

Jacques FRANÇOIS
Université de Caen & CRISCO, EA 4255

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

Chap.4 : Trois études de cas

4.1. Retour sur les chapitres précédents et exploitation de FRANTEXT-2 pour les 3 études de cas

Dans les trois premiers chapitres, j'ai abordé successivement

1. le choix d'un corpus présélectionné ou l'élaboration d'un corpus personnel et les résultats d'une recherche 'simple',
2. les résultats d'une recherche 'assistée' ou 'avancée' à l'aide de formules CQL
3. et le post-traitement de ces trois types de recherches à l'aide d'un tableur.

Ce quatrième et dernier chapitre du didacticiel est consacré à une question inéluctable : **À quoi bon ?** On peut estimer qu'il n'appartient pas à un didacticiel consacré à une base de données textuelles de passer du questionnement central sur le **comment ?** à un questionnement facultatif sur le **pourquoi ?**

Cependant la **LINGUISTIQUE DE CORPUS**, en tant que pratique associée à une théorie sous-jacente – issue des premiers travaux de J.R. Firth dans les années 1950 et développée ultérieurement, notamment en Grande-Bretagne par J. Sinclair, D. Biber, M. Hoey, etc. et en France B. Habert¹ – ambitionne de répondre à ce **pourquoi ?** et la publication de FRANTEXT-2 est destinée à offrir aux chercheurs et aux étudiants une variété de corpus textuels par auteur, par genre textuel et par époque de l'histoire du français.

C'est pourquoi il n'est pas indifférent d'élever le questionnement à ce niveau en proposant des études de cas qui tirent parti de cet éventail de corpus, sans délivrer pour autant des analyses abouties, laissant aux éventuels intéressés le soin d'aller plus loin à partir des corpus proposés ou de corpus personnels mieux affinés.

¹ Cf. Benoît Habert, Adeline Nazarenko, André Salem (1997), *Les linguistiques de corpus*, Armand Colin, Paris, 1997,

Benoît Habert (2005), *Instruments et ressources électroniques pour le français*. Paris /Gap : Ophrys

Benoît Habert (2009), *Construire des bases de données pour le français : Tome 1, Notions*: Paris /Gap : Ophrys

4.1.1. Dépens : du mot autonome au constituant dégradé d'une locution

La première étude portera sur une illustration de la théorie de la grammaticalisation : en français moderne, le n.m. *dépens* n'est plus un mot autonome, il se présente uniquement dans les locutions *aux dépens de N* / *à DetPoss dépens* / [droit] *condamn* aux dépens*. Mais en français médiéval c'était un mot autonome, une variante du n.f. *dépense*, tous deux provenant du participe passé du v. lat. *dispendere* au neutre (*dispensum* > *despens* > *dépens*) ou au fém. (*dispensa* > *despense* > *dépense*).

L'exploration de FRANTEXT-2 permet d'élucider les conditions dans lesquelles les deux termes se sont complètement dissociés, l'un restant un mot autonome, l'autre se réduisant à un constituant d'une locution partiellement figée.

Compte tenu du nombre relativement réduit d'occurrences de *dépens* (incluant la graphie *de(s)pens*) dans l'ensemble de la base FRANTEXT intégrale, la méthode la plus appropriée consiste à enregistrer toutes ses occurrences d'une part dans les corpus d'Ancien français, de Moyen français et de Français pré-classique (période allant au total de ±900 à 1650) et d'autre part dans le corpus du 20^e siècle et celui du 21^e siècle.

Il est ensuite facile de repérer dans quelles occurrences *dépens* se présente à la suite de *aux* ~, de *à DetPoss* {*mes* / *tes* / *ses* / *nos* / *vos* / *leur* ~} ou dans un autre contexte et d'éclairer ainsi la perte d'autonomie du n.m. *dépens* (► §4.2).

Exemples d'emplois autonomes de *depens* en Moyen français

*ANONYME	Le Canarien, Pièces justificatives	1327	[avenir], sans jamais aller ne souffrir aller encontre comment ou par quelconque voye ou maniere que ce soit ; et rendre et pour en plein leur partie à l' autre tous consommé ,	<i>depens</i>	, <i>dommages</i> , <i>interests</i> qui soyent [faits] et soustenus par l' une des dites parties , (...)
*ANONYME	Le Canarien, Pièces justificatives	1327	il feut sur ce presentement rabattu la somme de <i>trente cinq livres pour les dommages, mises, despens, interestz et salaires des</i>	<i>depens</i>	, comme dessus est fait mention .
DESCHAMPS Eustache	Le Miroir de mariage	1385	Et fay d' elle ton espousée : Pran lors argent, or et finance, Et ton corps en honeur advance . Honoure la selon son temps ; Elle fera <i>petis</i>	<i>depens</i>	. Quant d' un voyage revendras, Finance preste trouveras.
*ANONYME	Ysaÿe le Triste	1400	- Et comment , fait Ysaÿe , ne passeroit on point par la sans avoir le bataille ? - Nennil , fait Menet , car il y a <i>LX hommes armés a ses</i>	<i>depens</i>	qui ne servent d'aultre cose fors de gaitier ceulx qui la passent .

4.1.2. Permanence et renouvellement dans le lexique de la souffrance

La seconde étude de cas aura une base empirique impliquant une certaine familiarité avec le français médiéval : elle applique à l'histoire du lexique français de la souffrance la thèse défendue par le germaniste Jost Trier en 1931 sur la reconfiguration de champs sémantiques au fil des siècles (► §4.3). Certains des mots à prendre en compte ont plus ou moins complètement disparu du français actuel (*dol*, *doloir*), ou sont d'un usage restreint (*pâtir*),

d'autres ne sont apparus qu'à la fin du Moyen-âge (*souffrance, langueur, languissant*), d'autres enfin figurent à toutes les époques, avec une fréquence modeste (*languir*) ou élevé (*souffrir, douleur*). Quant à *passion*, dérivé de *pâtir*, je n'en ai pas tenu compte, car en dehors du discours des religions chrétiennes (cf. *la passion du Christ, la Passion selon Saint-Mathieu de J.S. Bach*), ce nom est sorti du champ sémantique de la souffrance physique. La méthode employée pour constituer le corpus approprié consiste à

- sélectionner successivement les différents corpus historiques proposés par FRANTEXT-2 (ancien français, moyen français, français pré-classique, français classique),
- y adjoindre trois corpus personnels du français du 19^e, du 20^e et du 21^e siècle
- et rechercher les occurrences de tous ces termes (et éventuellement d'autres comme *dolent, souffreteux, souffrant, douloureux, indolore*, etc.) dans chaque sous-corpus.

On peut explorer ainsi le profil diachronique de chaque mot et la limitation progressive des contextes de certains d'entre eux, comme dans les tableaux ci-dessous répartissant les 34 occurrences de *languir* dans le corpus du 21^e siècle en fonction de la sous-catégorisation du sujet [H : humain vs NH : non-humain] et de la construction intransitive [Ø], réfléchie [S] ou factitive [F] :

Sujet	Ø / faire / se	Nb	Sujet	Ø / faire / se	Nb	Sujet	Ø / faire / se	Nb
H	Ø	5	NH	Ø	5	conversation	Ø	5
H	F	5	NH	F	0	conversation	F	0
H	S	13	NH	S	1	conversation	S	0

Quand le sujet est [humain] le sens est affectif, la construction la mieux représentée est réfléchie et elle est majoritairement accompagnée d'un complément introduit par la préposition *de* (en revanche quand la construction est intransitive, le sens est plutôt pathologique).

LAZARD Christian	Journal : 1930	2012	Déjeuner chez Paul Leroy , les Max [Lazard] nous avaient chargé de lui parler de leur sœur qui se	<i>languit</i>	d'amour pour lui .
GARAT Anne-Marie	L'enfant des ténèbres	2008	Elle l' aimait , nous le savions tous, et il ne la voyait pas se	<i>languir</i>	de lui . C'est si triste de passer à côté d'un amour , mais ce qui vous manque , le cherchez -vous vraiment ?
GARAT Anne-Marie	L'enfant des ténèbres	2008	En attendant , Gabrielle se	<i>languit</i>	au Mesnil , te sachant ici , en rupture de famille , sans nouvelles de toi
GARAT Anne-Marie	L'enfant des ténèbres	2008	il (...) s' était présenté en destination providentielle , à trois sauts de puce de Digne , chez sa grand-mère Louvain , qui se	<i>languissait</i>	de lui en sa grande maison sur la route de Gap .

JABLONKA Ivan	Histoire des grands- parents que je n'ai pas eus	2012	Toute la semaine, j'ai compté des bateaux qui arrivent ici de toute l' Europe, mais je me	languissais	de votre lettre .
JABLONKA Ivan	Histoire des grands- parents que je n'ai pas eus	2012	De notre côté, tout va comme d'habitude, nous sommes en bonne santé et ensemble comme avant. Raquel se	languit	déjà un peu moins .
GARAT Anne-Marie	Dans la main du diable	2006	Millie s' ennuyait de cette pantomime automate , malgré le cercle des domestiques applaudissant à la prouesse, malgré les encouragements de Gabrielle . Elle se	languissait	de Tout Roux , des leçons de musique , des promenades à cheval, et de ses cahiers
GARAT Anne-Marie	Dans la main du diable	2006	Maximilien Jamais rédige les éditoriaux du Temps ; il délaisse Justine qui se	languit	de lui
AYMÉ Marcel	Nouvelles complètes	2002	C'est vrai , je ne me sens point de courage à travailler , mais si vous voulez mon idée , c'est parce que je me	languis	de cette petite Léontine.
GARAT Anne-Marie	Nous nous connais- sons déjà	2003	il faut dire que dans ce coin l' Occupation n'avait pas été méchante , c'étaient de braves pères soldats qui se	languissaient	de leur pays, qu' on logeait ici et là, ils venaient casser la croûte aux cuisines et se gobergeaient de bon vin et de tricandilles grillées, ils avaient du tabac.
GUIBERT Hervé	Le Mausolée des amants	2001	Je me	languis	de lui et d'eux.

Quand le sujet est [non humain], la construction est majoritairement intransitive :

GAVALDA Anna	La Consolante	2008	— Tous les mois , il se précipite sur cette page -là pour savoir quel petit génie forcément moins malin que lui a gagné les 1 000 euros ... — 1000 euros ...	languit	l' écho , et elles sont toujours nulles leurs inventions ...
LEVY Marc	L'étrange voyage de Monsieur Daldry	2011	Un portail en fer forgé perçant un mur d' enceinte , une cour carrée où	languissait	un vieux figuier , des bancs qui vieillissaient sous un préau .
GARAT Anne-Marie	L'enfant des ténèbres	2008	On ne sait ce que retiennent des récits les enfants réfugiés sous la table des cuisines , abandonnés aux sommolences d' été , fin des repas , le temps	languit	. Les hommes sont absents , partis à la sieste ou à l' écurie , dehors le soleil brûle .

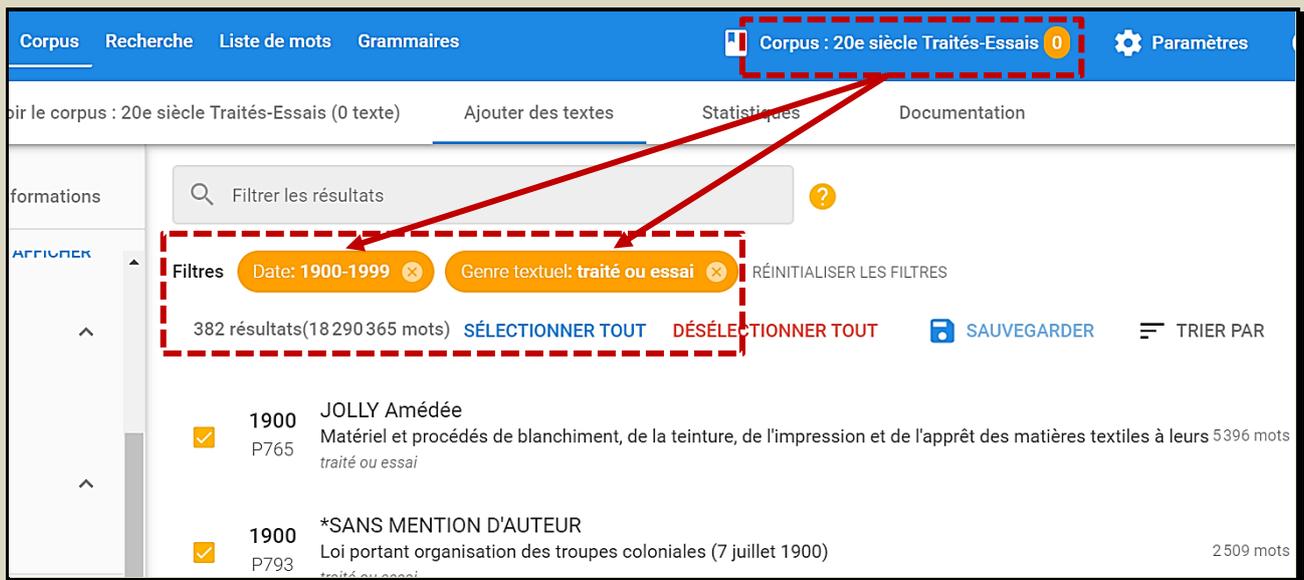
AYMÉ Marcel	Nouvelles complètes	2002	. Depuis que la panthère	languissait	au coin du feu , la maison était devenue triste .
MILLET Catherine	La vie sexuelle de Catherine M.	2001	Dans quelle divagation est -ce que je laissais alors partir ma pensée quand mon excitation commençait à	languir	?
GARAT Anne-Marie	Hongrie : blason	2009	Nul jour ici pour expliquer comment l' ombre des rêves mort-nés migre et s' attarde , se tord , et	languit	. — Je regrette mon pays d' avant-naître , un endroit invécu , inécrit , sinon que saurais -je de Hongrie ?

Enfin on découvre une collocation, **la conversation languit*** :

NOURISSIER François	À défaut de génie	2000	Clara tenait à rester sur les sommets . Elle y remontait d' un coup d' aile si la conversation	languissait	dans les sujets subalternes .
NOURISSIER François	À défaut de génie	2000	Le temps était médiocre et la conversation	languissait	. Parti dans les bois d' alentour , un fusil sous le bras que j' avais trouvé dans le vestibule , je tirai au jugé sur une ombre grise qui volait
GARAT Anne-Marie	Tranquille	2013	J' ai tenté de mettre ses photos de monstres sur le tapis , mais la conversation	languissait	, le cœur n' y était pas comme quand il s' excitait sur les curiosités de la nature , à se tordre de rire .
GARAT Anne-Marie	Dans la main du diable	2006	...Mme Victor feignit l'enjouée, se fit raconter des nouvelles de sa fille, le train dont allait la maison, là-bas , et décrire les magasins de Paris, mais la conversation	languissait	. Alors elle s' était mise à donner des nouvelles du pays , à tort et à travers .
AYMÉ Marcel	Nouvelles complètes	2002	Je les écoutais sans plaisir , sans ennui non plus , en m' appliquant courtoisement à paraître présent à leur propos . Je n' y réussissais pas toujours et , comme la conversation	languissait	, je les informai que je venais de purger huit mois de prison pour vol .

4.1.3. L'effet des genres et sous-genres textuels sur la sélection du sens du v. **investir**

J'ai brièvement évoqué au début du chap.2 l'esquisse d'une corrélation entre les genres textuels et les sens du verbe **investir**. C'est dans le genre désigné dans FRANTEXT-2 comme « Traités et Essais » que cette corrélation se laisse tester le plus aisément. La recherche assistée du lemme **investir** dans le corpus personnel croisant deux « métadonnées », la période du 20^e siècle et le genre des Traités et Essais, délivre 436 occurrences.



Tester la corrélation présumée est assez coûteux en temps, car cela implique le croisement de trois opérations :

1. la répartition du genre « Traités et Essais » en domaines thématiques ou sous-genres,
2. l'attribution d'une structure morphosyntaxique à chacune des 436 occurrences et
3. l'association à chaque occurrence d'un synonyme d'*investir* en fonction du contexte morphosyntaxique, avec le souci de limiter drastiquement l'éventail de ces synonymes.

Sur le premier point j'ai distingué neuf sous-genres :

Adm(inistration)
Dro(it)
Éco(nomie)

His(toire)
Ind(ustrie)
LitArt(littérature-art)

Phi(losophie)
Sci(ences)
Soc(iété)

Sur le second point, j'ai retenu 5 formats morpho-syntaxiques

qqn ~ N<bien|argent> (Lieu)
qqn ~ N<lieu>
qqn ~ qqn
qqn ~ qqn (Lieu)
qqn ~ qqn de N<bien|fonction>

Et sur le troisième point, j'ai réuni 7 synonymes :

assiéger
doter
engager
installer

intégrer
occuper
soigner

Ces trois opérations préalables une fois réalisées, le croisement des données présenté plus loin [► §4.4] s'effectue aisément.

4.2. Comment un lexème est sorti de l'usage autonome : dépens

La rubrique Étymologie et histoire de l'article *dépens* du TLF accorde peu de place à l'emploi autonome (§1, daté de 1170) au profit de la « locution prépositionnelle » apparue au tout début du 14^e siècle (§2). *Le dépens* est défini comme « ce que l'on dépense » et *la dépense* comme « l'action de dépenser de l'argent », mais aussi comme « somme dépensée, frais », ce qui signifie qu'à l'origine les deux termes sont mutuellement substituables dans certains contextes.

- **DÉPENS**, subst. masc. plur.

Étymol. et Hist.

1. 1170 « ce que l'on dépense » (CHR. DE TROYES, *Erec et Enide*, éd. M. Roques, 3498); en partic, ca1260 *justice despens et dommages* (*Liv. de Just. et de Plet*, éd. Rapetti, IX, XII, p. 176);
 2. loc. prép. *aux dépens de*,
 - a) 1306 « aux frais de » *aus despens du roy* (JOINVILLE, *St Louis*, éd. N. de Wailly, § 400);
 - b) 1580 « au détriment de » (MONTAIGNE, *Essais*, éd. A. Thibaudet, livre 1, chap. 22 : nos souhaits intérieurs pour la plupart naissent et se nourrissent aux **despens** d'autrui);
 - c) 1673, 27 oct. *au despens de* + inf. (M^{me}DE SEVIGNE, *Lettres*, éd. M. Monmerqué, t. 3, p. 254).
- Du lat. *dispensum*, part. passé, subst. neutre de *dispendere* proprement « peser en distribuant », d'où « distribuer » en lat. impér.; cf. le dér. *dispendium* « frais, dépense »; maintien du -n- prob. en raison du caractère juridique du mot et de sa relation toujours sentie avec l'a. fr. *dispendre* « dépenser » (XII^es., T.-L.), issu de *dispendere*.

- **DÉPENSE**, subst. fém.

- A. 1. 1176 « action de dépenser de l'argent; somme dépensée, frais » *a grant despance* (CHR. DE TROYES, *Cliges*, éd. A. Micha, 4533);
 2. 1693 « emploi, utilisation d'une chose quelconque » ici en parlant du temps (LA BRUY., *Disc. sur Theophr. ds LITRE*);
 3. 1851 « quantité de matière utilisée » (COURNOT, *Fondem. conaiss.*, p. 531 : Le rapport qui subsiste entre la dépense de la machine et son effet utile).
 - B. Ca 1180 « lieu où l'on conserve les provisions » *despenses... celiers* (M. DE FRANCE, *Fables*, 9, 21 ds T.-L.); fin XII^es.-début XIII^es. « nourriture » (*Jourdain de Blaye*, éd. P.-F. Dembowski, 2363).
- Du lat. *dispensa* part. passé fém. subst. de *dispendere*, v. *dépens*. B lat. médiév. « ce qui est nécessaire à l'entretien, provisions » (819 ds NIERM.), d'où « garde-manger » (*ibid.*).

Progressivement, *dépens* entre dans la classe des locutions, lesquelles présentent, dans les termes de la linguistique cognitive des « degrés de schématicité » variables. On peut distinguer trois degrés de schématicité croissante avec deux variantes pour le troisième :

Degré de schématicité croissant	exemple
1. Construction purement substantielle	<i>prou</i> [peu ou ~]
2. Construction substantielle extensible	<i>fur</i> [au ~ et à mesure (d*/qu*)]
3. Construction partiellement schématique	
3a. par variation temporelle sur le composant verbal	<i>lurette</i> [il y AVOIR belle ~ (qu*)]
3b. par variation sur la détermination possessive	<i>dépens</i> [à DetPoss / aux ~ (d*)]

L'exploration du corpus du MOYEN FRANÇAIS délivre sept occurrences de **dépens** dont seulement 2 introduits par **aux ~** ou **à mes/ses ~** :

~, dommages, interests
despens, interestz et salaires des ~
paya tous mes ~
payés mes ~
petis ~

hommes armés à ses ~
soudoyé à mes ~

En revanche, le corpus du FRANÇAIS CLASSIQUE ne délivre plus que 2 occurrences d'un emploi autonome contre 63 de l'emploi figé :

diminue les depens (1551)	1
les depens passent le principal (1646)	1
à mes ~	14
à ses ~	7
à tes ~	6
à nos ~	2
aux ~ de ma vie	2
aux ~ de mon cœur	2
à leurs ~	1
à mes propres ~	1
à vos ~	1
au ~ d'un sot	1
aux ~ d'autrui	1
aux ~ de la mienne	1
aux ~ de la sienne	1
aux ~ de la vie	1
aux ~ de leurs têtes	1
aux ~ de l'honneur	1
aux ~ de ma tête	1

aux ~ de mes larmes	1
aux ~ de mon sang	1
aux ~ de nos peines	1
aux ~ de nos pleurs	1
aux ~ de qui il apartiendra	1
aux ~ de sa foi	1
aux ~ de sa propre vie	1
aux ~ de sa vie	1
aux ~ de son honneur	1
aux ~ de son pauvre mineur	1
aux ~ de ton aise	1
aux ~ de ton sang	1
aux ~ de vos jours	1
aux ~ du barbier	1
aux ~ du fermier	1
aux ~ du Roi	1
aux ~ du sien	1
aux ~ d'un Empire	1
soudoyer à ses ~	1

Les 7 extraits datés du moyen français et du français pré-classique présentant **dépens** sans la préposition **à** sont :

Moyen français	ANONYME	Le Canarien, Pièces justificatives	1327	et rendre et pour en plein leur partie à l' autre tous consommé,	depens	, dommages , interests qui soyent [faits] et soustenus par l' une des dites parties ,
	ANONYME	Le Canarien, Pièces justificatives	1327	il feut sur ce presentement rabattu la somme de trente cinq livres pour les dommages, mises, despen , interestz et salaires des	depens	, comme dessus est fait mention .
	DESCHAMPS Eustache	Le Miroir de mariage	1385	Honoure la selon son temps ; Elle fera petis	depens	. Quant d'un voyage revendras , Finance preste trouveras .

	Lannoy Ghillebert de	Voyages et ambassades 1399-1450	1450	Item, au départir de laditte armée , alay devers le roy de Portugal, lequel me recueilla grandement et paya tous mes	dépens	parmy son royaume .
	ANONYME	La Confession Riffart	1480	RIFFLART. Cuidez vous doncques que je voise, Se vous ne payés mes	dépens	? LE PRESTRE . Se Dieu m' aïst , je me repens
Français préclassique	Cotereau Claude	Les douze livres de Lucius ...	1551	ce qui augmente le pris des fruicts, qui y sont serrez, et diminue les	dépens	de ce qu'on y veult amener
	DU LORENS Jacques	Satires	1646	Mais souvent les	dépens	passent le principal , et qui ne peut ne peut , vertu - bieu , c'est le mal .

L'examen des contextes de **dépens** dans le corpus du 20^e siècle à l'aide de la fonction « Clusters / N-Gram » du concordancier AntConc (► Chap.2, §2.1) ne révèle aucun contexte droit significatif, mais une variété de contextes gauches qu'on peut répartir en trois classes en fonction des seuils de fréquence absolue 5, 10 et 50 : à ses dépens figure seul dans la 1^{ère} classe avec 51 occurrences, les autres déterminants possessifs constituent la 2^{ème} classe avec 10 à 33 occ. et la préposition amalgamée **aux** la 3^{ème} classe avec 7 occ. La 4^{ème} est constituée de contextes plus spécifiques avec des fréquences négligeables.

Fréq	Contexte gauche	Pivot
51	à ses	dépens
33	mes	dépens
32	à mes	dépens
31	à nos	dépens
23	à leurs	dépens
10	à tes	dépens
10	à vos	dépens
7	aux	dépens
3	et à leurs	dépens
3	ont appris à leurs	dépens

2	ai appris à mes	dépens
2	amuser à ses	dépens
2	appris à ses	dépens
2	condamné aux	dépens
2	est à ses	dépens
2	l' humour à ses	dépens
2	me pourvoir à vos	dépens
2	pourvoir à vos	dépens
2	s' amuser à ses	dépens
2	vivre à ses	dépens

CONCLUSIONS

1. En moyen français (1350-1549) **dépens** présente déjà un figement morphologique en nombre (uniquement au pluriel), mais pas de figement syntaxique.
2. En français pré-classique (1550-1650), **dépens** ne se présente plus qu'exceptionnellement en dehors du contexte à Det(Poss) ~ (de N) : 2/65 = 3,1%.
3. En français du 20^e siècle, **dépens** ne se présente plus que dans ce contexte, dans un format semi-substantiel, semi-schématique.

En outre on rencontre 3 cas de coordination interne, le 1^{er} sans répétition de la préposition **à** :

à mon N et mes ~

BERGER Yves <i>Le Sud</i> 1962	je sentais ma mémoire chargée , mécanique perfide elle enregistrait à mon insu et mes	dépens	, la vie comme elle est , vieillissante et mortelle .
--------------------------------	---	--------	--

contrairement aux deux autres :

aux ~, et Adv à Det N de N

WEIL Simone	Œuvres	1929	Si le système capitaliste se trouve plus ou moins endommagé dans l'affaire , ce ne peut être qu' aux	dépens	et non au profit des valeurs humaines et du prolétariat , si loin que puisse peut-être aller en certains cas la démagogie.
CHAZELLE Jacques	La Diplo- matie	1962	à côté , aux	dépens	, et souvent à la place des organes purement politiques , les organes techniques en viennent aujourd'hui à jouer un rôle prépondérant .

4.3. L'évolution du lexique de la souffrance entre l'ancien français et le français du 19^e siècle

Le lexique de la souffrance a beaucoup évolué entre l'ancien français représentés par les premiers écrits enregistrés dans FRANTEXT-2 et le français du 19^e siècle. Le tableau ci-dessous représente la fréquence brute de 4 verbes (*pâtir*, *souffrir*, *languir*, *doloir*), 4 noms (*souffrance*, *langueur*, *dol*, *do(u)leur*) et 1 adjectif (*languissant*) au fil de cinq corpus définis par leur datation qui subdivise cette histoire. Ces 9 termes appartiennent à 4 familles lexicales

1. *pâtir* (je n'ai pas tenu compte de *passion* en raison de l'évolution de son sens)
2. *souffrir* ; *souffrance* (on peut ajouter *souffrant* ; *souffreteux*)
3. *languir* ; *langueur* ; *languissant*
4. *doloir* ; *dol* ; *do(u)leur* (on peut ajouter *dolent*)

Fréquences brutes	<i>pâtir</i>	<i>souffrir</i>	<i>souffrance</i>	<i>languir</i>	<i>langueur</i>	<i>languissant</i>	<i>doloir</i>	<i>dol</i>	<i>do(u)leur</i>	TOTAL
ancien fr.	0	322	4	39	4	0	39	114	192	714
moyen fr.	14	2149	104	303	83	15	62	29	1847	4606
fr. pré-classique	150	4695	288	542	452	226	1	32	4038	10424
fr. classique	43	12692	860	1323	1241	664	0	7	11842	28672
fr. du 19 ^e s.	41	14011	5192	748	1063	404	0	77	12095	33631

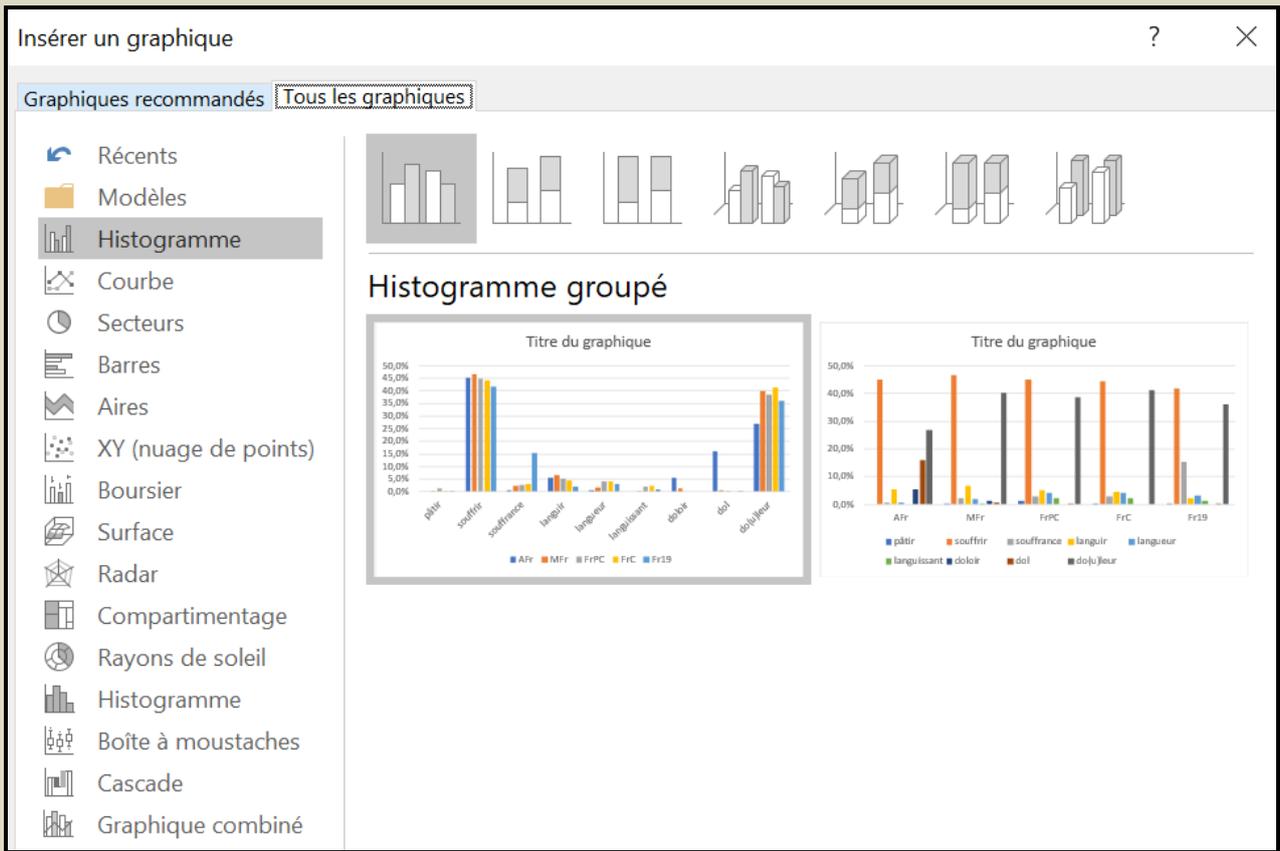
Les fréquences brutes enregistrées dans le tableau ci-dessus ne sont pas interprétables en raison du nombre régulièrement croissant des extraits enregistrés à chaque époque (47 fois plus d'extraits dans le corpus du 19^e siècle que dans celui de l'ancien français). Il faut donc remplacer les **fréquences brutes** par des **fréquences relatives** en divisant pour chaque époque le nombre d'occurrences de chaque lemme par le nombre total des occurrences des 9 lemmes, comme dans le tableau ci-dessous.

Fréquences relatives	<i>pâtir</i>	<i>souffrir</i>	<i>souffrance</i>	<i>languir</i>	<i>langueur</i>	<i>languissant</i>	<i>doloir</i>	<i>dol</i>	<i>do(u)leur</i>
Ancien fr.	0,0%	45,1%	0,6%	5,5%	0,6%	0,0%	5,5%	16,0%	26,9%
Moyen fr.	0,3%	46,7%	2,3%	6,6%	1,8%	0,3%	1,3%	0,6%	40,1%
Fr. pré-classique	1,4%	45,0%	2,8%	5,2%	4,3%	2,2%	0,0%	0,3%	38,7%
Fr. classique	0,1%	44,3%	3,0%	4,6%	4,3%	2,3%	0,0%	0,0%	41,3%
Fr. du 19 ^e s.	0,1%	41,7%	15,4%	2,2%	3,2%	1,2%	0,0%	0,2%	36,0%

Les pourcentages surlignés en jaune sont supérieurs à 10%, ceux surlignés en gris sont inférieurs à 1%. L'adoption de ces deux seuils permet d'associer à chaque terme un profil évolutif.

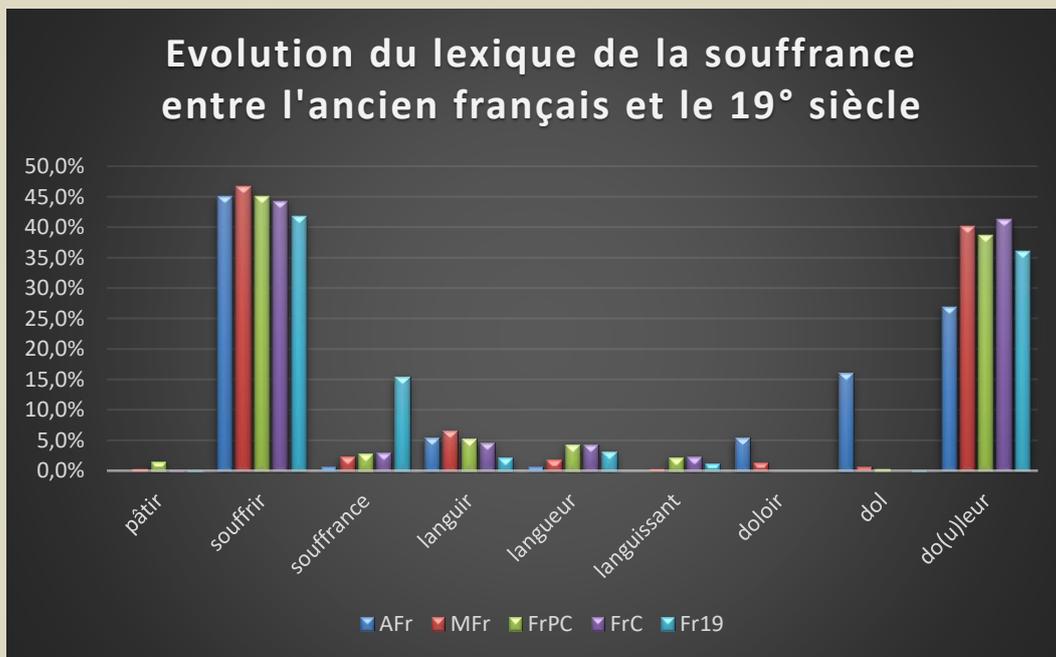
- Le verbe *souffrir* et le nom *do(u)leur* se distinguent clairement par leur présence largement au-dessus de 10% à toutes les époques. Ils constituent donc deux invariants.
- Parmi les verbes, la permanence de *souffrir* diffère de la régression de *doloir* qui n'est plus enregistré en français pré-classique (1550-1650) et du profil de *pâtir* qui ne dépasse 1% qu'à cette époque avant de régresser.
- Parmi les noms, la permanence de *do(u)leur* diffère également de la progression brutale de *souffrance* au 19^e s., peut-être en compensation de la régression tout aussi brutale de *dol* entre l'ancien et le moyen français.
- Quant à la famille de *languir* / *langueur* / *languissant*, elle évolue dans une bien moindre mesure entre 0% et 6,6%.

Ces résultats gagnent en clarté une fois reportés dans un graphique, mais de quel type ? EXCEL en propose une multitude :



J'en ai sélectionné trois, dont le format est respectivement un histogramme groupé, un histogramme empilé et des courbes renseignées.

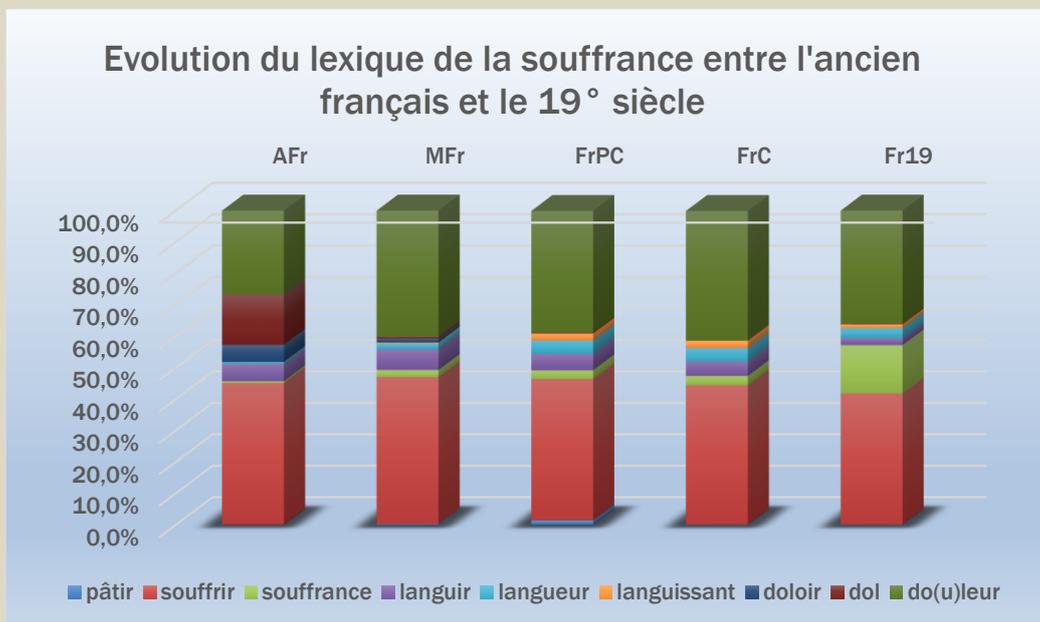
Graphique 1 par histogramme groupé



L'histogramme groupé rassemble pour chaque mot sa présence relative dans les cinq périodes spécifiées. Il a l'avantage de mettre en valeur la forte présence de *souffrir* et de *do(u)leur* avec

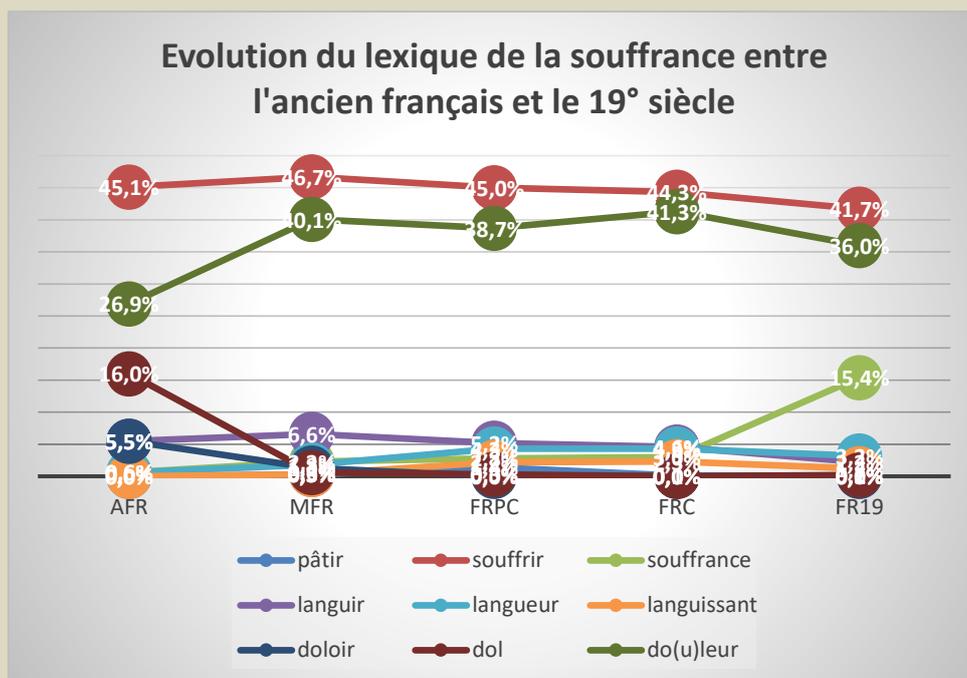
une remarquable permanence sur plus de mille ans. On lit aussi clairement le profil inverse de **souffrance** qui progresse brutalement au 19^e siècle et de **dol** qui disparaît pratiquement dès le moyen-français. Mais les autres profils sont peu lisibles sur ce graphique.

Graphique 2 par histogramme empilé



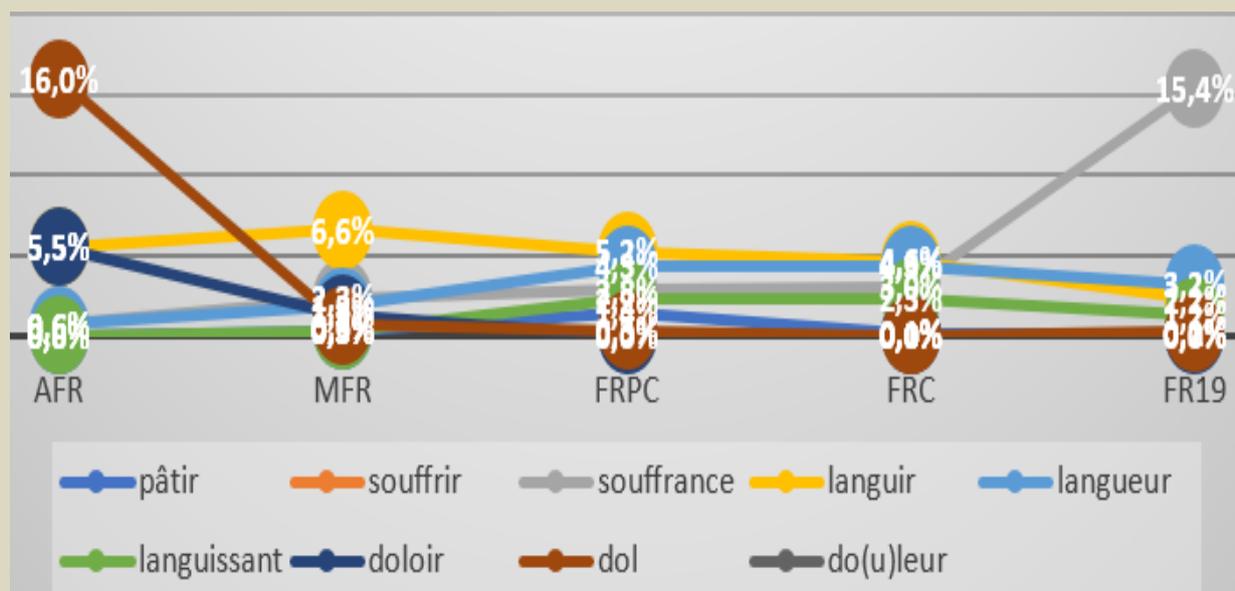
Le graphique par histogramme empilé regroupe sur une même colonne les valeurs des 9 mots examinés. Ici aussi la place dominante de **souffrir** (en brun-rouge) et de **douleur** (en vert) dans chacune des 5 périodes est bien lisible, ainsi que la présence de **dol** (en rouge sang) limitée à l'ancien français, mais le poids exact de chaque mot reste opaque en raison de leur empilement.

Graphique 3 par courbes renseignées



Le 3^{ème} graphique vise à compenser les défaillances du 2^{ème} en conservant la disposition des époques en abscisses. On y voit clairement que **souffrir** occupe la 1^{ère} place à toutes les époques et **douleur** la seconde. On distingue aussi que **dol** occupe la 3^{ème} en ancien français avant de s'écrouler et que **souffrance** fait une percée en français du 19^e siècle, mais ici aussi le profil des 7 autres mots reste opaque à partir du moyen français.

Toutefois, il est possible d'agrandir la partie difficilement lisible du graphique pour mieux distinguer ces 7 profils :



4.4. Le sens privilégié du verbe **INVESTIR** dans les différents domaines thématiques des « *Traités & essais* »

Comme nous l'avons vu plus haut (► §4.1.3) il existe en français actuel une corrélation au moins partielle entre les domaines thématiques du genre « *Traités & essais* » et les 4 sens principaux du verbe investir (respectivement caractérisés par les synonymes [1] ASSIEGER / S'EMPARER, [2] DOTER, [3] ENGAGER et [4] INSTALLER).

La partie gauche du tableau ci-après fournit les fréquences brutes de chacun de ces 4 sens dans les 11 domaines distingués au §4.1.3 et la partie droite fournit les fréquences relatives par domaine thématique.

Essais-traités 20° s.	TOTAL	assiéger-s'emparer	doter	engager	installer	TOTAL	assiéger-s'emparer	doter	engager	installer
TOTAL	436	50	152	206	26	100,00%	11,47%	34,86%	47,25%	5,96%
ADM	53	0	39	11	3	12,16%	0,0%	73,6%	20,8%	5,7%
AGR	6	0	0	6	0	1,38%	0,0%	0,0%	100,0%	0,0%
DRO	46	0	29	0	17	10,55%	0,0%	63,0%	0,0%	37,0%
ECO	151	1	5	144	0	34,63%	0,7%	3,3%	95,4%	0,0%
HIS	27	5	16	4	2	6,19%	18,5%	59,3%	14,8%	7,4%
IND	24	0	8	16	0	5,50%	0,0%	33,3%	66,7%	0,0%
LIT-ART	17	7	10	0	0	3,90%	41,2%	58,8%	0,0%	0,0%
PHI	47	17	25	5	0	10,78%	36,2%	53,2%	10,6%	0,0%
SCI	21	8	6	7	0	4,82%	38,1%	28,6%	33,3%	0,0%
SOC	26	6	11	8	1	5,96%	23,1%	42,3%	30,8%	3,8%
STRA	6	5	1	0	0	1,38%	83,3%	16,7%	0,0%	0,0%

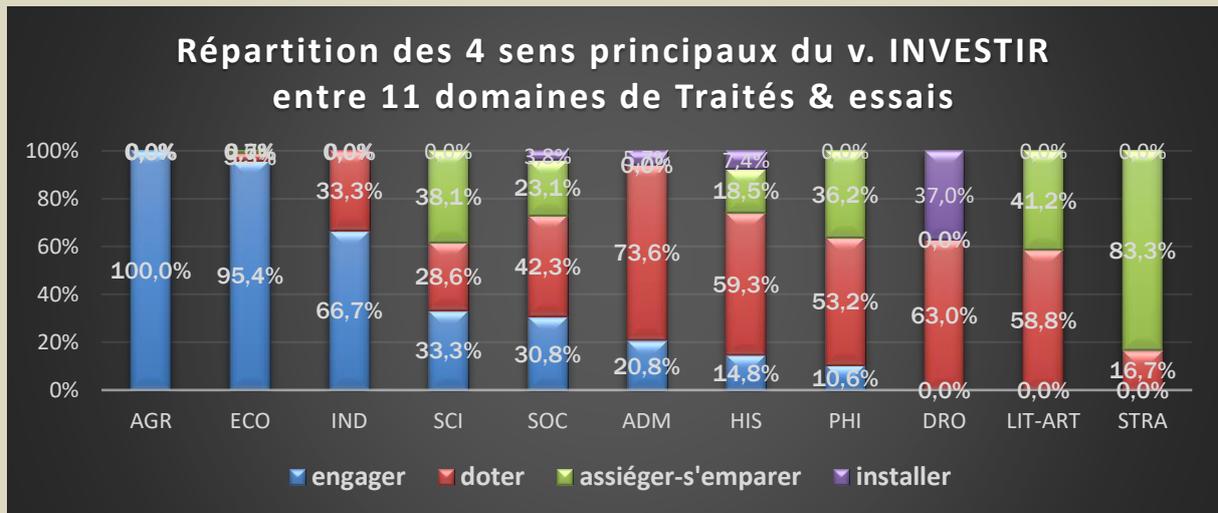
On peut distinguer trois types de profils :

1. Dans les domaines AGR(onomie) et ÉCO(nomie), un seul sens s'impose, **ENGAGER**.
2. Dans les domaines ADM(inistration), DRO(it), IND(ustrie), LIT(térature)-ART, PHI(losophie) et STRA(tégie), le 1^{er} sens partage l'essentiel des occurrences avec un 2^e sens. En droit p.ex., il s'agit en premier de **DOTER** (63%) et en second d'**INSTALLER** (37%).
3. Enfin dans les 3 autres domaines, HIS(toire), SCI(ences) et SOC(iété), trois sens se partagent le domaine. P.ex. en histoire, le 1^{er} sens est **DOTER** (59%), suivi d'**ASSIEGER/S'EMPARER** (19%) et d'**ENGAGER** (15%).

La représentation de ces corrélations partielles entre domaines thématiques et sens lexicaux est difficile car le tableau source (ci-dessous) croise 4 sens et 11 domaines dans 44 cases.

Base de calcul des 2 graphiques	engager	doter	assiéger-s'emparer	installer
AGR	100,0%	0,0%	0,0%	0,0%
ECO	95,4%	3,3%	0,7%	0,0%
IND	66,7%	33,3%	0,0%	0,0%
SCI	33,3%	28,6%	38,1%	0,0%
SOC	30,8%	42,3%	23,1%	3,8%
ADM	20,8%	73,6%	0,0%	5,7%
HIS	14,8%	59,3%	18,5%	7,4%
PHI	10,6%	53,2%	36,2%	0,0%
DRO	0,0%	63,0%	0,0%	37,0%
LIT-ART	0,0%	58,8%	41,2%	0,0%
STRA	0,0%	16,7%	83,3%	0,0%

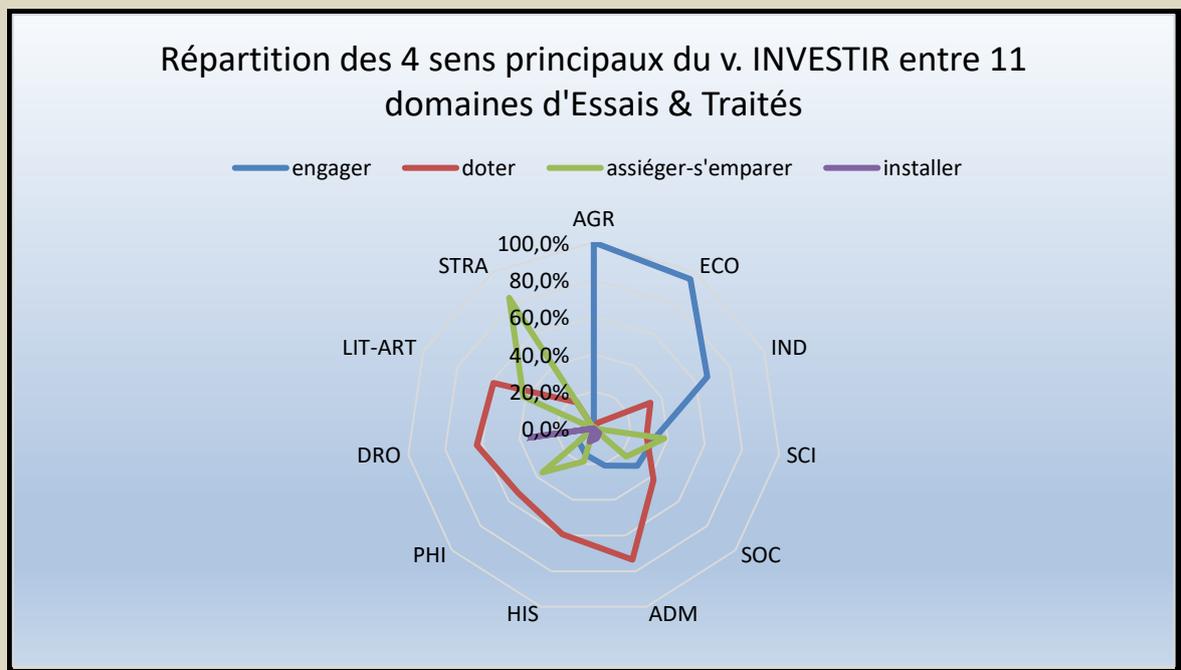
Graphique 1 : histogramme empilé et renseigné



Il est possible de renseigner la part de chaque sens dans chaque domaine de manière lisible à l'aide d'un histogramme empilé. L'ordre des domaines a été sélectionné en fonction de la part du sens **ENGAGER** qui décroît de gauche à droite. On constate que simultanément la part du sens **ASSIEGER-S'EMPARER** croît inversement, tandis que celle du sens **DOTER** reste fluctuante.

Le sens le plus largement distribué est **DOTER**, présent avec au moins 30% dans 7 domaines, suivi d'**ENGAGER**, présent avec la même fréquence relative dans 4 domaines. **ASSIEGER-S'EMPARER** est présent dans 3 domaines et **INSTALLER** (le sens le plus ancien) dans un seul.

Graphique 2 : 'Radar'



La présentation 'radar' met en évidence les domaines spécialisés dans un sens : **AGR** et **ÉCO** pour **ENGAGER**, **STRA** pour **ASSIEGER-S'EMPARER** et **ADM** pour **DOTER**. Le **DROIT** est le domaine privilégié du sens **INSTALLER**, lequel est faiblement représenté dans l'ensemble.

CONCLUSION GÉNÉRALE

Ce didacticiel s'écarte du format classique dans la mesure où seuls les chapitres 1 et 2 concernent spécifiquement les techniques d'exploitation de la base de données textuelles FRANTEXT-2 (le « *comment ?* »), tandis que les chapitres 3 et 4 portent sur le post-traitement et la recherche de corrélations entre le résultat d'une recherche et les « métadonnées » des extraits enregistrés (de date, de genre, de domaine thématique et d'auteur), étendant le questionnement au « *pourquoi ?* » ou au « *pour quoi faire ?* ».

FRANTEXT-2 est une base de données essentiellement littéraires, et pour l'étude du français actuel, il est utile de comparer différents types de corpus, notamment littéraires, politiques et journalistiques. Dans cette perspective, je me contenterai de renvoyer au Cahier du CRISCO 34 (Jacques François et Yacoub Gherissi, *Pour une linguistique orientée outils : la polysémie du verbe compter et les genres textuels*, 2012, téléchargeable à l'adresse <https://hal.archives-ouvertes.fr/hal-01811292>).

SOMMAIRE DES QUATRE CHAPITRES

Chap. 1 : Sélection d'un corpus et recherche « simple »	§1.1 <i>Les corpus prédéfinis et les corpus personnels</i>	p.I–2
	§1.2. <i>Intérêt et limites des recherches « simples »</i>	p.I–5
Chap.2 : Au-delà de la recherche « simple », les modalités de recherche « assistée et « avancée »	§2.1 <i>L'exploitation du concordancier AntConc</i>	p.2–1
	§2.2. <i>La recherche « assistée » : des formes de mots aux lemmes et aux catégories syntaxiques</i>	p.II–3
	§2.3. <i>La recherche « avancée » à l'aide du codage CQL (Corpus Query Language)</i>	p.II–8
Chap.3 : Le « post-traitement : pour un usage raisonné des algorithmes d'exploration lexicale	§3.1. <i>Pré-traitement vs post-traitement : L'exemple de [qqn tient à qqn / qqc / VINP]</i>	p.III–1
	§3.2. <i>Le post-traitement des collocations du N conscience</i>	p.III–4
	§3.2.1. <i>Exemples des cooccurrences les plus fréquentes dans l'ensemble du corpus</i>	p.III–5
	§3.2.2. <i>Fréquence relative de bonne conscience et mauvaise conscience dans l'ensemble du corpus</i>	p.III–8
	§3.2.3. <i>Prendre / reprendre / prise de conscience dans le contexte gauche, par auteur et par fréquence relative décroissante par rapport aux occurrences de conscience</i>	p.III–12
§3.2.4. <i>Fréquence relative de que P dans le contexte droit (par rapport aux occurrences de conscience)</i>	p.III–12	
	§3.3. <i>Conclusion</i>	p.III–13

Chap.4 : Trois études de cas	<i>§4.1. Retour sur les chapitres précédents et exploitation de FRANTEXT-2 pour les 3 études de cas</i>	p.IV-1
	<i>§4.1.1. Dépens : du mot autonome au constituant dégradé d'une locution</i>	p.IV-2
	<i>§4.1.2. Permanence et renouvellement dans le lexique de la souffrance</i>	p.IV-2
	<i>§4.2.3. L'effet des genres et des sous-genres textuels sur la sélection du sens du v. investir</i>	p.IV-5
	<i>§4.2. Comment un lexème est sorti de l'usage autonome : dépens</i>	p.IV-7
	<i>§4.3. L'évolution du lexique de la souffrance entre l'ancien français et le français du 19^e siècle</i>	p.IV-10
	<i>§4.4. Le sens privilégié du verbe investir dans les différents domaines thématiques des « Traités et essais »</i>	p.IV-14
Conclusion générale	p.IV-17	