

## La base de données textuelles FRANTEXT-2 : Comment en tirer parti dans les études linguistiques et littéraires

Les chapitres 1 et 2 ont été consacrés respectivement :

- (chap.1) à la sélection d'un **corpus** et à des recherches « **simples** », portant sur une forme de mot ou une chaîne de formes de mots,
- (chap.2) et aux deux autres modalités de recherche, « **assistée** » (avec lemmatisation et classes syntaxiques) et « **avancée** » (avec l'emploi de formules de la métalangue de recherche CQL, *Corpus Query Language*).

Le chapitre 3 se concentre sur le « **post-traitement** », c'est-à-dire sur l'exploitation, à l'aide d'un tableur employé comme concordancier, du résultat du traitement effectué par l'algorithme de recherche de FRANTEXT-2.

### Chap.3 : Le « post-traitement » : pour un usage raisonné des algorithmes d'exploration lexicale

#### 3.1. Pré-traitement vs post-traitement : L'exemple de [qqn tient à qqn / qqc / VINF]

À titre d'illustration de ce qu'on peut raisonnablement attendre d'un **pré-traitement** en recherche avancée et de ce que le linguiste a encore à compléter pour une analyse pertinente, je choisis la construction [qqn tient à qqn / qqc / VINF] (la variante [qqn tient à ce que P<subj>] est volontairement ignorée pour ne pas alourdir l'illustration).

L'intérêt de cette construction tient au fait qu'à l'origine la préposition **à** prend une valeur locative (ex. *Le crochet tient au plafond*) sans variante du complément à l'infinitif, avant de prendre un sens d'attachement par **dépendance** si le sujet est non-animé (ex. *Son succès tient à un heureux hasard*) ou d'attachement **affectif** à une personne, une chose ou une activité si le sujet est animé (ex. *Paul tient à ses enfants / à son image de père attentif / à protéger ses enfants*). Bien que de même origine (la notion d'attachement), les deux variantes à sujet animé vs non-animé ont des sens parfaitement distincts.

La question est ici : **Jusqu'à quel point la métalangue CQL permet-elle de distinguer ces sens ?** Nous allons tester l'adéquation de trois formules de pré-traitement en recherche avancée. Pour chacune des trois, le test portera sur 100 extraits aléatoires (classés par ordre alphabétique du nom des auteurs) de FRANTEXT-2, corpus 21<sup>e</sup> siècle.

✚ 1<sup>ère</sup> recherche par la formule CQL ⇒

[lemma="tenir"] [lemma="à"] [pos="VINF | NC | DET"]

Cette formule se compose du lemme "*tenir*", suivi du lemme "à" lui-même suivi d'un mot caractérisé par un éventail de classes syntaxique **VINF** ou **NC** ou **DET**.

On obtient 7 occurrences de la structure **N<qn> TENIR à N<qqc/qqn>** et 12 de la structure **N<qn> TENIR à VINF**

Exemples de résultats

GROULT Benoîte	Mon évasion	2008	<i>L' idée qu'elle survit à son amant , qu'elle trouve encore plaisir à vivre, même sans lui, c'est ça qui a choqué, beaucoup plus que le vocabulaire . Mais je</i>	<i>tenais à cette</i>	<i>fin parce que j' étais écœurée (...)</i>	N
BLOCH- DANO Évelyne	Porte de Cham- perret	2013	<i>Une fille coquette, qui portait des bracelets en or et un sac en cuir élégant qu'elle avait</i>	<i>tenu à garder</i>	<i>à la main.</i>	I

✚ 2<sup>e</sup> recherche par la formule CQL ⇒

[lemma!="se"] [lemma="tenir"] [lemma="à"] [pos="VINF | NC | DET"]

La formule est identique à la précédente, moyennant l'ajout de la composante [lemma!="se"] destinée à écarter l'essentiel des emplois pronominaux réfléchis de 3<sup>e</sup> personne du v. *tenir*. On obtient 14 occurrences de la structure **N<qn> TENIR à N<qqc/qqn>** et 18 de la structure **N<qn> TENIR à VINF**

Exemples de résultats

GARAT Anne-Marie	Pense à demain	2010	<i>...elle n' est pas très sûre qu' Antoine appréciera le projet . Elle redoute son jugement , son sens ombrageux des choses droites , de passer à ses yeux pour une écervelée .</i>	<i>Elle tient à son</i>	<i>estime . Elle tient à ce qu' il ignore cette histoire scabreuse .</i>	N
BÉNABOU Marcel	Écrire sur Tamara	2002	<i>Un des rares moments où s'atténuait sa colère était celui qu il consacrait aux lettre , le dimanche matin. Manuel</i>	<i>avait tenu à garder</i>	<i>des liens avec sa petite bande demeurée au Maro , et il envoyait de temps en temp , à l' un ou à l' autre, de brefs messages .</i>	I

3<sup>e</sup> recherche par la formule CQL ⇒

[lemma!="se"] [lemma="tenir"] [lemma="à"]

La formule est identique à la précédente, moyennant la suppression de la composante [pos="VINF|NC|DET"] afin de vérifier si cette composante a un effet sélectif efficace. On obtient cette fois 14 occurrences de la structure N<qn> TENIR à N<qqc/qqn> et 25 de la structure N<qn> TENIR à VINF.

Exemples de résultats

JENNI Alexis	L'Art français de la guerre	2011	« En montant vers le poste ... — Le bordj , capitaine, le bord , coupa Chambol .	<i>Je tiens à</i>	ce terme. En arabe il désigne la tour, et c' est un mot très fort en leur langue .	N
BÉNABOU Marcel	Écrire sur Tamara	2002	La fête ne put se faire dans les formes que Violetta avait prévues J' avais	<i>donc tenu à</i>	lui offrir, pour me faire pardonner, un tête-à-tête d une exceptionnelle longueur.	I

Le tableau ci-dessous synthétise ces trois résultats et montre la décroissance progressive du « bruit », c'est-à-dire des extraits qui ne présentent aucune des deux structures recherchées : 81 ⇒ 68 ⇒

	1	2	3
	∅ [lemma="tenir"] [lemma="à"] [pos="VINF NC DET"]	[lemma!="se"] [lemma="tenir"] [lemma="à"] [pos="VINF NC DET"]	[lemma!="se"] [lemma="tenir"] [lemma="à"] ∅
Structure		fréq.	fréq.
N<qn> TENIR à N<qqc/qqn>		7	14
N<qn> TENIR à VINF		12	18
bruit		81	68

## ➤ Que peut-on en conclure ?

1. Entre les recherches 1 et 2, on constate que **l'ajout de la composante visant à écarter les formes du pronom objet réfléchi a un effet positif**, augmentant le nombre des occurrences repérées de 19 à 32.
2. En revanche, entre les recherches 2 et 3, le constat est contraire aux attentes : **la suppression de la composante de catégorisation du complément indirect a un effet positif** et non négatif, faisant passer le nombre des occurrences de 32 à 44.
3. Toutefois, même avec le résultat de la recherche 3, **le « bruit » continue à dépasser 50%**. Il est donc clair que pour arriver à un résultat plus satisfaisant il faudrait disposer d'une métalangue de recherche plus sophistiquée et que les formules ainsi composées atteindraient une ampleur difficilement gérable.
4. Les 44 occurrences enregistrées correspondent bien aux structures recherchées. La défaillance de la 3<sup>e</sup> formule CQL relève donc exclusivement de la surgénération. Celle-ci tient notamment à **l'indistinction entre un sujet [+animé] et [-animé] avec la différence de sens qui en découle (« vouloir garder » vs « être lié à »)**.
5. La solution que je préconise consiste donc à considérer le résultat de la recherche 3 comme **un tremplin qui doit être complété par un post-traitement manuel**, destiné à compenser les effets de la surgénération.
6. Cette solution ne doit pas être considérée comme un pis-aller, car il me paraît préférable que les linguistes « reprennent la main » là où les algorithmes patinent. L'objectif n'est pas de consacrer une énergie et un temps déraisonnables pour améliorer les résultats de la recherche de quelques points, mais de **comprendre les finesses de nos propres raisonnements quand il s'agit de décider si un énoncé valide bien une structure syntactico-sémantique**.

## 3.2. Le post-traitement des collocations du N *conscience*

Le corpus 21<sup>e</sup> siècle délivre **1323** occurrences de *conscience* et **34** de *conscienc*s. La méthode de post-traitement proposée se décline en trois phases :

- a) Des cooccurrences (a priori intéressantes) à gauche et à droite sont recherchées dans les **100 premières occurrences** classées par ordre chronologique inverse (de 2013 à 2000).
- b) Sur la base de ce premier test, toutes les cooccurrences figurant au minimum 2 fois dans les 100 premières occurrences sont recherchées dans le reste du corpus pour vérifier leur fréquence absolue, afin de **dégager des collocations**.
- c) On enregistre le nombre de cooccurrences présentant une fréquence supérieure à un seuil à déterminer chez les 6 auteurs qui emploient le plus *conscience(s)* relativement

au volume (en nombre de mots) de leur œuvre enregistrée, afin de dégager les **préférences de chaque auteur(e) pour l'une ou l'autre collocation**.

L'un des intérêts majeurs du post-traitement sur tableur-concordancier est – si le nombre des occurrences enregistrées est suffisant – d'inciter à un **éventail de recherches locales** (par auteur, voire par œuvre, par genre, par date, etc.).

### 3.2.1. Exemples des cooccurrences les plus fréquentes dans l'ensemble du corpus

Le tableau ci-dessous enregistre – pour les 100 premiers extraits – les cooccurrences séparément dans les deux contextes gauche et droit et dans le contexte gauche et droit (classement par fréquence décroissante). Les **cooccurrences associées par la morphologie et le sens** (soulignées) sont regroupées.

Contexte gauche		Contexte droit		Contextes gauche et droit	
<i>avoir</i>	18	<i>de N</i>	16	<i>avoir _ de N</i>	7
<u><i>prendre / reprendre / prise de</i></u>	11	<u><i>que P / de ce que P</i></u>	10	<i>avoir _ que P</i>	6
<i>mauvaise</i>	5	<i>de VINF</i>	4	<i>avoir _ tranquille</i>	3
<i>en son âme et</i>	4	<i>tranquille</i>	4	<u><i>prendre / prise de _ que P</i></u>	3
<u><i>objecteur/objection de</i></u>	3	<i>professionnelle</i>	2	<i>avoir _ de INF</i>	2
<i>par acquit de</i>	3				
<i>bonne</i>	2				
<i>cas de</i>	2				
<i>liberté de</i>	2				
<u><i>perdre / perte de</i></u>	2				
<i>religion de la</i>	2				
<i>suivre</i>	2				

*mauvaise ~*

5 / 51

CHALANDON Sorj	Retour à Killybegs	2011	Paris , plus de deux millions d' habitants et je tombais sur le seul que je ne devais pas rencontrer. Dieu , le hasard, ma <b>mauvaise</b>	<b>conscience</b>	ou sa naïveté nous ont fait prendre ce jour -là et à cette heur , non seulement le métro, mais la même ligne, la même direction et le même wagon .
----------------	--------------------	------	--	-------------------	--

*en son âme et ~*

4 / 8

GARAT Anne-Marie	Dans la main du diable	2006	— Il vous revient de prendre la décision que vous jugerez bonne, <b>en votre âme et</b>	<b>conscience</b>	, au sujet de ce garçon ... Mais mon avis est que Grandrieux es , pour l' instan , moins dangereux pour nous dans l'Institut qu' au-dehors .
------------------	------------------------	------	---	-------------------	--

*objecteur / objection de ~*

3 / 4

LINHART Virginie	Le jour où mon père s'est tu	2008	J' ai fait mon service militaire comme <b>objecteur de</b>	<b>conscience</b>	— ça a pris vingt mois de ma vie , mais je n' aurais pas pu me faire passer pour fou .
LAZARD Christian	Journal : 1937	2012	Nous allons au Globe voir une très bonne pièce « Robert's wife » qui pose tous les problèmes d' actualité : « birth control » , <b>objection de</b>	<b>conscience</b>	, vie indépendante des époux, et fait régler tout ça avec sagesse et pondération par un excellent « bishop » qui personnifie la vieille sagesse et la bonne humeur anglaise.

*par acquit de ~*

3 / 20

GARAT Anne-Marie	L'enfant des ténèbres	2008	... elle déposa sur le tapis, comme d' habitude , sa paire de chaussures. Ce soir , très mouillées . Puis , <b>par acquit de</b>	<b>conscience</b>	, elle revint examiner sous la lampe les revues controversée . Polémiques, éloges et commentaires, blablabla, de l' anglais pur sucre .
------------------	-----------------------	------	--	-------------------	---

➤ Contexte droit

~ de N<sub>[Patient]</sub>

(236 : de N<sub>Agent/Patient</sub>)

JENNI Alexis	L'Art français de la guerre	2011	<i>Pour se transformer en art , le talent doit prendre</i>	<b>conscience</b>	<i>de lui -même , et de ses limites , et être aimanté d' un but , qui l' oriente dans une direction indiscutable .</i>
--------------	-----------------------------	------	--	-------------------	--

~ de N<sub>[Agent]</sub>

(idem)

OSMONT Stéphane	Éléments incontrôlés	2012	<i>« Celui qui ne connaît pas l'Histoire est condamné à la revivre. » Ou plus sophistiqué : « Ce n'est pas la</i>	<b>conscience</b>	<i>des hommes qui détermine leur existence, c' est au contraire leur existence sociale qui détermine leur conscience. »</i>
-----------------	----------------------	------	---	-------------------	---

~ de VINF

CARRÈRE Emmanuel	D'autres vies que la mienne	2009	<i>...maintenant je suis contrarié par notre fâcherie et par la</i>	<b>conscience</b>	<i>de n' avoir pas été à la hauteur de la situation .</i>
------------------	-----------------------------	------	---	-------------------	---

➤ Contexte gauche et droit

avoir ~ de N

(6)

SIMONET Mathieu	Marc Beltra : roman autour d'une disparition	2013	<i>Si on n' avait pas évoqué un crime , les autorités locales ne nous auraient jamais autorisés à revenir . Le problème ( je ne sais pas si vous <b>en avez eu</b></i>	<b>conscience</b>	<i>) , le problème ... c' est que ça n' arrive jamais d' ouvrir une instruction pour crime , lorsqu' on n' a pas de cadavre .</i>
-----------------	--	------	--	-------------------	---

avoir ~ que P

(6)

PICQUET Pauline	Sans illustration	2013	<i>« Je n'ai eu aucune difficulté . Tout s'est bien passé . Mais j'<b>ai bien</b></i>	<b>conscience</b>	<i><b>que</b> ce sera beaucoup plus difficile avec les élèves . »</i>
-----------------	-------------------	------	---	-------------------	---

**avoir la ~ tranquille** (3)

LAZARD Christian	Journal : 1930	2012	Le matin j'ai téléphoné à Durvis pour lui dire que je tiens à revoir avec lui les cultures de Vallée. Sans cela, je n'aurais pas la	<b>conscience</b>	<b>tranquill</b> . Nous prenons rendez-vous pour jeudi à 8h .
---------------------	-------------------	------	---	-------------------	---

**prendre / prise de \_ de N** (3)

LINHART Virginie	La vie après	2012	Le jour même de l'opération , j' ai croisé un de mes copains, un Noir américain qui faisait ses études à Paris, un type fantastique . Devant lui, j'ai brutalement pris	<b>conscience</b>	<b>de ma lâcheté</b> ; Jo était noir et ne pourrait jamais le cacher, alors que moi je voulais cacher le fait d' être juif.
---------------------	-----------------	------	---	-------------------	---

**avoir \_ de INF** (2)

LINHART Virginie	La vie après	2012	À l'heure de l'Europe , j'ai	<b>conscience</b>	<b>d'être</b> complètement démodé , mais j'avais reçu une éducation très patriotique (...)
---------------------	-----------------	------	------------------------------	-------------------	--

**3.2.2. Fréquence relative de *bonne conscience* et *mauvaise conscience* dans l'ensemble du corpus**

- **Bonne conscience** et **mauvaise conscience** sont deux collocations dont la fréquence diffère sensiblement dans notre corpus (33 vs 51). Sans doute la mauvaise conscience est-elle un état psychologique plus intéressant pour les littérateurs que son inverse, la bonne conscience.
- Sur 122 auteurs figurant dans le corpus, seuls 34 présentent au moins une occurrence de bonne et/ou mauvaise conscience. Cependant pour rendre compte de l'ampleur du corpus pour cette période se limitant aux 14 premières années du 21<sup>e</sup> siècle, je mentionnerai également les 88 autres auteurs dans le tableau ci-dessous.
- La 2<sup>e</sup> colonne fournit la fréquence absolue de **conscience**, la 3<sup>e</sup> celle de **bonne ~** et la 5<sup>e</sup> celle de **mauvaise ~**. La 4<sup>e</sup> délivre la fréquence relative de **bonne conscience** par rapport à celle de **conscience** chez chaque auteur et la 6<sup>e</sup> en fait autant pour **mauvaise conscience**. En dehors des deux auteurs (Clémence BOULOUQUE et Claude CROCQ) qui n'emploient qu'une fois **conscience** dans la collocation **bonne conscience**, l'auteur qui fait proportionnellement le plus grand usage de **bonne/mauvaise conscience** est Jean-Luc LAGARCE avec 2 occurrences de **bonne conscience** et 1 de **mauvaise conscience** sur 4 occurrences de **conscience** (75%).

	fréq.abs.	bonne ~	fréq.rel bonne ~	mauvaise ~	fréq.rel mauvaise ~
ANGOT Christine	12	0	0,0%	0	0,0%
ARNAUD Claude	15	0	0,0%	1	6,7%
ARTIÈRES Philippe	5	0	0,0%	0	0,0%
AUBRY Gwenaëlle	4	0	0,0%	0	0,0%
AUDIN Michèle	2	0	0,0%	0	0,0%
AUROY Berthe	4	0	0,0%	0	0,0%
AYMÉ Marcel	133	4	3,0%	2	1,5%
BAILLY Jean-Christophe	4	0	0,0%	0	0,0%
BEIGBEDER Frédéric	2	0	0,0%	0	0,0%
BÉNABOU Marcel	7	0	0,0%	0	0,0%
BENOZIGLIO Jean-Luc	2	0	0,0%	0	0,0%
BERGOUNIOUX Pierre	3	0	0,0%	0	0,0%
BERNFELD Karin	1	0	0,0%	0	0,0%
BERR Hélène	46	0	0,0%	0	0,0%
BILLETDOUX Marie	8	0	0,0%	0	0,0%
BLANC Jean-Noël	1	0	0,0%	0	0,0%
BLOCH-DANO Évelyne	2	0	0,0%	1	50,0%
BOBIN Christian	1	0	0,0%	0	0,0%
BOLTANSKI Christian	3	0	0,0%	0	0,0%
BOULLIER Grégoire	6	0	0,0%	0	0,0%
BOULOUQUE Clémence	1	1	100,0%	0	0,0%
BRIÈRE-BLANCHET Claire	6	0	0,0%	0	0,0%
BROCHET Anne	1	0	0,0%	0	0,0%
CALLE Sophie	1	0	0,0%	0	0,0%
CARRÈRE Emmanuel	26	0	0,0%	2	7,7%
CASTEL Hélène	16	0	0,0%	0	0,0%
CÉSAIRE Aimé	5	1	20,0%	0	0,0%
CHAIX Marie	3	0	0,0%	1	33,3%
CHALANDON Sorj	5	0	0,0%	1	20,0%
CROCQ Claude	1	1	100,0%	0	0,0%
DELAUME Chloé	2	0	0,0%	0	0,0%
DEPUSSÉ Marie	1	0	0,0%	0	0,0%
DEVILLE Patrick	1	0	0,0%	0	0,0%
DOMENACH-LALLICH Denise	1	0	0,0%	0	0,0%
DOUBROVSKY Serge	10	0	0,0%	1	10,0%
DURAS Marguerite	11	0	0,0%	1	9,1%
ECHENOZ Jean	6	0	0,0%	0	0,0%
ERIBON Didier	42	0	0,0%	2	4,8%
ERNAUX Annie	42	0	0,0%	1	2,4%
FALLET René	1	0	0,0%	0	0,0%
FILIPPETTI Aurélie	2	0	0,0%	0	0,0%

	fréq.abs.	bonne ~	fréq.rel bonne ~	mauvaise ~	fréq.rel mauvaise ~
FLEM Lydia	4	0	0,0%	0	0,0%
FOREST Philippe	11	2	18,2%	0	0,0%
FOURNEL Paul	5	0	0,0%	0	0,0%
FRANÇOIS Annie	1	0	0,0%	0	0,0%
GARAT Anne-Marie	154	7	4,5%	2	1,3%
GARRÉTA Anne	8	0	0,0%	0	0,0%
GARY Alexandre Diego	3	0	0,0%	0	0,0%
GAULT Vanessa	13	2	15,4%	0	0,0%
GAVALDA Anna	7	0	0,0%	3	42,9%
GENETTE Gérard	23	1	4,3%	2	8,7%
GOSCINNY Anne	1	0	0,0%	0	0,0%
GRANGAUD Michelle	7	0	0,0%	0	0,0%
GRENIER Roger	3	0	0,0%	0	0,0%
GROULT Benoîte	9	0	0,0%	0	0,0%
GUIBERT Hervé	26	0	0,0%	0	0,0%
GUILLEBAUD Catherine	4	0	0,0%	0	0,0%
GUYOTAT Pierre	13	0	0,0%	0	0,0%
HAVET Mireille	14	0	0,0%	0	0,0%
HEINICH Nathalie	1	0	0,0%	0	0,0%
HOPPENOT Hélène	7	1	14,3%	0	0,0%
JABLONKA Ivan	7	0	0,0%	0	0,0%
JENNI Alexis	15	0	0,0%	2	13,3%
JONQUET Thierry	8	0	0,0%	1	12,5%
LAGARCE Jean-Luc	4	1	25,0%	2	50,0%
LANG Luc	6	0	0,0%	0	0,0%
LAURENS Camille	1	0	0,0%	0	0,0%
LAZARD Christian	28	0	0,0%	0	0,0%
LE CLÉZIO Jean-Marie Gustave	2	0	0,0%	0	0,0%
LEGENDRE Claire	1	0	0,0%	0	0,0%
LEJEUNE Philippe	11	0	0,0%	0	0,0%
LEVÉ Édouard	10	0	0,0%	0	0,0%
LEVY Marc	1	0	0,0%	0	0,0%
LINHART Virginie	26	0	0,0%	2	7,7%
LITTELL Jonathan	15	2	13,3%	0	0,0%
LUCA Ghérasim	1	0	0,0%	0	0,0%
LUCOT Hubert	7	0	0,0%	0	0,0%
MANCHETTE Jean-Patrick	56	0	0,0%	0	0,0%
MARIN Claire	4	0	0,0%	0	0,0%
MARTIN Lionel-Édouard	2	0	0,0%	0	0,0%
MASPERO François	22	2	9,1%	4	18,2%
MATHEWS Harry	2	1	50,0%	0	0,0%

	fréq.abs.	bonne ~	fréq.rel bonne ~	mauvaise ~	fréq.rel mauvaise ~
MAXIMIN Daniel	4	0	0,0%		0,0%
MÉRIL Macha	4	0	0,0%	1	25,0%
MILEWSKI Françoise	4	0	0,0%		0,0%
MILLET Catherine	22	0	0,0%		0,0%
MODIANO Patrick	2	1	50,0%		0,0%
MORGIÈVE Richard	7	0	0,0%		0,0%
MRÉJEN Valérie	1	0	0,0%		0,0%
NDIAYE Marie	3	0	0,0%		0,0%
NOURISSIER François	22	1	4,5%	1	4,5%
OLLAGNIER Jeanne	6	0	0,0%		0,0%
OSMONT Stéphane	14	0	0,0%	1	7,1%
OZOUF Mona	19	1	5,3%	1	5,3%
PACHET Pierre	29	0	0,0%		0,0%
PÉCHEROT Patrick	2	0	0,0%		0,0%
PEREC Georges	25	1	4,0%	2	8,0%
PERRUT Dominique	9	0	0,0%	2	22,2%
PICQUET Pauline	4	0	0,0%	1	25,0%
PIERRAT Emmanuel	5	0	0,0%		0,0%
PODALYDÈS Denis	7	0	0,0%	1	14,3%
PONTALIS Jean-Bertrand	16	0	0,0%		0,0%
POUQUET Jeanne	1	0	0,0%		0,0%
PUECH Jean-Benoît	1	0	0,0%		0,0%
ROBIN Régine	6	0	0,0%		0,0%
ROLIN Olivier	2	0	0,0%	1	50,0%
ROUBAUD Jacques	31	1	3,2%	5	16,1%
ROUX Annelise	17	0	0,0%		0,0%
SCHROEDER Liliane	1	0	0,0%		0,0%
SIMONET Mathieu	10	0	0,0%		0,0%
SIZUN Marie	1	0	0,0%	1	100,0%
SONNET Martine	1	0	0,0%		0,0%
STORTI Martine	17	0	0,0%		0,0%
SZCZUPAK-THOMAS Yvette	9	0	0,0%		0,0%
TENENBAUM Gérald	6	0	0,0%		0,0%
TORRÈS Tereska	4	0	0,0%		0,0%
VARGAS Fred	13	1	7,7%		0,0%
VEIL Simone	28	1	3,6%	2	7,1%
VIGAN Delphine de	1	0	0,0%		0,0%
VIOLET Lydie	4	0	0,0%		0,0%
WEIL Sylvie	1	0	0,0%		0,0%
WINOCK Michel	10	0	0,0%		0,0%
	1357	33	2,4%	50	3,7%

### 3.2.3. *Prendre / reprendre / prise de conscience* dans le contexte gauche, par auteur et par fréquence relative décroissante par rapport aux occurrences de *conscience*

En dehors des calculs portant sur l'ensemble du corpus, il est intéressant de comparer la fréquence de collocations entre les auteurs les mieux représentés. On constate ainsi qu'un seul auteur, Didier ERIBON, fait un usage abondant de *prendre / reprendre / prise de conscience*.

	Nb textes	fréq.abs.	fréq.rel *10 <sup>-3</sup>	Nb mots	<i>prendre</i> ~	<i>reprendre</i> ~	<i>prise de</i> ~	TOTAL	Fréq.rel. à <i>conscience</i>
ERIBON Didier	1	42	0,647	64940	4	0	1	5	11,9%
AYMÉ Marcel	1	133	0,257	518110	6	2	0	8	6,0%
GARAT Anne-Marie	8	154	0,128	1207137	3	3	1	7	4,5%
BERR Hélène	1	46	0,625	73612	2	0	0	2	4,3%
MANCHETTE Jean-Patrick	1	56	0,299	187431	2	0	0	2	3,6%
ERNAUX Annie	6	42	0,289	145368	1	0	0	1	2,4%

### 3.2.4. Fréquence relative de *que P* dans le contexte droit (par rapport aux occurrences de *conscience*)

On constate qu'un seul auteur - à nouveau Didier ERIBON - fait un usage abondant de *conscience que P* (généralement précédé de *avoir / prendre / prise de / la*), ce qui atteste d'un emploi centré sur le contenu propositionnel « conscientisé ». Un complément d'enquête consisterait à comparer cet emploi avec (*avoir / prendre / DET*) *conscience de VINF* pour vérifier si les auteurs font un usage analogue des deux types de constructions.

[ <i>que P</i> en subordonnée conjonctive et non relative]	Nb textes	fréq.abs.	fréq.rel *10 <sup>-3</sup>	Nb mots	<i>que P</i>	Fréq.rel. à <i>conscience</i>
ERIBON Didier	1	42	0,647	64940	7	16,7%
MANCHETTE Jean-Patrick	1	56	0,299	187431	5	8,9%
BERR Hélène	1	46	0,625	73612	5	7,8%
ERNAUX Annie	6	42	0,289	145368	2	4,7%
AYMÉ Marcel	1	133	0,257	518110	4	3,0%
GARAT Anne-Marie	8	154	0,128	1207137	1	0,6%

## 3.4. Conclusion

- Ce chapitre 3 du didacticiel a insisté sur l'importance d'un **post-traitement approfondi**, lequel présuppose un pré-traitement effectué à l'aide de formules de la métalangue CQL (en recherche avancée). Les formules CQL sont destinées à « préparer le terrain » : une simple catégorisation peut suffire à distinguer p.ex. entre un complément à tête N ou VINF mais pas entre un sujet animé et un sujet non-animé, distinction souvent associée à des sens différents du verbe. Dans de nombreux cas, il est aussi nécessaire de lire attentivement l'extrait pour vérifier si la forme **que** est une conjonction de subordination ou un pronom relatif.
- Il en résulte qu'on ne peut pas faire l'économie d'un post-traitement manuel, extrait par extrait. Cet effort fastidieux a toutefois une contrepartie : certains questionnements ne se font jour qu'à la lecture attentive d'une succession d'extraits.
- En étudiant les 1357 extraits du corpus FRANTEXT-2 du 21<sup>e</sup> siècle contenant **conscience**, on s'aperçoit par exemple que la séquence **DET + conscience** ne figure que très rarement en tête de phrase :

. <b>La conscience</b>	<b>11</b>
. <b>Ma conscience</b>	<b>5</b>
. <b>Sa conscience</b>	<b>4</b>
. <b>Cette conscience</b>	<b>1</b>
. <b>Une conscience</b>	<b>1</b>

. <b>Chaque conscience</b>	<b>0</b>
. <b>Leur conscience</b>	<b>0</b>
. <b>Toute conscience</b>	<b>0</b>
<b>TOTAL</b>	<b>22 : 1,6%</b>

Il y a peu de chances que cette tendance syntaxique transparaisse sans une lecture attentive des extraits enregistrés, ce qui ne met cependant pas en doute la valeur des formules CQL comme base de départ.

Le **quatrième** et dernier chapitre sera consacré à des **études de cas plus développées** que celles qui ont été proposées dans les trois premiers chapitres. Il s'agira de montrer que l'exploitation pertinente d'un corpus littéraire vise en dernier ressort à **dégager des tendances** morphosyntaxiques (les plus faciles à repérer) et indirectement stylistiques, sémantiques et discursives, quels que soient les critères de sélection du corpus qu'on se donne, **par date, par genre, par auteur ou par oeuvre**.