



**HAL**  
open science

# L'attestation des combinaisons lexicales à l'aide de la base de données textuelles FRANTEXT

Jacques François

► **To cite this version:**

Jacques François. L'attestation des combinaisons lexicales à l'aide de la base de données textuelles FRANTEXT. Cahiers du CRISCO (Univ. Caen), 2010, 29, pp.1-50. hal-01834424

**HAL Id: hal-01834424**

**<https://hal.science/hal-01834424>**

Submitted on 10 Jul 2018

**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.



**Cahier du CRISCO**  
**n°29**

**Janvier 2010**



Centre de  
Recherches  
Inter-langues  
sur la Signification  
en COntexte  
E. A. 4255

**L'ATTESTATION DES  
COMBINAISONS LEXICALES  
À L'AIDE DE  
LA BASE DE DONNÉES TEXTUELLES  
*FRANTEXT***

**Jacques FRANÇOIS**

***CRISCO***

Université de Caen (Bât. Sciences Porte SA S13), 14032 CAEN CEDEX  
Tél. : 02 31 56 56 27 — Fax : 02 31 56 54 27 — Site web : [www.crisco.unicaen.fr](http://www.crisco.unicaen.fr)  
Courriel direction : [franck.neveu@unicaen.fr](mailto:franck.neveu@unicaen.fr)  
Courriel secrétariat : [liliane.docquiert@unicaen.fr](mailto:liliane.docquiert@unicaen.fr)

Le *Centre de Recherches Interlangues sur la Signification en Contexte* (CRISCO) contribue au débat linguistique à l'aide de trois outils éditoriaux :

## I

### *Syntaxe & Sémantique*

Revue de linguistique créée en 2000 aux Presses Universitaires de Caen

*Responsables de la revue* : J. FRANÇOIS & N. LE QUERLER

*Comité de rédaction* : E. DELENTE, J. FRANÇOIS, E. GILBERT, C. GUIMIER, I. HAIK, M. KRAUSE, D. LEGALLOIS, V. LENEPVEU, N. LE QUERLER, F. NEVEU, R. RENAULT.

*Comité scientifique* : P. BLUMENTHAL (Cologne), A. BORILLO (Toulouse-le-Mirail, Toulouse II), M. CHAROLLES (Sorbonne Nouvelle, Paris III), B. COMBETTES (Nancy II), J.P. DESCLES (Paris-Sorbonne, Paris IV), D. GEERAERTS (Louvain), B.N. GRUNIG (Paris VIII, Vincennes – Saint-Denis), P. HIRSCHBUHLER (Ottawa), G. KLEIBER (Sciences Humaines, Strasbourg II), P. KOCH (Tübingen), B. LAMIROY (Louvain), A. LEMARECHAL (Paris-Sorbonne, Paris IV), R. MARTIN (Paris-Sorbonne, Paris IV), S. MEJRI (Villetaneuse, Paris XIII), C. VET (Groningen), B. VICTORRI (CNRS, Paris).

n°1	<i>Connecteurs et marqueurs de connexions</i> , sous la direction de Claude GUIMIER (mars 2000)
n°2	<i>Sémantique du lexique verbal</i> , sous la direction de Françoise CORDIER, Jacques FRANÇOIS et Bernard VICTORRI (avril 2001)
n°3	<i>Les grammaires du français et les "mots outils"</i> , sous la direction de Gudrun LEDEGEN et Nathalie ROSSI (mars 2002)
n°4	<i>La valence vue d'Allemagne</i> , sous la direction de Peter BLUMENTHAL et Peter KOCH (mars 2003)
n°5	<i>Polysémie et polylexicalité</i> , sous la direction de Salah MEJRI (avril 2004)
n°6	<i>Aux marges de la prédication</i> , sous la direction de Irmtraud BEHR, Jacques FRANÇOIS, Anne LACHERET-DUJOUR & Florence LEFEUVRE (décembre 2005)
n°7	<i>La terminologie linguistique – Problèmes épistémologiques, conceptuels et traductionnels</i> , sous la direction de Franck NEVEU (mars 2007)
n°8	<i>Éléments de relation : de la phrase au texte</i> , sous la direction de Claude GUIMIER (décembre 2007)
n°9	<i>Texte, documents numériques, corpus. Pour une science des textes instrumentée</i> , sous la direction de Mathieu VALETTE (décembre 2008)

## II

### *Bibliothèque de Syntaxe & Sémantique*

collection de linguistique créée en 2002 aux Presses Universitaires de Caen

<i>Traits d'union</i> , sous la direction de Nicole LE QUERLER & Georges KLEIBER [Actes des <i>Journées de Linguistique Rhénane</i> de novembre 2000, organisées par SCOLIA et le CERLICO] (mai 2002)
<i>L'adjectif en français et à travers les langues</i> , sous la direction de Jacques FRANÇOIS [Actes du colloque de Caen, 28-30 juin 2001] (janvier 2005)
<i>Composition syntaxique et figement lexical</i> , sous la direction de Jacques FRANÇOIS & Salah MEJRI [Actes de l'atelier franco-tunisien de Caen, septembre 2002] (printemps 2006)
<i>Autour de la préposition</i> , sous la direction de Jacques FRANÇOIS, Eric GILBERT, Claude GUIMIER et Maxi KRAUSE (novembre 2009)

# L'ATTESTATION DES COMBINAISONS LEXICALES À L'AIDE DE LA BASE DE DONNÉES TEXTUELLES *FRANTEXT*

**Résumé :** La base de données textuelles FRANTEXT constitue un outil précieux pour mesurer la fréquence relative des occurrences d'un item lexical par rapport au nombre de mots graphiques d'un corpus délimité en termes d'époque, de genre et éventuellement d'auteur ou d'œuvre, et celle d'un segment répété interprétable, c'est-à-dire d'une chaîne de mots graphiques autosuffisante incluant l'item à l'étude par rapport à la fréquence de ce dernier. Le Cahier a un objectif pratique : fournir une méthode d'investigation des occurrences et des cooccurrences des items lexicaux dans différents corpus issus de la base FRANTEXT. Le chapitre I examine particulièrement les occurrences des items *humanité* et *inhumanité* dans l'ensemble de la base au 20<sup>e</sup> siècle, le chapitre II examine les cooccurrences de l'item *conscience* dans le même corpus et dans une œuvre particulière et le chapitre III teste une hypothèse sur la variation des cooccurrences de l'item *mœurs* entre 1600 et 2000. La récurrence des segments repérés n'implique pas nécessairement leur figement, mais atteste une certaine 'information mutuelle'. Le chapitre IV offre une ouverture vers la lexicométrie combinatoire, c'est-à-dire l'analyse statistique de cette information mutuelle en proposant une illustration élémentaire de la théorie des « collostructions » avec les syntagmes adjectivaux du type **Adj<couleur> de N<émotion>** (ex. **rouge de colère**).

**Abstract :** Textual data bank FRANTEXT turns out to be an efficient tool for measuring the relative frequency of the occurrences of a lexical item with regard to the amount of graphic words in a corpus divided according to period of production, type of text as well as according to author or work. FRANTEXT can also be used to measure the relative frequency of an interpretable iterated segment, i.e. a self-sufficient string of graphic words including the lexical item under study with regard to its frequency. The first three chapters investigate the occurrences of the items *humanité* and *inhumanité* in the 20<sup>th</sup> century, the cooccurrences of the item *conscience* in the same period and those of the item *moeurs* from the 17<sup>th</sup> to the 20<sup>th</sup> century. Finally chapter four delivers an illustration of the mutual information between two lexical items associated in a 'collostruction'.

**Mots-clés :** combinaison lexicale ; lexicologie combinatoire ; cooccurrence ; collostruction

L'objectif de ce 29<sup>e</sup> Cahier du CRISCO est avant tout pratique : il vise, spécialement à l'intention des étudiants désireux de se familiariser avec la linguistique de corpus, à illustrer l'aptitude de la base de données textuelles FRANTEXT (sous ses deux formes, intégrale et catégorisée) à fournir pour tout item lexical des cotextes récurrents dont la fréquence peut être évaluée pour le corpus choisi relativement à celle de cet item.

J'emprunte à Lebard & Salem (1994 : 319) la notion de « segment répété » pour désigner une combinaison lexicale formée d'un item déterminé et d'un cotexte récurrent. Ce terme présente l'avantage de couvrir à la fois des chaînes présentant un certain degré de figement (locutions, mots composés, phrasèmes ou collocations) et des chaînes dont la répétition n'est patente que dans un certain type de texte (par exemple **adopter / repousser un amendement** dans les débats parlementaires) sans qu'aucune manifestation de figement syntaxique soit repérable.

Dans le chapitre I j'aborderai la question de la constitution de différents corpus de travail et du calcul de la fréquence relative d'un item lexical dans chacun des corpus. Ce premier point sera illustré par les items **humanité(s)** et **inhumanité** au 20<sup>e</sup> siècle dans les genres Essais et Traités comparés à l'ensemble de la base.

Le chapitre II sera consacré au repérage des segments répétés autour d'un item, en l'occurrence **conscience**, et au calcul de la fréquence de ces segments relativement à celle de l'item étudié. La période d'investigation est le 20<sup>e</sup> siècle.

Le chapitre III introduit une perspective historique : j'y examinerai les segments répétés autour de l'item **mœurs** de 1600 à 2000. La raison de ce choix est que cet item présente une fréquence relative élevée durant approximativement un siècle (1750-1850), le sommet de la courbe étant atteint dans la décennie de la Révolution française. L'hypothèse testée est celle du recouvrement partiel de deux types d'emplois successifs, d'abord normatif et à fondement religieux jusqu'au milieu du 19<sup>e</sup> siècle, puis descriptif, voire sociologique à partir du milieu du 18<sup>e</sup> siècle.

Enfin le chapitre IV vise à donner une illustration élémentaire d'une analyse lexicométrique des segments répétés, c'est-à-dire de « l'information mutuelle » entre deux items lexicaux associés dans une construction. L'exemple choisi est celui du remplissage des positions N1 et N2 dans la construction **N1<adjectif de couleur> de N2<nom d'émotion>**.

La notion de **combinaison de mots** a fourni la matière du dictionnaire de D. Lefur & al. (2007) et celle apparentée de **lexicologie combinatoire** qui dégage des **fonctions lexicales** est au centre des travaux d'I. Mel'čuk et A. Polguère (cf. Mel'čuk & al. 1984, Mel'čuk, Clas & Polguère, 1995, Mel'čuk, 2003, Mel'čuk & Polguère 2006)<sup>1</sup>. Elle croise celle de **collocation** développée par l'école contextualiste britannique. Sur cette terminologie et les débats qu'elle suscite, je renvoie au Cahier 21 du CRISCO (2006) sous la direction de D. Legallois & J. François et plus particulièrement au chapitre II « La phraséologie dans la linguistique contextualiste » (p.28-48), ainsi qu'à François & Manguin (2006), Grossmann & Tutin (2002) et Tutin & al. (2006).

---

<sup>1</sup> Les fondements du *Dictionnaire des combinaisons de mots* et ceux des fonctions lexicales de la lexicologie combinatoire de I. Mel'čuk et A. Polguère diffèrent : l'ouvrage des éditions Le Robert exploite un logiciel de sélection statistique des combinaisons tandis que les fonctions lexicales sont définies comme une d'une paire de constituants fonctionnant dans un rapport de base à collocatif sans support statistique.

# SOMMAIRE

	page
<b>Chapitre I : La fréquence des occurrences lexicales dans FRANTEXT illustrée par les items <i>humanité</i> et <i>inhumanité</i></b>	4
<i>I-1. Comparaison des deux bases FRANTEXT intégrale et catégorisée</i>	6
<i>I-2. Calcul de fréquence relative des occurrences d'un item dans un corpus (application à <b>humanité</b> et <b>inhumanité</b>)</i>	8
<i>I-3. Comparaison des fréquences absolues et relatives de <b>humanité</b> entre auteurs (1980-2000)</i>	12
<i>I-4. Bilan</i>	14
<b>Chapitre II : La recherche des segments répétés illustrée par les cooccurents de l'item <i>conscience</i></b>	15
<i>II-1. Rapatriement des résultats de la recherche dans un fichier de traitement de texte et transfert dans un fichier de tableur</i>	15
II-1.1. Mode de rapatriement des résultats dans la base intégrale	15
II-1.2. Mode de rapatriement des résultats dans la base catégorisée	18
<i>II-2. Les œuvres et les auteurs représentés dans le corpus</i>	19
<i>II-3. Méthode de recherche des cooccurents à gauche et à droite de l'item <b>conscience</b></i>	22
<i>II-4. Les cooccurents de l'item <b>conscience</b> dans les Œuvres complètes de Pierre Mendès-France</i>	26
<i>II-5. Bilan</i>	26
<b>Chapitre III : La variation des segments répétés à travers les siècles illustrée par les cooccurents de l'item <i>mœurs</i></b>	28
<i>III-1. L'hypothèse à tester sur l'évolution de la fréquence relative de l'item <b>mœurs</b> depuis le 17<sup>e</sup> siècle</i>	28
<i>III-2. Fréquence absolue et relative de l'item <b>mœurs</b> à travers la base FRANTEXT</i>	29
III-2.1. La variation des fréquences à travers les siècles	29
III-2.2. La variation des fréquences à travers les oeuvres	30
<i>III-3. La variation des cooccurences de l'item <b>mœurs</b> entre 1600 et 2000 dans la base FRANTEXT</i>	35
<i>III-4. Bilan</i>	39
<b>Chapitre IV : Une incursion en lexicométrie combinatoire : Adj&lt;couleur&gt; de N&lt;émotion&gt;</b>	40
<i>IV-1. La démarche 'collostructionnelle'</i>	40
<i>IV-2. L'analyse des 'collexèmes covariants'</i>	42
<i>IV-3. Essai d'application élémentaire</i>	44
<i>IV-4. Bilan</i>	47
<b>RÉFÉRENCES</b>	49

## Chapitre I

### La fréquence des occurrences lexicales dans FRANTEXT illustrée par les items *humanité* et *inhumanité*

La base de données textuelles FRANTEXT se présente sous deux versions, intégrale et catégorisée, toutes deux accessibles par abonnement institutionnel (la plupart des ordinateurs de la salle de travail du CRISCO bénéficient de cet abonnement) ou par abonnement privé obtenu auprès de l'ATILF ► <http://www.atilf.fr/>, cf. Figure I-1.

**BASE TEXTUELLE FRANTEXT** <http://www.frantext.fr>

---

### Accueil

■ **Avertissement !**

La base Frantext est conçue pour permettre des recherches de mots, de lemmes ou d'expressions régulières dans un corpus donné. Elle ne permet ni le téléchargement ni la lecture des textes pleins. Les versions numériques des textes libres de droits sont téléchargeables sur le site du CNRTL.

■ **Frantext intégral**

La base la plus complète avec 3 952 références (avril 2009). Consultation sur abonnement.

**Accès à 'Frantext intégral'**

Accès par mot de passe

Bibliographie  
Nouveautés avril 2009  
Quoi de neuf dans FRANTEXT ?  
Abonnement

■ **Frantext catégorisé**

Un extrait de 1 940 textes catégorisés de Frantext. Consultation sur abonnement.

**Accès à 'Frantext catégorisé'**

Accès par mot de passe

Figure I-1 : Accueil de la base de données textuelles FRANTEXT

La sélection du corpus souhaité s'effectue dans la base catégorisée par la page suivante qui offre quatre types de sélection : par auteur, par titre, par date et par genre. Dix genres sont proposés, huit sont proprement littéraires, les deux derniers sont les essais et les traités (cf. Figure I-2).

<input type="button" value="Aide"/>	<b>Sélection par auteur</b> <input type="text"/>	<input type="button" value="Aide"/>	<b>Sélection par titre</b> <input type="text"/>
<b>Sélection par date</b> <input checked="" type="radio"/> Date indifférente <input type="radio"/> A la date précise DATE1 <input type="radio"/> Avant DATE1 <input type="radio"/> Après DATE1 <input type="radio"/> Entre DATE1 et DATE2 DATE1: <input type="text" value="0"/> DATE2: <input type="text" value="2000"/>		<input type="button" value="Aide"/>	
		<b>Sélection par genre</b> <input type="checkbox"/> correspondance <input type="checkbox"/> récit de voyage <input type="checkbox"/> éloquence <input type="checkbox"/> roman <input type="checkbox"/> mémoires <input type="checkbox"/> théâtre <input type="checkbox"/> pamphlet <input type="checkbox"/> traité <input type="checkbox"/> poésie <input type="checkbox"/> essai	
<input type="button" value="ENREGISTRER LA SELECTION"/>		<input type="button" value="EFFACER LE FORMULAIRE"/>	

Figure I-2 : Mode de sélection du corpus examiné dans la version catégorisée

Pour la version intégrale, deux modes « ancien » et « nouveau » de sélection du corpus sont proposés. Le mode ancien est le même que dans la version catégorisée. Le mode nouveau comporte des options nouvelles (cf. Figure I-3) et en particulier une révision drastique (inspirée par le développement récent de la textométrie) du classement des genres textuels (cf. Figure I-4)

### Définition du corpus de travail

- Il n'y a pas de texte dans le corpus de travail**
- Recherche dans un élément bibliographique**  

**Options**  
 sensible à la casse  
 sensible aux diacritiques  
 sous-chaine  
 bibliographie détaillée

Combiner plusieurs éléments
- Formulaires spécialisés**
- Autres actions**

Figure I-3 : Le nouveau mode de sélection du corpus examiné dans la version intégrale



3	argument
1	articles de journaux
77	autobiographie
10	autofiction
1	autres formes brèves fictionnelles
2	avant-propos
1	ballet
1	biographie
9	chanson
1	chronique historique
1	chroniques
1	cinéma
166	comédie
1	comédie : dialogue
4	comédie héroïque
1	comédie historique
2	commentaire
37	conte
157	correspondance
11	critique

Figure I-4 : La première fenêtre de la nouvelle liste des genres textuels proposée dans la version intégrale

Cette nouvelle liste très sophistiquée présente un intérêt indiscutable pour certains types d'études, cependant – dans la mesure où notre intérêt pour les genres peut se limiter au moins provisoirement à la distinction entre « belles-lettres » et « essais et traités » – je me contenterai du mode de sélection classique qui présente l'avantage d'être partagé par les deux versions de FRANTEXT.

### ***I-1. Comparaison des deux bases FRANTEXT intégrale et catégorisée***

La base FRANTEXT, dans ses deux versions intégrale et catégorisée, offre la possibilité, non seulement d'obtenir le nombre d'occurrences d'un segment quelconque, forme lexicale, unité lexicale lemmatisée ou chaîne de formes ou d'unités lemmatisées dans le corpus sélectionné – c'est-à-dire leur fréquence absolue – mais aussi de calculer leur fréquence relative par rapport à ce corpus.

La base catégorisée constitue un sous-ensemble de la base intégrale. J'effectuerai dans le chapitre III des recherches sur une période de quatre siècles (1600-1999) dans la base intégrale. La base catégorisée ne permet d'effectuer des recherches qu'entre la décennie 1830-39 et la décennie 1990-99. Le tableau I-1 donne pour les 40 décennies comprises entre 1600 et 1999 et la période 2000-2007 (derniers textes enregistrés dans cette base) le nombre de mots graphiques de la base intégrale et pour les 17 décennies comprises entre 1830 et 1999 ce même nombre dans les deux bases intégrale et catégorisée.

FRANTEXT	intégral	catégorisé	FRANTEXT	intégral	catégorisé
<b>1600-1609</b>	1 555 323	<b>0</b>	<b>1800-09</b>	3 455 052	<b>0</b>
<b>1610-1619</b>	1 299 933	<b>0</b>	<b>1810-19</b>	2 737 383	<b>0</b>
<b>1620-1629</b>	3 033 453	<b>0</b>	<b>1820-29</b>	4 489 937	<b>0</b>
<b>1630-1639</b>	2 479 361	<b>0</b>	<b>1830-39</b>	8 185 083	<b>7 617 212</b>
<b>1640-1649</b>	1 553 584	<b>0</b>	<b>1840-49</b>	14 633 104	<b>12 244 543</b>
<b>1650-1659</b>	2 256 022	<b>0</b>	<b>1850-59</b>	7 780 740	<b>6 024 638</b>
<b>1660-1669</b>	1 831 879	<b>0</b>	<b>1860-69</b>	7 737 367	<b>6 657 757</b>
<b>1670-1679</b>	2 150 130	<b>0</b>	<b>1870-79</b>	7 412 556	<b>5 503 474</b>
<b>1680-1689</b>	2 996 396	<b>0</b>	<b>1880-89</b>	7 491 507	<b>5 430 589</b>
<b>1690-1699</b>	2 684 459	<b>0</b>	<b>1890-99</b>	6 651 236	<b>6 034 102</b>
<b>1700-1709</b>	978 656	<b>0</b>	<b>1900-09</b>	5 279 283	<b>5 229 841</b>
<b>1710-1719</b>	2 401 759	<b>0</b>	<b>1910-19</b>	6 042 569	<b>5 986 761</b>
<b>1720-1729</b>	1 120 153	<b>0</b>	<b>1920-29</b>	12 990 467	<b>11 495 246</b>
<b>1730-1739</b>	4 934 740	<b>0</b>	<b>1930-39</b>	15 016 545	<b>13 155 370</b>
<b>1740-1749</b>	3 081 036	<b>0</b>	<b>1940-49</b>	11 541 676	<b>10 144 130</b>
<b>1750-1759</b>	4 176 849	<b>0</b>	<b>1950-59</b>	11 176 559	<b>9 403 724</b>
<b>1760-1769</b>	563 616	<b>0</b>	<b>1960-69</b>	14 916 188	<b>11 759 333</b>
<b>1770-1779</b>	5 429 458	<b>0</b>	<b>1970-79</b>	6 992 004	<b>3 132 477</b>
<b>1780-1789</b>	5141 173	<b>0</b>	<b>1980-89</b>	9 827 796	<b>5 431 608</b>
<b>1790-1799</b>	3 426 086	<b>0</b>	<b>1990-99</b>	6 654 688	<b>1 668 155</b>
			<b>2000-07</b>	2 553 206	<b>0</b>

Tableau I-1 : Nombre de mots graphiques dans la base FRANTEXT intégrale et catégorisée entre 1600 et 2007

La figure I-5 représente par périodes de 20 ans le volume comparé des deux bases pour la période comportant des textes catégorisés (1830-1999). Jusqu'à la période 1950-69 la base catégorisée représente environ 80% de la base intégrale. Cette proportion diminue dans les 40 dernières années (69% pour 1970-89 et 44% pour 1990-2007, la chute étant liée à l'absence de textes catégorisés à partir de 2000).

A titre d'illustration, je vais montrer comment comparer les occurrences des items lexicaux **humanité** et **inhumanité** rencontrées au long du 20<sup>e</sup> siècle par époque, par genre et accessoirement par auteur. La période 1900-1999 est retenue parce que le nombre des occurrences de **inhumanité** rencontrées durant cette période s'élève à 104, ce qui permet *a priori* un examen instructif.

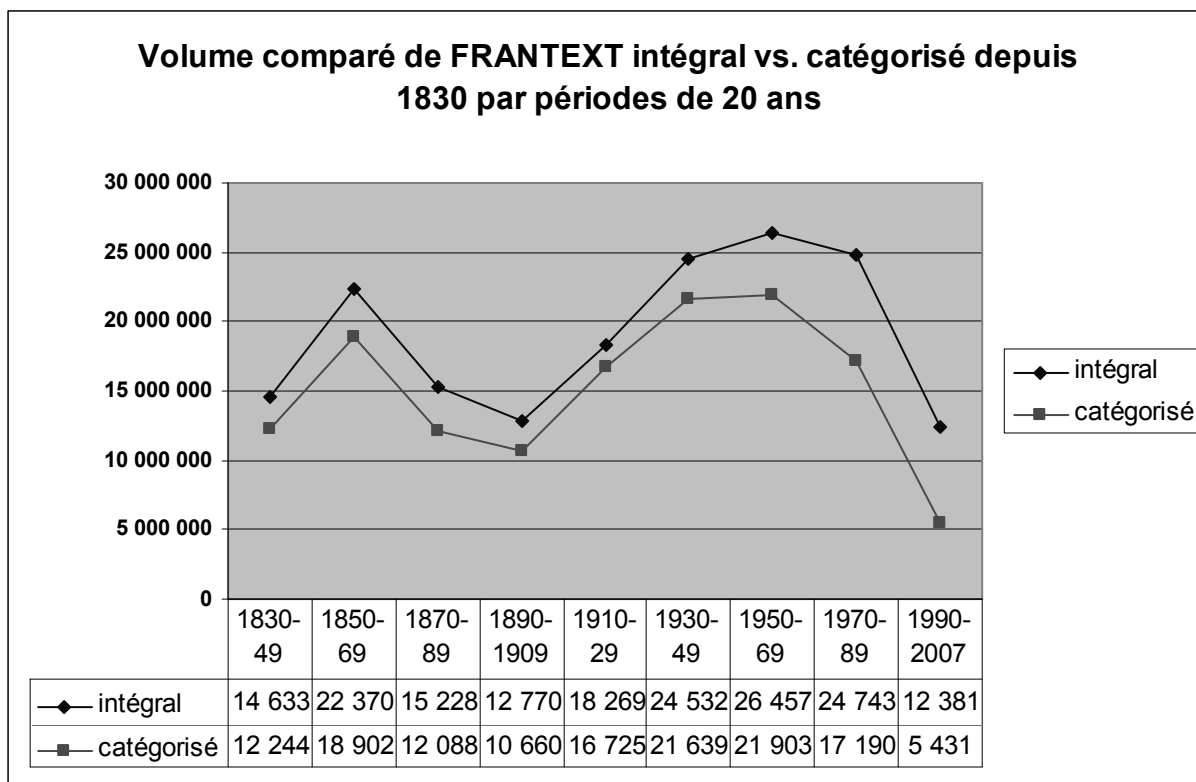


Figure I-5

***I-2. Calcul de fréquence relative des occurrences d'un item dans un corpus (application à 'humanité' et 'inhumanité')***

Le tableau I-2 fournit pour les 10 décennies de 1900 à 1999 le nombre d'occurrences, c'est-à-dire la fréquence absolue, des items **humanité** et **inhumanité** rencontrés dans la base intégrale et dans le sous-ensemble de cette base constitué par les deux genres Essais et Traités qui se distinguent des huit autres genres proprement littéraires (selon le classement « ancien » des genres que partagent les deux bases). On voit clairement pour chaque décennie le rapport entre la fréquence absolue des deux items.

décennies	<i>humanité</i>		<i>inhumanité</i>		Volume en mots des 2 bases	
	intégral	Essais-traités	intégral	Essais-traités	intégral	Essais-traités
1900-09	381	140	1	0	<b>5 279 283</b>	<b>1 178 964</b>
1910-19	414	73	5	1	<b>6 042 569</b>	<b>890 725</b>
1920-29	800	467	23	10	<b>12 990 467</b>	<b>4 013 254</b>
1930-39	1222	743	8	5	<b>15 016 545</b>	<b>4 986 616</b>
1940-49	676	436	18	7	<b>11 541 676</b>	<b>4 383 538</b>
1950-59	660	491	6	1	<b>11 176 559</b>	<b>4 871 808</b>
1960-69	513	382	9	8	<b>14 916 188</b>	<b>8 235 538</b>
1970-79	304	45	17	16	<b>6 992 004</b>	<b>580 424</b>
1980-89	284	57	3	0	<b>9 827 796</b>	<b>969 436</b>
1990-99	234	49	14	9	<b>6 654 688</b>	<b>655 910</b>
Total :	<b>5488</b>	<b>2883</b>	<b>104</b>	<b>57</b>	<b>100 437 775</b>	<b>30 766 213</b>

Tableau I-2 : Nombre des occurrences des items *humanité* et *inhumanité* par décennie dans l'ensemble de la base FRANTEXT intégrale et dans les genres Essais et Traités

La disparité est criante à certaines périodes, par exemple en 1900-1909 : 1 occurrence de *inhumanité* contre 381 de *humanité* (0,26%), mais moins marquée à d'autres, par exemple en 1990-99 avec 14 occurrences de *inhumanité* contre 234 de *humanité* (5,98%). Mais comparer par exemple pour la période 1900-09 les 381 occurrences de *humanité* dans FRANTEXT intégral et ses 140 occurrences dans les essais et traités n'est pas pertinent, parce que d'une décennie à l'autre le volume relatif (en mots graphiques) des Essais et Traités n'est pas constant. C'est pourquoi il faut ajouter les deux colonnes intitulées « volume en mots des 2 bases », lesquelles fournissent les données numériques permettant le calcul des fréquences relatives à chaque corpus, cf. Tableau I-3. Les fréquences relatives indiquées s'entendent au cent millième ( $E^{-5}$ ), c'est-à-dire par exemple pour la fréquence « 7,22 » de *humanité* dans la base intégrale en 1900-09 : 7,22 /100 000.

	Fréquence relative ( $E^{-5}$ )			Fréquence relative ( $E^{-5}$ )	
	intégral	Essais & Traités		<i>inhumanité</i>	intégral
<i>humanité</i>			<i>inhumanité</i>		
1900-09	7,22	11,87	1900-09	0,02	0,00
1910-19	6,85	8,20	1910-19	0,08	0,11
1920-29	6,16	11,64	1920-29	0,18	0,25
<b>1930-39</b>	<b>8,14</b>	<b>14,90</b>	1930-39	0,05	0,10
1940-49	5,86	9,95	1940-49	0,16	0,16
1950-59	5,91	10,08	1950-59	0,05	0,02
1960-69	3,44	4,64	1960-69	0,06	0,10
1970-79	4,35	7,75	<b>1970-79</b>	<b>0,24</b>	<b>2,76</b>
1980-89	2,89	5,88	1980-89	0,03	0,00
1990-99	3,52	7,47	<b>1990-99</b>	<b>0,21</b>	<b>1,37</b>

Tableau I-3 : Fréquence relative des items *humanité* et *inhumanité* au long du 20<sup>e</sup> siècle par décennie dans la base FRANTEXT intégrale et dans les Essais et Traités

La fréquence relative (au corpus sélectionné) de l'item *humanité* se révèle à chaque décennie supérieure dans les Essais et Traités par rapport à la base intégrale (cf. Figure I-6) avec une disparité maximale au profit des Essais et Traités en 1930-39.

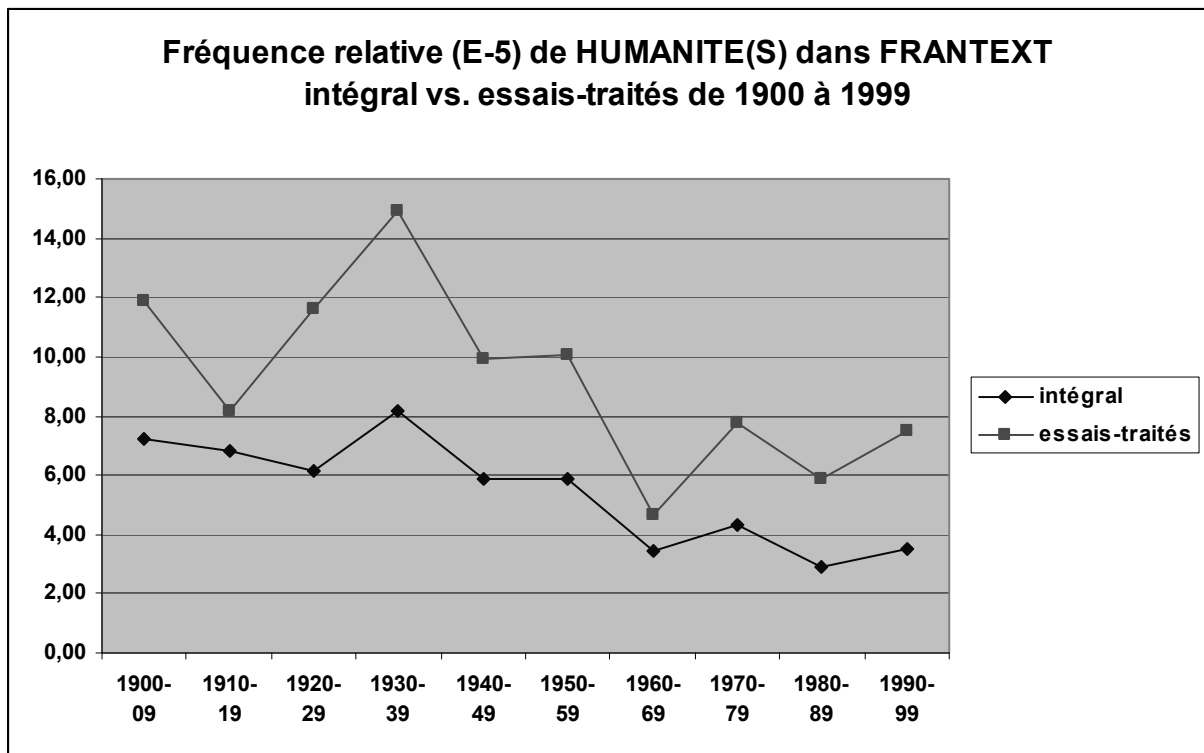


Figure I-6

En revanche celle de l'item *inhumanité* est souvent comparable dans les deux corpus et présente deux disparités exceptionnelles au profit des Essais et Traités en 1970-79 et 1990-99 (cf. Figure I-7).

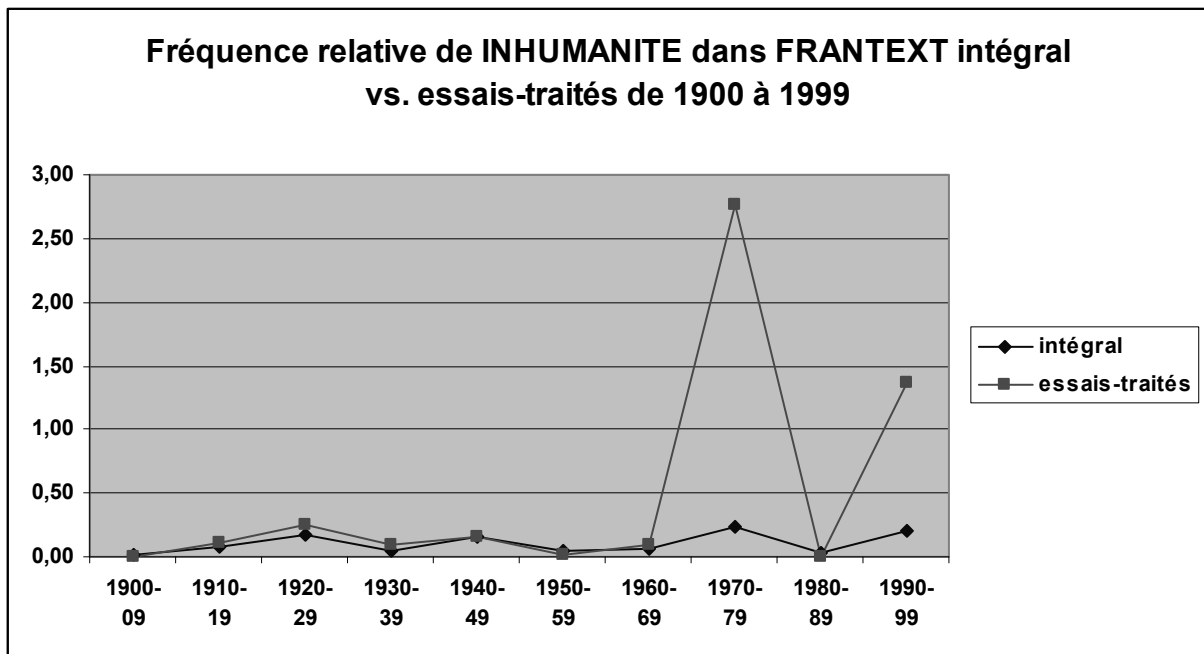


Figure I-7

Comme je l'ai déjà montré ailleurs<sup>2</sup>, ces deux disparités laissent supposer la présence de ce que j'appelle un « effet-auteur », c'est-à-dire que l'item recherché figure abondamment dans une œuvre sélectionnée dans le corpus, ce qui produit une fréquence relative aberrante. Cette supposition se confirme, car pour la période 1970-79, les 15 occurrences de **inhumanité** inventoriées (et reproduites ci-dessous) proviennent toutes de la même œuvre : Raoul VANEIGEM, *Traité de savoir-vivre à l'usage des jeunes générations* (1972).

- 1) *ils s'en font les chiens de garde de la manière la plus habile qui soit : en aboyant contre toutes les séquelles de l'**inhumanité** ancienne.*
- 2) *Sans doute s'efforçait-elle par là de modeler et d'envelopper l'**inhumanité** foncière de sa domination dans une humanité de liens idylliques.*
- 3) *Certes, le sacrifice exclut le don pur mais tel fut bien souvent l'empire du jeu, du gratuit, de l'humain, que l'**inhumanité**, la religion, le sérieux purent passer pour accessoires dans des préoccupations comme la guerre, l'amour, l'amitié, le service d'hospitalité.*
- 4) *C'est que l'identification prend, comme toute **inhumanité**, sa source dans l'humain.*
- 5) *Comme si la destruction de l'homme sous l'originalité faisandée du nom ne valait pas l'**inhumanité** d'une série de chiffres.*
- 6) *ses personnages à tête vide dressent bien le bilan accusateur de l'**inhumanité**.*
- 7) *La pire **inhumanité** n'est jamais qu'une volonté d'émancipation cédant aux compromis et se fossilisant sous la couche de ses renoncements successifs.*
- 8) *Ils vont vers l'homme total par le biais de l'**inhumanité** totale.*
- 9) *Un tel projet implique le blocage absolu de l'histoire, donc le refus extrême du mouvement d'émancipation, c'est-à-dire l'**inhumanité** totale. '*
- 10) *L'humanité du maître tend vers zéro tandis que l'**inhumanité** du pouvoir désincarné tend vers l'infini.*
- 11) *L'**inhumanité** du seigneur est une humanité qui se cherche ;*
- 12) *l'**inhumanité** de l'exploiteur tente de se dédouaner par la séduction qu'exercent sur l'humain le progrès technique, le confort, la lutte contre la faim et la maladie ;*
- 13) *l'**inhumanité** du cybernéticien est une inhumanité qui s'accepte. Ainsi, l'inhumanité du maître est devenue de moins en moins humaine.*
- 14) *l'inhumanité du cybernéticien est une **inhumanité** qui s'accepte.*
- 15) *Ainsi, l'**inhumanité** du maître est devenue de moins en moins humaine.*

Et plus curieusement encore, les 9 occurrences inventoriées en 1990-99 sont toutes extraites d'une autre œuvre du même auteur : Raoul VANEIGEM, *Nous qui désirons sans fin* (1996). Il s'agit donc bien d'un « effet-auteur », comme le montre la figure I-8 où la « fréquence relative corrigée » est la fréquence obtenue après élimination des deux œuvres en cause.

---

<sup>2</sup> cf. François (2007 : 203-4) à propos de la disparité exceptionnelle dans le quart de siècle 1950-74 de la chaîne « ce qui / cela exig- » dans FRANTEXT intégral qui révèle une quantité écrasante d'occurrences extraites des *Mémoires de guerre* du Général De Gaulle.

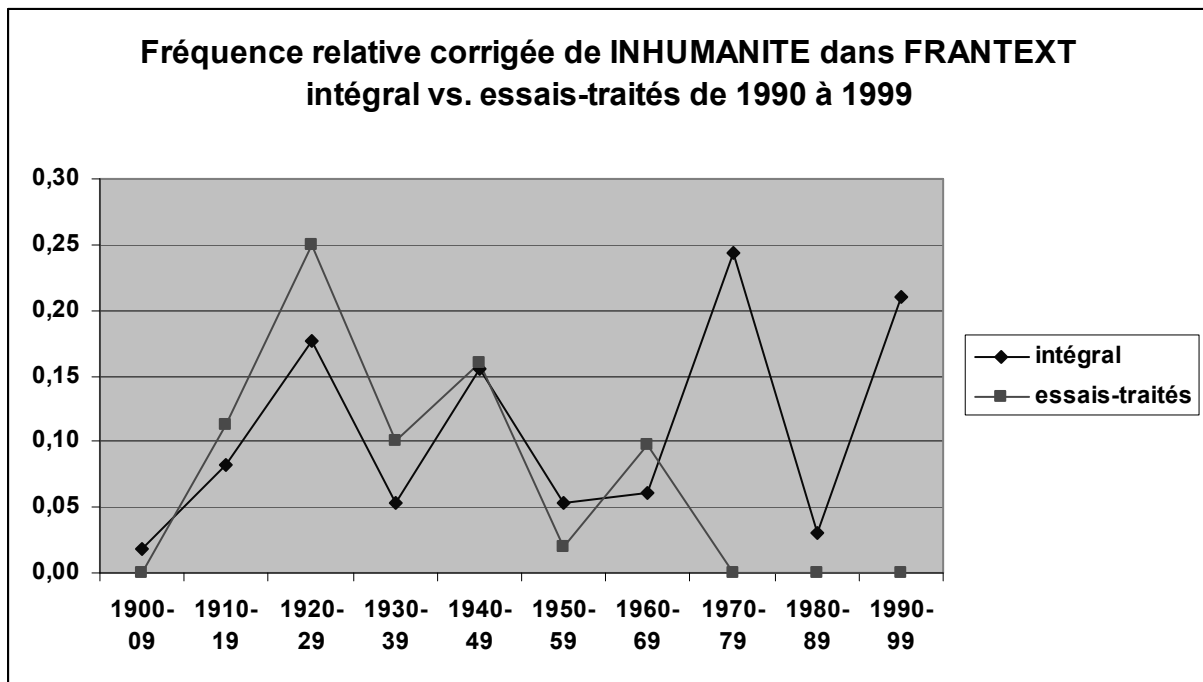


Figure I-8

Cependant il faut noter qu'une telle élimination n'est pas réellement légitime, car d'autres recherches pourraient délivrer des fréquences relatives aberrantes à cause de préférences lexico-syntaxiques exceptionnelles chez certains auteurs. La procédure la plus adaptée consisterait sans doute à fixer un nombre plafond d'occurrences retenues pour chaque œuvre.

### ***I-3. Comparaison des fréquences absolues et relatives de humanité entre auteurs (1980-2000)***

La constatation de cet « effet-auteur » conduit à examiner de plus près la composition du corpus étudié pour **humanité**, à savoir la totalité des textes de la version intégrale de FRANTEXT entre 1980 et 2000, soit 200 textes. Le tableau I-4 indique, pour chaque auteur auquel au moins 5 occurrences de l'item sont associées, le nombre d'occurrences de la forme **humanité** (au singulier) par fréquence décroissante<sup>3</sup>.

<sup>3</sup> La méthode d'obtention de ces chiffres est expliquée plus bas en section II-2.

MENDÈS-FRANCE Pierre	69	KRISTEVA Julia	8
ROUBAUD Jacques	28	PENNAC Daniel	8
DOLTO Françoise	26	THOREZ Paul	8
ORMESSON Jean d'	23	WEYERGANS François	8
VANEIGEM Raoul	18	ARAGON Louis	7
NAVARRÉ Yves	17	BIANCIOTTI Hector	7
DURAS Marguerite	16	L'HÔTE Jean	7
QUENEAU Raymond	16	MERTENS Pierre	7
POZZI Catherine	15	POIROT-DELPECH Bertrand	7
BONNEFOY Yves	14	BILLETDOUX Marie	6
CHAMOISEAU Patrick	14	ERNAUX Annie	6
JULIET Charles	14	GUYOTAT Pierre	6
SOLLERS Philippe	14	LABRO Philippe	6
FOREST Philippe	13	SCHREIBER Boris	6
HAGÈGE Claude	13	VERNE Jules	6
DEL CASTILLO Michel	11	VIOLET Lydie & DESPLECHIN Marie	6
LIPOVETSKY Gilles	11	BOUILLIER Grégoire	5
PONTALIS Jean-Bertrand	11	GRACQ Julien	5
HALIMI Gisèle	10	GROULT Benoîte & GROULT Flora	5
PEREC Georges	10	GUIBERT Hervé	5
SERRES Michel	10	HUGUENIN Jean-René	5
HAVET Mireille	9	KOLTÈS Bernard-Marie	5
CHANDERNAGOR Françoise	8	TOURNIER Michel	5

Tableau I-4 : Nombre d'occurrences de la forme **humanité** par auteur entre 1980 et 2000 (classement par fréquence décroissante)

L'auteur qui en emploie le plus grand nombre est Pierre Mendès-France (pour la partie de son œuvre comprise entre 1980 et 2000). Peut-on en déduire qu'il est celui qui emploie **LE PLUS FREQUEMMENT** la forme **humanité** ? Pour le savoir, il faut calculer la fréquence relative de cette forme par rapport au volume des œuvres concernées. C'est ce que révèle le tableau I-5 qui classe les auteurs par fréquence relative décroissante de l'emploi de la forme **humanité**.

Auteur	Fréq. abs.	Volume des œuvres	Proportion E-5
<b>VANEIGEM Raoul</b>	<b>18</b>	<b>35 666</b>	<b>50,47</b>
BONNEFOY Yves	14	106 448	<b>13,15</b>
DURAS Marguerite	16	122 650	<b>13,05</b>
POZZI Catherine	15	277 600	<b>5,40</b>
ORMESSON Jean d'	23	546 126	<b>4,21</b>
NAVARRÉ Yves	17	430 794	<b>3,95</b>
ROUBAUD Jacques	28	712 658	<b>3,93</b>
QUENEAU Raymond	16	421 817	<b>3,79</b>
MENDÈS-FRANCE Pierre	69	2 157 390	<b>3,20</b>
DOLTO Françoise	26	2 157 390	<b>1,21</b>

Tableau I-5 : Classement des auteurs de la période 1980-2000 par fréquence relative décroissante de la forme **humanité**

Le résultat est sensiblement différent. Pierre Mendès-France régresse en 9<sup>e</sup> position et son second Jacques Roubaud régresse lui aussi en 7<sup>e</sup> position. En revanche Raoul Vaneigem, déjà rencontré précédemment, passe de la 5<sup>e</sup> à la 1<sup>ère</sup> position pour son œuvre de 1996, *Nous qui désirons sans fin* mentionnée dans la section précédente. Le « phénomène Vaneigem » se confirme donc pour la période 1980-2000 : avec 18



occurrences de **humanité** dans une œuvre relativement courte (35 666 mots graphiques), il devance de très loin tous les autres auteurs. Et pour l'item **inhumanité** dans la même période 1980-2000 ? Et bien les 18 occurrences rencontrées de **inhumanité** se répartissent en 9 occurrences partagées entre 9 auteurs et ...9 occurrences extraites du même ouvrage de Raoul Vaneigem, ce qui correspond à une fréquence relative de  $25,23 \text{ E}^{-5}$  (soit environ la moitié de celle de la forme **humanité**).

<b>VANEIGEM Raoul</b>	<b>9</b>
GARAT Anne-Marie	1
GRÈCE Michel de	1
IZZO Jean-Claude	1
LANZMANN Jacques	1
MENDÈS-FRANCE Pierre	1

PENNAC Daniel	1
PONTALIS Jean-Bertrand	1
QUENEAU Raymond	1
THOREZ Paul	1

#### **I-4. Bilan**

Dans le chapitre I, j'ai cherché à montrer l'intérêt de calculer la fréquence relative d'un item lexical par rapport au volume du corpus choisi en nombre de mots-occurrences. Après avoir comparé la fréquence relative des items **humanité** et **inhumanité** dans la base intégrale et dans la paire de genres Essais et Traités au long du 20<sup>e</sup> siècle et plus spécialement durant sa dernière décennie, je suis passé à une comparaison par auteur entre 1980 et 2000 qui a fait apparaître un « effet-auteur » constituant l'un des biais de l'examen quantitatif des occurrences dans la base FRANTEXT.

Sur cette base, nous pouvons maintenant passer à l'objet principal du Cahier, l'examen des segments répétés interprétables, constitués d'un item lexical et d'un cotexte suffisant pour leur attribuer une signification (sans que les segments correspondent nécessairement à un syntagme).

## Chapitre II

### La recherche des segments répétés illustrée par les cooccurrents de l'item *conscience*

#### **II-1 Rapatriement des résultats de la recherche dans un fichier de traitement de texte<sup>4</sup> et transfert dans un fichier de tableur**

Comme nous l'avons vu dans la section I-1, les deux bases, intégrale et catégorisée, diffèrent dans le mode de sélection du corpus examiné, et elles diffèrent également dans le mode de rapatriement des résultats, c'est-à-dire des contextes d'occurrences.

##### **II-1.1. Mode de rapatriement des résultats dans la base intégrale**

Je recherche toutes les occurrences de l'item ***conscience(s)*** entre 1975 et 2000 dans la base intégrale. Pour le rapatriement des résultats, différentes options sont proposées (cf. Figure II-1).

J'ai opté dans la rubrique « Tri des résultat »s pour l'ordre chronologique, dans la rubrique Contexte pour « Mise en évidence du mot »<sup>5</sup>, dans la rubrique « Informations bibliographiques » je n'ai retenu que l'auteur, le titre et la date et enfin dans la rubrique « Format de sortie » j'ai opté pour « texte » et non « colonnes pour tableur ». La raison de ce choix est qu'il est nécessaire d'effectuer une transformation sur la chaîne à l'aide de l'outil « Remplacer » de WORD avant de transférer les résultats dans le tableur.

Sur la page suivante on reçoit l'instruction : « pour rapatrier les résultats sur votre disque dur, cliquer sur 'Télécharger avec le bouton droit de la souris' et ensuite sur 'Enregistrer la cible sous...' » On enregistre ensuite le fichier en format Texte (~.txt) et on l'ouvre à l'aide de WORD (Microsoft Office) ou de WRITER (Open Office). Pour les résultats 1-5, on obtient la liste suivante qu'il s'agit de transformer en colonnes :

---

4 Parmi les logiciels de traitement de texte, certaines transformation de chaînes applicables par la fonction « Remplacer » de Microsoft WORD s'effectuent à l'aide d'une expression régulière dans le logiciel WRITER de Open Office. Parmi les tableurs, EXCEL présente des fonctionnalités absentes dans le logiciel CALCUL de Open office.

5 Deux remarques : d'abord en cas de segment composé de plusieurs mots graphiques, chaque mot est mis en évidence séparément, ce qui fournit autant de colonnes dans le tableur ; ensuite un « bug » informatique en attente de règlement empêche d'utiliser l'option « contexte élargi » qui fournit certains résultats aberrants (sans le ou les mots recherchés). Mais en fait le contexte est en général suffisamment large.

## Rapatriement des résultats : étape 1/2

### ■ Formulaire de paramétrage

#### Tri des résultats

- par ordre chronologique croissant
- par ordre chronologique décroissant
- par ordre alphabétique des auteurs
- par ordre alphabétique des titres
  
- numéroter les résultats

#### Informations bibliographiques

- afficher les droits
- afficher la cote
- afficher l'auteur
- afficher le titre
- afficher la date
- afficher la page
- afficher la sous-référence

#### Contexte

- contexte élargi
- mise en évidence du mot
- respecter la fin de ligne

#### Format de sortie

- texte
- colonnes pour tableur
- XML / TEI

Construire le fichier de résultats

Figure II-1 : Mode de rapatriement des résultats de la requête dans la version intégrale de FRANTEXT

#### Résultat 1

AJAR pseud. Romain GARY Émile/La Vie devant soi/1975  
bonne nouvelle. Madame Rosa peut encore vivre pas mal de temps, même si elle ne le saura plus vraiment. Elle évolue très rapidement. Mais elle aura des moments de <\*>conscience\*> et elle sera heureuse de regarder autour d'elle et de voir qu'elle est chez elle. Dis à sa famille de passer me voir, je ne bouge plus, tu sais. Il me posa la main sur la

#### Résultat 2

AJAR pseud. Romain GARY Émile/La Vie devant soi/1975  
, tu sais. Il me posa la main sur la tête. C'est dingue ce qu'il y a comme personnes qui me mettent la main sur la tête. ça leur fait du bien. - Si Madame Rosa reprend <\*>conscience\*> avant son départ, tu lui diras que je la félicite. - C'est ça, je lui dirai mazltov. Le docteur Katz me regarda avec fierté. - Tu dois être le seul Arabe au monde à Résultat parler

#### Résultat 3

BARTHES Roland/Roland Barthes par Roland Barthes/1975  
/détruire Admettons que la tâche historique de l'intellectuel (ou de l'écrivain), ce soit aujourd'hui d'entretenir et d'accentuer la décomposition de la <\*>conscience\*> bourgeoise. Il faut alors garder à l'image toute sa précision ; cela veut dire qu'on feint volontairement de rester à l'intérieur de cette conscience et qu'on

#### Résultat 4

BARTHES Roland/Roland Barthes par Roland Barthes/1975  
de la conscience bourgeoise. Il faut alors garder à l'image toute sa précision ; cela veut dire qu'on feint volontairement de rester à l'intérieur de cette <\*>conscience\*> et qu'on va la délabrer, l'affaïsser, l'effondrer, sur place, comme on ferait d'un morceau de sucre en l'imbibant d'eau. La décomposition s'oppose donc ici

#### Résultat 5

BARTHES Roland/Roland Barthes par Roland Barthes/1975  
, l'effondrer, sur place, comme on ferait d'un morceau de sucre en l'imbibant d'eau. La décomposition s'oppose donc ici à la destruction : pour détruire la <\*>conscience\*> bourgeoise, il faut s'en absenter, et cette extériorité n'est possible que dans une situation révolutionnaire : en Chine, aujourd'hui, la conscience de classe

La méthode<sup>6</sup> permettant d'effectuer le transfert de ces résultats dans des colonnes de tableur opère comme suit (par une série d'emplois de la fonction REMPLACER du menu EDITION). La méthode à l'aide de WORD est indiquée dans la colonne du milieu, celle à l'aide de WRITER dans la colonne de droite. Les deux premières opérations sont destinées à attribuer à l'item conscience une colonne propre. Les trois suivantes regroupent d'abord tous les résultats en un seul paragraphe avant de séparer chaque résultat. On obtient ainsi un tableau où chaque ligne représente un résultat et où les 5 colonnes hébergent respectivement le numéro du résultat, les références, le cotexte gauche, l'item **conscience** et le cotexte droit.

Fonction de remplacement	dans WORD	dans WRITER
1. Remplacer <* par le signe de tabulation	^t	\t
2. Remplacer *> par le signe de tabulation	idem	idem
3. Remplacer le signe de saut de ligne par celui de tabulation	^p → ^t	\n → \t
4. Supprimer la 1 <sup>ère</sup> mention de « Résultat »		
5. Remplacer la chaîne {----- + tabulation + Résultat} par le signe de saut de ligne	^p	\n
6. Fermer le fichier Texte en le laissant dans ce format.		
7. Ouvrir le fichier Texte à l'aide d'EXCEL et l'enregistrer comme classeur EXCEL		

Tableau II-1 : Opérations de transfert des résultats du fichier Texte dans un fichier Word ou Writer et de là dans un classeur EXCEL

On obtient la disposition suivante (Tableau II-2) :

A	B	C	D	E
1	AJAR pseud. Romain GARY Émile/La Vie devant soi/1975	bonne nouvelle. Madame Rosa peut encore vivre pas mal de temps, même si elle ne le saura plus vraiment. Elle évolue très rapidement. Mais elle aura des moments de	<b>conscience</b>	et elle sera heureuse de regarder autour d'elle et de voir qu'elle est chez elle. Dis à sa famille de passer me voir, je ne bouge plus, tu sais. Il me posa la main sur la
2	AJAR pseud. Romain GARY Émile/La Vie devant soi/1975	, tu sais. Il me posa la main sur la tête. C'est dingue ce qu'il y a comme personnes qui me mettent la main sur la tête. ça leur fait du bien. - Si Madame Rosa reprend	<b>conscience</b>	avant son départ, tu lui diras que je la félicite. - C'est ça, je lui dirai mazltov. Le docteur Katz me regarda avec fierté. - Tu dois être le seul Arabe au monde à parler
3	BARTHES Roland/Roland Barthes par Roland Barthes/1975	/détruire Admettons que la tâche historique de l'intellectuel (ou de l'écrivain), ce soit aujourd'hui d'entretenir et d'accentuer la décomposition de la	<b>conscience</b>	bourgeoise. Il faut alors garder à l'image toute sa précision ; cela veut dire qu'on feint volontairement de rester à l'intérieur de cette conscience et qu'on
4	BARTHES Roland/Roland Barthes par Roland Barthes/1975	de la conscience bourgeoise. Il faut alors garder à l'image toute sa précision ; cela veut dire qu'on feint volontairement de rester à l'intérieur de cette	<b>conscience</b>	et qu'on va la délabrer, l'affaïsser, l'effondrer, sur place, comme on ferait d'un morceau de sucre en l'imbibant d'eau. La décomposition s'oppose donc ici
5	BARTHES Roland/Roland Barthes par Roland Barthes/1975	, l'effondrer, sur place, comme on ferait d'un morceau de sucre en l'imbibant d'eau. La décomposition s'oppose donc ici à la destruction : pour détruire la	<b>conscience</b>	bourgeoise, il faut s'en absenter, et cette extériorité n'est possible que dans une situation révolutionnaire : en Chine, aujourd'hui, la conscience de classe

Tableau II-2 : Les 5 premières occurrences de **conscience** dans le corpus

<sup>6</sup> Cette méthode a été initialement conçue par Jean-Luc Manguin, à l'époque ingénieur de recherche au CRISCO, actuellement membre du GREYC (CNRS-Université de Caen)

## II-1.2. Mode de rapatriement des résultats dans la version catégorisée

Dans la version catégorisée de FRANTEXT, le rapatriement des résultats s'effectue d'une manière légèrement différente. En premier lieu la page de sélection du format de rapatriement est plus simple, sans possibilité de sélectionner tout ou partie des données bibliographiques (voir Figure II-2). Cette fois, l'option « contexte élargi » peut être sélectionnée, il n'y a pas de bug. Ensuite un fichier de sortie de format Texte est proposé sous le nom provisoire **browserc.exe** qu'il suffit de renommer.

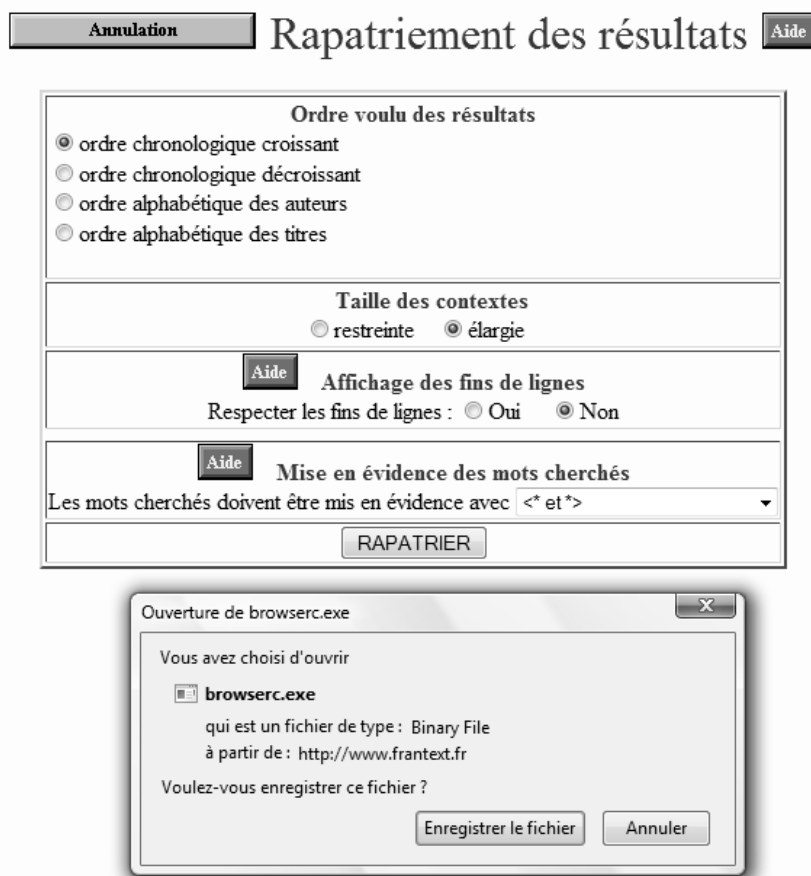


Figure II-2 : Les deux fenêtres successives du rapatriement des résultats dans la base catégorisée

Enfin les données bibliographiques sont disposées sur trois lignes au lieu de deux parce que la procédure de téléchargement des résultats dans la base catégorisée fournit d'office toutes les « informations bibliographiques » (cf. Figure II-1, case supérieure droite).

1 (Texte sous droits)

R616/ AJAR.E /LA VIE DEVANT SOI/1975

Pages 253-254 / \*

c'est une bonne nouvelle. Madame Rosa peut encore vivre pas mal de temps, même si elle ne le saura plus vraiment. Elle évolue très rapidement. Mais elle aura des moments de <\* conscience \*> et elle sera heureuse de regarder autour d'elle et de voir qu'elle est chez elle. Dis à sa famille de passer me voir, je ne bouge plus, tu sais. Il me posa la main sur la tête.

2 (Texte sous droits)  
R616/ AJAR.E /LA VIE DEVANT SOI/1975  
Pages 253-254 / \*

bouge plus, tu sais. Il me posa la main sur la tête. C'est dingue ce qu'il y a comme personnes qui me mettent la main sur la tête. ça leur fait du bien. - Si Madame Rosa reprend <\* conscience \*> avant son départ, tu lui diras que je la félicite. - C'est ça, je lui dirai mazltov. Le docteur Katz me regarda avec fierté. - Tu dois être le seul Arabe au monde à parler

3 (Texte sous droits)  
R342/ CHAR.R /AROMATES CHASSEURS/1975  
Pages 513-514 / I , AROMATES CHASSEURS

n' est que mise en demeure et croissance menacée. C' est notre désespoir insurgé, intensément vécu, qui le constate, notre lucidité, notre besoin d' amour. Et tant de <\* conscience \*> finit par tapisser l' éphémère. Chère roulotte ! Le présent-passé, le présent-futur. Rien qui précède et rien qui succède, seulement les offrandes de l' imagination

4 (Texte sous droits)  
R534/ DELAY.F /LE AIE AIE DE LA CORNE DE BRUME/1975  
Pages 81-82 / II BAISERS

fermer le magasin, et la rousse était revenue voler la robe. Mais il ne comprenait pas pourquoi Charlie s'obstinait à mentir. Cela dit, comme il n'avait pas non plus la <\* conscience \*> tranquille et que Charlie avait l'air accablé, il proposa de n'en rien dire à Bob. Charlie ne sortit pas pour autant de son accablement. à sept heures, après s'être cent

5 (Texte sous droits)  
R534/ DELAY.F /LE AIE AIE DE LA CORNE DE BRUME/1975  
Pages 141-142 / III VISIONS

quelqu'un de la famille ?, répondait : du même sang. Rapport du médecin : « choc émotif violent survenu en terrain primaire ». Traumatisme étrange du volcan pareil à la <\* conscience \*> inconsciente. Mais vous connaissiez le mot politique ? répondait : non car on n'a pas le droit d'y toucher. - à la politique ? - Oui, c'est comme nous les femmes. - Non, pas

La méthode de transfert des résultats dans un tableur est identique à celle présentée dans la section précédente, mais le tableau obtenu comporte 6 colonnes : 1 colonne de numérotation, deux de données bibliographiques et trois colonnes de concordancier (voir Tableau II-3).

## **II-2. Les œuvres et les auteurs représentés dans le corpus**

Le corpus sélectionné n'est limité qu'en période (1975-2000) pour obtenir un nombre d'occurrences suffisant (2055 dont 1971 de la forme **conscience** et 84 de la forme **consciencés**) pour dégager des segments répétés. Le transfert des 2055 extraits sélectionnés et référencés par la base FRANTEXT dans un classeur EXCEL a, entre autres avantages, de permettre un examen détaillé des œuvres et des auteurs concernés par un emploi non négligeable de l'item **conscience(s)** durant la période choisie. Dans la section I-3, j'ai présenté (tableau I-4) le nombre d'occurrences de la forme **humanité** par auteur entre 1980 et 2000. Cette présentation était méthodologiquement un peu prématurée dans la mesure où je n'ai pas pu expliquer le mode d'obtention des chiffres de ce tableau.

Il faut d'abord disposer de la totalité des occurrences dans une feuille de tableur et transférer dans une feuille spéciale toute la colonne B, celle des références (simplifiées, comme nous l'avons vu plus haut). On obtient donc une liste de 2055 références qui sont classées par ordre chronologique des œuvres (classement proposé par défaut par la base FRANTEXT). En conservant ce classement, on peut obtenir le nombre d'occurrences par œuvre de la manière suivante. Si la liste est disposée en colonne A à partir de la cellule A2, on entre en B2 la formule « =SI(A2<>A1;1;B1+1) ». Appliquée à l'ensemble de la liste (A2:A2056), cette formule augmente de 1 le nombre associé à chaque occurrence d'une même œuvre (par exemple de 1 à 5 pour une œuvre présentant 5 occurrences de l'item recherché) et introduit le chiffre 1 dès que l'œuvre change. Ensuite dans la cellule C2 on introduit la formule « =SI(B2>B3;B2;"" ) ».

1 (Texte sous droits)	R616/ AJAR.E /LA VIE DEVANT SOI/1975	Pages 253-254 / *	c'est une bonne nouvelle. Madame Rosa peut encore vivre pas mal de temps, même si elle ne le saura plus vraiment. Elle évolue très rapidement. Mais elle aura des moments de	<b>conscience</b>	et elle sera heureuse de regarder autour d'elle et de voir qu'elle est chez elle. Dis à sa famille de passer me voir, je ne bouge plus, tu sais. Il me posa la main sur la tête.
2 (Texte sous droits)	R616/ AJAR.E /LA VIE DEVANT SOI/1975	Pages 253-254 / *	bouge plus, tu sais. Il me posa la main sur la tête. C'est dingue ce qu'il y a comme personnes qui me mettent la main sur la tête. ça leur fait du bien. - Si Madame Rosa reprend	<b>conscience</b>	avant son départ, tu lui diras que je la félicite. - C'est ça, je lui dirai mazltov. Le docteur Katz me regarda avec fierté. - Tu dois être le seul Arabe au monde à parler
3 (Texte sous droits)	R342/ CHAR.R /AROMATES CHASSEURS/ 1975	Pages 513-514 / I , AROMATES CHASSEURS	n' est que mise en demeure et croissance menacée. C' est notre désespoir insurgé, intensément vécu, qui le constate, notre lucidité, notre besoin d' amour. Et tant de	<b>conscience</b>	finit par tapisser l' éphémère. Chère roulotte ! Le présent-passé, le présent-futur. Rien qui précède et rien qui succède, seulement les offrandes de l' imagination
4 (Texte sous droits)	R534/ DELAY.F /LE AIE AIE DE LA CORNE DE BRUME/1975	Pages 81-82 / II BAISERS	fermer le magasin, et la rousse était revenue voler la robe. Mais il ne comprenait pas pourquoi Charlie s'obstinait à mentir. Cela dit, comme il n'avait pas non plus la	<b>conscience</b>	tranquille et que Charlie avait l'air accablé, il proposa de n'en rien dire à Bob. Charlie ne sortit pas pour autant de son accablement. à sept heures, après s'être cent
5 (Texte sous droits)	R534/ DELAY.F /LE AIE AIE DE LA CORNE DE BRUME/1975	Pages 141-142 / III VISIONS	quelqu'un de la famille ?, répondait : du même sang. Rapport du médecin : « choc émotif violent survenu en terrain primaire ». Traumatisme étrange du volcan pareil à la	<b>conscience</b>	inconsciente. Mais vous connaissiez le mot politique ? répondait : non car on n'a pas le droit d'y toucher. - à la politique ? - Oui, c'est comme nous les femmes. - Non, pas

Tableau II-3 : Présentation des résultats 1-5 de la recherche de l'item **conscience** dans la base catégorisée entre 1975 et 2000

Appliquée à l'ensemble du pavé B2:B2056, cette seconde formule ne retient plus que la valeur la plus élevée pour chaque œuvre. On trie ensuite la totalité des cellules A2 :C2056 par tri décroissant dans la colonne C. On obtient pour les 2055 occurrences 206 œuvres différentes, soit une moyenne de 10 occurrences par œuvre. Cependant le nombre des occurrences par œuvre est extrêmement variable, compris entre 1 et 89. On trouvera en tableau II-4 la liste des 61 œuvres qui présentent au moins 10 occurrences de **conscience(s)** :

<b>1</b>	<b>MENDÈS-FRANCE Pierre/OEuvres complètes. 4. Pour une République moderne. 1955-1962./1987</b>	<b>89</b>
2	CARRÈRE D'ENCAUSSE Hélène/L'Empire éclaté : la révolte des nations en U.R.S.S./1978	66
<b>3</b>	<b>MENDÈS-FRANCE Pierre/OEuvres complètes. 6. Une vision du monde. 1974-1982./1990</b>	<b>66</b>
<b>4</b>	<b>MENDÈS-FRANCE Pierre/OEuvres complètes. 5. Préparer l'avenir. 1963-1973./1989</b>	<b>61</b>
<b>5</b>	<b>MENDÈS-FRANCE Pierre/OEuvres complètes. 3. Gouverner, c'est choisir. 1954-1955./1986</b>	<b>58</b>
6	JULIET Charles/Accueils. Journal IV 1982-1988/1994	55
7	LIPOVETSKY Gilles/L'Ère du vide : essais sur l'individualisme contemporain/1983	44
8	DOLTO Françoise/La Cause des enfants/1985	43
9	LIPOVETSKY Gilles/L'Empire de l'éphémère : la mode et son destin dans les sociétés modernes/1987	42
10	NAVARRÉ Yves/Biographie/1981	40
11	MALRAUX André/La Corde et les souris. VI : [Lazare]/1976	39
12	BLONDIN Antoine/Ma vie entre les lignes/1982	38
13	HUGUENIN Jean-René/Journal/1993	36
14	VANEIGEM Raoul/Nous qui désirons sans fin/1996	35
<b>15</b>	<b>MENDÈS-FRANCE Pierre/OEuvres complètes. 2. Une politique de l'économie. 1943-</b>	<b>33</b>

	<b>1954./1985</b>	
16	POZZI Catherine/Journal : 1913-1934/1997	31
17	CHANDERNAGOR Françoise/L'Allée du Roi/1981	30
18	FOREST Philippe/Toute la nuit/1999	29
19	<b>MENDÈS-FRANCE Pierre/OEuvres complètes. 1. S'engager. 1922-1943./1984</b>	<b>28</b>
20	HAGÈGE Claude/L'Homme de paroles : contribution linguistique aux sciences humaines/1985	25
21	CHANDERNAGOR Françoise/L'Enfant des Lumières/1995	24
22	BOSQUET Alain/Une mère russe/1978	22
23	QUENEAU Raymond/Journaux 1914-1965/1996	22
24	SCHREIBER Boris/Un silence d'environ une demi-heure/1996	21
25	GRÈCE Michel de/La Nuit du sérail/1982	20
26	MALRAUX André/Antimémoires/1976	20
27	BIANCIOTTI Hector/Sans la miséricorde du Christ/1985	19
28	BONNEFOY Yves/Rue Traversière et autres récits en rêve/1987	19
29	PENNAC Daniel/La Petite marchande de prose/1989	18
30	CAVANNA François/Les Russkoffs/1979	16
31	CHABROL Jean-Pierre/La Folie des miens/1977	16
32	DEGAUDENZI Jean-Louis/Zone/1987	16
33	FOREST Philippe/L'enfant éternel/1997	16
34	GRACQ Julien/En lisant, en écrivant/1980	16
35	GROULT BenoîteGROULT Flora/Journal à quatre mains/1994	16
36	HALIMI Gisèle/La Cause des femmes/1992	16
37	PACHET Pierre/Autobiographie de mon père/1987	16
38	MERTENS Pierre/Les Éblouissements/1987	15
39	ROUBAUD Jacques/Poésie : récit/2000	15
40	TOURNIER Michel/Le Vent Paraclet/1977	15
41	TOURNIER Michel/Les Météores/1975	15
42	BONNEFOY Yves/La Tempête/1997	14
43	CHAR René/Recherche de la base et du sommet/1981	14
44	DEL CASTILLO Michel/La Nuit du décret/1981	14
45	KRISTEVA Julia/Les Samourais/1990	14
46	LEIRIS Michel/La Règle du jeu : 4 : Frêle bruit/1976	14
47	PAGE Alain/Tchao pantin/1982	14
48	BÂ Amadou Hampâté/Oui mon commandant !/1991	13
49	MATZNEFF Gabriel/Ivre du vin perdu/1981	12
50	ROMILLY Jacqueline de/Les OEufs de Pâques/1993	12
51	SIMON Claude/L'Acacia/1989	12
52	THOREZ Paul/Les Enfants modèles/1982	12
53	BARTHES Roland/Roland Barthes par Roland Barthes/1975	11
54	BOUDARD Alphonse/Mourir d'enfance/1995	11
55	EGEN Jean/Les Tilleuls de Lautenbach/1979	11
56	ORMESSON Jean d'/La Douane de mer/1993	11
57	BARTHES Roland/La chambre claire. Note sur la photographie/1980	10
58	BIANCIOTTI Hector/Le Pas si lent de l'amour/1995	10
59	DUPEREY Annie/Je vous écris/1993	10
60	L'HÔTE Jean/Le Mécréant ou les preuves de l'existence de Dieu/1981	10
61	ROUBAUD Jacques/La Boucle/1993	10

Tableau II-4 : Classement par nombre d'occurrences décroissant des 61 œuvres du corpus FRANTEXT intégral (1975-1000) présentant au moins 10 occurrences de **conscience(s)**

Dans le tableau II-4, j'ai mis en évidence les œuvres 1, 3, 4, 5, 15 et 19 qui sont 6 tomes des *Œuvres complètes* de Pierre Mendès-France, car elles montrent le poids considérable de cet auteur dans le corpus. Comment peut-on s'en assurer et repérer d'autres auteurs très présents à travers différentes œuvres ? La méthode consiste pour cela à transférer la liste des œuvres dans un fichier WORD (ou WRITER) et d'y remplacer les « / » par des tabulations (^t pour WORD, \t pour WRITER). On peut convertir le pavé ainsi obtenu en tableau et vérifier que – sauf cas exceptionnel – les références entrent dans trois colonnes (auteur – œuvre – année de parution). On renvoie alors ce pavé dans la feuille EXCEL après avoir ajouté les colonnes nécessaires. En pratiquant alors un tri dans la colonne des auteurs on peut regrouper toutes les œuvres d'un même auteur, quelle que soit leur année de parution. Ce tri permet de vérifier<sup>7</sup> qu'à travers l'ensemble

<sup>7</sup> On aurait évidemment obtenu le même résultat en effectuant une requête dans FRANTEXT précisant, en dehors de la période 1975-2000 le nom de Pierre Mendès-France (cf. Figure I-2, « sélection par auteur »)



de ses *Œuvres complètes*, P. Mendès-France n'emploie pas moins de 335 fois **conscience(s)** :

Auteur	Œuvre	Année	Occ.
MENDÈS-FRANCE Pierre	Œuvres complètes. 4. Pour une République moderne. 1955-1962.	1987	89
MENDÈS-FRANCE Pierre	Œuvres complètes. 6. Une vision du monde. 1974-1982.	1990	66
MENDÈS-FRANCE Pierre	Œuvres complètes. 5. Préparer l'avenir. 1963-1973.	1989	61
MENDÈS-FRANCE Pierre	Œuvres complètes. 3. Gouverner, c'est choisir. 1954-1955.	1986	58
MENDÈS-FRANCE Pierre	Œuvres complètes. 2. Une politique de l'économie. 1943-1954.	1985	33
MENDÈS-FRANCE Pierre	Œuvres complètes. 1. S'engager. 1922-1943.	1984	28
<b>TOTAL</b>			<b>335</b>

Tableau II-5 : Volume des occurrences de **conscience(s)** dans les Œuvres complètes de Pierre Mendès-France

Le tri permet également de repérer André Malraux comme le second auteur employant le plus l'item **conscience(s)** dans la période considérée, mais le nombre total d'occurrences est beaucoup moins élevé (82) :

MALRAUX André	La Corde et les souris. VI : [Lazare]	1976	39
MALRAUX André	Antimémoires	1976	20
MALRAUX André	Inauguration de l'exposition "André Malraux et le musée imaginaire"	1976	5
MALRAUX André	La Corde et les souris. IV : [Les Chênes qu'on abat]	1976	5
MALRAUX André	La Corde et les souris. V : [La Tête d'obsidienne]	1976	5
MALRAUX André	La Corde et les souris. I, II, III : [Hôtes de passage]	1976	4
MALRAUX André	Le Règne du malin	1976	3
MALRAUX André	Fragments d'une réflexion sur les "Antimémoires"	1976	1
<b>TOTAL</b>			<b>82</b>

Tableau II-6 : Volume des occurrences de **conscience(s)** dans les oeuvres d'André Malraux (toutes parues en 1976)

Il est à noter que pour apprécier l'importance réelle de cet item chez ces auteurs, c'est-à-dire sa fréquence relative au nombre de mots graphiques des œuvres en question, il faudrait encore rechercher dans la fonction « visualiser le corpus » le nombre de mots de chaque œuvre, comme je l'ai fait à propos de la forme **humanité** dans le tableau I-5 (section I-4).

### **II-3. Méthode de recherche des cooccurrents à gauche et à droite de l'item CONSCIENCE**

Sur un ensemble de plus de 2000 occurrences, sauf dans le cas où l'on ne recherche que certains segments supposés répétés par une analyse préalable, on peut effectuer un tri alphabétique du cotexte droit (colonne E), mais il est impossible d'effectuer un tri alphabétique à partir de la fin du cotexte gauche, ce qui interdit la recherche de cooccurrences à gauche.

Pour cette raison nous allons tirer profit d'un concordancier en usage libre, AntConc<sup>8</sup>, qui présente une fonction « clusters » (que je traduirai par « agrégats ») permettant de dégager rapidement les cooccurrents à gauche et à droite de l'item examiné. Pour chaque agrégat, AntConc fournit son rang et sa fréquence en fonction des paramètres de la requête (agrégat à gauche / à droite / à gauche et à droite avec un empan de n mots graphiques). L'objectif de la recherche des agrégats est de mettre en

<sup>8</sup> Concordancier téléchargeable librement sur le site ► <http://www.antlab.sci.waseda.ac.jp/software.html>

évidence des chaînes syntaxiquement autosuffisantes et interprétables. Nous retiendrons donc en priorité les chaînes constituant un syntagme potentiel et nous consulterons les occurrences dans le fichier EXCEL afin de vérifier la validité de l'interprétation effectuée sur la base du seul agrégat.

La figure II-3 montre le début des résultats de la recherche des agrégats avec l'item **conscience** (au singulier) à droite, la taille des agrégats étant fixée à 2 et leur fréquence minimale à 3. Avec 511 occurrences, le premier rang est occupé **par la conscience**. La première paire ADJ + **conscience** est **mauvaise conscience** (4<sup>e</sup> rang, fréquence 96), qui devance **bonne conscience** (7<sup>e</sup> rang, fréquence 73). La première paire V + **conscience** est **prendre conscience** (5<sup>e</sup> rang, fréquence 82).

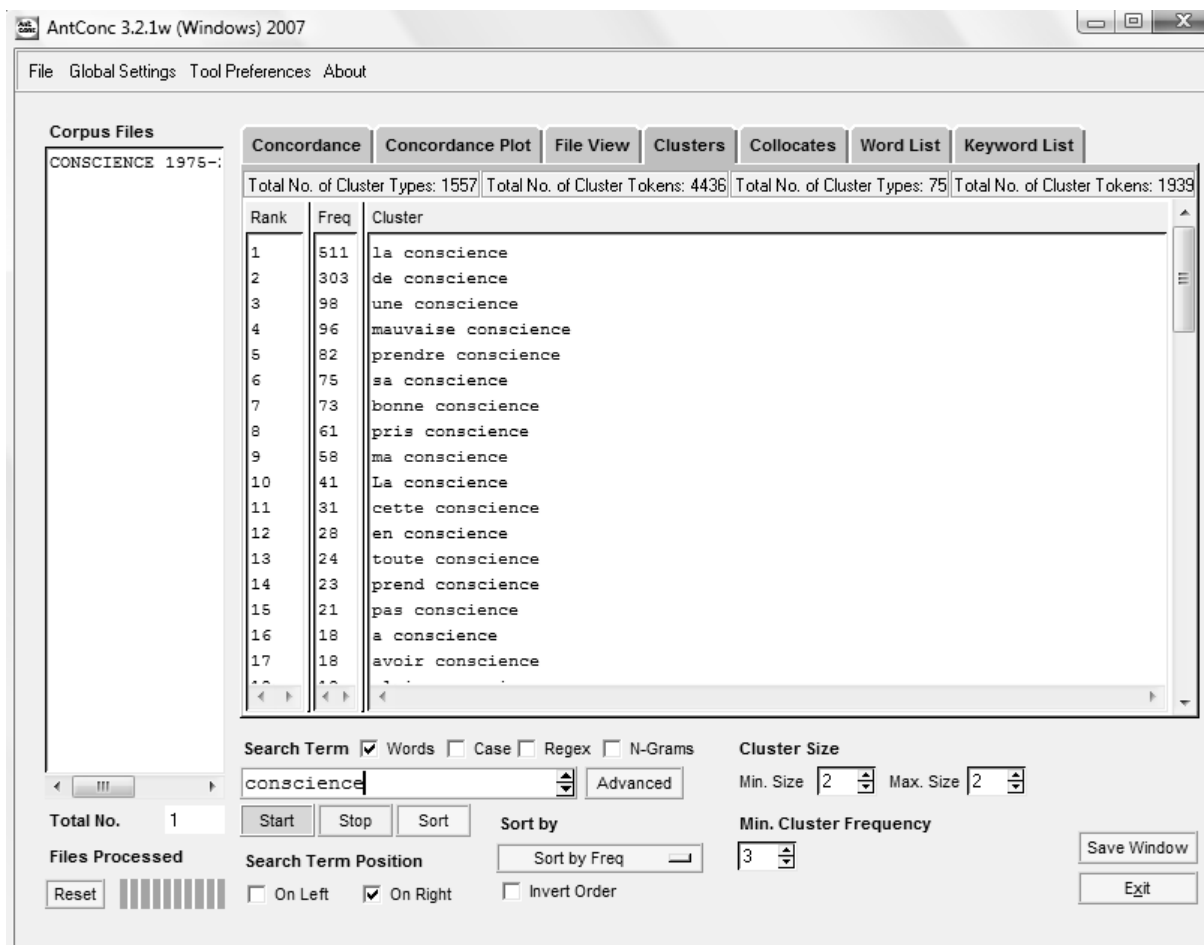


Figure II-3 : Fenêtre du concordancier AntConc en recherche d'agrégats avec l'item **conscience** à droite

Du côté gauche, on enregistre les paires suivantes :

► **ADJ + conscience**

Les paires **mauvaise conscience** et **bonne conscience** distancent largement les cinq paires suivantes.

4	96	mauvaise ~	27	10	claire ~	74	3	ultime ~
7	73	bonne ~	45	5	autre ~			
18	18	pleine ~	46	5	nouvelle ~			

► **V + conscience**

- avec V = **avoir**

9 formes du verbe **avoir** figurent directement à gauche de **conscience** pour un total de 84 cooccurrences.

16	18	a ~	23	11	avait ~	60	4	ont ~
17	18	avoir ~	35	7	avais ~	61	3	ait ~
22	13	ai ~	37	7	eut ~	62	3	ayant ~

- avec V = **prendre**

12 formes du verbe **prendre** figurent immédiatement à gauche de **conscience** pour un total de 229 cooccurrences.

5	82	prendre ~	25	11	prit ~	51	5	prenait ~
8	61	pris ~	31	9	prennent ~	52	5	prenant ~
14	23	prend ~	33	8	prenne ~	71	3	prenions ~
21	14	prends ~	50	5	prenais ~	72	3	prennes ~

- avec V = **reprendre**

3 formes du verbe **reprendre** figurent immédiatement à gauche de **conscience** pour un total de 15 cooccurrences.

40	7	reprit ~	53	5	reprendre ~	73	3	reprend ~
----	---	----------	----	---	-------------	----	---	-----------

- avec V = **perdre**

Enfin 2 formes du verbe **perdre** figurent immédiatement à gauche de **conscience** pour un total de 13 cooccurrences.

39	7	perdre ~	42	6	perdu ~
----	---	----------	----	---	---------

Du côté droit les seules paires autosuffisantes sont de la forme **conscience** + ADJ . 3 agrégats présentent au moins 20 occurrences, 3 en présentent entre 8 et 13 et 13 entre 3 et 5. Les paires **conscience** + Adjectif se répartissent essentiellement entre celles comportant un adjectif psychologique : *conscience tranquille, aiguë, claire, sensible, morale, scrupuleuse*, et celles comportant un adjectif exprimant une collectivité : *conscience civique, nationale, professionnelle, politique, collective, universelle, sociale, bourgeoise*. Seules les paires *conscience fondamentale* et *conscience nouvelle* n'entrent dans aucune de ces deux classes.

8	32	~ nationale	68	4	~ sociale
10	29	~ professionnelle	82	3	~ bourgeoise
13	20	~ tranquille	83	3	~ civique
18	13	~ politique	84	3	~ commune
24	11	~ aiguë	87	3	~ fondamentale
33	8	~ collective	89	3	~ morale
51	5	~ claire	90	3	~ nouvelle
52	5	~ individuelle	92	3	~ scrupuleuse
56	5	~ universelle	108	3	~ collective
67	4	~ sensible			

La recherche d'agrégats constitués de trois mots graphiques délivre à gauche 13 triplets de la forme **N de conscience** avec un emploi dominant de **prise(s) de conscience** (58 occurrences au singulier et 6 au pluriel). **Directeur(s) de conscience** apparaît au singulier et au pluriel (20 occurrences au total) ainsi que **objecteur(s) de conscience** (9 occurrences au total).

► **N de conscience**

11	58	prise de ~	53	11	liberté de ~	135	4	crise de ~
18	26	examen de ~	81	7	objection de ~	152	4	objecteurs de ~
23	23	acquies de ~	97	6	prises de ~	158	4	problème de ~
26	21	cas de ~	107	5	directeurs de ~			
40	14	directeur de ~	114	5	objecteur de ~			

Du côté droit on rencontre 15 triplets de la forme **conscience de N/ProN/INF** avec **conscience de classe** en tête (18 occurrences) :

► **conscience de N/Pron/INF**

6	18	~ de classe	37	5	~ de tout	67	3	~ de quoi
20	7	~ du fait	38	5	~ du danger	68	3	~ de toute
21	7	~ du temps	46	4	~ d'exister	69	3	~ des maisons
27	6	~ de ceux	49	4	~ de vivre	70	3	~ du corps
28	6	~ de soi	50	4	~ du monde	71	3	~ du vivant

La recherche d'agrégats d'un empan supérieur délivre deux expressions figées : **se donner bonne conscience** (7 occ.) et **en son âme et conscience** (3 occ.).

Les syntagmes repérés, en général figés, peuvent comporter des extensions. Ainsi **examen de conscience** est qualifié deux fois par l'adjectif **politique** et une fois par les adjectifs **collectif**, **mental** ou **poétique** :

▪ **examen de conscience politique**

s'était forgé une certaine conception du pouvoir et de son bon usage comme exécutant prépondérant du général de Gaulle. Mai 1968, en l'obligeant à un **examen de conscience politique**, sa mise sur la touche en lui consentant le temps de la réflexion, l'amènèrent à redéfinir clairement cette conception et sa morale de l'action. [625 BLONDIN Antoine/Ma vie entre les lignes/1982]

J'ai longuement hésité à me lancer dans la description condensée que je viens de faire pour l'inclure dans ces pages : d'un **examen de conscience politique**. Ce n'est pas parce que j'avais été tenté d'omettre ce qu'il valait mieux oublier. [2051 ROUBAUD Jacques/Poésie : récit/2000]

▪ **examen de conscience mental**

dans la vie ! Vendredi 30 décembre Que dire ? Que je suis en plein travail et que je n'ai pas l'intention de me déranger pour faire un retour sur moi-même. Mon **examen de conscience mental**, je le fais à peu près tous les jours, et souvent plusieurs fois par jour, me demandant si ce que j'ai dit était vrai, si ce que j'ai fait était bien. Selon les [1602 HUGUENIN Jean-René/Journal/1993]

▪ **examen de conscience poétique**

À intervalles réguliers, faire son examen de conscience **poétique**. (À cette consigne esthétique, je donnais une formulation morale, inspirée d'une pratique qui s'associait dans mon esprit aux plus sévèrement moraux [2043 ROUBAUD Jacques/Poésie : récit/])

#### ▪ **examen de conscience collectif**

Un grand examen de conscience **collectif** n'a jamais été plus nécessaire. Mais c'est aux hommes d'État, aux vrais hommes d'État, de prendre l'initiative et de donner l'exemple. [930 MENDÈS-FRANCE Pierre/OEuvres complètes. 2. Une politique de l'économie. 1943-1954./1985]

### **II-4. Les cooccurrents de l'item CONSCIENCE dans les Oeuvres complètes de Pierre Mendès-France**

Nous avons vu précédemment (section II-2) que sur les 2055 occurrences de l'item **conscience(s)** dans notre corpus, 335 sont extraits des *Œuvres complètes* de P. Mendès-France. A priori, ce volume semble suffisant pour dégager des cooccurrences immédiates régulières. Cependant l'examen de la partie des cooccurrences régulières préalablement repérées pratiquées par ce seul auteur se révèle assez décevant. Parmi les cooccurrents à gauche, seul un petit groupe figure plus de 5 fois :

prendre pris	53	examen de	9	en toute	5
prise de	16	cas de	8	mauvaise	5
pleine	13	claire	6		

Et les cooccurrences régulières à droite sont inexistantes à l'exception de **conscience de + DetPoss + N** qui figure 17 fois. Quant à l'examen des cooccurrences régulières chez le second auteur (André Malraux avec 82 occurrences), il ne livre aucun résultat.

La conclusion est donc que l'examen des segments répétés chez un auteur n'a des chances d'être instructif qu'à plusieurs conditions : il doit s'agir d'un auteur prolifique, il faut envisager l'ensemble de son œuvre sans restriction de période (quitte ultérieurement à identifier des segments répétés propres à une époque ou une œuvre particulière), il faut s'appuyer sur des fréquences relatives et enfin l'examen sera plus pertinent s'il est comparatif, soit avec un autre auteur, soit avec un ensemble d'auteurs de la même époque et présentant des caractéristiques similaires. De ce point de vue, le nouveau classement des œuvres par genres détaillés proposé par la base FRANTEXT intégrale (voir la figure I-4) peut constituer un outil efficace.

### **II-5. Bilan**

Cette étude constitue un exercice d'entraînement au repérage des cooccurrents récurrents d'un item examiné, en l'occurrence ici **conscience(s)** rencontré 2055 fois dans le corpus de la base FRANTEXT intégrale entre 1975 et 2000. La collecte des segments répétés effectuée à l'aide du concordancier AntConc peut être remaniée en regroupant les cotextes comportant des formes différentes d'un même lemme. De la sorte les segments répétés constitué d'un verbe immédiatement suivi de l'item conscience présentent **prendre** en premier avant **avoir**, tandis que **reprendre** et **perdre** sont marginalisés :

prendre ~	229
avoir ~	84
reprendre ~	15
perdre ~	13

Un dernier point d'ordre sémantique : on distingue traditionnellement entre la conscience psychologique (en all. **Bewußtsein**) et la conscience morale (en all. **Gewissen**). Certains cooccurrents induisent clairement l'une des valeurs, par ex. **directeur de conscience, cas de conscience** ou **examen de conscience** induisent la valeur morale, inversement **prise de conscience** ou **état(s) de conscience** (4 résultats seulement) orientent vers la valeur psychologique. Cependant il apparaît clairement au fil de certains extraits que la conscience morale est un cas particulier de la conscience psychologique : les **cas de conscience** (moraux) proviennent de la **prise de conscience** (psychologique) d'une faute, qui conduit les plus nombreux à un **examen de conscience** et certains (surtout à l'époque du français classique) à se rechercher un **directeur de conscience**. Il y a donc manifestement des cas de passage d'une valeur à l'autre, comme par exemple l'extrait suivant, qui réclamerait un cotexte plus large pour une interprétation assurée : c'est manifestement un cas de ségrégation antisémite qui est évoqué, l'homme âgé en question portant l'étoile juive. L'auteur joue successivement sur quatre contextes de **cas** : **lui exposer notre cas** → **quel cas ?** → **pas même un cas de conscience** → **Un cas d'inconscience, plutôt,** et où **conscience** à valeur morale dans **cas de conscience** entre en opposition avec la valeur psychologique d'**inconscience**.

Nous n'osions le suivre de trop près. Ou pire encore, lui exposer notre cas. Quel cas ? Pas même un cas de **conscience**. Un cas d'inconscience, plutôt. Boris sans moi devant l'horreur de ce contraste : un homme âgé, élégant, pensif, étiqueté Juif, visiblement. [SCHREIBER Boris/Un silence d'environ une demi-heure/1996]

## Chapitre III

### La variation des segments répétés à travers les siècles illustrée par les cooccurrents de l'item *mœurs*

Dans ce troisième chapitre, je me propose de montrer que l'examen des occurrences d'un item lexical à travers une longue période de la base FRANTEXT intégrale (en l'occurrence quatre siècles de 1600 à 1999 étudiés par décennies) peut se révéler instructif pour l'histoire des idées. L'item choisi est *mœurs*, qui a pour particularité d'avoir connu une recrudescence de ses emplois dans la base entre la deuxième moitié du 18<sup>e</sup> et la première moitié du 19<sup>e</sup> siècle.

Dans la section III-1 je développerai l'hypothèse qu'il s'agira de tester au long du chapitre. J'évoquerai ensuite dans la section III-2 la question de l'enregistrement de la fréquence absolue et relative de l'item à travers les siècles et du poids particulier de certaines des œuvres appartenant au corpus en examinant particulièrement deux périodes, la décennie 1750-59 et les deux décennies 1820-29 et 1830-39 en raison d'un décrochement dans la fréquence relative de l'item entre ces deux décennies. La section III-3 sera consacrée à l'étude de la variation des segments répétés typiquement normatifs ou descriptifs durant la période-cadre (1600-2000). Les segments examinés seront sélectionnés à partir de trois sondages au 17<sup>e</sup>, 18<sup>e</sup> et 19<sup>e</sup> siècle. nous verrons que l'hypothèse est partiellement validée dans la mesure où il y a bien eu un effet de recouvrement entre emplois normatifs et descriptifs entre approximativement 1750 et 1850, mais d'une part les emplois descriptifs repérés ont un caractère fugace (ils tendent à disparaître dès la fin du 19<sup>e</sup> siècle) et d'autre part de nombreux segments répétés ne peuvent pas être classés comme typiquement normatifs ou descriptifs.

#### **III-1. L'hypothèse à tester sur l'évolution de la fréquence relative de l'item MŒURS depuis le 17<sup>e</sup> siècle**

Durant cette période longue de quatre siècles, on n'observe pas de changement de désignation de l'item *mœurs*. En d'autres termes, on n'est pas en présence d'un problème classique de polysémie, mais d'un changement de point de vue ou de type de discours sur les mœurs. Le premier type de discours sera qualifié de « normatif » : il s'agit de louer les bonnes mœurs, de vilipender les mauvaises mœurs et de développer le thème éternel de la dégradation des mœurs caractéristique du temps présent. Le second peut être qualifié de « descriptif ». Sur la base de la progression des connaissances sur les mœurs d'un nombre de plus en plus grand de peuples à travers le monde (indiens d'Amérique, peuples d'Océanie, etc.) et de la sécularisation de la société française, le discours devient comparatif, la référence religieuse judéo-chrétienne perd de l'importance et le thème de la dégradation des mœurs régresse rapidement.

En se basant sur la fréquence relative et non absolue de l'item à travers les œuvres saisies dans chacune des 40 décennies, on observe une croissance de ses emplois à partir de 1730-1750 et une décroissance symétrique à partir de 1840-60. L'hypothèse que je chercherai à tester est que cette recrudescence temporaire serait due à un effet

d'empilement des deux types de discours au moment où (avant la sécularisation produite par la Révolution Française, mais aussi sous la restauration<sup>9</sup>) le discours normatif reste présent, tandis que le discours nouveau, largement associé aux Lumières et tout particulièrement à l'entreprise universalisante de Montesquieu dans *L'esprit des lois*, progresse dans les esprits éclairés et va s'imposer à partir de la Monarchie de Juillet.

Mais comment cette hypothèse peut-elle être validée par une analyse lexicale ? L'idée que je défends est que le profil lexical de l'item **mœurs**, tel qu'on peut l'identifier à partir de la variation de ses cooccurrents, a varié au long de ces quatre siècles. Il est vraisemblable qu'à partir de l'époque où le second type d'emploi descriptif s'impose, l'emploi normatif de **mœurs** laisse progressivement la place à celui de **savoir-vivre** (qui perd tout fondement religieux), mais cette seconde hypothèse reste à tester.

L'hypothèse de l'empilement transitoire des deux types de points de vue, normatif et descriptif, peut se synthétiser ainsi :

point de vue / type de discours	jusqu'à ≈ 1750	de ≈ 1750 à ≈ 1850	depuis ≈ 1850
normatif	<b>OUI</b>	<b>OUI</b>	
descriptif		<b>OUI</b>	<b>OUI</b>

### **III-2. Fréquence absolue et relative de l'item MŒURS à travers la base FRANTEXT**

Nous avons vu dans le chapitre I comment calculer la fréquence relative d'un segment à partir de sa fréquence absolue fournie par le traitement de la requête et du nombre de mots du corpus sélectionné. Sur cette base, j'ai enregistré par décennies sur quatre siècles le nombre d'occurrences de l'item **mœurs** et j'ai calculé sa fréquence relative en fonction du volume de chaque corpus décennal (section III-2.1) avant d'examiner de plus près dans quelles oeuvres de périodes-clés l'item **mœurs** est surreprésenté en fréquence absolue et en fréquence relative (section III-2.2).

#### **III-2.1. La variation des fréquences à travers les siècles**

Il ressort du tableau III-1 (p.30) et de la figure III-1 (p.31) que la fréquence relative de **mœurs** a connu une croissance relativement régulière à partir de la décennie 1710-1719 avec un pic<sup>10</sup> exceptionnel en 1750-1759, puis un premier décrochement en 1830-1839 et un second en 1850-1859. A partir de cette décennie, les valeurs redeviennent similaires à celles de la période antérieure à 1710 et régressent peu à peu jusqu'à la valeur la plus faible ( $1,67 \cdot 10^{-5}$ ) pour la décennie la plus récente. Les fréquences absolues sont mentionnées dans le tableau III-1 puisque c'est sur leur base qu'on peut calculer les fréquences relatives en les divisant par le volume du corpus en mots graphiques, mais en elles-mêmes, elles sont non significatives, puisque le volume de chaque décennie est extrêmement variable, compris entre ≈ 563 000 mots graphiques

<sup>9</sup> Parmi les valeurs d'Ancien Régime à restaurer, les mœurs constituent évidemment une pièce de choix.

<sup>10</sup> L'observation d'un pic de fréquence est largement dépendante de la durée des périodes comparées. Si l'on remplace les périodes de 10 ans par des périodes de 20 ans, la courbe est lissée. Pour le pic de 1750-59, que l'on prenne l'une des doubles décennies 1740-59 ou 1750-69, l'effet de pic disparaît.



entre 1760 et 1769 et  $\approx 14\,633\,000$  mots graphiques entre 1840 et 1849. Les 77 occurrences de l'item en 1760-69 représentent donc une fréquence relative bien supérieure ( $13,66 \cdot 10^{-5}$ ) aux 1189 occurrences de 1840-49 ( $8,13 \cdot 10^{-5}$ ).

Décennie	Corpus	Fréquence		Décennie	Corpus	Fréquence	
		absolue	relative			absolue	relative
<b>1600-1609</b>	1555323	96	<b>6,17</b>	<b>1800-1809</b>	3 455 052	425	<b>12,30</b>
<b>1610-1619</b>	1299933	81	<b>6,23</b>	<b>1810-1819</b>	2 737 383	414	<b>15,12</b>
<b>1620-1629</b>	3033453	96	<b>3,16</b>	<b>1820-1829</b>	4 489 937	644	<b>14,34</b>
<b>1630-1639</b>	2479361	70	<b>2,82</b>	<b>1830-1839</b>	8 185 083	669	<b>8,17</b>
<b>1640-1649</b>	1553584	75	<b>4,83</b>	<b>1840-1849</b>	14 633 104	1189	<b>8,13</b>
<b>1650-1659</b>	2256022	112	<b>4,96</b>	<b>1850-1859</b>	7 780 740	406	<b>5,22</b>
<b>1660-1669</b>	1831879	130	<b>7,10</b>	<b>1860-1869</b>	7 737 367	331	<b>4,28</b>
<b>1670-1679</b>	2150130	91	<b>4,23</b>	<b>1870-1879</b>	7 412 556	255	<b>3,44</b>
<b>1680-1689</b>	2996396	138	<b>4,61</b>	<b>1880-1889</b>	7 491 507	288	<b>3,84</b>
<b>1690-1699</b>	2684459	179	<b>6,67</b>	<b>1890-1899</b>	6 651 236	210	<b>3,16</b>
<b>1700-1709</b>	978656	32	<b>3,27</b>	<b>1900-1909</b>	5 279 283	195	<b>3,69</b>
<b>1710-1719</b>	2401759	180	<b>7,49</b>	<b>1910-1919</b>	6 042 569	142	<b>2,35</b>
<b>1720-1729</b>	1120153	108	<b>9,64</b>	<b>1920-1929</b>	12 990 467	353	<b>2,72</b>
<b>1730-1739</b>	4934740	539	<b>10,92</b>	<b>1930-1939</b>	15 016 545	438	<b>2,92</b>
<b>1740-1749</b>	3081036	402	<b>13,05</b>	<b>1940-1949</b>	11 541 676	257	<b>2,23</b>
<b>1750-1759</b>	4176849	1102	<b>26,38</b>	<b>1950-1959</b>	11 176 559	238	<b>2,13</b>
<b>1760-1769</b>	563616	77	<b>13,66</b>	<b>1960-1969</b>	14 916 188	336	<b>2,25</b>
<b>1770-1779</b>	5429458	1243	<b>22,89</b>	<b>1970-1979</b>	6 992 004	151	<b>2,16</b>
<b>1780-1789</b>	5141173	905	<b>17,60</b>	<b>1980-1989</b>	9 827 796	248	<b>2,52</b>
<b>1790-1799</b>	3 426 086	666	<b>19,44</b>	<b>1990-1999</b>	6 654 688	111	<b>1,67</b>

Tableau III-1 : Fréquence absolue et relative(\*E<sup>-5</sup>) de la lexie *mœurs* par décennies entre 1600 et 1999

### III-2.2. La variation des fréquences à travers les œuvres

Chacun des corpus décennaux est constitué d'un nombre variable d'œuvres relevant de genres différents (la base catégorisée distingue 10 genres différents). On peut donc supposer que l'intérêt des différents auteurs pour l'item *mœurs* est très variable et lorsqu'on enregistre une brusque recrudescence on s'attend à ce que l'intérêt particulier de certains auteurs pour cet item ressorte de la fréquence relative de cet item dans l'œuvre en cause. Nous allons pouvoir le vérifier à deux périodes clé, celle du pic d'occurrences de 1750-59 et celle du décrochement autour de 1830 (1820-29 vs. 1830-39).

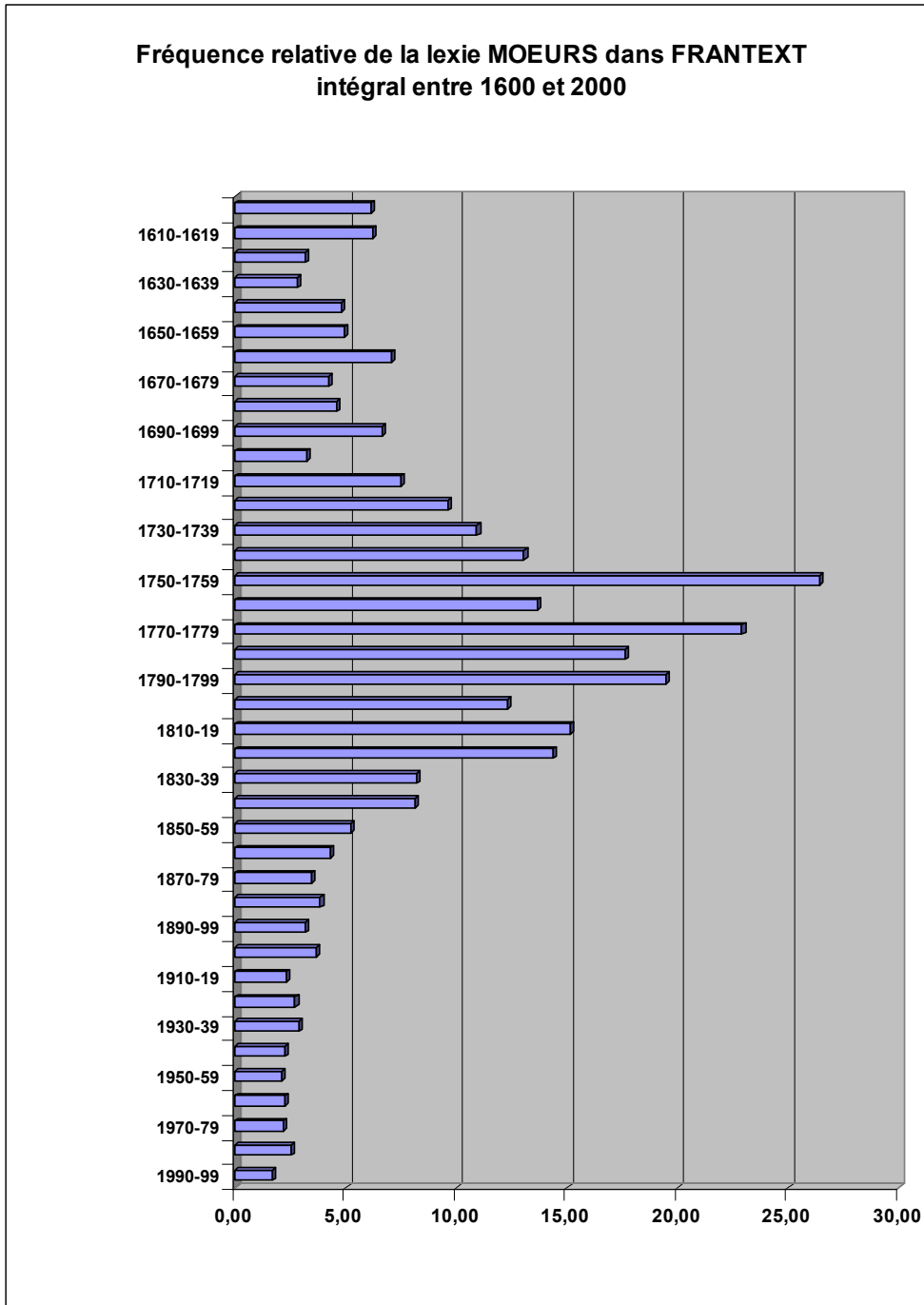


Figure III-1

### I. Le pic d'occurrences de 1750-59

Le tableau III-2 liste dans l'ordre chronologique toutes les œuvres enregistrées comme publiées entre 1750 et 1759. Certaines œuvres présentent un nombre d'occurrences relativement élevé et pour les principales j'ai indiqué la fréquence relative ( $\times 10^{-4}$ ). Chronologiquement, on rencontre d'abord les *Considérations sur les mœurs de ce siècle*

de Charles Pinot Duclos<sup>11</sup> (1751), puis *L'ami des hommes ou Traité de la population* (1755) du marquis Victor de Mirabeau (physiocrate, le père du révolutionnaire)<sup>12</sup> et la même année *L'esprit des loix* en quatre tomes de Montesquieu, œuvre majeure du libéralisme économique et de l'approche descriptiviste et comparative en termes de mœurs. Un an plus tard paraît *l'Essay sur l'histoire générale et sur les mœurs et sur l'esprit des nations* de Voltaire qui relève de la philosophie de l'histoire et enfin en 1758 *De l'esprit* du philosophe matérialiste Claude-Adrien Helvétius et la *Lettre à M. d'Alembert sur les spectacles* de Jean-Jacques Rousseau qui relève du genre du pamphlet.

Auteur	Œuvre	Date	Fréq. abs	Volume œuvre	Fréq. rel.
FOUGERET DE MONBRON Jean-Louis	Le Cosmopolite ou le Citoyen du monde	1750	2		
FOUGERET DE MONBRON Jean-Louis	Margot la ravaudeuse	1750	2		
ROUSSEAU Jean-Jacques	Discours sur les sciences et les arts	1750	25		
ALEMBERT Jean d'	Discours préliminaire de l'Encyclopédie	1751	3		
CRÉBILLON fils	Ah quel conte !	1751	5		
DIDEROT Denis	Lettre sur les sourds et muets à l'usage de ceux qui entendent et qui parlent	1751	3		
<b>DUCLOS Charles PINOT</b>	<b>Considérations sur les mœurs de ce siècle</b>	<b>1751</b>	<b>58</b>	<b>42947</b>	<b>13,51</b>
DUCLOS Charles PINOT	Mémoires pour servir à l'histoire des mœurs du XVIIIe siècle	1751	16		
PRÉVOST L'Abbé	Lettres angloises ou Histoire de miss Clarisse Harlove [trad.] : t. 1	1751	18		
PRÉVOST L'Abbé	Lettres angloises ou Histoire de miss Clarisse Harlove [trad.] : t. 2	1751	17		
PRÉVOST L'Abbé	Lettres angloises ou Histoire de miss Clarisse Harlove [trad.] : t. 3	1751	11		
PRÉVOST L'Abbé	Lettres angloises ou Histoire de miss Clarisse Harlove [trad.] : t. 4	1751	9		
PRÉVOST L'Abbé	Lettres angloises ou Histoire de miss Clarisse Harlove [trad.] : t. 5	1751	3		
PRÉVOST L'Abbé	Lettres angloises ou Histoire de miss Clarisse Harlove [trad.] : t. 6	1751	10		
BONNET Charles	Essai de psychologie	1754	4		
<b>MIRABEAU marquis Victor de</b>	<b>L'Ami des hommes ou Traité de la population</b>	<b>1755</b>	<b>212</b>	<b>269538</b>	<b>7,87</b>
MONTESQUIEU	Correspondance : 1716-1755	1755	3		
<b>MONTESQUIEU<sup>13</sup></b>	<b>De l'esprit des loix : t. 1</b>	<b>1755</b>	<b>58</b>	<b>264390</b>	<b>8,25</b>
<b>MONTESQUIEU</b>	<b>De l'esprit des loix : t. 2</b>	<b>1755</b>	<b>45</b>		
<b>MONTESQUIEU</b>	<b>De l'esprit des loix : t. 3</b>	<b>1755</b>	<b>97</b>		
<b>MONTESQUIEU</b>	<b>De l'esprit des loix : t. 4</b>	<b>1755</b>	<b>18</b>		
PALISSOT DE MONTENOY Charles	Le Cercle, ou les Originaux	1755	2		

<sup>11</sup> (1704-1772) Secrétaire perpétuel de l'Académie Française. Il est vu d'abord comme un moraliste et malgré certains traits de modernisme, il eut des relations difficiles avec les encyclopédistes.

<sup>12</sup> Victor de Mirabeau est avant tout un économiste et son discours relève donc du point de vue « descriptif »

<sup>13</sup> Fréquence absolue totale pour *De l'esprit des loix* : 218. Le volume total de 264 390 mots graphiques est étonnamment bas et suggère soit une erreur d'enregistrement, soit une sélection.

PRÉVOST L'Abbé	Nouvelles lettres anglaises ou Histoire du chevalier Grandisson [trad.]	1755	25		
ROUSSEAU Jean-Jacques	Discours sur l'origine et les fondements de l'inégalité parmi les hommes	1755	11		
VOLTAIRE	L'Orphelin de la Chine	1755	7		
VOLTAIRE	La Pucelle d'Orléans	1755	2		
HOLBACH Paul-Henri d'	Le Christianisme dévoilé ou Examen des principes et des effets de la religion chrétienne	1756	23		
<b>VOLTAIRE</b>	<b>Essay sur l'histoire générale et sur les moeurs et sur l'esprit des nations</b>	<b>1756</b>	<b>125</b>	<b>419389</b>	<b>2,98</b>
VOLTAIRE	La Religion naturelle	1756	4		
ARGENSON René-Louis d'	Considérations sur le gouvernement ancien et présent de la France	1757	14		
ARGENSON René-Louis d'	Journal et mémoires : t. 9 (1755)	1757	2		
ARGENSON René-Louis d'	Journal et mémoires : t. 9 (1756-1757)	1757	4		
DIDEROT Denis	Entretiens sur le Fils naturel	1757	7		
DIDEROT Denis	Le Fils naturel ou les Épreuves de la vertu	1757	5		
PALISSOT DE MONTENOY Charles	Petites lettres sur de grands philosophes	1757	7		
DIDEROT Denis	De la poésie dramatique	1758	20		
DIDEROT Denis	Le Père de famille	1758	7		
<b>HELVÉTIUS Claude-Adrien</b>	<b>De l'Esprit</b>	<b>1758</b>	<b>74</b>	<b>183 939</b>	<b>4,02</b>
<b>ROUSSEAU Jean-Jacques</b>	<b>Lettre à M. d'Alembert sur les spectacles</b>	<b>1758</b>	<b>82</b>	<b>53 410</b>	<b>15,35</b>
ALEMBERT Jean d'	Lettre à M. Rousseau	1759	10		
DIDEROT Denis	Essais sur la peinture	Salon de 1759	5		
Salon de 1761	Salon de 1763	1759-1766	12		
LACURNE DE SAINTE-PALAYE	Mémoires sur l'ancienne chevalerie considérée comme un établissement politique et militaire	1759	6		
VOLTAIRE	Candide, ou l'Optimisme	1759	18		

Tableau III-2. Liste des œuvres appartenant au corpus de FRANTEXT intégral pour la décennie 1750-59 et mise en évidence de la fréquence relative de l'item *mœurs* dans 6 de ces œuvres

Le classement de ces six œuvres par fréquence décroissante de l'item *mœurs* diffère selon qu'on prend comme base la fréquence absolue ou relative. Dans le classement par fréquence absolue, c'est Montesquieu qui vient en premier avec 218 occurrences dans l'ensemble des quatre tomes de *De l'esprit des lois*, suivi de Mirabeau avec 212 occurrences dans *L'Ami des hommes ou Traité de la population*.

Auteur	Œuvre	Date	Fréq.abs.
<b>MONTESQUIEU</b>	<b>De l'esprit des lois : t. 1-4</b>	<b>1755</b>	<b>218</b>
<b>MIRABEAU marquis Victor de</b>	<b>L'Ami des hommes ou Traité de la population</b>	<b>1755</b>	<b>212</b>
VOLTAIRE	Essay sur l'histoire générale et sur les moeurs et sur l'esprit des nations	1756	125
ROUSSEAU Jean-Jacques	Lettre à M. d'Alembert sur les spectacles	1758	82
HELVÉTIUS Claude-Adrien	De l'Esprit	1758	74
DUCLOS Charles PINOT	Considérations sur les moeurs de ce siècle	1751	58

En revanche, en fréquence relative, Montesquieu ( $8,25 \cdot 10^{-4}$ ) et Mirabeau (7,87) passent derrière le pamphlet de Jean-Jacques Rousseau (15,35) et le traité de morale de Charles Pinot Duclos (13,51).

Auteur	Œuvre	Date	Fréq.rel.
<b>ROUSSEAU Jean-Jacques</b>	<b>Lettre à M. d'Alembert sur les spectacles</b>	<b>1758</b>	<b>15,35</b>
DUCLOS Charles PINOT	Considérations sur les moeurs de ce siècle	1751	13,51
MONTESQUIEU	De l'esprit des loix : t. 1-4	1755	8,25
MIRABEAU marquis Victor de	L'Ami des hommes ou Traité de la population	1755	7,87
HELVÉTIUS Claude-Adrien	De l'Esprit	1758	4,02
VOLTAIRE	Essay sur l'histoire générale et sur les moeurs et sur l'esprit des nations	1756	2,98

Pour établir assurément le type de discours représenté par chacune de ces œuvres, il faudrait identifier les cotextes gauche et droit (immédiats ou à distance) répétitifs dans chacune de ces œuvres.

## II. Le premier décrochement autour de 1830

Pour la décennie immédiatement antérieure au décrochement observé, qui comporte au total 644 occurrences de la lexie *mœurs*, cinq auteurs l'emploient plus de 4 fois. En fréquence absolue, Emmanuel de Las Cases occupe la première place avec 130 occurrences, suivi de Joseph Joubert, François Guizot et Paul-Louis Courier.

Auteur 1820-1829	Fréq.abs.
<b>LAS CASES Emmanuel de</b>	<b>130</b>
JOUBERT Joseph	77
GUIZOT François	76
COURIER Paul-Louis	60
STENDHAL	41

En fréquence relative c'est Joseph Joubert<sup>14</sup> qui arrive en tête :

Auteur 1820-1829	Volume corpus	Fréq.rel.
<b>JOUBERT Joseph</b>	<b>105 515</b>	<b>7,3 E<sup>-4</sup></b>
GUIZOT François	117 473	6,4 E <sup>-4</sup>
STENDHAL	118917	3,4 E <sup>-4</sup>
COURIER Paul-Louis	223 066	2,7 E <sup>-4</sup>
LAS CASES Emmanuel de	750 191	1,7 E <sup>-4</sup>

Pour la décennie immédiatement postérieure au décrochement (1830-39), qui comporte au total 669 occurrences, c'est Alexis de Tocqueville, avec *De la démocratie en Amérique* qui arrive en tête, aussi bien en fréquence absolue (159 occurrences) qu'en fréquence relative ( $8,34 \cdot 10^{-4}$ ). Dans le premier cas, il précède Honoré de Balzac, dans le second, Alphonse de Lamartine vient s'insérer entre eux deux.

- Classement par fréquence absolue décroissante :

Auteur 1830-1839	Fréq.abs.
<b>TOCQUEVILLE Alexis de</b>	<b>159</b>
BALZAC Honoré de	124
LAMARTINE Alphonse de	79
MICHELET Jules	35
SAND George	38
STENDHAL	37

<sup>14</sup> Moraliste et essayiste (1754-1824). L'œuvre en cause est *Pensées, essais, maximes et correspondance* (1824)

- Classement par fréquence relative décroissante :

Auteur 1830-1839	Volume du corpus	Fréq.rel E <sup>-4</sup>
<b>TOCQUEVILLE Alexis de</b>	<b>190 678</b>	<b>8,34</b>
MICHELET Jules	193 899	1,80
BALZAC Honoré de	917 948	1,35
LAMARTINE Alphonse de	759 620	1,04
STENDHAL	928 716	0,40
SAND George	1 020 132	0,37

### **III-3. La variation des cooccurrences de l'item MOEURS entre 1600 et 1999 dans la base FRANTEXT**

Je vais maintenant chercher à tester l'hypothèse selon laquelle deux types d'emploi de l'item **mœurs** se succèdent avec une période de superposition depuis le 17<sup>e</sup> siècle : un emploi normatif (à base religieuse) et un emploi descriptif lié aux Lumières et à l'émergence des sciences sociales.

Afin de dégager des illustrations de ces deux types d'emploi d'une fréquence suffisante, je procède d'abord à trois sondages dans trois intervalles de temps présentant entre 600 et 800 occurrences de l'item, à savoir la deuxième moitié du 17<sup>e</sup> siècle (1650-1699), la décennie de la Révolution française (1790-1799) et la décennie suivant la révolution de 1848 (1850-1859).

Dans chacun de ces trois sondages, je répartiss les cooccurrences de l'item **mœurs** entre trois groupes : « normatif », « descriptif » et « autres », c'est-à-dire ni clairement normatif, ni clairement descriptif.

#### ▪ **Sondage 1 : 1650-1699**

##### **Emplois normatifs**<sup>15</sup>

bonnes ~	42
corruption des ~   corrompre* les ~	16
règle règlement des ~   régler leurs ~	10
mauvaises ~	9
réformation réforme des ~   réformer* les ~	8
pureté des ~	5

##### **Emplois descriptifs**

qui regarde(nt) les ~	8
rapport se rapporter aux ~	5

##### **Autres emplois**

facilité de (ses) ~	6
suivre* les ~ (de son pays)	6
~ de l'église	5
~ des chrétiens	5
~ des israélites   juifs	5
~ des païens	5
~ de son pays (suivre les ~)	5
~ des habitants	5

<sup>15</sup> L'astérisque note le statut de lemme de l'infinitif qui couvre différentes formes conjuguées.

▪ **Sondage 2 : 1790-99**

**Emplois normatifs**

bonnes ~	30
mauvaises ~	14
corruption des ~	11
changer* les ~	5
adoucir* les ~	4
la pureté de leurs ~	4
~ austères	4
~ sauvages	4

**Emplois descriptifs**

peindre* les ~	6
~ publiques	6
~ nationales	4

**Autres emplois**

sans ~	9
la simplicité de ses ~	4
~ de l'autre	5

▪ **Sondage 3 : 1850-1859**

**Emplois normatifs**

bonnes ~	26
mauvaises ~	4

**Emplois descriptifs**

peinture des ~	4
----------------	---

**Autres emplois**

anciennes ~	4
~ du peuple	4

Ces trois sondages dégagent plusieurs segments répétés dominants à l'une ou l'autre de ces trois périodes :

**Emplois normatifs**

*bonnes ~*  
*corruption des ~*  
*corrompre\* les ~*  
*mauvaises ~*

**Emplois descriptifs**

*regarder\* les ~*  
*peindre\* les ~*  
*connaître\* les ~*  
*peinture des ~*

On recherche ensuite ces segments répétés dans la période-cadre considérée, 1600-2000. Les cooccurrences de type normatif les mieux représentées sont (sur un total de 13622 occurrences de **mœurs**) :

<i>bonnes ~</i> :	451
<i>mauvaises ~</i> :	169

*pureté PREP DET ~* : 118  
*corrompre\* DET ~* : 99  
*corruption des ~* : 85

Les premières cooccurrences de type descriptif sont pour la même période-cadre :

*peindre\* les ~* : 89  
*~ publiques* : 88  
*connaître\* DET ~* : 58  
*peinture des ~* : 50

En fréquence relative du segment par rapport aux occurrences de **mœurs**, la distribution du segment **bonnes mœurs** accuse une baisse régulière de la première moitié du 17<sup>e</sup> siècle à la première moitié du 19<sup>e</sup>, avant de remonter légèrement dans la deuxième moitié du 20<sup>e</sup>. Elle valide l'hypothèse de l'emploi décroissant d'un segment à valeur normative. La fréquence relative de **bonnes mœurs** étant beaucoup plus élevée que celle des quatre autres segments du même type, je lui consacre un graphique indépendant (Fig.).

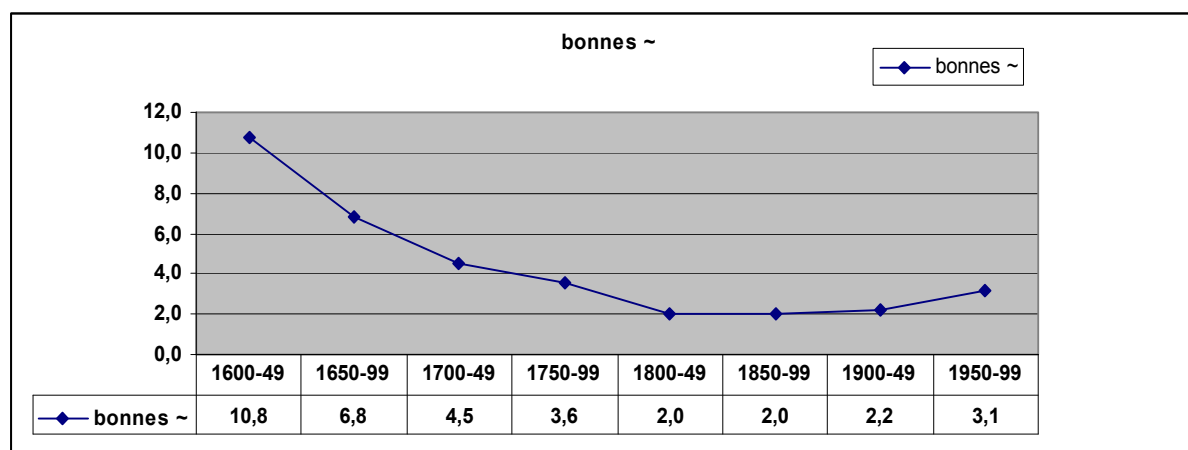


Figure III-3 : Évolution de la fréquence relative du segment répété **bonnes mœurs**

Pour les quatre autres segments du type normatif, les fréquences relatives varient de 2,4% à 0,1%. Le segment **mauvaises mœurs** commence par présenter une chute rapide entre la première moitié du 17<sup>e</sup> et la première moitié du 18<sup>e</sup>, avant de connaître des fluctuations entre 0,7 et 1,7%. Seule la période 1600-1649 présente une fréquence relative un peu plus élevée. La courbe ne dégage donc pas d'évolution significative.

La distribution du segment **corruption des mœurs** varie entre 1,3 et 0,1 avec une première moitié de stagnation autour de 1% entre 1600-1649 et 1750-1759 et une seconde moitié en baisse régulière de 0,4 à 0,1%. La distribution est conforme à l'hypothèse.

Le segment **corrompre\* DET mœurs** part d'1% pour atteindre son maximum de 1,7% autour de 1700, il décline dans la deuxième moitié du 18<sup>e</sup> siècle et stagne entre 0,1 et 0,5% depuis lors. Cette distribution est conforme à l'hypothèse.

Enfin le segment **pureté PREP DET mœurs** a une fréquence relative très basse (0,2%) dans la période 1600-1649 avant d'atteindre son maximum entre la deuxième moitié du 17<sup>e</sup> et la deuxième moitié du 18<sup>e</sup> siècle, puis décline autour de 6%. Cette distribution est également conforme à l'hypothèse à l'exception du premier demi-siècle où elle est plus basse que prévu.



Au total quatre des cinq segments retenus pour le type normatif ont, en fréquence relative, une distribution par demi-siècle globalement conforme à l'hypothèse d'un déclin progressif de ce type d'emploi, mais seul le segment le plus représenté, **bonnes mœurs**, se caractérise par une baisse régulière et importante (de 10,8% à 2,0% jusqu'à la fin du 19<sup>e</sup> siècle (cf. Figure II-3).

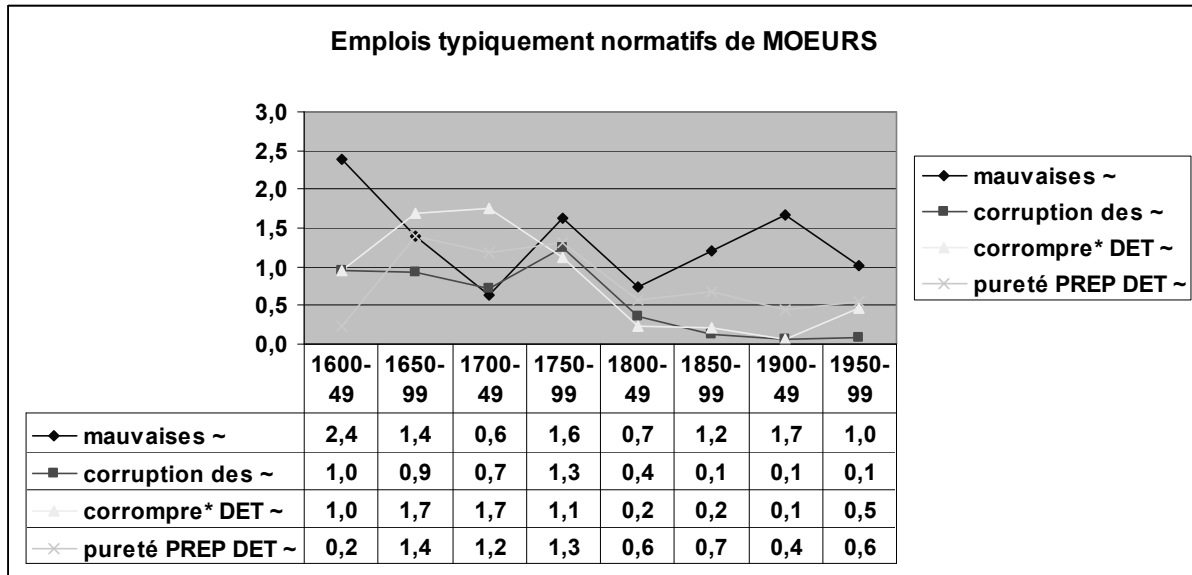


Figure II-3

Les quatre segments représentant le type d'emploi descriptif présentent en revanche des profils très distincts (cf. Figure II-4). La fréquence relative du segment **peinture des mœurs** stagne au plus bas jusqu'à la première moitié du 18<sup>e</sup> siècle et présente ensuite une courbe en cloche dont le maximum se situe dans la première moitié du 19<sup>e</sup> siècle pour retomber au niveau-plancher au 20<sup>e</sup> siècle. On peut y voir un effet de mode centré sur la première moitié du 19<sup>e</sup> siècle.

Le segment verbal correspondant **peindre\* DET mœurs** présente une courbe en cloche de la première moitié du 17<sup>e</sup> à la deuxième moitié du 19<sup>e</sup> avec un maximum dans la deuxième moitié du 18<sup>e</sup> siècle (0,9%) et connaît un léger rebond dans la première moitié du 20<sup>e</sup> siècle. Les deux courbes ont donc leur apogée à la même époque.

En dehors de la première période (1600-1649) où il a une fréquence relative très basse, le segment **connaître\* les mœurs** fluctue autour de 0,4% (maximum 0,7%, minimum 0,1%). Ce segment ne valide pas l'hypothèse d'une progression significative à partir de la fin du 18<sup>e</sup> siècle.

Enfin le segment **mœurs publiques** est totalement absent jusqu'à la première moitié du 18<sup>e</sup> siècle avant de connaître une envolée à 1,1% (44 occurrences) et de retomber à 0,2% en 1900-1949 et 0,3% en 1950-59. Cette collocation est donc particulièrement prisée durant le demi-siècle pré-révolutionnaire et révolutionnaire.

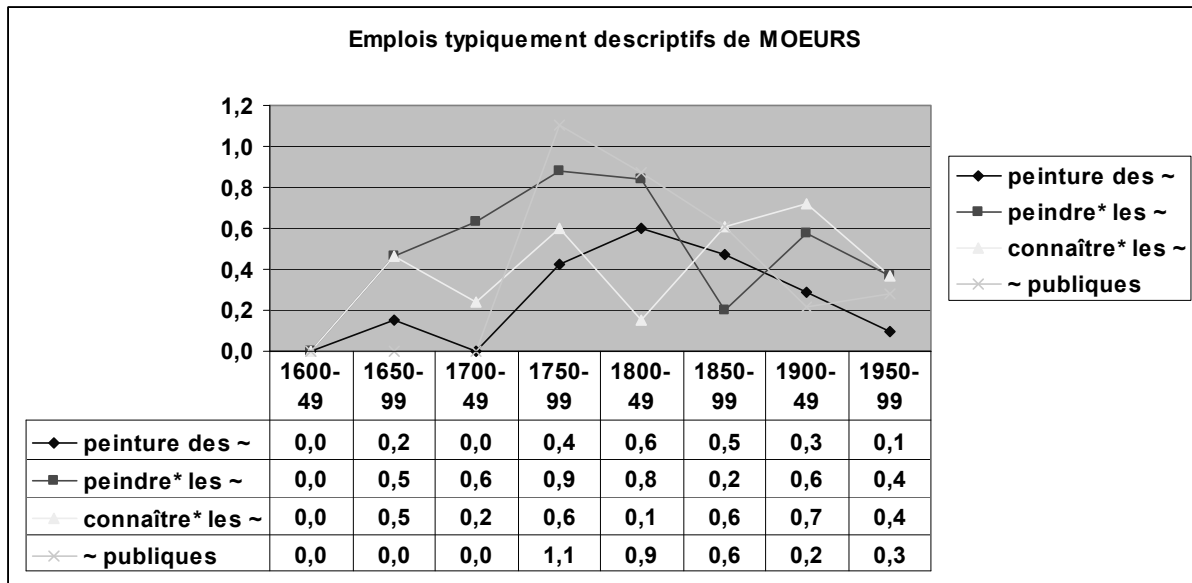


Figure II-4

### III-4. Bilan

Globalement, les segments à valeur normative **bonnes mœurs**, **corruption des mœurs**, **corrompre\* DET mœurs** et **pureté des mœurs** déclinent à partir au plus tard de la première moitié du 19<sup>e</sup> siècle et valident donc l'hypothèse de la désaffection à l'égard de ce type d'emploi à partir du 19<sup>e</sup> siècle. Pour les segments à valeur descriptive, aucun ne progresse jusqu'au 20<sup>e</sup> siècle, mais **peinture des mœurs**, **peindre\* les mœurs** et **mœurs publiques** atteignent leur maximum entre 1750 et 1850, ce qui valide l'hypothèse de l'apparition d'une nouvelle vision des mœurs autour de la Révolution française. Mais on observe en même temps un effet de mode pour ces trois segments, puisqu'en fréquence relative ils déclinent rapidement à partir de la moitié du 19<sup>e</sup> siècle. Il s'avère également que certains auteurs ont joué un rôle déterminant dans cette évolution du discours normatif vers le discours descriptif. Cependant, même dans l'œuvre de Tocqueville, qui ne compte pas moins de 159 occurrences de **mœurs**, les segments répétés autosuffisants sont quasiment inexistantes. Il n'est donc pas possible de vérifier si le discours de Tocqueville dans *De la démocratie en Amérique* (1835) sur les mœurs présente des traits distinctifs par rapport au corpus du même demi-siècle (1800-1849). La même observation s'applique à *L'esprit des lois* de Montesquieu malgré ses 218 occurrences de **mœurs**.

## Chapitre IV

### Une incursion en lexicométrie combinatoire : **Adj<sub>couleur</sub> de N<sub>émotion</sub>**

Dans ce dernier chapitre je vais examiner, sans entrer dans les détails, la notion d'attraction ou de répulsion statistiquement significative entre deux items lexicaux ou entre un item lexical et une construction. Je le ferai à partir d'un cas élémentaire, la relation (d'attraction, de répulsion, ou de composition neutre) entre différents adjectifs de couleur et différents noms d'émotion dans le cadre de la construction {Adj<couleur> de N<émotion>}. Ce type de relation entre deux types d'items lexicaux coconstituants d'une construction a été examiné pour l'anglais par Anatol Stefanowitsch & Stefan Gries dans un article paru en 2005 sous le titre « *Covarying collexemes* » comme un cas particulier de leur théorie statistique des constructions connue sous le nom d'analyse des « collostructions »<sup>16</sup>. Je vais donc commencer par résumer brièvement la démarche « collostructionnelle » (section IV-1), puis le mode d'analyse des « co-lexèmes covariants » (section IV-2) et je proposerai finalement (section IV-3) l'application au français indiquée ci-dessus.

#### **IV-1. La démarche 'collostructionnelle'**

Stefanowitsch & Gries (2003) commencent par rappeler que dans une approche constructionnelle du langage, la grammaire est constituée de signes conçus comme des paires forme-sens et ne diffère donc pas fondamentalement du lexique (p.212). Sur cette base, une construction grammaticale est dotée d'un sens propre qu'il s'agit d'identifier<sup>17</sup>. Les composants particuliers de la construction (les *slots*) ont un effet d'attraction sur un ensemble particulier de lexèmes ou se combinent uniquement avec cet ensemble. Le principe de l'analyse « collostructionnelle » est d'examiner dans une construction particulière quels lexèmes sont fortement attirés ou repoussés par un composant particulier dans la construction (c'est-à-dire figurent plus ou moins fréquemment qu'on est en droit de l'attendre sur la base d'un test statistique). Les lexèmes attirés par une construction particulière sont appelés « collexèmes » de cette construction ; inversement, une construction associée à un lexème particulier peut être appelée un « collostruit » (*collostruct*) ; la combinaison d'un « collexème » et d'un « collostruit » sera appelée une « collostruction ». (p.215)

Les auteurs illustrent ce rapport entre un « collostruit » et un « co-lexème » entre autres à l'aide de la construction **N waiting to happen** (fr. *N ne demandant qu'à se produire*). Ils montrent que les décomptes bruts de fréquence combinatoire présentent un biais : ils notent essentiellement des co-occurrences avec des mots fonctionnels, lesquels sont souvent de peu d'utilité par exemple pour l'identification de différences

---

<sup>16</sup> Le *Dictionnaire des combinaisons de mots – Les synonymes en contexte* de D. Lefur et al. aux éditions du Robert fonde ses combinaisons sur des calculs apparentés.

<sup>17</sup> cf. François & Sénéchal 2008 pour une application au français

subtiles entre quasi-synonymes. Leur méthode consiste en revanche à prendre en compte dans la combinaison de **N waiting to happen** avec un N particulier, par exemple (*an*) *accident*, des données suivantes :

- i. le nombre d'occurrences (dans une partie catégorisée et analysée syntaxiquement du *British National Corpus*) de la construction particulière ***an accident waiting to happen*** (14) ;
- ii. le nombre d'occurrences de **N<≠ *an accident*> waiting to happen** (21), ce qui donne au total 35 occurrences de **N waiting to happen** ;
- iii. le nombre des autres constructions dont (*an*) *accident* est un constituant (8606), ce qui donne un total de 8620 constructions auxquelles est associé (*an*) *accident* ;
- iv. le nombre des constructions auxquelles sont associés<sup>18</sup> – sauf erreur d'interprétation – les N différents de *accident* co-lexèmes de **N waiting to happen** (10 197 659), ce qui au total donne un total de référence de 10 206 300 constructions particulières<sup>19</sup>.

Le résultat est représenté sur le tableau croisé IV-1

constructions	<b><i>accident</i></b>	<b>~<i>accident</i></b>	Totaux bruts
<b><i>N waiting to happen</i></b>	<b>14</b>	<b>21</b>	<b>35</b>
autres constructions	8 606	10 197 659	10 206 265
Totaux de colonnes	8 620	10 197 680	10 206 300

Tableau IV-1 : co-occurrences entre la construction ***N waiting to happen*** et l'item (*an*) *accident*

Ensuite à l'aide d'un test statistique (le test dit « Fisher exact »), ils montrent que seuls deux items constituent une collocation (c'est-à-dire entretiennent une relation d'attraction mutuelle) avec ***N waiting to happen*** : ***accident*** (14 occ., 2,12 E-34) et ***disaster*** (12 occ., 1,36 E-33), tous les autres co-lexèmes ne figurant qu'une fois dans la construction (≈E-5/E-2). Les deux points intéressants sont évidemment d'une part la relation de cohyponymie entre les deux items et d'autre part la connotation négative des deux types d'événement en question.

Cette connotation négative se retrouve (comme l'avait préalablement suggéré Stubbs 1995) avec les co-lexèmes du verbe *to cause*. A priori ce verbe pourrait régir des compléments désignant des situations évaluées aussi bien comme positives que comme négatives. Mais tel n'est pas le cas : les cinq premiers co-lexèmes objets de *to cause* sont par ordre de fréquence brute décroissante : *problem* (22 occ.) ; *damage* (9) ; *harm* (3) ; *havoc* (3) et *distress* (3), tous à connotation négative

Stefanowitsch & Gries terminent leur article par deux observations particulièrement intéressantes :

- i. Contrairement à A. Goldberg (1999) qui examine l'acquisition des constructions en termes de fréquences brutes, les auteurs estiment que la force d'attraction collocationnelle est plus prometteuse que les fréquences brutes sur ce point, car la

<sup>18</sup> C'est en tout cas ainsi que j'interprète leur calcul.

<sup>19</sup> Ces dernières valeurs sont obtenues par des évaluations qui ne sont elles-mêmes possibles que parce que le corpus utilisé est non seulement catégorisé mais aussi analysé syntaxiquement ; il faut y voir seulement des ordres de grandeur.

méthode collostructionnelle calcule le degré d'association entre le co-lexème et le 'collostruit', ce qui correspond à la notion de *cue validity* (validité comme signal) en psychologie cognitive, déterminant décisif de la formation d'un prototype, en l'occurrence d'une association co-lexème / collostruit. L'analyse collostructionnelle fournit à l'analyste les expressions qui sont hautement caractéristiques du sémantisme de la construction et qui, de ce fait, sont aussi pertinentes pour l'apprenant. (p.237)

- ii. Ils proposent également d'inverser la perspective pour déterminer dans quelles constructions un verbe particulier figure avec une fréquence significativement distinctive. Hanks (1996) appelle cet éventail de constructions le profil comportemental d'un verbe et on peut ajouter que les travaux récents de Peter Blumenthal sur le 'profil lexical' en français (2006, 2007, 2008, 2009 ; Blumenthal et al. 2005), menés à l'aide du test dit « log-likelihood », vont dans le même sens.

#### **IV-2. L'analyse des 'collexèmes covariants'**

Dans Stefanowitsch & Gries (2005 :9), les auteurs font valoir qu'une construction délivre fréquemment deux ou plusieurs positions associables à des ensembles d'items dont on souhaite examiner les propriétés sémantiques relativement l'une à l'autre. Il cherche donc à identifier la force de l'association entre des paires d'items lexicaux figurant dans deux différentes positions de la même construction, c'est-à-dire qu'ils examinent de quelle manière les items lexicaux figurant dans une position covarient avec ceux de l'autre position.

La méthode est de même nature que celle de la section IV-1. Le tableau IV-2 donne le cadre général du calcul.

		<b>A</b>	<b>B</b>
		M<pos2> (mot M en position 2)	~M<pos2> (tous les autres mots en position 2)
<b>1</b>	L<pos1> (mot L en position 1)	Fréq (L<pos1> & M<pos2>)	Fréq (L<pos1> & ~M<pos2>)
<b>2</b>	~L<pos1> (tous les autres mots en pos1>)	Fréq (~L<pos1> & M<pos2>)	Fréq (~L<pos1> & ~M<pos2>)

Tableau IV-2 : Principe général de la table des fréquences attendues (ou table de contingence)

Cela implique de déterminer pour chaque co-lexème potentiel L figurant dans la position 1 quel co-lexème potentiel en position 2 cooccure avec celui-ci plus souvent qu'attendu. Cela est effectué en comparant les fréquences réelles de cooccurrence avec les fréquences attendues sur la base d'une table de distribution 2 à 2 (cf. Tableau IV-3).

<i>X fools Y into thinking</i>		A	B	C
		<i>think</i>	autres verbes	Total des lignes
<b>1</b>	<i>fool</i>	<b>46</b> (7)	31 (70)	<b>77</b>
<b>2</b>	autres verbes	101 (140)	1408 (1369)	1509
<b>3</b>	Total des colonnes	<b>147</b>	1439	<b>1586</b>

Tableau IV-3 : Table de comparaison des fréquences observées et attendues pour la combinaison entre les items *to fool* en V1 et *to think* en V2 dans la construction *X fools Y into thinking*

L'analyse porte sur la covariance entre V1 et V2 dans le cadre d'une famille de constructions particulière de l'anglais : *X V1 Y into V2-ing* qui s'analyse sémantiquement comme « X<+hum> accomplit un type d'acte décrit par V1 sur Y<+hum> de telle sorte que Y accomplit un type d'action décrit par V2 »<sup>20</sup>. La covariance maximale est observée entre le verbe *fool* en position 1 (V1) et le verbe *think* en position 2 (V2-ing). Le sens global de la construction *X fools Y into thinking Z* est donc « X dupe Y avec pour effet que Y pense Z ».

Les chiffres en italique reproduisent les valeurs observées et les chiffres entre parenthèses les valeurs attendues. La cellule A1 spécifie que 46 instances de *X fools Y into thinking* ont été enregistrées dans le corpus étudié contre 7 instances attendues. Le total des occurrences de *fool* en position 1 (V1) de la construction est 77 (cellule C1). Si le nombre des instances de la construction *X fools Y into thinking* avait été 7 comme attendu, celui des occurrences d'autres verbes que *think* en position 2 aurait donc été 77-7 = 70 (cellule B1). Comme le nombre effectif est 46, le nombre des autres verbes que *think* est 77-46 = 31 (cellule B1). Le total des occurrences de *think* en position 2 (V2-ing) est 147 (cellule A3). Le nombre des occurrences d'autres verbes en position 2 serait 140 si le nombre des instances de la construction était 7, mais comme il s'élève à 46, il n'est plus que 101 (cellule A2).

Sachant que le nombre total des occurrences de la construction (quels que soient V1 et V2) est 1586 (cellule C3), il en résulte d'une part que le nombre des occurrences de la construction avec *fool* en V1 et un autre verbe que *think* en V2 est 1586-147 (nombre des occurrences de *think* en position V2) = 1439 (cellule B3) et d'autre part que le nombre des occurrences de la construction avec *think* en V2 et un autre verbe que *fool* en V1 est 1596 - 77 (nombre des occurrences de *fool* en position V1) = 1509 (cellule C2) et de ce fait le nombre des occurrences de la construction avec un autre verbe que *fool* en V1 et un autre verbe que *think* en V2 est 1408, résultat de la soustraction de 1439 (cellule B3) - 31 (cellule B1) ou de la soustraction de 1509 (cellule C2) - 101 (cellule A2). Ce nombre diffère du nombre attendu (1369). Les 4 nombres attendus indiqués dans les cellules A1, A3, C1 et C3 sont obtenus par la méthode présentée dans la section IV-1.

L'affirmation selon laquelle la construction *X fools Y into thinking Z* est celle qui présente la covariance maximale entre V1 et V2 résulte de l'application à la table XX du « Test exact de Fisher-Yates », le test que les auteurs considèrent comme le plus adapté (cf. Wiechmann 2008 pour un jugement nuancé sur ce test). Les trois constructions suivantes sont *X misleads Y into thinking Z*, *X misleads Y into believing Z* et *X deceives Y into thinking Z* qui présentent une remarquable unité sémantique : dans chacune de ces constructions V1 spécifie une tromperie et V2 une activité cognitive

<sup>20</sup> « De telle sorte » est préféré à « afin » car à l'aspect perfectif la construction affirme l'accomplissement de l'action décrite par V2.

(penser ou croire), ce qui conforte le principe de 'compatibilité sémantique' entre les verbes occupant les deux positions défendu par les auteurs.

### IV-3. Essai d'application élémentaire

Nous allons maintenant appliquer la méthode des 'collexèmes covariants' de Stefanowitsch & Gries (2005) à un objet d'étude très élémentaire, la famille des constructions de la forme {Adj<couleur> de N<émotion>}. La première question qui se pose est celle des chaînes à rechercher et du corpus dans lequel on a des chances de les trouver. La seconde question est celle du test de significativité de la combinaison entre Adj et N qu'on peut leur appliquer.

Sur le premier point la solution que j'ai adoptée consiste à effectuer des requêtes par 'expressions régulières' (c'est-à-dire par des formules comportant des informations d'ordre grammatical) dans la base FRANTEXT catégorisée. Plusieurs adjectifs de couleur intuitivement susceptibles de se combiner avec un nom d'émotion sont retenus ainsi que plusieurs noms d'émotion susceptibles de se combiner avec un adjectif de couleur. Dans le premier cas les requêtes ont la forme « &mADJECTIF de &e(g=S) », par exemple « &mblême de &e(g=S) » où S symbolise toute forme d'un substantif. Dans le second cas elle ont la forme &e(g=A) de NOM, par exemple « e&(g=A) de rage » où A symbolise toute forme d'un adjectif.

Sur le second point j'ai retenu un test d'un maniement simple, celui du  $\chi^2$  disponible sur tableur EXCEL, dont le seul inconvénient est qu'il est inopérant en-dessous d'un nombre d'occurrences d'environ 6.

Les requêtes sont effectuées dans l'ensemble de la base FRANTEXT catégorisée, soit de 1830 à 2000. Le nombre des occurrences obtenu pour certains adjectifs de couleur ou nom d'émotion est trop faible pour qu'ils soient retenus, par exemple *livide* et *frayeur*. Ensuite pour chaque requête on repère le nombre de cooccurrences avec un des noms d'émotion ou un des adjectifs de couleur retenus. Ainsi par exemple la requête « e&(g=A) de colère » délivre 318 occurrences parmi lesquelles on compte 49 occurrences de *rouge(s) de colère*, 25 de *pâle(s) de colère*, 8 de *blême(s) de colère* et 4 de *vert(e)(s) de colère*. Inversement la requête « &mrouge de &e(g=S) » délivre les mêmes 50 occurrences de *rouge(s) de colère* et 5 occurrences de *rouge(s) de fureur*.

Le tableau IV-4 donne le nombre de chacune des cooccurrences pour les 16 combinaisons des 4 adjectifs de couleur, *blême*, *pâle*, *rouge* et *verte* et pour les 4 noms d'émotion *colère*, *fureur*, *peur* et *rage*. Le total est pour chaque ligne ou chaque colonne le total des cooccurrences rencontrées pour les 4 adjectifs et les 4 noms.

	<i>blême de ~</i>	<i>pâle de ~</i>	<i>rouge de ~</i>	<i>vert de ~</i>	Total
<i>~ de colère</i>	8	25	50	4	<b>87</b>
<i>~ de fureur</i>	1	6	5	0	<b>12</b>
<i>~ de peur</i>	5	4	0	7	<b>16</b>
<i>~ de rage</i>	8	10	0	7	<b>25</b>
<b>Total</b>	<b>22</b>	<b>45</b>	<b>55</b>	<b>18</b>	<b>140</b>

Tableau IV-4 : Combinaisons observées d'items de la construction {Adj<couleur> de N<émotion>}

L'ordre des combinaisons à partir des adjectifs en valeur décroissante est **rouge de ~** (50) > **pâle de ~** (25) > **blême de ~** (8) > **vert de ~** (5). A partir des noms d'émotion c'est **~ de colère** (87) > **~ de rage** (25) > **~ de peur** (16) > **~ de fureur** (12). Trois combinaisons ne sont pas représentées : **vert de fureur**, **rouge de peur** et **rouge de rage** et cinq ont un nombre d'occurrences insuffisant pour appliquer le test : **blême de fureur**, **blême de peur**, **pâle de peur**, **rouge de fureur** et **vert de colère**. Les combinaisons effectivement testables sont donc finalement au nombre de huit : **blême de colère / de rage**, **pâle de colère / de fureur / de rage**, **rouge de colère**, **vert de peur / de rage**.

Pour chacune de ces 8 combinaisons, le test compare les quatre valeurs produites par la combinaison {Adj<sub>i</sub> vs. Adj<sub>≠i</sub> x Nom<sub>j</sub> vs. Nom<sub>≠j</sub>} aux valeurs attendues calculées par la formule  $L_{1/2} * C_{1/2} / N$  où L<sub>1</sub> symbolise la ligne du nom d'émotion examiné, L<sub>2</sub> la ligne de tous les autres noms d'émotion, C<sub>1</sub> la colonne de l'adjectif de couleur examiné et C<sub>2</sub> la colonne de tous les autres adjectifs de couleur (cf. Tableau IV-5).

		<b>Adj<sub>i</sub></b>		<b>Adj<sub>≠i</sub></b>		
		observations	attendues	observations	attendues	Total
<b>Nom<sub>j</sub></b>	<b>VAL(i,j)</b>	$[L_1 * C_1 / N]$		<b>VAL(≠i,j)</b>	$[L_1 * C_2 / N]$	L <sub>1</sub>
<b>Nom<sub>≠j</sub></b>	<b>VAL(i,≠j)</b>	$[L_2 * C_1 / N]$		<b>VAL(≠i,≠j)</b>	$[L_2 * C_2 / N]$	L <sub>2</sub>
		C <sub>1</sub>		C <sub>2</sub>		<b>N</b>

Tableau IV-5 : Double table des observations et des attendues destinée au calcul du test du  $\chi^2$

Je commencerai par la valeur observée la plus élevée, celle de **rouge de colère** (50 occurrences), cf. Tableau IV-6.

#### **rouge de colère**

		<b>rouge</b>		<b>~rouge</b>		
		observations	attendues	observations	attendues	Total
<b>colère</b>	<b>50</b>		34,18	<b>37</b>	52,82	87
<b>~colère</b>	<b>5</b>		20,82	<b>48</b>	32,18	53
<b><math>\chi^2 = 1,65 \text{ E-}08</math></b>		55		85		<b>140</b>

Tableau IV-6 : Double table des observations et des attendues pour la construction **rouge de colère**

On s'attend en fonction du nombre total d'occurrences de **rouge de colère / de fureur / peur / rage** et du nombre total d'occurrences de **blême / pâle / rouge / vert de colère** à ce que la combinaison **rouge de colère** présente 34,18 occurrences sur 140, or elle apparaît beaucoup plus fréquemment (50 occurrences). Le test renvoie une valeur de 1,65 E-08 qui signifie un risque négligeable de proximité de la moyenne, donc une valeur significative.



**blême de colère**

	<b>blême</b>		<b>~blême</b>		Total
	observations	attentes	observations	attentes	
<b>colère</b>	<b>8</b>	13,67	<b>79</b>	73,33	87
<b>~colère</b>	<b>14</b>	8,33	<b>39</b>	44,67	53
$\chi^2 = 0,0066$	22		118		<b>140</b>

Tableau IV-7 : Double table des observations et des attentes pour la construction **blême de colère**

Dans le cas de **blême de colère**, le nombre d'occurrences observées est sensiblement inférieur au nombre des occurrences attendues (8 contre 13,37). La valeur renvoyée par le test représente donc une relation significative, mais cette fois-ci non pas d'attraction, mais au contraire de répulsion. C'est-à-dire que **blême** et **colère** sont deux items lexicaux qui tendent à se repousser dans la construction Adj de N.

Il est intéressant de comparer ce résultat avec celui de **blême de rage** : le nombre des combinaisons observées est le même (8), mais le rapport est très différent entre le nombre des occurrences de **blême** (22) et celui de **rage** (25) vs. **colère** (87). De ce fait les attentes sont inverses (on attend seulement 3,93 occurrences de **blême de rage** contre les 8 observées. La valeur renvoyée par le test est significative et cette fois dans le sens de l'attraction et non de la répulsion. Donc **blême** et **rage** sont deux items lexicaux qui s'attirent significativement dans la construction Adj de N.

**blême de rage**

	<b>blême</b>		<b>~blême</b>		Total
	observations	attentes	observations	attentes	
<b>rage</b>	<b>8</b>	3,93	<b>17</b>	21,07	25
<b>~rage</b>	<b>14</b>	18,07	<b>101</b>	96,93	115
$\chi^2 = 0,01356$	22		118		<b>140</b>

Tableau IV-8 : Double table des observations et des attentes pour la construction **blême de rage**

Il reste encore deux combinaisons significatives dans le sens de l'attraction, c'est **vert de rage** et **vert de peur** :

**vert de rage**

	<b>vert</b>		<b>~vert</b>		Total
	observations	attentes	observations	attentes	
<b>rage</b>	<b>7</b>	3,21	<b>18</b>	21,79	25
<b>~rage</b>	<b>11</b>	14,79	<b>104</b>	100,21	115
$\chi^2 = 0,0256$	18		122		<b>140</b>

Tableau IV-9 : Double table des observations et des attentes pour la construction **vert de rage**

Le nombre des combinaisons attestées (7) est sensiblement supérieur au nombre attendu (3,21) et la valeur renvoyée par le test est significative dans le sens de l'attraction.

**vert de peur**

	<b>vert</b>		<b>~vert</b>		Total
	observations	attentes	observations	attentes	
<b>peur</b>	<b>7</b>	2,06	<b>17</b>	13,94	16
<b>~peur</b>	<b>11</b>	15,94	<b>101</b>	108,06	124
$\chi^2 = 8,76 \text{ E-}05$	18		122		<b>140</b>

Tableau IV-10 : Double table des observations et des attentes pour la construction **vert de peur**

Le nombre total des occurrences de **peur** étant encore plus limité (16) que celui de **rage** (25), le nombre des combinaisons observées (7) est largement supérieur au nombre attendu (2,06). La valeur du test renvoyée (8,76 E-05) est donc significative dans le sens d'une attraction entre l'adjectif et le nom. Les trois autres combinaisons ne renvoient pas un résultat significatif :

- pour **pâle de colère** : 25 occurrences contre 27,9 attendues
- pour **pâle de fureur** : 6 occurrences contre 3,85 attendues
- pour **pâle de rage** : 10 occurrences contre 8,04 attendues

### **III-4. Bilan**

En conclusion on constate donc que sur les huit combinaisons testables, quatre présentent une attraction significative , une présente une répulsion significative et trois ne présentent ni attraction ni répulsion significative :

<b>attraction</b>	<b>répulsion</b>	<b>non significatif</b>
<i>rouge de colère</i> <i>blême de rage</i> <i>vert de rage</i> <i>vert de peur</i>	<i>blême de colère</i>	<i>pâle de colère</i> <i>pâle de fureur</i> <i>pâle de rage</i>

On peut s'étonner que le test révèle une attraction entre **blême** et **rage** et inversement une répulsion entre **blême** et **colère** avec un même nombre de combinaisons, mais cela signifie seulement que le nombre des occurrences de **colère** est plus de trois fois supérieur à celui de **rage**. De ce fait, les 8 occurrences de **blême de colère** sont largement inférieures aux 13,67 occurrences attendues (d'où l'interprétation comme répulsion) alors que les 8 occurrences de **blême de rage** sont largement supérieures aux 3,93 occurrences attendues (d'où l'interprétation comme attraction).

## Références

- Blumenthal (2006), *Wortprofil im Französischen*. Tübingen : Niemeyer.
- Blumenthal (2007), « De la logique des mots à l'analyse de la synonymie ». *Langue Française* 150, p.14-31.
- Blumenthal (2008), « Histoire de mots : affinités (s)électives ». Congrès Mondial de Linguistique Française-CMLF08. Téléchargeable sur le site <http://www.linguistiquefrancaise.org/> (rubrique « Table des matières »)
- Blumenthal, P. (2009), « Éléments d'une théorie de la combinatoire des noms ». *Cahiers de lexicologie* 94, 2009-1, p.11-29.
- Blumenthal, P., Siwersy, S. & Mielebacher, J. (2005), « Kombinatorische Wortprofile und Profilkontraste. Berechnungsverfahren und Anwendungen ». *Zeitschrift für romanische Philologie* 121-1, p.49-83.
- François, J. (2007), « L'émergence d'une construction : comment plusieurs verbes français de requête ont acquis l'aptitude à exprimer l'implication depuis le 17<sup>ème</sup> siècle ». *Bulletin de la Société de Linguistique de Paris*, Tome CTT, Fascicule 1, p.193-215.
- François, J. & Manguin, J.L. (2006), « *Dispute théologique, discussion oiseuse et conversation téléphonique* : Les collocations adjectivo-nominales au cœur du débat », *Langue Française* 150, p.50-65.
- François, J. & Sénéchal, M. (2008), « Le sémantisme propre des cadres prédicatifs et la polysémie des verbes de production de parole ». *Faits de Langues* 31/32 (J.M. Merle, dir. *La prédication*), p.325-336.
- Goldberg, A. (1999), "The emergence of the semantics of argument structure constructions" In : W. McWhinney ed., *The emergence of language*. Mahwah, New-Jersey, Chapter 7, p.197-212.
- Grossmann, F. & Tutin, A. (2002), « Collocations régulières et irrégulières ». *Revue Française de Linguistique Appliquée* VII-1, p.7-25
- Hanks, P. (1996), "Contextual dependency and lexical sets". *International Journal of Corpus Linguistics*, I (1), p.75-98.
- Heid, U. (2009), « La saisie automatique des collocations : objectifs et approches ». *Cahiers de Lexicologie* 94, 2009-1, p.71-89.
- Lebart, L. & Salem, A. (1994), *Statistique textuelle*. Paris : Dunod.
- Lefur, D. & alii (2007), *Dictionnaire des combinaisons de mots – Les synonymes en contexte*. Paris : Editions Le Robert.
- Legallois, D. & François, J. (2006), *Autour des grammaires de constructions et de patterns*. Cahier 21 du CRISCO. Téléchargeable sur le site <http://www.crisco.unicaen.fr>
- Mel'čuk, I. (2003), « Collocations : définition, rôle et utilité ». In : Grossmann, F. & Tutin, A. (dir.) *Les collocations : analyse et traitement*. Travaux de recherche en linguistique appliquée E-1, Amsterdam : De Wereldt, p.23-44.
- Mel'čuk, I. & alii (1984), *Dictionnaire explicatif et combinatoire du français contemporain – Recherches lexico-sémantiques I*. Montréal : Presses de l'Université de Montréal

- Mel'čuk, I., Clas, A. & Polguère, A. (1995), *Introduction à la lexicologie explicative et combinatoire*. Louvain-la-Neuve : Duculot.
- Perek, F. (à paraître), « Identification de constructions grammaticales en corpus : une approche quantitative de l'augmentation de valence ». Travaux linguistiques du CERLICO 23. Rennes : Presses Universitaires de Rennes.
- Stefanowitsch, A. & Gries, S. Th (2003), "Collostructions : Investigating the interaction of words and constructions". *International Journal of Corpus Linguistics* 8:2, p.209-243
- Stefanowitsch, A. & Gries, S. Th (2005), "Covarying collexemes". *Corpus Linguistics and Linguistic Theory* 1-1, p.1-43.
- Tutin, A., Novakova, I, Grossmann, F. & Cavalla, Ch. (2006), « Esquisse de typologie des noms d'affect à partir de leurs propriétés combinatoires ». *Langue Française* 150, p.50-65.
- Wiechmann, D. (2008), "On the computation of collostruction strength : Testing measures of association as expressions of lexical bias", *Corpus Linguistics and Linguistic Theory* 4-2, p.253-290.

### III

#### *Cahiers du CRISCO*

Cahiers de recherche linguistique diffusés par le CRISCO depuis 2000. Les *Cahiers du CRISCO* sont distribués gratuitement.

##### **A. Cahiers 1-12 en format papier**

n°1	Jacques FRANÇOIS, <i>Sur la grammaire de van Valin</i> (septembre 2000)
n°2	Véronique LENEPVEU, <i>Adjectifs et adverbes : une corrélation syntactico-sémantique</i> (janvier 2001)
n°3	Jacques FRANÇOIS & Hansjakob SEILER, <i>Gespräche über Hansjakob Seilers Universalienforschung</i> (mai 2001, version originale en allemand)
n°4	Denis RAMASSE, <i>L'intonation des phrases présentant un détachement à gauche en français : une première approche</i> (juin 2001)
n°5	Heike BALDAUF, <i>Voyage au bout de la langue : quelques remarques sur des énoncés minimaux</i> (mars 2002)
n°6	Jacques FRANÇOIS (avec la participation de Hansjakob SEILER), <i>La faculté de langage et les universaux : perspectives fonctionnalistes actuelles</i> (mars 2002)
n° 7	Catherine CAMUGLI GALLARDO, <i>La langue des jeunes en Italie — Guide à une enquête de terrain</i> (mars 2002)
n° 8	Jacques FRANÇOIS, <i>La représentation des structures verbo-nominales et du figement verbal dans deux formalismes de grammaire fonctionnelle</i> (mai 2002)
n° 9	Valeriano BELLOSTA von COLBE & Jacques FRANÇOIS, <i>Windowing of attention and the core-periphery boundary</i> (septembre 2002)
n° 10/1	Maxi KRAUSE, <i>AUS, élément prépositionnel, élément constitutif de particules diverses et particule verbale</i> (septembre 2002)
n° 10/2	Maxi KRAUSE, <i>VON, préposition et élément constitutif d'adverbes et de tournures adverbiales</i> (octobre 2002)
n° 10/3	Maxi KRAUSE, <i>BEI, préposition, élément constitutif de particules diverses et particule verbale</i> (novembre 2002)
n° 10/4	Maxi KRAUSE, <i>MIT, préposition, élément constitutif de particules diverses et particule adverbiale</i> (novembre 2002)
n° 11	Richard RENAULT, <i>Morphosyntaxe des participes finnois</i> (décembre 2002)
n° 12	Dominique LEGALLOIS (coordination), <i>Unité(s) du texte</i> [Journée Scientifique du CRISCO, 6 décembre 2002] (juin 2003)

**B. Cahiers 13-28 téléchargeables sur le site du CRISCO : [www.crisco.unicaen.fr](http://www.crisco.unicaen.fr)**

- |      |   |
|------|---|
| n°13 | Jacques FRANÇOIS (dir.), <i>Aspects de la « Role and Reference Grammar »</i> [Journée Scientifique du CRISCO, 14 mars 2003] (septembre 2003)  |
| n°14 | Jacques FRANÇOIS, Jean-Luc MANGUIN & Bernard VICTORRI, <i>La réduction de la polysémie adjectivale en cotexte nominal : une méthode de sémantique calculatoire</i> (octobre 2003)               |
| n°15 | Emmanuelle ROUSSEL, <i>Le commentaire associé au passif anglais</i> (février 2004)  |
| n°16 | Corinne OZOUF, <i>Caractère différentiel et relation d'équivalence entre <b>voir</b> et <b>regarder</b></i> (juin 2004)   |
| n°17 | Jean-Luc MANGUIN & Jacques FRANÇOIS (coordination), <i>Le Dictionnaire Electronique des Synonymes du CRISCO – Un mode d'emploi à trois niveaux</i> (juillet 2004)                               |
| n°18 | Dominique LEGALLOIS & Song-Nim KWON, <i>Sémantique lexicale et examen écologique de la co-occurrence</i> (janvier 2005)   |
| n°19 | Jacques FRANÇOIS, <i>Les bases épistémologiques des grammaires cognitives et 'néo-fonctionnelles'</i> (juin 2005)   |
| n°20 | Jacques FRANÇOIS, <i>Le fléchage synonymique de la polysémie verbale</i> (octobre 2005)   |
| n°21 | Dominique LEGALLOIS & Jacques FRANÇOIS (coordination), <i>Autour des grammaires de constructions et de patterns</i> (janvier 2006) [avec la participation de Philippe GREA et Morgane SENECHAL] |
| n°22 | Odile BLANVILLAIN, <i>Le marqueur SUCH en anglais contemporain – Anaphore qualitative et identification notionnelle</i> (mai 2006)  |
| n°23 | Jacques FRANÇOIS & Ahmed BRAHIM (coordination), <i>Morphosyntaxe et sémantique du verbe</i> (novembre 2007)   |
| n°24 | Jacques FRANÇOIS, <i>Une approche diachronique quantitative de la polysémie verbale</i> (janvier 2008)  |
| n°25 | Mayu SHINTANI, <i>Une vision rhétorique des expressions BOUT et HASHI – universalité cognitive et comparaison entre français et japonais</i> (février 2008)                                     |
| n°26 | Jacques FRANÇOIS, <i>Les grammaires de construction, un bâtiment ouvert aux quatre vents</i> (septembre 2008)   |
| n°27 | Lise HAMELIN, <i>L'alternance entre construction ditransitive et complémentation prépositionnelle : étude des marqueurs <b>to</b> et <b>for</b></i> (mai 2009)                                  |
| n°28 | <i>Travaux des doctorants du CRISCO 2008-2009</i> (octobre 2009)  |