**? Est-ce que le *texte* impose au cours des siècles sa présence massive, indépendamment des variétés récentes que les discours à la mode ont greffées sur le vieux tronc? La réponse est dans les deux graphiques ci-dessous qui rendent compte de la distribution à travers les siècles des deux formes *texte* et *textes*, le singulier (5433 occurrences) représentant le double du pluriel (2677 occ.). Ces deux histogrammes - qui sont issus directement de la nouvelle version de *Frantext* - ont un profil parallèle et croissant, qui ne laisse guère place au doute. Visiblement la fortune du mot est attachée au XX<sup>ème</sup> siècle.

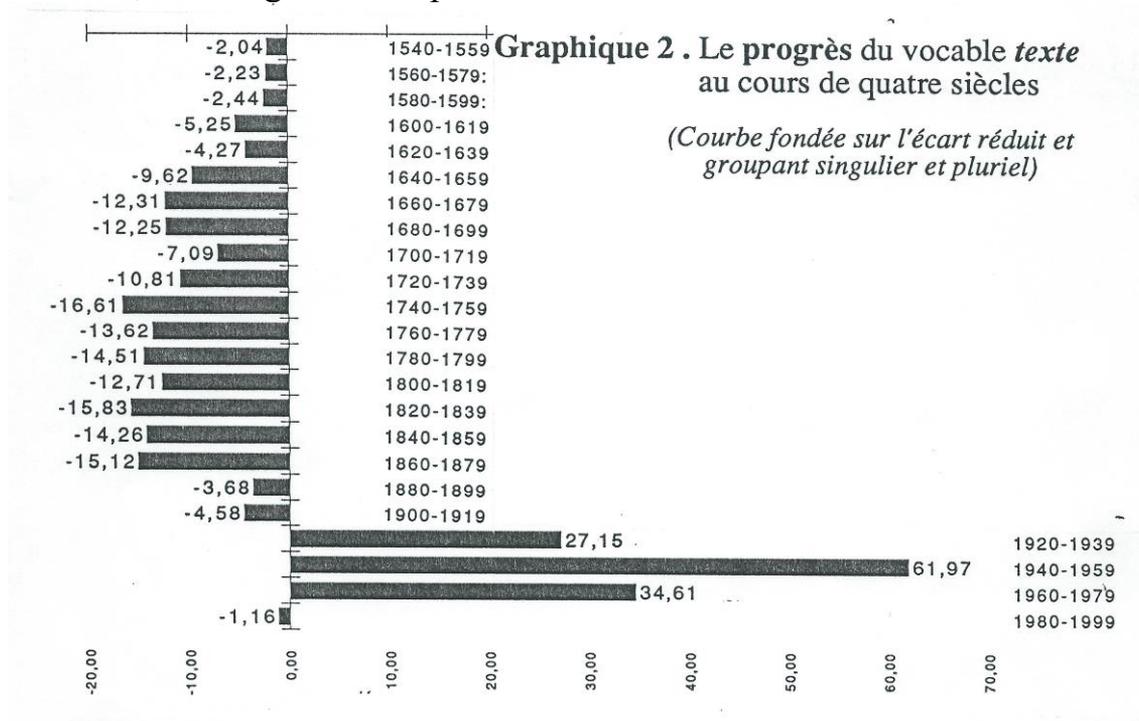
Graphique 1. L'évolution des formes *texte* et *textes*



<sup>1</sup> Même les composés non agglutinés font défaut et deux éléments seulement ont été rencontrés qui mettent en oeuvre le trait d'union: *petit-texte* (2 occ.) et *hors-texte* (28 occ.).

Le mot *texte*

On prendra garde que le calcul est fondé sur la fréquence relative. Si l'on applique aux données observées (colonne des fréquences absolues) le calcul plus sûr des écarts réduits, la progression soudaine du mot dans les dernières tranches apparaît plus nettement encore, comme en témoigne le graphique 2<sup>1</sup>. Au reste les effectifs absolus sont si manifestement inégaux que l'on pourrait presque se dispenser de calcul et de pondération. Quand les effectifs relevés sont de l'ordre de la centaine dans les premières tranches, ils atteignent et dépassent le millier dans les dernières.



<sup>1</sup> Pour calculer l'écart réduit, on a besoin de connaître l'étendue de chaque tranche ou de chaque sous-ensemble. Or cette donnée, qui n'est pas fournie directement par FRANTEXT, est restituée sans peine si l'on prend soin de diviser l'un par l'autre les deux nombres fournis par le programme. On sait en effet que la fréquence relative est le quotient de la fréquence absolue par l'étendue :

$$\text{Frel} = \frac{\text{Fabs}}{\text{N}} \quad \text{d'où} \quad \text{N} = \frac{\text{Fabs}}{\text{Frel}}$$

On veillera toutefois à restituer l'échelle convenable, la fréquence relative ayant été délivrée avec un facteur  $10^6$  pour plus de lisibilité. Pour une précision plus grande, il est souvent utile de se fonder sur un mot très fréquent (par exemple la préposition *de*). Lorsque l'étendue est connue pour l'ensemble et les sous-ensembles, on calcule la probabilité  $p$  (étendue du sous-ensemble / étendue de l'ensemble) et la fréquence théorique du mot dans le sous-ensemble (fréquence totale \*  $p$ ). La formule de l'écart réduit s'applique alors sans difficulté:

$$r = \frac{(\text{Fabs} - \text{Fthéo})}{\sqrt{\text{Fthéo} * (1-p)}}$$

2 - À quoi tient cette place croissante du *texte* dans les textes? Il est prudent de différer un peu l'explication. Voyons d'abord quelle est la distribution du mot dans les **genres** littéraires. Car les genres ne sont pas également représentés dans la chronologie et il faut se garder d'attribuer à l'une des causes possibles ce qui peut-être appartient à l'autre.

Il se trouve que les milliers de textes réunis dans la base *Frantext* ont été catalogués à plusieurs niveaux, le premier distinguant seulement la prose et les vers, le second classant les oeuvres par genre littéraire, et le troisième, réservé aux essais ou textes dits techniques, indiquant le domaine ou la discipline intéressée. Nous avons fait figurer dans le même tableau 3 et dans le même graphique (figure 4) les niveaux 2 et 3, le genre "essai" se trouvant détaillé dans ses éléments constituants. On a toutefois négligé les genres ou disciplines trop faiblement représentés. Et la pondération de l'écart réduit a été appliquée<sup>1</sup>.

**Tableau 3. Le mot *texte* dans des genres**

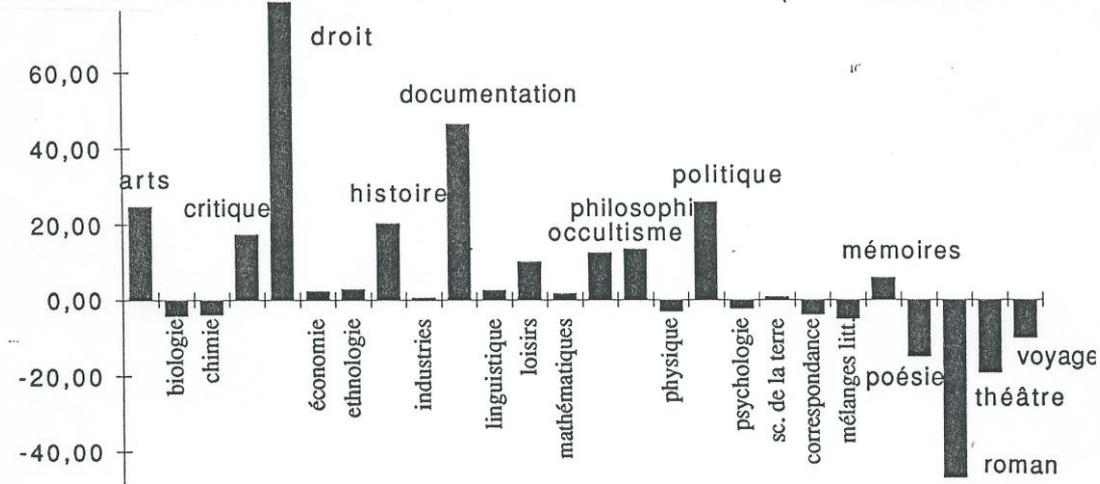
	nb oeuvres	étendue	effectif	théorique	réduit
<i>arts</i>	80	3116723	469	160	24,68
<i>biologie</i>	52	700542	9	36	-4,51
<i>chimie</i>	15	340581	0	17	-4,19
<i>critique</i>	67	3404337	401	175	17,30
<i>droit</i>	23	690852	502	35	78,52
<i>économie</i>	34	1728612	110	89	2,27
<i>ethnologie</i>	21	1179089	83	61	2,90
<i>histoire</i>	95	6071202	663	312	20,30
<i>industries</i>	30	1083993	59	56	0,45
<i>information</i>	31	1173817	419	60	46,39
<i>linguistique</i>	9	263608	23	14	2,58
<i>loisirs</i>	16	600514	88	31	10,32
<i>mathématiques</i>	15	240255	18	12	1,62
<i>occultisme</i>	12	396207	77	20	12,58
<i>philosophie</i>	131	8096085	684	416	13,52
<i>physique</i>	23	467714	8	24	-3,27
<i>politique</i>	33	1497369	303	77	25,92
<i>psychologie</i>	12	760765	23	39	2,57
<i>sc. de la terre</i>	54	1993352	111	102	0,86
<b><i>essai</i></b>	<b>1035</b>	<b>50208835</b>	<b>5239</b>	<b>2577</b>	<b>63,48</b>
<i>correspondance</i>	157	8502767	352	436	-4,16
<i>mélanges litt.</i>	111	5316524	186	273	-5,35
<i>mémoires</i>	241	16581985	1015	851	5,94
<i>poésie</i>	279	7136477	83	366	-15,15
<i>roman</i>	873	54163610	764	2780	-47,17
<i>théâtre</i>	418	8963391	55	460	-19,44
<i> récits voyage</i>	55	3516123	43	180	-10,35

<sup>1</sup> *Frantext*, même dans sa nouvelle version, ne rend pas l'approche du genre aussi aisée ni aussi rapide que celle de la chronologie et celle des auteurs. Pour constituer l'effectif d'un mot ou d'une série de mots dans un genre, une interrogation spécifique est nécessaire. Cette opération a dû être renouvelée pour chacune des lignes du tableau 3, au lieu qu'un seul passage suffit à dresser la courbe des époques ou des auteurs.

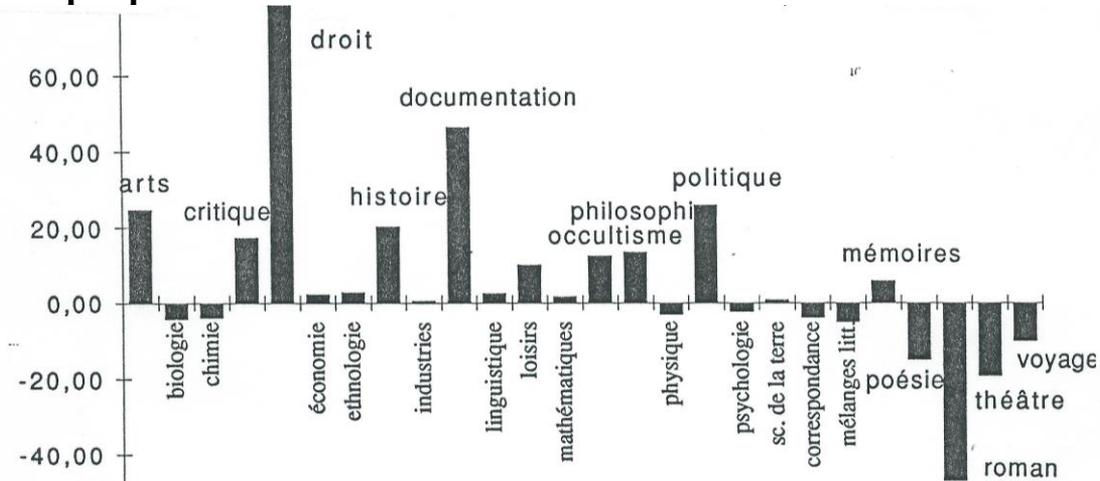
## Le mot *texte*

Le fait le plus frappant, quoique le moins inattendu, est l'abondance du mot *texte* dans les oeuvres techniques. Les quelque 1000 titres réunis sous la rubrique "essai" produisent 5239 occurrences du mot, soit deux fois plus que l'espérance mathématique: 2577. Le test de l'écart réduit ( $z = 63$ ) invite à conclure formellement à une liaison fonctionnelle entre le mot et le genre. Mais ce genre n'est pas homogène et là encore il convient de distinguer entre les sciences dures et les activités littéraires. Les disciplines qui reposent sur les faits et les observations ont assez peu recours au témoignage des textes, qu'il s'agisse de sciences exactes comme la physique, la chimie, la biologie ou les sciences de la terre (qui se situent dans la zone négative du graphique 4), ou de sciences humaines comme l'économie et l'industrie, l'ethnologie, les loisirs, la psychologie et la linguistique (qui se maintiennent dans la zone médiane de l'indifférence). Mais quand l'expérimentation fait défaut, le texte tend à devenir l'objet du débat et la matière même de la discipline. Cela est vrai dans les arts et les lettres (esthétique, philosophie, religion, critique littéraire, histoire), et plus encore en matière documentaire, politique ou juridique. Au haut de l'histogramme 4 culmine le droit, qui en France a une tradition romaine et écrite et s'inscrit dans des codes, des décrets, des lois, c'est-à-dire des textes.

Un affaissement profond se lit au contraire dans la partie droite de l'histogramme, qui est consacrée aux textes véritablement littéraires et qu'on peut répartir en deux groupes, selon qu'il s'agit ou non de fiction. Les genres littéraires qui reposent sur le réel (correspondance, récits de voyage, mémoires) s'éloignent moins du mot *texte* que les oeuvres fictives (poésie, théâtre, roman). Si le déficit est ici généralisé, les mémoires y échappent pourtant parce que leurs auteurs sont souvent des généraux ou des hommes politiques soucieux d'établir, par le témoignage non douteux des textes et des documents, la vérité historique par quoi se trouve justifiée leur action. La fiction au contraire écarte tout ce qui peut nuire à la crédibilité. Car les mots par quoi l'on caractérise le projet littéraire sont précisément absents du texte littéraire, où l'on rencontrera rarement les termes de la critique, comme *style*, *écriture*, *littérature*, *roman* ou *texte*. On a déjà observé que les oeuvres d'imagination parlent à l'imagination, mais n'en parlent pas et ne citent pas le terme.

**Graphique 4. Courbe d'emploi du mot *texte* à travers les disciplines**

La lumière éclatante de ce dernier résultat n'est pas sans porter de l'ombre au précédent. Le dosage des textes techniques est très divers dans le corpus et très favorable aux tranches les plus proches de notre époque. Les excédents constatés dans celles-ci pour le mot *texte* ne viennent-ils pas alors de la part prépondérante qu'y jouent les essais? L'étude chronologique doit donc être reprise en neutralisant le facteur indésirable. Il suffit de choisir un corpus purement littéraire d'où les essais sont exclus. La souplesse du logiciel d'interrogation permet l'élargissement du champ ou sa focalisation, et toutes sortes de sélections et de mises au point. Le graphique ci-dessous représente deux réductions successives: la première ne s'intéresse qu'à la littérature, la seconde rétrécit encore l'espace et n'envisage que les oeuvres de fiction. Or l'orientation de la courbe reste croissante, même si son amplitude diminue. Le matériau plus pur qu'on vient d'isoler livre donc les mêmes enseignements que la vue d'ensemble et le progrès du mot *texte* s'en trouve confirmé.

**Graphique 5. L'évolution du mot *texte* dans la littérature et la fiction**

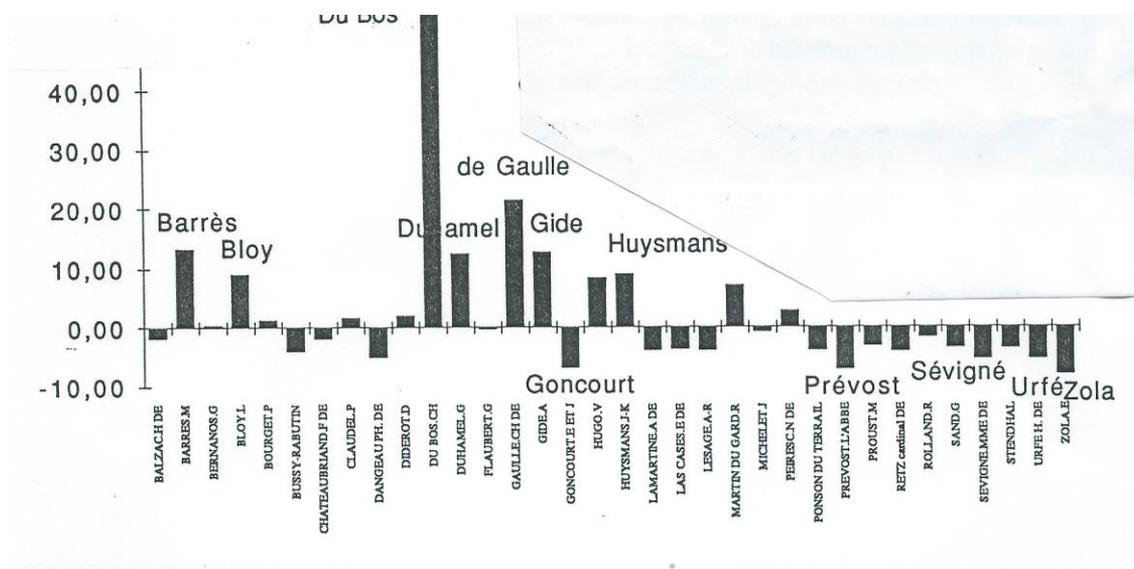
Le mot *texte*Tableau 6. Le mot *texte* chez les écrivains

AUTEUR	réel	étendue théorique	réduit	
DU BOS.CH	253	911031	17,06	57,34
GAULLE.CH DE	122	1165181	21,81	21,55
BARRES.M	118	2001574	37,47	13,26
GIDE.A	112	1935966	36,24	12,68
DUHAMEL.G	66	865264	16,20	12,42
HUYSMANS.J-K	50	805633	15,08	9,02
BLOY.L	62	1121512	21,00	8,99
HUGO.V	105	2548735	47,72	8,38
MARTIN DU GARD.R	50	1017283	19,05	7,12
PEIRESC.N DE	47	1694026	31,71	2,73
DIDEROT.D	23	818328	15,32	1,97
CLAUDEL.P	26	1014617	19,00	1,61
BOURGET.P	24	1004781	18,81	1,20
BERNANOS.G	17	835041	15,63	0,35
FLAUBERT.G	43	2439744	45,68	-0,40
MICHELET.J	21	1354102	25,35	-0,87
ROLLAND.R	9	862667	16,15	-1,79
BALZAC.H DE	39	2784547	52,13	-1,84
CHATEAUBRIAND.F DE	35	2595990	48,60	-1,97
PROUST.M	20	2187157	40,95	-3,30
SAND.G	8	1411115	26,42	-3,60
STENDHAL	13	1856400	34,75	-3,72
LAS CASES.E DE	5	1223060	22,90	-3,76
LESAGE.A-R	5	1294897	24,24	-3,93
PONSON DU TERRAIL	0	822111	15,39	-3,94
LAMARTINE.A DE	11	1801495	33,73	-3,94
BUSSY-RABUTIN	0	863011	16,16	-4,03
RETZ cardinal DE	0	915694	17,14	-4,16
DANGEAU PH. DE	0	1383796	25,91	-5,12
URFE H. DE	0	1566683	29,33	-5,45
SEVIGNE.MME DE	5	2078153	38,91	-5,48
GONCOURT.E ET J	22	4473711	83,75	-6,87
PREVOST.L'ABBE	7	3401694	63,68	-7,20
ZOLA.E	8	4187845	78,40	-8,09

3 - Les distinctions que l'on opère parmi les genres ou parmi les siècles et que nous venons d'explorer paraissent pourtant devoir s'effacer derrière celles qui règnent parmi les **auteurs**. Car jamais un genre littéraire ou une époque n'ont produit par eux-mêmes une oeuvre. C'est l'écrivain qui en dernier ressort appose sa signature et le signe exclusif de l'appropriation et de l'identification. Or *Frantext* permet aisément l'étude séparative des auteurs, quel que soit leur nombre. Sans dresser ici leur liste complète, nous nous contenterons de citer ceux qui sont le mieux représentés et dont l'oeuvre soumise au dépouillement dépasse 800 000 occurrences. Ici le logiciel d'interrogation délivre une série très longue et pourtant incomplète. Il cite tous les écrivains sous la plume desquels le mot étudié a été rencontré, ne serait-ce qu'une fois, mais il passe sous silence ceux qui précisément ont passé le mot sous silence. Or la fréquence 0 peut être très révélatrice si la taille du sous-corpus correspondant est élevée. C'est pourquoi nous avons repêchés et mentionnés dans le tableau 6 les écrivains qui ont ignoré le mot *texte*, dès lors que cette absence est largement significative. Il s'agit surtout des écrivains des siècles passés,

comme Ponson du Terrail, Dangeau, Bussy-Rabutin, le cardinal de Retz et H. d'Urfé. Presque tous les écrivains que le graphique 6 situe dans la zone négative sont antérieurs au XXe siècle (Balzac, Chateaubriand, Goncourt, Lamartine, Lesage, abbé Prévost, G. Sand, Mme de Sévigné, Stendhal et Zola). Ceux que le graphique 7 désigne à l'attention dans la zone positive appartiennent au contraire à l'époque contemporaine. Mais la prédilection pour le *texte* est surtout le fait des critiques ou mémorialistes, de ceux qui se confient à leur *Cahiers*, comme Barrès, ou à leur *Journal* comme Du Bos, Gide, Bloy. On retrouve ici l'influence du genre, la réflexion historique (*Mémoires* de de Gaulle), religieuse (Huysmans), spiritualiste (Duhamel) ou plus largement littéraire s'accompagnant de références multipliées au *texte*.

Figure 7. Courbe selon les écrivains



- II -

1 - On vient d'explorer la base et ses grandes divisions en isolant le mot *texte* sans vouloir considérer aucun autre terme. Mais les mots prennent leur sens à partir de leur environnement dans la chaîne du discours. Est-il possible de tenir compte de l'entourage lexical du mot qu'on retient pour pôle? Sans aller jusqu'à exécuter l'analyse syntaxique des documents, *Frantext* offre pourtant un nouveau moyen, purement documentaire, d'étudier les associations de mots. Il ne s'agit pas ici de recherche de cooccurrences ou d'expressions, fonction disponible depuis longtemps. Cette fonction essentielle rend de grands services mais elle réclame l'initiative et les présupposés du chercheur, lequel doit préciser de son propre chef les séquences qu'il souhaite retrouver - ce qui est

Le mot *texte*

nécessairement long, subjectif et incomplet. Or la commande nouvellement implantée dans *Frantext* échappe à ces limites puisque pour un mot donné (ou une liste de mots) elle repère les formes qui lui sont le plus souvent associées dans la chaîne du discours, après qu'on a fixé les limites souhaitées pour l'exploration de cette chaîne (c'est généralement le cadre naturel de la phrase). Le résultat est une liste de formes, triées selon le nombre des cooccurrences observées, et reproduites à l'état brut dans l'extrait ci-dessous:

**Tableau 8. L'environnement lexical des formes *texte* et *textes***

Nombre d'occurrences du pivot : 8110		
Taille des voisinages explorés : 340729 occurrences		
Nombre de graphies trouvées dans ces voisinages : 24827		
5947 <i>texte</i>	159 <i>part</i>	127 <i>aucun</i>
2891 <i>textes</i>	156 <i>auteur</i>	127 <i>lettre</i>
328 <i>sens</i>	153 <i>jour</i>	126 <i>question</i>
257 <i>assemblée</i>	152 <i>histoire</i>	125 <i>monde</i>
243 <i>conseil</i>	151 <i>ensemble</i>	125 <i>projet</i>
227 <i>*dieu</i>	150 <i>général</i>	125 <i>république</i>
221 <i>temps</i>	148 <i>donner</i>	125 <i>voir</i>
217 <i>faut</i>	148 <i>mots</i>	124 <i>commission</i>
215 <i>livre</i>	147 <i>première</i>	124 <i>critique</i>
198 <i>saint</i>	146 <i>article</i>	124 <i>pris</i>
192 <i>état</i>	146 <i>certains</i>	123 <i>oeuvre</i>
183 <i>cas</i>	142 <i>écrit</i>	122 <i>pensée</i>
183 <i>doit</i>	141 <i>gouvernement</i>	122 <i>raison</i>
176 <i>forme</i>	141 <i>grand</i>	120 <i>doute</i>
175 <i>lieu</i>	138 <i>nationale</i>	120 <i>langue</i>
174 <i>vie</i>	132 <i>constitution</i>	117 <i>partie</i>
173 <i>homme</i>	132 <i>effet</i>	113 <i>lequel</i>
169 <i>droit</i>	131 <i>exemple</i>	113 <i>ordre</i>
169 <i>esprit</i>	131 <i>lire</i>	113 <i>selon</i>
168 <i>mot</i>	129 <i>abord</i>	112 <i>siècle</i>
168 <i>premier</i>	129 <i>trouve</i>	112 <i>sujet</i>
164 <i>français</i>	128 <i>seul</i>	111 <i>donné</i>

Il faut se garder de commenter trop précipitamment une telle liste, même si l'on y découvre une lisibilité immédiate. Que les mots *livre* ou *sens*, *assemblée* ou *conseil* soient souvent associés au *texte*, est hautement attendu et la statistique donne la confirmation à l'intuition linguistique. Mais le trouble surgit très tôt dans l'esprit lorsqu'apparaissent des compagnons inattendus comme le *temps* ou l'*homme*, dont on ne voit pas a priori la liaison avec le *texte*. Or précisément cette liaison n'existe pas d'un point de vue probabiliste. C'est la très grande fréquence de ces formes qui les pousse mécaniquement dans l'entourage du *texte*, sans avoir avec ce mot la moindre attache particulière. Ici comme ailleurs on doit se méfier des chiffres absolus. Une pondération est nécessaire qui prenne en compte

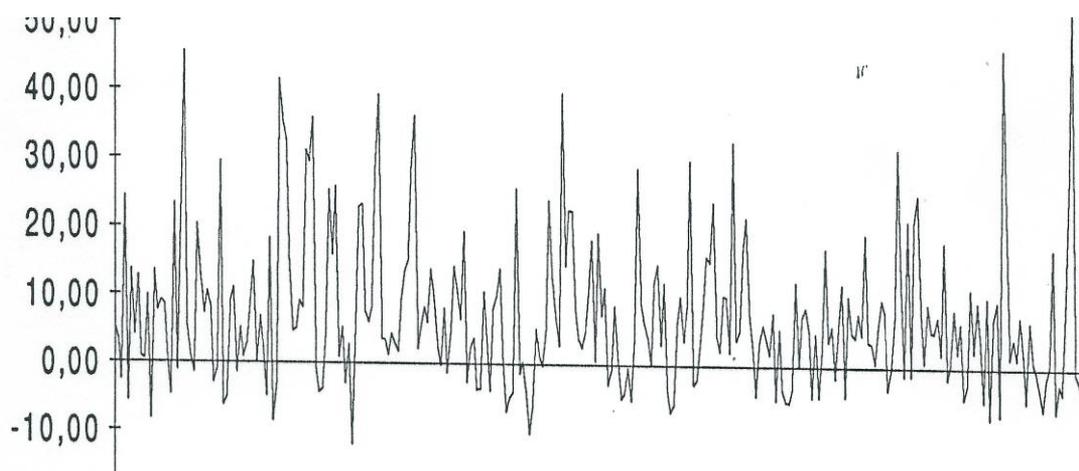
la fréquence des mots<sup>1</sup>. Pour réaliser cette pondération, il suffit de reprendre la liste brute en demandant la fréquence absolue de ses éléments indépendamment de toute contrainte - ce qu'une commande spécifique de *Frantext* exécute parfaitement. Reste alors à comparer pour chaque mot les deux valeurs et à calculer l'écart réduit.

On voit alors que certains des mots de la liste, même placés en tête, sont loin d'être attirés par le mot pôle (ici le *texte*), puisque les cooccurrences observées sont déficitaires. Ainsi compte tenu de la fréquence du mot *homme* dans le corpus (171619), de la taille globale des phrases examinées qui contiennent les mots *texte* ou *textes* (340729) et de l'étendue du corpus (158 millions de mots), le nombre théorique des cooccurrences s'élève à:

$$171619 * (340729 / 158000000) = 370$$

Or l'effectif relevé est de 173. Il s'agit donc d'un déficit que l'écart réduit ( $z = -10,26$ ) souligne avec force. Ce cas, qui n'est pas isolé, constitue une série dont les éléments s'agglomèrent sur le pôle négatif de l'aimant: *coeur, amour, tête, grand, monde, yeux, voir, hommes, bon, vie, mort, petit, main, temps, peine, choses, soir, corps, heure, etc.* On en a pris la mesure d'ensemble dans le graphique ci-dessous qui traite les premiers mots du tableau 8 et établit pour chacun le calcul de l'écart réduit.

**Figure 9. Excédents et déficits des cooccurents de *texte***



<sup>1</sup> La commande de *Frantext* écarte d'ailleurs les mots-outils de haute fréquence, pour lesquels les résultats sont présumés inintéressants.

Le mot *texte*Tableau 10. Liste triée de l'environnement privilégié du *texte*

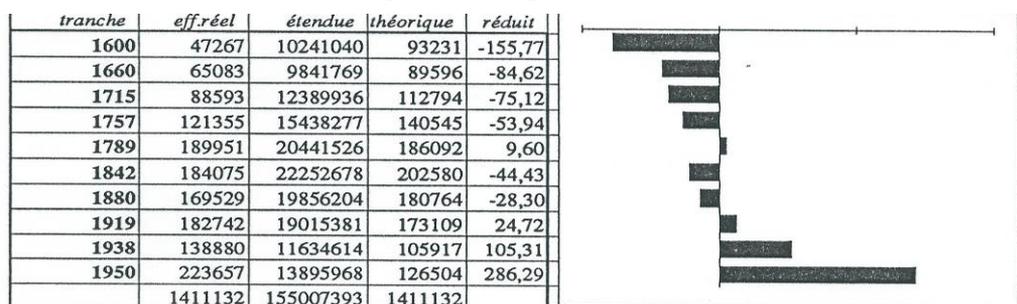
mot	réel	fréq.	théo.	réduit	mot	réel	fréq.	théo.	réduit
<i>voité</i>	54	445	0,96	54,20	<i>président</i>	82	10842	23,38	12,14
<i>traduction</i>	111	2384	5,14	46,74	<i>ministre</i>	89	12330	26,59	12,12
<i>assemblée</i>	257	11888	25,64	45,74	<i>soumis</i>	51	5321	11,47	11,68
<i>commentaire</i>	61	933	2,01	41,63	<i>lois</i>	101	16005	34,51	11,33
<i>interprétation</i>	101	2665	5,75	39,78	<i>règles</i>	52	5820	12,55	11,15
<i>documents</i>	105	2914	6,28	39,42	<i>examen</i>	53	6058	13,06	11,06
<i>édition</i>	88	2421	5,22	36,27	<i>cas</i>	183	38303	82,60	11,06
<i>constitutionnel</i>	53	933	2,01	35,98	<i>forme</i>	176	37502	80,87	10,59
<i>commentaires</i>	54	999	2,15	35,36	<i>avis</i>	81	12329	26,59	10,56
<i>original</i>	84	2646	5,71	32,81	<i>principe</i>	101	17194	37,08	10,51
<i>commission</i>	124	5547	11,96	32,43	<i>termes</i>	75	11111	23,96	10,44
<i>rédaction</i>	54	1213	2,62	31,80	<i>oeuvre</i>	123	23099	49,81	10,38
<i>conseil</i>	243	19452	41,95	31,08	<i>mots</i>	148	30220	65,17	10,27
<i>nationale</i>	138	7676	16,55	29,88	<i>lignes</i>	63	8785	18,94	10,13
<i>constitution</i>	132	7272	15,68	29,40	<i>oeuvres</i>	63	8785	18,94	10,13
<i>bible</i>	64	1929	4,16	29,37	<i>générale</i>	88	14687	31,67	10,02
<i>manuscrits</i>	54	1447	3,12	28,83	<i>question</i>	126	24635	53,13	10,01
<i>document</i>	55	1504	3,24	28,77	<i>agit</i>	99	17659	38,08	9,88
<i>écriture</i>	105	4991	10,76	28,76	<i>titre</i>	78	12523	27,01	9,82
<i>vote</i>	56	1736	3,74	27,04	<i>suivant</i>	105	19488	42,03	9,72
<i>décret</i>	79	3552	7,66	25,80	<i>volume</i>	53	7146	15,41	9,59
<i>grec</i>	84	3987	8,60	25,74	<i>droit</i>	169	38282	82,56	9,52
<i>critique</i>	124	8186	17,65	25,34	<i>nombreux</i>	54	7506	16,19	9,41
<i>sacrés</i>	58	2099	4,53	25,16	<i>scène</i>	73	11837	25,53	9,41
<i>traduire</i>	52	1777	3,83	24,63	<i>conditions</i>	86	15227	32,84	9,29
<i>*guide</i>	59	2301	4,96	24,28	<i>caractère</i>	110	21645	46,68	9,28
<i>imprimé</i>	53	1922	4,14	24,02	<i>anciens</i>	81	14002	30,20	9,26
<i>notes</i>	84	4486	9,67	23,92	<i>ensemble</i>	151	33602	72,46	9,24
<i>dispositions</i>	95	5802	12,51	23,35	<i>matière</i>	80	14371	30,99	8,81
<i>application</i>	93	5621	12,12	23,26	<i>points</i>	64	10426	22,48	8,76
<i>discussion</i>	75	3912	8,44	22,94	<i>lu</i>	56	8569	18,48	8,74
<i>article</i>	146	12572	27,11	22,86	<i>musique</i>	78	14024	30,24	8,69
<i>latin</i>	77	4165	8,98	22,72	<i>siècle</i>	112	23534	50,75	8,61
<i>lecture</i>	109	7819	16,86	22,46	<i>anglais</i>	69	11936	25,74	8,54
<i>pages</i>	95	6458	13,93	21,75	<i>église</i>	98	19769	42,63	8,49
<i>république</i>	125	10499	22,64	21,53	<i>base</i>	54	8521	18,38	8,32
<i>sacré</i>	88	5739	12,38	21,52	<i>doctrine</i>	51	7874	16,98	8,26
<i>auteur</i>	156	16301	35,15	20,40	<i>questions</i>	57	9327	20,11	8,23
<i>proposé</i>	54	2805	6,05	19,52	<i>connaissance</i>	73	13417	28,93	8,20
<i>explication</i>	78	5423	11,69	19,41	<i>général</i>	150	36127	77,91	8,18
<i>livre</i>	215	28929	62,39	19,34	<i>époque</i>	79	15084	32,53	8,16
<i>sens</i>	328	57327	123,63	18,40	<i>certaines</i>	84	16477	35,53	8,14
<i>citer</i>	51	2793	6,02	18,35	<i>passage</i>	69	12731	27,45	7,94
<i>lire</i>	131	14026	30,25	18,34	<i>problème</i>	56	9430	20,34	7,92
<i>vigueur</i>	59	3887	8,38	17,50	<i>lequel</i>	113	25535	55,07	7,82
<i>précis</i>	56	3560	7,68	17,46	<i>ancien</i>	65	11992	25,86	7,70
<i>page</i>	76	6222	13,42	17,10	<i>unique</i>	64	11763	25,37	7,68
<i>notamment</i>	60	4528	9,76	16,09	<i>saint</i>	198	53930	116,30	7,58
<i>déclaration</i>	51	3433	7,40	16,04	<i>selon</i>	113	26161	56,42	7,54
<i>note</i>	61	5023	10,83	15,26	<i>papier</i>	61	11156	24,06	7,54
<i>écrits</i>	56	4369	9,42	15,19	<i>divers</i>	69	13364	28,82	7,49
<i>composition</i>	65	5687	12,26	15,08	<i>plan</i>	75	15151	32,67	7,41
<i>certaines</i>	146	21123	45,55	14,90	<i>autorité</i>	71	14102	30,41	7,37
<i>ministres</i>	70	6559	14,14	14,87	<i>rapport</i>	88	18982	40,93	7,36
<i>langue</i>	120	15996	34,50	14,57	<i>livres</i>	98	22117	47,70	7,29
<i>étude</i>	108	13852	29,87	14,31	<i>théâtre</i>	62	11775	25,39	7,27
<i>gouvernement</i>	141	21333	46,00	14,02	<i>chapitre</i>	62	12185	26,28	6,98
<i>enseignement</i>	74	7812	16,85	13,94	<i>intelligence</i>	69	14185	30,59	6,95
<i>*paul</i>	73	7846	16,92	13,65	<i>valeur</i>	85	18898	40,75	6,94
<i>écrit</i>	142	22158	47,78	13,64	<i>exemple</i>	131	34122	73,58	6,70
<i>analyse</i>	66	6701	14,45	13,58	<i>simplement</i>	57	11189	24,13	6,70
<i>auteurs</i>	66	7160	15,44	12,88	<i>présente</i>	62	12833	27,67	6,53
<i>accord</i>	77	9230	19,90	12,81	<i>mot</i>	168	47425	102,27	6,51
<i>in</i>	52	4950	10,67	12,66	<i>membres</i>	54	10897	23,50	6,30
<i>phrase</i>	66	7449	16,06	12,47	<i>éléments</i>	54	11049	23,83	6,19
<i>mise</i>	80	10382	22,39	12,19					

Le tableau 10 ne retient que les mots statistiquement associés au *texte*, rangés par signification décroissante. L'aspect juridique et formel est ce qui frappe le plus. Le *texte* apparaît d'abord comme la manifestation de la *loi*. Il indique le *droit*, le *principe*, les *règles*. Soumis au *vote* à l'*assemblée nationale* dans le cadre de la *constitution* et de la *république*, il entre en vi-

gueur sous l'autorité du président, du gouvernement et des ministres qui fixent les conditions d'application des articles, des dispositions, des décrets. Le texte revêt aussi un caractère religieux, marqué par les adjectifs saint, sacré, sacrés et les substantifs église, doctrine, bible. Juridique ou religieux, le sens et la valeur du texte sont l'objet d'examen ou d'exégèse. Les documents sont sujets à étude, à analyse, à discussion, à explication, à interprétation, à commentaire. On propose des cas, des exemples, des questions, des problèmes et l'on obtient des avis, des rapports, des notes, des accords. Enfin le texte est le lieu de l'édition et de la littérature. Puisant sa source dans l'Antiquité (ancien, anciens, grec, latin, traduction, traduire), le mot *texte* accompagne toutes les étapes qui conduisent au livre, de l'écriture (écrits, écrit, manuscrits, auteurs, auteur), à la composition (volume, livre, livres, imprimé, scène, chapitre, pages, page, papier, phrase, lignes, mot, mots, orme, termes, langue) et à la lecture (lire, lu, citer, critique). En épuisant le tableau 10, on pourrait fouiller plus profondément encore le champ sémantique dont le *texte* fait l'unité. On n'insistera que sur un point: les éléments rencontrés dans le champ sont précisément localisés, nettoyés, classés. On dirait un champ de fouilles, où tout respire l'exactitude et la propreté et où chaque objet est précis, original, unique.

2 - Mais ce champ thématique, dont l'unité apparaît si forte dans le terrain clos de l'expérience, en présence du mot *texte*, maintient-il encore sa cohérence lorsque cesse la contrainte et que le mot fédérateur a disparu? A-t-on affaire à un magnétisme provisoire qu'un autre aimant peut abolir? Voyons par exemple si l'évolution chronologique du champ tout entier est parallèle ou non au mouvement croissant du mot *texte*. Cette fois nous avons choisi des tranches plus larges, dont les jalons sont mieux accommodés aux tournants perceptibles de l'histoire. Effectifs et calculs apparaissent dans la figure 11. Près d'un million et demi d'occurrences sont ici rassemblées et distribuées selon l'axe chronologique dont la direction est aisée à reconnaître. L'évolution de l'ensemble suit de très près celle du mot *texte* que le graphique 2 avait mise en relief.

Figure 11. L'évolution du champ sémantique établi autour du mot *texte*





Le tableau, trop volumineux, ne sera pas reproduit ici, non plus que les modifications auxquelles il a donné matière. Car les tranches étant d'étendue inégale, il a bien fallu pondérer et réduire à la même échelle les observations. Seul sera montré dans la figure 12 le résultat de l'analyse factorielle appliquée au tableau.

Précisons à ceux des lecteurs qui demeurent peu familiers de ces procédures mathématiques que les mots (les lignes du tableau) et les tranches (les colonnes) sont représentés sur un plan de telle sorte que les alliances et les conflits soient respectés qui s'établissent entre les mots, ou entre les périodes, ou entre les mots et les périodes. Le dessin des tranches est le plus facile à suivre puisqu'il décrit un arc de cercle caractéristique des données sérielles, passant de la droite du graphique (17<sup>e</sup> et 18<sup>e</sup> siècle), au haut de la figure (19<sup>e</sup> siècle) avant de s'étendre sur la gauche (20<sup>e</sup> siècle). Or que voit-on à sur la marge droite dans l'entourage des tranches 1600, 1660, 1715, 1757 et 1789? Des *Anciens*, du *grec*, du *latin* et une *écriture sainte* (*sacré, sacrés, bible, église*). Le champ se modernise et se laïcise, au haut de la figure, après la Révolution, et les aspects techniques de l'édition y prennent le pas. Enfin la politique et le droit campent dans la moitié droite réservée au 20<sup>e</sup> siècle. Voilà dessinée à grands traits la dynamique du champ dont le tableau 10 ne montrait que la structure statique. Certes il n'y a pas là de découverte majeure et l'histoire des mentalités que la statistique devine à travers celle des mots est bien celle dont on a profondément - mais vaguement - conscience. Mais on a gagné un peu de sûreté et un peu de précision et l'on tient en mains ce qu'on envie tant aux sciences exactes: un commencement de preuve.

### - III -

Resterait à étudier les contextes du *texte*, à voir les choses "dans le texte", pour reprendre une expression souvent rencontrée dans nos textes. Car ce qui précède n'épuise pas la combinatoire qui gouverne l'environnement lexical du *texte*. À aucun moment les mots qui entourent le pôle n'ont été considérés **dans l'ordre**, c'est-à-dire dans la séquence linéaire et syntaxique qui donne un sens à la chaîne du discours. Les mots relevés par la statistique avaient seulement de statut de cooccurrents. Seule la proximité était prise en compte, et non pas les liens organiques que la syntaxe et la sémantique établissent entre les éléments du discours.

1 - Un pas reste à faire que *Frantext* permet de franchir, avec la commande très classique de "*recherche dans les textes*". La seule difficulté vient de la masse considérable des contextes à explorer. En se contentant

Le mot *texte*

d'un contexte minimum (une seule phrase), il faudra pourtant du temps et de l'argent pour capter et enregistrer les 8110 contextes que la base fournit pour les formes *texte* et *textes*. Certes le temps de la recherche est très court, mais celui du transfert des résultats (même à 2400 bauds) risque d'être rédhibitoire<sup>1</sup>. Aussi avons-nous réduit le volume des données et celui des demandes. D'une part on s'est contenté de la forme du singulier et d'autre part on a réduit le corpus aux seuls textes littéraires, en excluant les essais techniques - qui sont fort accueillants au mot *texte*. Encore fallait-il quelque moyen de traiter cet amas d'information, les quelque 1788 contextes récoltés par la voie télématique. Notre logiciel *Hyperbase* a été mis à profit pour ce corpus d'un type particulier dont toutes les phrases contiennent le mot *texte*. Le principal traitement a consisté à trier les contextes à partir de l'expansion droite du mot-pôle, puis à partir de l'expansion gauche. Dans le premier cas on peut prendre la mesure de certaines constructions répétitives où le mot *texte* est suivi d'une détermination. Dans le second on peut se livrer à l'étude des déterminants qui précèdent le *texte*.

Considérons d'abord la construction *texte* + adjectif (et la construction inverse, qui est plus rare, adjectif + *texte*). Le qualificatif sert alors à préciser la langue utilisée, l'antiquité venant en tête (*latin* 19 occ., *grec* 15) devant *l'anglais* 11, *le français* 7, *l'allemand* 5 (*hébreu*, *hébraïque*, *arabique*, *biblique* et *oriental* ont 2 occurrences). L'adjectif sert parfois aussi à rehausser l'éclat du texte qui est souvent dit *admirable*, *auguste*, *superbe*, *magnifique*, *capital* et *inépuisable*. Jamais le texte n'apparaît pris en mauvaise part et si 16 occurrences de *petit* répondent aux 7 de *grand*, il ne s'agit pas d'une antinomie car un *petit texte* est un texte court, non un texte méprisable, alors qu'un *grand texte* est un texte de valeur, sans être nécessairement long. Mais la vertu la plus prisée dans un texte est son authenticité et son intégralité. Les adjectifs soulignent à l'envi cette précieuse qualité: *original* 10, *exact* 9, *définitif* 8, *intégral* 5, *complet* 4, *entier* 3, *authentique* 4, *officiel* 4, *formel* 3. L'adverbe *même* (27 emplois) (ou *lui-même* 4) participe souvent à l'expression de la garantie, qui se réclame parfois de l'autorité divine (*sacré* 13 occ., *saint* 4). Au vu des adjectifs qui l'entourent le texte semble empreint de quelque solennité religieuse. Il paraît garder le souvenir de ses premiers emplois, dans les temps mythiques où l'on gravait les commandements de Dieu dans la pierre.

La construction *texte* + *de* est fort répandue, soit pour introduire un nom propre - en ce cas il s'agit de l'amorce d'une citation -, soit pour

---

<sup>1</sup> On ne peut se contenter des résultats sur papier qui sont disponibles sur l'imprimante locale car ce qui est délivré ici en sortie doit entrer de nouveau comme donnée pour les traitements ultérieurs.

préciser la référence du propos. Les auteurs cités sont en majorité des évangélistes ou des pères de l'Église (*Jean, Luc, Mathieu, Marc, Paul, Thomas* ont 3 occurrences chacun), des Anciens (*Hérodote 2, Tacite 2*) ou des classiques (*Pascal 6, Racine 2, Fénelon 4, Corneille 2, Richelieu 2*). Il n'y a guère qu'un auteur moderne à être cité plus d'une fois: Gide. Quant aux références, on voit que tout ne peut pas être porteur de message et contenir un texte. La nature, l'univers, l'animal peuvent receler un mystère ou quelque vérité à déchiffrer. On ne parle pas de *texte* en de tels cas, même de façon métaphorique. Le mot est réservé à l'expression consciente, publique, écrite et matérielle d'une pensée humaine. Mais les manifestations en sont multiples: religieuses (*Évangile 4 occ., Vulgate 4, méditation 2*), éditoriales (*livre 5, édition 4, volume 2, auteur 4, note 3*), communicationnelles (*télégramme 7, lettre 8, communication 4, réponse 5, message 5, journal 5, document 4, écran 2*) et surtout juridiques et politiques: on a recours au *texte* dès que quelque convention régit les rapports humains. C'est le *texte* en effet qui cimente la société et établit les relations codifiées entre les hommes et entre les peuples (*accord 7, traité 4, convention 2, déclaration 4, proclamation 2*). Et c'est à travers le *texte* que prennent forme la *loi* (13 occurrences), les *décrets* (4), le *code* (4), les *procès* (3) et les *jugements* (6). Si quelque accident fatal survenait à la civilisation et qu'un explorateur extérieur découvrit notre planète, c'est dans les textes gravés et retrouvés qu'il chercherait la société disparue et dans les textes le mot *texte* lui livrerait une clé précieuse, une sorte de miroir réflexif où la pensée humaine jette sur elle-même son regard.

2 - On peut prêter attention aussi au contexte qui précède. Au voisinage immédiat d'un substantif, la place est quasiment réservée aux déterminants. On a trouvé quelques rares adjectifs, dont *petit* et *grand* déjà signalés. Plus en avant on rencontre généralement un verbe transitif ou quelque substantif dérivé suivi d'un génitif objectif, sur le modèle *publier un texte, publication d'un texte*. Les observations relevées ici ne paraissent guère dignes d'intérêt. Mais nous nous attacherons précisément aux déterminants qui précèdent le mot *texte*. Car la poussière des mots grammaticaux est trop souvent balayée d'un revers de manche dans les études lexicométriques, alors qu'elle contribue grandement à la spécificité de l'atmosphère qu'on respire dans une oeuvre. Ainsi parmi les cooccurrents du mot *texte*, on aurait dû faire mention des prépositions *suivant* et *selon* que le calcul distingue et qui participent à l'expression de l'intertextualité et de l'interprétation. Peut-on de même tirer quelque enseignement de l'étude des proclitiques qui accompagnent le mot *texte*? Procédons d'abord au relevé, lequel fournit 215 occurrences de l'article indéfini, et 1079 de l'article défini (en cumulant les formes *le* 859,

Le mot *texte*

au 75 et du 145). Le rapport du second au premier est donc de 5,02. Cherchons un terme de comparaison dans le corpus du TLF, où nous trouvons les effectifs suivants: *le* 2710723, *au* 687088, *du* 963466, *un* 1609877. Prenons garde toutefois à l'ambiguïté de *le* qui peut-être aussi un pronom personnel. Comme cet emploi est exclu dans la séquence *le texte*, il faut l'exclure aussi dans le modèle. Et plutôt que d'examiner près de 3 millions d'exemples, appliquons la proportion fournie par Gunnel Engwall dans un corpus finement analysé de la langue contemporaine. Le pourcentage de l'article étant de 85% ( $8636/10160 = 0,85$ )<sup>1</sup>, nous ne retiendrons dans le TLF que  $2710723 * 0,85 = 2301446$  occurrences de *le*. En fin de compte le rapport dans le TLF s'élève donc à :

$$\frac{39952000}{1609877} = 2,45$$

soit deux fois moins qu'au voisinage du mot *texte*. C'est dire la force de l'actualisation qui impose le mot *texte* dans le discours et dont l'ampleur est encore renforcée par l'abondance des possessifs *mon* (36) et *son* (56) - ce qui justifierait une approche pragmatique<sup>2</sup>.

**3** - Mais pour en finir avec l'approche statistique, tentons une dernière expérience où l'environnement lexical du mot *texte* puisse être appréhendé de manière plus fine. L'objectif serait d'obtenir une représentation des forces d'attraction et de répulsion qui s'exercent parmi les mots constitutifs d'un même champ thématique. Or le tableau 10 ne répond qu'imparfaitement à cet objectif, car les effectifs relevés n'envisagent qu'une seule éventualité: la cooccurrence avec le mot choisi, ici *texte*. On note par exemple que *l'assemblée* se trouve souvent en présence d'un *texte* (à discuter et à voter) et que de la même façon la *bible* fait volontiers intervenir un *texte* (à traduire ou à interpréter). Mais est-ce bien dans les mêmes phrases et les mêmes contextes? Il ne suffit pas de noter les rencontres de *texte-bible* et de *texte-assemblée*. Encore faudrait-il relever la cooccurrence *bible-assemblée*, dans le cas peu probable où l'Assemblée nationale aurait à se prononcer sur un verset biblique. C'est ce que réalise une fonction nouvelle d'Hyperbase. Une fois définis les éléments du champ étudié (par le biais des spécificités), toutes les phrases du corpus sont analysées successivement pour y rechercher tous les mots du champ thématique. Si deux d'entre eux sont présents on comptabilisera cette rencontre dans la case adéquate d'un tableau triangulaire où les

<sup>1</sup> Gunnel Engwall, *Vocabulaire du roman français* (1962-1968), Almqvist & Wiksell International, Stockholm, 1984, p. 249 et 258.

<sup>2</sup> Ajoutons un dernier indice statistique, qui est fourni par notre logiciel *Hyperbase* et qui est tiré du vocabulaire spécifique du corpus étudié, la référence étant le TLF. Parmi les spécificités du corpus bâti autour des contextes de *texte*, les formes *le* et *du* figurent au premier rang.

éléments du champ sont portés en ligne et en colonne. Si par exemple le mot *bible* occupe la ligne et la colonne 10, et le mot *grec* la ligne et la colonne 100, la cooccurrence dans la même phrase de *bible* et de *grec* sera relevée à l'intersection de la ligne 10 et de la colonne 100 (et pareillement au croisement de la ligne 100 et de la colonne 10, si l'on désire obtenir un tableau carré symétrique). On obtient une mesure généralisée de la proximité lexicale et sémantique, chaque case du tableau permettant d'apprécier les accointances qui lient la ligne et la colonne correspondantes. L'analyse factorielle semble le moyen approprié pour traiter de façon synthétique un tel tableau. Ainsi avons-nous procédé, à partir d'une liste de 53 mots spécifiques fournie par la comparaison du corpus précédent<sup>1</sup> à celui du TLF.

Les résultats offrent une excellente lisibilité. Leur fiabilité paraît garantie par la sensibilité du calcul aux associations naturelles et logiques qui unissent le singulier et le pluriel d'un même vocable (*note* et *notes*, *page* et *pages*, *livre* et *livres*, *lettre* et *lettres*, *commentaire* et *commentaires*) ou les différentes formes d'un même paradigme (*imprimer* et *imprimé*, *traduire* et *traduction*). Quand les évidences attendues sont confirmées par les chiffres, on est en droit d'accorder à ceux-ci un crédit raisonnable, lors même que la conscience linguistique manque d'appui. On a le soupçon qu'un champ thématique doit être sensible aux relations horizontales de synonymie et d'antonymie comme aux liaisons verticales et hiérarchiques et à d'autres rapports plus lointains qui s'appuient sur la métonymie et le symbole. On trouve en effet dans le graphique des exemples de ces différentes relations. S'il n'y a pas parmi nos 53 mots de véritables oppositions antonymiques du type *vice* et *vertu*, du moins y trouve-t-on des variétés complémentaires, comme *grec* et *latin*, *anglais* et *français*. Or ces couples sont nettement mis en évidence dans la moitié droite du graphique 13, le premier dans la partie basse, le second sur la ligne médiane. Les relations d'hostilité semblent être d'une grande stabilité, comme si les ennemis étaient pour les mots les compagnons les plus fidèles, comme l'ombre pour le soleil. Les accords synonymiques ne sont pas aussi franchement déclarés, mais, quoique tacites, ils suscitent souvent un partage à l'amiable des mêmes territoires. Ainsi observe-t-on sur le graphique le voisinage de *saint* et *sacré*, de *Dieu* et *Christ*, *accord* et *armistice*, *note* et *commentaire*. Peut-être pourrait-on trouver ici la trace du souci de variété lexicale qui proscriit les répétitions trop rapprochées et qui

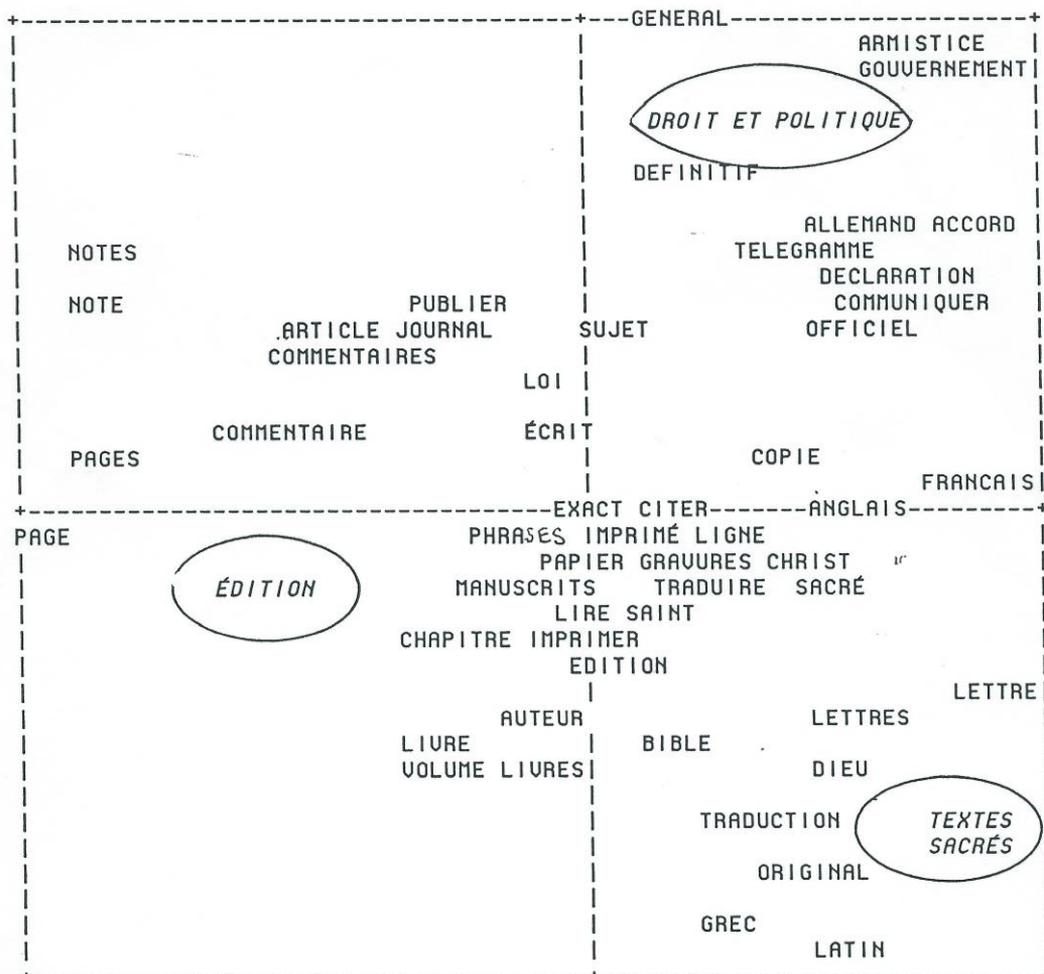
---

<sup>1</sup> Il s'agit des 1788 contextes contenant le mot "texte" et extraits de l'ensemble littéraire (textes techniques exclus). Quant à la toile de fond, sur laquelle se détachent les spécificités de ce corpus, elle est limitée aux textes postérieurs à 1789, qui constituent la base de documentation du *Trésor*.

Le mot *texte*

s'est maintenu dans la tradition rhétorique au point de susciter l'inclusion de dictionnaires de synonymes dans les logiciels de traitement de textes. Cependant les observations les plus nombreuses relèvent de la structure hiérarchique du lexique. Ce trait est particulièrement visible dans la moitié gauche du graphique où tous les éléments de l'édition, du plus large au plus menu, apparaissent emboîtés et solidaires: *auteur, édition, livre, volume, chapitre, page, commentaire, note, ligne, phrase*. De même trouve-t-on ensemble *l'article* et *le journal*, *l'original* et *la traduction*, *lire* et *manuscrits*. Dans bien des cas on est amené à sous-entendre une liaison syntaxique, déterminative (*l'article du journal, la traduction de l'original*) ou transitive (*lire un manuscrit*).

**Figure 13. Analyse factorielle des relations de proximité**



Mais comment s'organise ce réseau de relations? N'y a-t-il que des ententes bilatérales, ou bien plutôt un accord d'ensemble? Nous avons surajouté au résultat brut trois éléments d'interprétation qui nous semblent suggérés par les configurations observées. Le haut du champ paraît être en effet le lieu du droit et de la politique et s'oppose aux textes sacrés qui prennent position dans la partie basse. L'univers de la politique est celui de

l'actualité. Les textes qu'on y examine sont ceux des *déclarations*, des *télégrammes*, des *communiqués*, des *journaux*, des *armistices* et des *lois*. Celui de l'exégèse religieuse ou savante au contraire est plus proche de l'édition que de la communication, et tourné vers le passé. On y recherche le texte *original* et non la déclaration *officielle* (les deux adjectifs s'opposent sur l'axe vertical). Quant à la partie gauche de la figure, on n'y décèle aucun conflit de cette sorte et les points se rapprochent de l'axe horizontal. Les éléments qui se concentrent dans cette zone pacifique sont en effet peu marqués. Ils appartiennent à l'univers technique de l'édition et ne prennent pas parti entre société civile et la religion.

Si l'on se reporte à l'analyse factorielle qu'on a représentée dans la figure 12 et qui observait l'évolution chronologique du champ, on ne peut qu'être frappé par la correspondance des deux graphiques, pour peu qu'on inverse les facteurs (par une rotation de 90 degrés), et qu'on inverse le haut et le bas<sup>1</sup>. On trouve dans les deux cas une structure triangulaire dont le croissant caractéristique de la figure 12 visitait successivement les trois sommets: d'abord le sacré, puis le technique, enfin le politique. Ces trois pôles se retrouvent ici dans une structure apparemment synchronique qui n'est peut-être que le reflet d'une structure diachronique mal neutralisée. On peut se réjouir de la convergence, si on est optimiste. Quand plusieurs méthodes aboutissent au même cliché, avec l'infra-rouge, le radar ou les ultra-sons, le résultat en est généralement corroboré. Mais les pessimistes peuvent aussi s'en inquiéter: que mesure-t-on vraiment quand un train de phénomènes peut en cacher un autre? Et que valent les études synchroniques de langue quand on les mène avec un empan trop large et qu'on veut trop ignorer le feu rampant du changement? Qui trop embrase mal éteint.

Il n'est pas sûr que le mot *texte*, dont la richesse symbolique reste faible et dont l'intension l'emporte sur l'extension, soit le meilleur exemple pour aborder l'étude des champs sémantiques. Le *texte* n'est guère ici qu'un prétexte, proposé à brûle-pourpoint, presque à titre de gageure, au directeur de la présente revue. Mais les gageures ont une vertu démonstrative et méthodologique que les études plus approfondies et plus complexes ne recèlent pas toujours. Née d'une hypothèse hâtive - le progrès paradoxal du mot *texte* - l'expérience a été réalisée en quelques jours grâce aux moyens puissants fournis par la technologie moderne.

---

<sup>1</sup> Ces manipulations n'ont rien d'illégitime, les directions bas-haut et gauche-droite étant arbitraires et donc réversibles.

*Le mot texte*

En réalité toute recherche historique, celle des mots, celle des mentalités, ou celle de l'univers, inclut nécessairement l'éloignement et la perspective et tire profit de la convergence d'approches opposées, dont les unes relèvent de l'infiniment petit et les autres de l'infiniment grand. Claude Allègre montre excellemment dans sa récente *Histoire naturelle* que l'âge du monde, depuis le *big-bang* originel et les grandes étapes de l'évolution des espèces ont pu être déterminés par le feu croisé des observations de l'astronomie, de la physique corpusculaire et de la géologie de terrain. Les méthodes documentaires et statistiques, déployées à grande échelle, comme les télescopes géants, font parcourir de grands espaces. Mais la philologie de terrain, qui dispose d'outils précis et sûrs, analogues au carbone 14 des géologues, fait parcourir de grands progrès.