
Lexico-grammaire et textométrie : identification et visualisation de schémas lexico-grammaticaux caractéristiques dans deux corpus juridiques comparables en français

Lexicogrammar and textometrics: Identification and visualisation of typical lexico-grammatical patterns in two comparable corpora of legal French

Christopher Gledhill, Stéphane Patin and Maria Zimina



Electronic version

URL: <http://journals.openedition.org/corpus/2868>

ISSN: 1765-3126

Publisher

Bases ; corpus et langage - UMR 6039

Printed version

Date of publication: 15 January 2017

ISBN: 1638-9808

ISSN: 1638-9808

Brought to you by Université Paris Diderot Paris 7



Electronic reference

Christopher Gledhill, Stéphane Patin and Maria Zimina, « Lexico-grammaire et textométrie : identification et visualisation de schémas lexico-grammaticaux caractéristiques dans deux corpus juridiques comparables en français », *Corpus* [Online], 17 | 2017, Online since 17 January 2018, connection on 12 June 2018. URL : <http://journals.openedition.org/corpus/2868>

This text was automatically generated on 12 June 2018.

© Tous droits réservés

Lexico-grammaire et textométrie : identification et visualisation de schémas lexico-grammaticaux caractéristiques dans deux corpus juridiques comparables en français

*Lexicogrammar and textometrics: Identification and visualisation of typical
lexico-grammatical patterns in two comparable corpora of legal French*

Christopher Gledhill, Stéphane Patin and Maria Zimina

1. Introduction

- 1 Nous nous proposons ici d'explorer le phénomène phraséologique des « schémas lexico-grammaticaux » (LG) dans une perspective de linguistique de corpus appliquée aux langues de spécialité. Il a souvent été observé que les genres textuels des langues de spécialité (LSP) obéissent à des conventions linguistiques et culturelles très fortes. Les genres spécialisés sont caractérisés par des séquences récurrentes de signes servant typiquement à structurer les différentes « sections rhétoriques » (Swales 1990)¹. En tant qu'unité de base servant à leur caractérisation, le schéma LG pourrait alors s'avérer utile.
- 2 L'objectif de cette contribution n'est cependant pas de définir formellement des schémas LG. À l'instar de linguistes dits « contextualistes » (Hoey 1991, Berber-Sardinha 1997, Williams 2003a et 2003b, Gledhill 2000, Khalifa 2015), nous estimons que la caractérisation des unités textuelles doit passer par deux étapes préliminaires : 1) l'observation formelle des co-occurrences² dans un corpus de textes, et 2) l'analyse des fonctions communicatives, pragmatiques et rhétoriques de ces phénomènes lexicaux dans le contexte d'un genre textuel spécifique (nous emploierons plus loin le terme « fonction discursive » pour regrouper ces différentes notions). Or, s'il n'est pas possible de proposer

une définition préalable des schémas LG sans analyser des exemples en discours, il convient néanmoins d'établir un certain nombre de principes qui, à nos yeux, constituent les éléments essentiels des schémas LG observés dans des études antérieures (Gledhill & Frath 2007, Gledhill 2011a, 2011b) :

- 3 a) chaque schéma LG comporte à la fois une « macro-structure » prévisible (correspondant à une structure grammaticale et une séquence de signes lexicaux dont la co-occurrence est observable et reconnue) et une « micro-structure » productive (la séquence n'est pas totalement figée ou lexicalisée ; cette variabilité permet d'intégrer le schéma dans des co-textes différents) ;
- 4 b) chaque schéma LG a un « macro-sens » global (correspondant à une fonction interpersonnelle ou textuelle dont le sens dépend des conventions du discours : il s'agit dans la plupart des cas du sens « pragmatique » de la séquence) et en même temps un « micro-sens » unique (un sens « propositionnel » dont la référence précise dépend des variations sémantiques du co-texte local) ;
- 5 c) chaque schéma LG est composé d'un ou plusieurs signes lexicaux (lexèmes de classe ouverte, car innombrables et en création continue) et grammaticaux (lexèmes de classe fermée, car limités). Peuvent intervenir également des morphèmes et des constructions abstraites, par ex. une extraposition, un verbe support, etc. ;
- 6 d) chaque schéma LG est composé d'un ou plusieurs élément(s) obligatoire(s) (le « pivot ») et d'un ou plusieurs éléments variables (le « paradigme ») : tous ces éléments peuvent être discontinus et plus ou moins éloignés les uns des autres. La variabilité de ces éléments (le pivot et le paradigme) permet au schéma LG d'être entrelacé dans d'autres schémas LG : en discours, ces imbrications mènent à la création de chaînes discursives plus étendues.
- 7 Afin de mieux illustrer notre propos, nous estimons que les trois exemples attestés ci-dessous (1a, 1b, 1c) constituent des réalisations différentes du même **schéma lexico-grammatical** :
 - 1a) [...] *il n'y avait dès lors pas lieu de **procéder** à une estimation à court terme [...]*
 - 1b) [...] *il n'a pas été possible de **procéder** à une évaluation concluante d'un agent [...]*
 - 1c) [...] *il convient de **procéder** à un second examen de toutes les régions [...]*
- 8 Nous commenterons ces exemples plus en détail dans la prochaine section (1.2). Il suffit de noter ici que, si les trois séquences comportent des éléments lexicaux variables, chacune possède la même macro-structure grammaticale (contenant deux structures : une proposition impersonnelle et une construction composée d'un Verbe support et d'un Nom prédicatif). De même, chaque séquence exprime le même type de macro-sens (une conclusion légale, formulée selon les conventions du discours législatif européen). Nous verrons plus loin que les formes et les fonctions des différents schémas construits autour des verbes pivots « procéder », « effectuer » et « réaliser », sont légèrement différentes selon les corpus dans lesquels nous les observons. Il n'est donc pas toujours possible d'attribuer à un schéma LG les mêmes caractéristiques fonctionnelles d'un schéma LG similaire observé dans un contexte différent.
- 9 L'objectif général de cette contribution est donc d'examiner comment les fonctions et les formes des schémas LG peuvent varier selon leur contexte d'emploi. Afin de démontrer ce principe, nous nous proposons d'examiner des schémas LG attestés dans deux corpus comparables de textes spécialisés. Mais avant d'examiner ces corpus, il convient de rappeler les origines du terme « schéma lexico-grammatical ».

- 10 La notion de « **lexico-grammaire** » fait référence au modèle systémique fonctionnel (Halliday & Matthiessen 2014). Le linguiste britannique M.A.K. Halliday a choisi ce terme pour souligner explicitement le degré d'imbrication postulé entre les niveaux lexical et grammatical du système linguistique : selon Halliday, dans la construction du discours, chaque choix grammatical est soumis à des contraintes lexicales, et vice versa. Le lecteur constatera aussi que nous nous servons de la métalangue du modèle systémique fonctionnel (SF) dans la définition des rôles sémantiques. Le modèle SF distingue six différents types de procès sémantique sur une échelle : matériel > comportemental > mental > communicatif > relationnel > existentiel (nous rencontrerons certains exemples de ces procès plus loin). De même, notre emploi du terme « fonction discursive » a été influencé par cette approche (cf. Gledhill 2012) car le modèle SF ne différencie pas la pragmatique et la sémantique, mais identifie plutôt trois « métafonctions » qui agissent simultanément dans la structuration du discours (et donc dans chaque schéma lexico-grammatical) : 1) la métafonction idéationnelle (= le sens propositionnel ou représentatif), 2) la métafonction interpersonnelle, et 3) la métafonction textuelle.
- 11 Il convient aussi de justifier notre choix du mot « schéma » parmi les nombreux autres termes employés dans la tradition lexicologique. Ce terme correspond à notre traduction – sans doute imparfaite – de l'anglais « *pattern* ». La notion du *pattern* a une histoire riche et complexe dans la tradition linguistique anglophone (Legallois & François 2006). Par exemple, le phraséologue averti reconnaîtra que notre conception des schémas LG a une certaine parenté avec les « lexical patterns » de Hunston & Francis (2000), les « sentence stems » de Pawley & Syder (1983) et les « unités linéaires » de Sinclair & Mauranen (2006). Mais à la différence de ces auteurs, nous estimons que les schémas LG comportent dans leur co-texte des régularités formelles qui s'étendent au-delà des éléments pivots, pour se chevaucher avec d'autres structures. Par ailleurs, notre conception des schémas LG se distingue des « constructions » de Goldberg (1995) et des « collostructions » de Stefanowitsch & Gries (2003) : à la différence de ces auteurs, nous nous focalisons sur le comportement des schémas LG dans des genres textuels spécifiques, ainsi que sur le fait que chaque schéma LG est défini par une fonction discursive globale que l'on peut associer à un contexte de communication conventionnel (le « macro-sens »)³. En somme, il nous apparaît surtout important de souligner que les schémas LG suivent une évolution spécifique pour exprimer des fonctions communicatives particulières dans les contextes précis des langues de spécialité.

1.1 Contexte de l'étude

- 12 Notre travail prend appui sur le volet français du corpus comparable multilingue élaboré par la Faculté de traduction et d'interprétation de Rome dans le cadre du projet international *Observatoire sur l'eurolecte. Analyse interlinguistique et intralinguistique des variétés juridiques dans l'Union européenne* (2011-2016)⁴. Ce corpus est constitué des 660 directives de l'Union européenne de 1998 à 2008 (D-UE) et de leur transposition en France, dans 129 textes de lois et ordonnances (T-FR), comptabilisant, respectivement, 4 181 496 et 1 862 241 occurrences. Il s'agit de textes juridiques comparables qui ont plusieurs similitudes : ces deux corpus sont en français et représentent des types discursifs hautement spécialisés (des directives et des lois). Dans le cadre du projet de l'*Observatoire sur l'eurolecte*, le rapprochement de ces ensembles textuels vise à pointer les caractéristiques linguistiques des directives de l'Union européenne en les contrastant

avec les textes nationaux issus des transpositions. La problématique de caractérisation de ce qui est appelé parfois l'eurolecte, *i.e.* technolecte spécifique (Goffin 2005) aux institutions européennes, a rapidement suscité un intérêt grandissant, comme en témoignent non seulement certains travaux de linguistes, traducteurs ou juristes (Gontrand 1991, Goffin 1994, 2002, 2005, Ciostek 2014, Biel 2014, etc.) mais aussi les différentes désignations néologiques employées pour décrire le phénomène. D'ailleurs si les termes tels que *eurolangue* ou *eurolecte* s'avèrent relativement neutres, d'autres, tels que l'anglais *eurofog*, le français *eurobabillage* ou encore l'allemand *euronebel*, dénoncent le caractère fortement cryptique de cette « langue ».

1.2 Problématique

- 13 Le haut degré de similarité des corpus **D-UE** et **T-FR** constitue un véritable défi pour le linguiste dans la mesure où plus les textes présentent des caractéristiques communes, plus il est difficile de caractériser leurs divergences linguistiques, textuelles ou discursives. C'est dans cette perspective que nous nous sommes fixé comme objectif principal d'explorer la façon dont les schémas LG sont déployés dans des contextes très similaires tout en étant associés à des fonctions communicatives distinctes. Prenons par exemple le verbe **procéder**, qui est employé régulièrement dans une structure de type *Verbe support + Nom prédicatif*. On notera en particulier une différence entre les emplois tirés du corpus **D-EU** et **T-FR** :
- 14 **D-UE**
- 1a) [...] *on **procédera** à un contrôle périodique de la surface [...]*
 - 1b) [...] *il n'y avait dès lors pas lieu de **procéder** à une estimation à court terme [...]*
 - 1c) [...] *il n'a pas été possible de **procéder** à une évaluation concluante d'un agent [...]*
- 15 **T-FR**
- 2a) [...] *le Conseil supérieur de l'audiovisuel **procède** à une consultation publique [...]*
 - 2b) [...] *des caractéristiques de la personne qui **procède** à la déclaration [...]*
 - 2c) [...] *la direction générale des douanes et des droits indirects qui **procède** à la publication des tarifs [...]*
- 16 Dans le volet **D-UE**, on note que le N prédicatif exprime un procès mental ou cognitif (souvent introduit par un participant non-animé), alors que dans le volet **T-FR** ce N exprime un procès communicatif (introduit par un participant humain)⁵. Ces observations indiquent l'existence d'une construction V + N très régulière dans ce type de contexte : nous avons précédemment démontré que les constructions V + N sont assez typiques des discours administratifs et juridiques français (Gledhill & Todirascu 2008). Cependant, pouvons-nous postuler ici l'existence d'un schéma LG plus étendu ? Et comment confirmer que les régularités identifiées ici sont statistiquement « saillantes », alors qu'on peut trouver les deux variations dans le corpus **D-UE** et **T-FR** ?
- 17 Lorsqu'il s'agit d'identifier les schémas LG caractéristiques dans des corpus de taille importante, il est nécessaire de recourir à la « linguistique outillée ». Né des rencontres pluridisciplinaires entre philologues, linguistes, analystes du discours et statisticiens du texte, le courant que l'on appelle aujourd'hui **textométrie** offre des bases solides pour ce type d'analyse outillée sur corpus (Lafon 1984 ; Lebart, Salem 1994 ; Heiden 2006 ; Söze-Duval 2008, Pincemin 2011 ; Zimina & Fleury 2015). La textométrie regroupe l'ensemble des méthodes quantitatives permettant d'opérer des réorganisations formelles de la séquence textuelle et des analyses statistiques portant sur l'ensemble des unités

textuelles d'un corpus. La particularité de l'analyse textométrique réside dans l'existence d'un rapport étroit entre les données textuelles, les modes d'observation de ces données et les résultats visés.

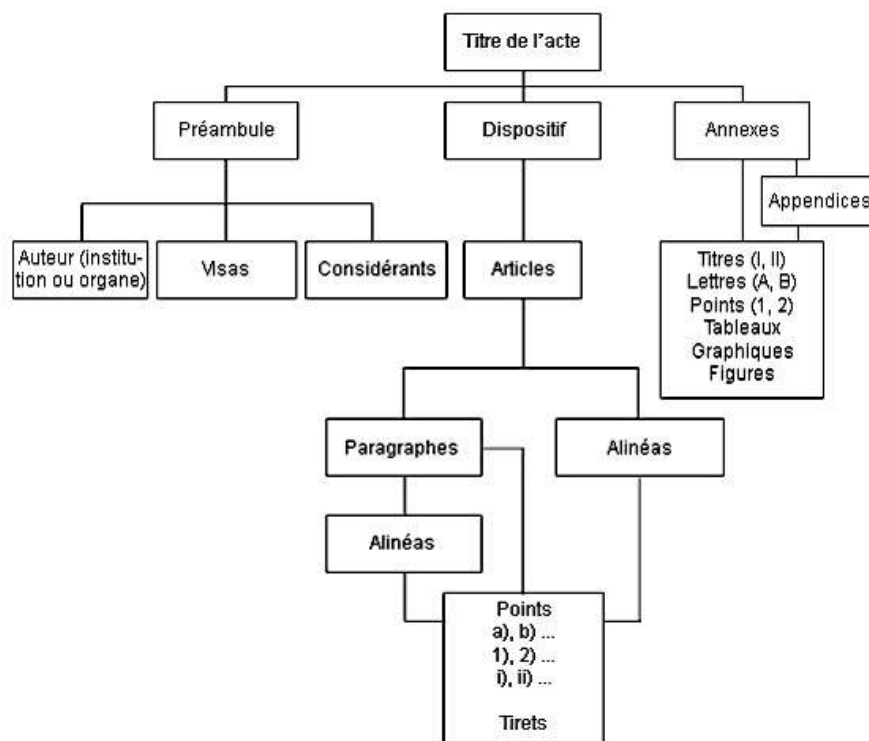
- 18 Notre étude repose sur les fonctionnalités du logiciel de textométrie *Le Trameur* (Fleury & Zimina 2014)⁶. Cet outil permet de mobiliser la Lecture Textométrique Différentielle : LTD (Patin *et al.* 2016) qui fournit des aides visuelles à la lecture contrastive de textes comparés, appuyées par l'affichage synchrone des résultats de leur analyse textométrique parallèle sous forme de surlignage au fil du texte.
- 19 Le fait d'employer des moyens informatiques pour chercher les schémas LG constitue cependant un défi non seulement pour le spécialiste du Traitement Automatique des Langues (TAL) et de la textométrie, mais aussi pour le phraséologue, qui se voit obligé de refondre ses systèmes de notation et même la métalangue utilisée. Ainsi, dans les lignes citées dans les exemples 1-2 ci-dessus, sommes-nous en présence d'un seul schéma LG ou de plusieurs ? Dans quelle mesure peut-on considérer des éléments du contexte à gauche (constructions passives, impersonnelles) ou du contexte à droite (déterminants et modificateurs du N prédicatif) comme faisant partie du même schéma ? Pouvons-nous par ailleurs postuler que ces deux schémas correspondent à une seule fonction discursive ?
- 20 Ces considérations nous mènent au deuxième objectif de cette contribution, qui consiste à explorer un certain nombre de pistes concernant le fonctionnement des schémas LG. En particulier, nous avançons l'hypothèse selon laquelle chaque schéma lexico-grammatical appartient à un type de discours reconnaissable selon les conventions discursives en question (Swales 1990, Hoey 1991, Gledhill 2012). Autrement dit, lorsqu'on extrait une séquence de signes d'un texte, il devrait être possible d'identifier le genre et le type discursifs auquel cette séquence appartient grâce à la comparaison avec d'autres schémas LG identifiés dans des corpus représentatifs du même genre textuel.

2. Corpus de travail

- 21 Les corpus de textes juridiques **D-EU** et **T-FR** sont extraits d'un corpus multilingue comparable composé de 660 directives de l'Union européenne de 1998 à 2008 (Corpus A de 29 398 122 occurrences), et de leurs transpositions dans 11 pays de l'UE, (Corpus B de 33 859 056 occurrences). Ce corpus multilingue a été conçu par l'*Observatoire sur l'eurolecte* pour mettre en évidence les différences entre le système linguistique employé dans les directives et les variétés juridiques nationales. Il n'est pas envisagé de comparer ces corpus avec un corpus de référence plus général. Les directives regroupées dans le volet **D-EU** font partie des textes du droit européen. Les directives fixent les objectifs à atteindre et laissent les États, *via* des textes nationaux, le choix des moyens pour les atteindre. La France doit donc mettre en œuvre des textes législatifs de « transposition » (regroupés dans notre corpus dans le volet **T-FR**).
- 22 Les corpus **D-EU** et **T-FR**, au format XML, sont structurés par des balises, répondant à la macro-structure des textes législatifs européens et nationaux : Titre, Préambule Disposition et Annexe, cf. Figure 1. Par « Préambule », on entend tous les éléments textuels se trouvant entre le Titre et le Dispositif de l'acte, c'est-à-dire, les Visas, les Considérants et les formules solennelles qui les entourent. Le Dispositif constitue la partie normative de l'acte. Il est composé d'Articles, éventuellement regroupés en parties, titres, chapitres et sections et peut être accompagné d'annexes. Enfin, l'Annexe d'un acte

contient généralement des règles ou des données techniques qui, pour des raisons d'ordre pratique, ne figurent pas dans le corps même du Dispositif (*Guide pratique commun pour la rédaction des textes législatifs de l'Union européenne* 2014, p. 45).

Figure 1. Macro-structure d'un acte juridique



3. Méthodologie

- 23 Les différences qui pourraient exister entre nos corpus de travail composés de textes très spécialisés et les corpus de textes de langue générale ne constituent pas l'objet de notre étude ; nous nous intéressons en revanche aux comparaisons internes de nos corpus de textes très proches sur le plan discursif (discours juridique européen), en étudiant des emplois caractéristiques de plusieurs phénomènes collocationnels présents dans les textes analysés. Nous démontrons notamment que les collocations type <V support + N prédicatif> intègrent des schémas LG plus étendus, typiques du français technico-administratif.

3.1 Fréquences générales

- 24 Pour une première approche quantitative du corpus **D-EU** et **T-FR**, nous avons retenu le lemme comme unité principale de dépouillement. Après l'étiquetage et la lemmatisation par *TreeTagger* (Schmid 1994) intégré dans le workflow du logiciel *Le Trameur*, l'analyse de la liste des cent verbes les plus fréquents des corpus **D-EU** et **T-FR** (cf. Tableau 1), a attiré notre attention sur la présence des lemmes « effectuer » (4 210 occurrences), « réaliser » (2 336 occurrences) et « procéder » (1 866 occurrences).

Tableau 1. Extrait de la liste des cent verbes les plus fréquents dans T-FR et D-EU

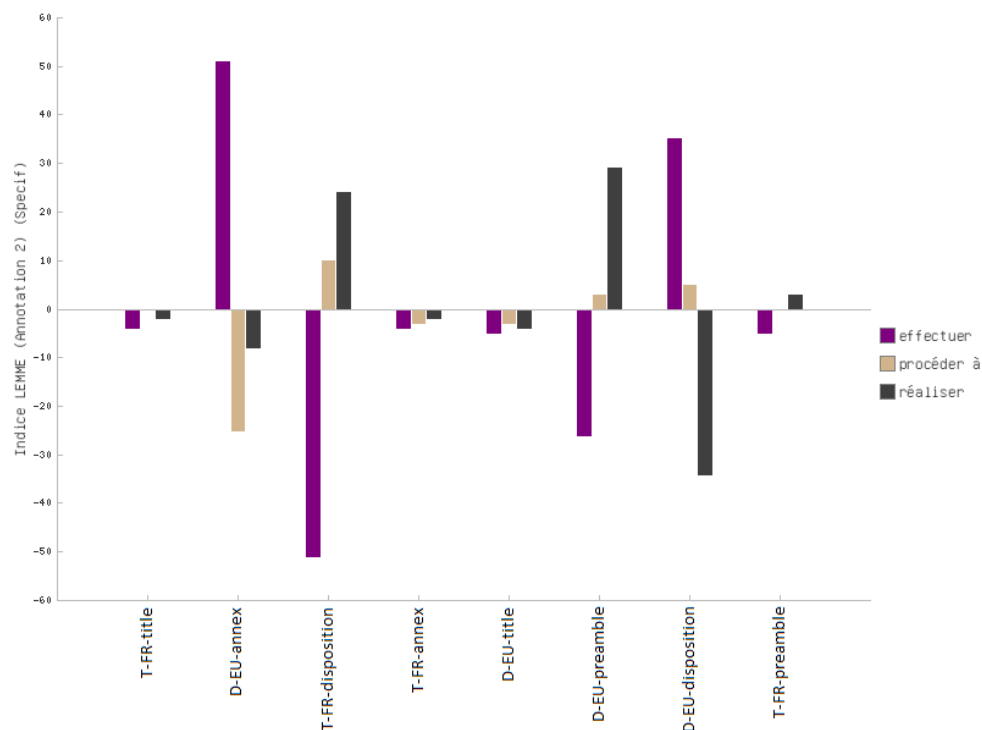
Verbe (lemme)	Freq T-FR	Freq D-EU	Freq Tot	Verbe (lemme)	Freq T-FR	Freq D-EU	Freq Tot
être	39498	52924	92422	adopter	299	3099	3398
pouvoir	7733	14896	22629	présenter	712	2580	3292
devoir	2303	18896	21199	contenir	386	2898	3284
prévoir	6648	7198	13846	autoriser	877	2355	3232
viser	1591	9733	11324	destiner	647	2553	3200
avoir	2463	7362	9825	compléter	2330	870	3200
modifier	3389	5820	9209	garantir	838	2123	2961
faire	3358	5200	8558	entreprendre	991	1961	2952
mentionner	6573	888	7461	durer	1282	1505	2787
rédiger	7327	120	7447	indiquer	221	2542	2763
concerner	1381	6052	7433	figurer	419	2303	2722
compter	3389	4032	7421	monter	1632	1041	2673
remplacer	4710	2627	7337	exercer	1560	1037	2597
mettre	1775	4851	6626	convenir	10	2569	2579
fixer	3128	3091	6219	arrêter	853	1647	2500
produire	1506	4325	5831	approprié	165	2299	2464
prendre	1736	3966	5702	informer	581	1869	2450
utiliser	631	4838	5469	réaliser	941	1395	2336
établir	1340	3978	5318	veiller	213	2051	2264
comprendre	1033	4230	5263	limiter	608	1600	2208
suivre	668	4505	5173	délivrer	811	1394	2205
appliquer	222	4829	5051	communiquer	490	1680	2170
demander	1779	3068	4847	ajouter	864	1274	2138
tenir	1266	3343	4609	constituer	753	1307	2060
définir	2245	2312	4557	désigner	815	1222	2037
effectuer	691	3519	4210	obtenir	379	1646	2025
permettre	1001	3094	4095	supprimer	1377	596	1973
insérer	3512	565	4077	situer	490	1478	1968
mesurer	572	3167	3739	lier	531	1410	1941
porter	1674	1984	3658	procéder	831	1035	1866
soumettre	1417	2214	3631	considérer	288	1573	1861
donner	958	2631	3589	placer	522	1311	1833
charger	2253	1273	3526	former	474	1343	1817
assurer	878	2613	3491	risquer	277	1530	1807
partir	880	2579	3459	relever	779	966	1745
échoir	953	2492	3445	traiter	126	1619	1745
fournir	540	2868	3408				
déterminer	1177	2221	3398				

- 25 Comme nous le verrons, ces verbes forment généralement des constructions type <V support + N prédicatif>. Dans la discussion suivante, nous démontrons comment identifier et visualiser les schémas LG associés à ces trois verbes dans les deux sous-corpus D-EU et T-FR.

3.2 Spécificités

- 26 L'analyse du corpus s'appuie sur la mise en opposition de certaines parties par rapport à d'autres, afin de comparer les variations de vocabulaire et les emplois caractéristiques.

Figure 2. Spécificités des lemmes « effectuer », « réaliser » et « procéder à » dans les parties de T-FR et D-EU



- 27 La comparaison des distributions des verbes lemmatisés « effectuer », « réaliser » et « procéder à » dans les différentes parties des documents juridiques regroupés dans **D-EU** et **T-FR** permet de constater que leurs profils d'emploi sont hétérogènes. L'histogramme de la Figure 2 s'appuie sur les résultats du calcul des *spécificités* (Lebart & Salem 1994) implémenté dans *Le Trameur*. Cette méthode permet de repérer les unités textuelles sur-employées (spécificités positives), ou au contraire sous-employées (spécificités négatives) dans chacune des parties des actes juridiques qui constituent le corpus : Titre, Préambules, Dispositions et Annexes. En fonction d'un seuil retenu, l'*indice de spécificité* indique si une unité textuelle est plutôt mal représentée ou plutôt abondante dans une partie donnée par rapport à ce que le modèle hypergéométrique laissait prévoir (Lafon 1984, Lebart & Salem 1994).
- 28 En termes de présence caractéristique en corpus **D-EU** et **T-FR**, on repère aisément sur la Figure 2 que le V « effectuer » déclenche une spécificité positive dans les Dispositifs des directives européennes (D-EU-disposition : +35) et dans les Annexes (D-EU-annex : +51). Le V « réaliser » est sur-employé dans les Préambules des directives européennes (D-EU-preamble : +29) et, dans une moindre mesure, dans les Préambules des lois nationales (T-FR-preamble : +3) et leurs dispositifs (T-FR-disposition : +24). Le V « procéder à » déclenche un diagnostic de spécificité positive dans les Dispositifs des lois nationales (T-FR-disposition : +9) et dans les Préambules et Dispositifs des directives européennes (D-EU-preamble : +3, D-EU-disposition : +5).
- 29 Pour les emplois qui relèvent des spécificités positives, peut-on identifier les fonctions discursives de ces trois verbes et déterminer si elles sont similaires ou distinctes dans les différentes parties des corpus comparables ? Par exemple, comme nous le verrons dans la discussion des résultats, le V « réaliser » est plutôt associé aux formules solennelles que

l'on trouve souvent dans les Préambules des deux sous-corpus. Par contre, le V « effectuer » est employé dans une variété de schémas LG que l'on associe aux directives. Pour démontrer ces régularités, nous faisons appel au calcul des *cooccurrences spécifiques*. Ce calcul utilise la méthode des spécificités pour analyser les contextes d'apparition d'une unité textuelle (forme, lemme, catégorie) utilisée en tant que « pôle » d'exploration. Selon notre hypothèse, l'analyse du voisinage contextuel des pôles « effectuer », « réaliser » et « procéder à » dans les différentes parties du corpus pourrait donner des indices sur le fonctionnement de ces verbes en discours. Ainsi, pour chaque partie du corpus où la présence caractéristique du pôle a été démontrée par le calcul des spécificités, on repère les éléments caractéristiques du voisinage contextuel immédiat correspondant à la phrase.

4. Résultats

30 Dans cette section nous examinons un échantillon de schémas LG potentiels associés aux verbes support « effectuer », « réaliser » et « procéder à » repérés au cours de nos explorations. Nous adopterons les conventions de notation suivantes afin de souligner la relation entre exemples attestés et schémas LG supposés :

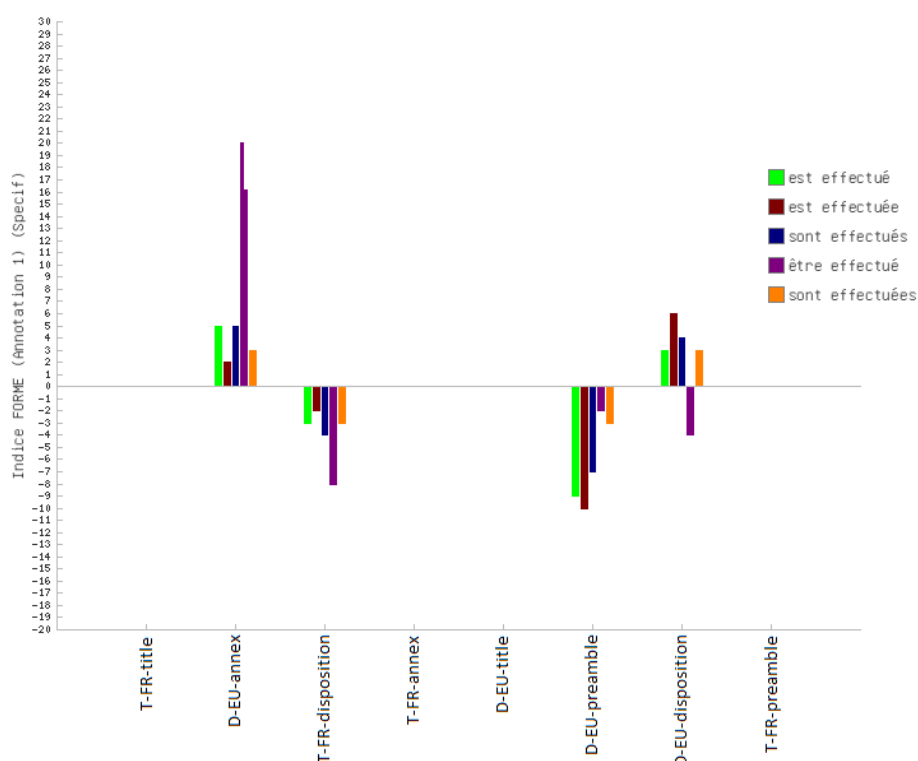
- /</ début du schéma
- />/ fin du schéma
- /.../ discontinuité
- /**gras**/ élément obligatoire (dit « pivot »)
- /*italique souligné*/ élément obligatoire (mais variable)
- /*italique*/ élément variable (dit « paradigmatique »)

4.1 Effectuer

4.1.1 Cooccurrences spécifiques de « effectuer »

31 Le calcul des cooccurrences spécifiques tient compte des segments répétés (Salem 1987). Pour le verbe « effectuer », nous avons identifié la cooccurrence des lemmes « effectuer » + « être » dans D-EU-disposition et D-EU-annexe, reflétée par la présence caractéristique des segments répétés « est effectué(e) », « sont effectué(e)s », et du segment « être effectué » qui déclenche une forte spécificité dans les Annexes des directives (88 occurrences), cf. Figure 3 ci-dessous.

Figure 3. Segments répétés avec « effectuer » + « être » dans les parties de T-FR et D-EU



4.1.2 Schémas LG associés à « effectuer »

- 32 Comme nous l'avons vu, « effectuer » est V support statistiquement saillant dans les Dispositifs et les Annexes du corpus **D-UE**. Dans ces contextes, « effectuer » est le plus souvent employé au passif : cette structure met des éléments circonstanciels en position finale de la proposition, et donc en position dite « focale » (Halliday & Matthiessen 2014).
- 33 L'analyse des contextes permet d'observer trois sous- schémas (E1-E3) pour ce verbe dans le corpus. Le premier (E1) établit les conditions législatives dans lesquelles doivent se dérouler des processus cognitifs (contrôles, inspections) mais aussi des activités matérielles industrielles (émission, traitement, transmission, etc.). Cet emploi a la particularité de comporter régulièrement deux groupes circonstanciels différents, portant sur la manière et sur le lieu. Le circonstanciel spatial accomplit une fonction de renvoi, tout acte juridique devant s'inscrire dans un ordonnancement juridique établi :
- E1.1 <Le traitement est effectué dans les conditions fixées [...] au paragraphe 6> [...]
- E1.2 <Les essais d'émission sont effectués conformément aux procédures ... décrites dans la présente directive> [...]
- E1.3 <La transmission des données est effectuée suivant les spécifications [...] figurant à l'annexe V> [...]
- 34 Le deuxième sous-schéma (E2) prend la même forme, tout en présentant un sémantisme plus restreint : cette construction comporte en position post-verbale un circonstanciel qui précise une personne ou un organisme chargé d'effectuer des « contrôles, examens, mesures », etc. :
- E2.1 <Un examen clinique des volailles et autres oiseaux captifs présents dans l'exploitation d'origine et, en particulier, de ceux à transporter est effectué par le vétérinaire officiel> [...]

E2.2 Article 81. Les États membres veillent à ce que <des inspections **soient effectuées** sous la responsabilité de l'autorité compétente> pour vérifier le respect des dispositions de la présente directive [...]

E2.3 <La communication **est effectuée** via les autorités compétentes de l'État membre> [...]

- 35 Le dernier sous-schéma (E3) présente les mêmes caractéristiques formelles, cette fois avec un modifieur circonstanciel servant à préciser la périodicité ou la durée d'application du procès exprimé par le N prédicatif :

E3.1 Au besoin, <un lavage pourra **être effectué** à l'issue de ce délai> [...]

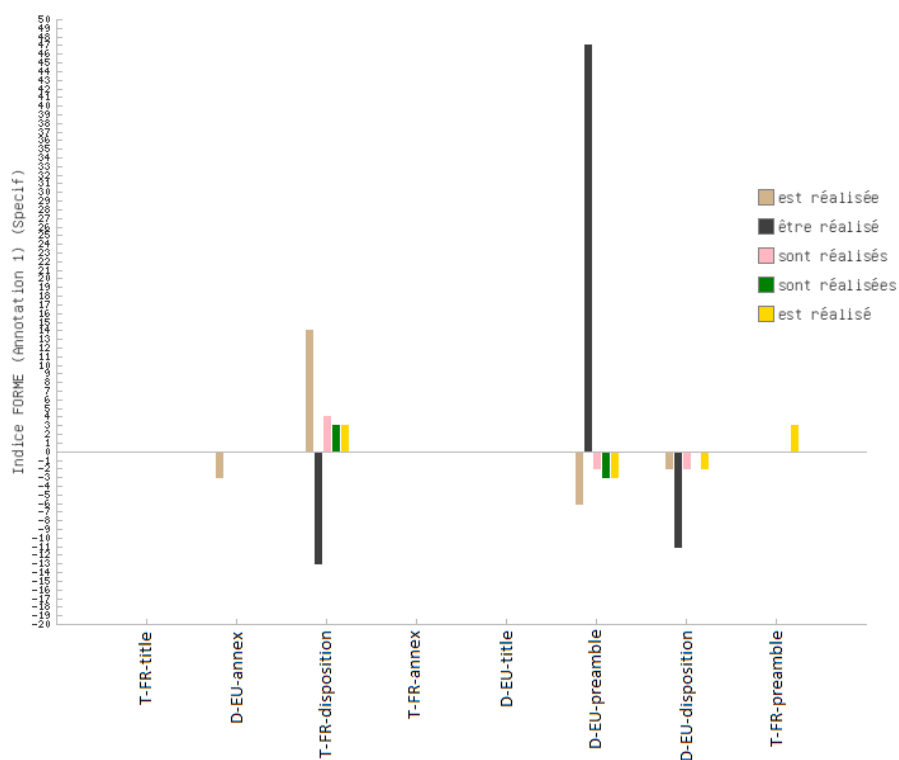
E3.2 <Ces mesures devraient **être effectuées** tous les trois mois> [...]

E3.3 Le coordinateur veille à ce que <le calcul visé au premier alinéa **soit effectué** au moins une fois par an> [...]

4.2 Réaliser

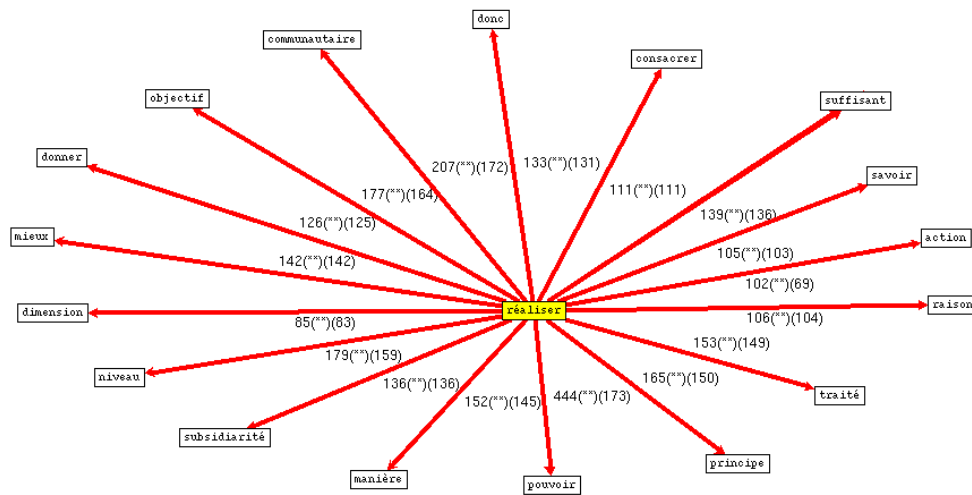
4.2.1 Cooccurrences spécifiques de « réaliser »

Figure 4. Segments répétés avec « réaliser » + « être » dans les parties de T-FR et D-EU



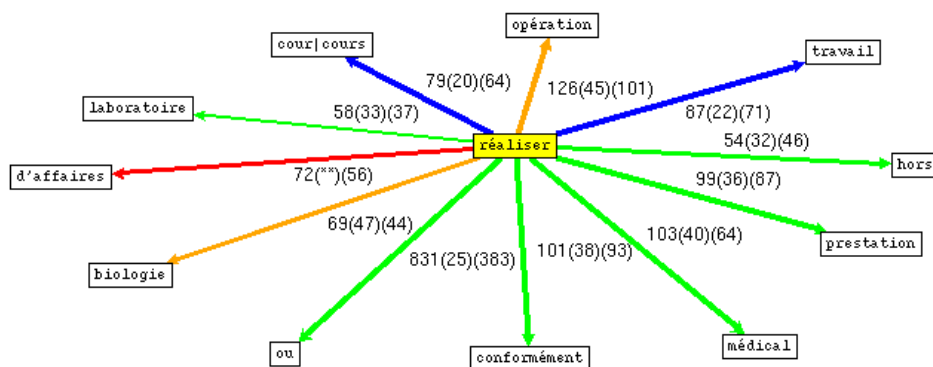
- 36 Pour le verbe « réaliser », nous sommes face à des profils très différents dans les deux corpus **D-EU** et **T-FR**. Dans les Préambules des directives européennes, on constate une très forte présence caractéristique du segment « être réalisé » (83 occurrences, indice de spécificité : +47), cf. Figure 4. De même, parmi les premiers cooccurents de « réaliser » triés par le nombre de contextes partagés on trouve, par exemple, « pouvoir » (173 contextes), « communautaire » (172 contextes), « objectif » (164 contextes), cf. Figure 5.

Figure 5. Cooccurents spécifiques du pôle « réaliser » dans les Préambules des directives (partie D-UE preamble)



- 37 Dans les Préambules et le corps des Dispositifs des lois nationales, on note la présence caractéristique des segments répétés « est réalisé(e) », « sont réalisé(e)s ». Le voisinage contextuel de « réaliser » dans les Dispositifs des lois est caractérisé par l'utilisation de N prédicatifs : « opération » (101 contextes), « prestation » (87 contextes), « travail » (71 contextes), etc. (cf. Figure 6). L'épaisseur des traits représentés sur la Figure 6 reflète le nombre de contextes partagés, tandis que la couleur correspond à la force de l'attraction lexicale définie en fonction des seuils de spécificité. Il s'agit de l'environnement lexical du pôle. Ainsi, les traits rouges (qui reflètent les attractions les plus fortes, dont l'indice de spécificité est supérieur à +50), permettent d'identifier les collocations N + V (ex. : « chiffre d'affaire » + « réaliser »). Les autres couleurs (orange, vert et bleu) montrent d'autres associations lexicales typiques des schémas LG. Comme nous le verrons plus loin, ces éléments correspondent aux éléments « paradigmatiques » (variables) que l'on observe dans les trois sous-schémas LG bâtis autour du V « réaliser ».

Figure 6. Cooccurents spécifiques du pôle « réaliser » dans les dispositifs des lois nationales (partie T-FR-disposition)



4.2.2 Schémas LG associés à « réaliser »

38 Nous avons indiqué plus haut que « réaliser » est sur-employé dans les Préambules du corpus **D-EU** et dans les Dispositifs du corpus **T-FR**. Comme « effectuer », « réaliser » est typiquement employé au passif.

39 Le premier sous-schéma (R1) est comparable au sous- schéma E1. Mais à la différence de E1, les syntagmes prépositionnels employés en position post-verbale (focale) sont toutes bâties autour du même adverbe <conformément... aux bonnes pratiques, à la présente directive>, etc. De même, par contraste avec E1, où les participants principaux de la proposition sont des instances administratives, dans le sous-schéma R1, les agents impliqués dans ces processus plutôt techniques sont des fabricants ou des entreprises (ce qui est souvent explicité dans le co-texte) :

R1.1 <Le fabricant s'assure que les opérations de fabrication **sont réalisées conformément aux bonnes pratiques** de fabrication et à l'autorisation de fabrication> [...]

R1.2 Dans le cas de médicaments expérimentaux, le fabricant s'assure que <toutes les opérations de fabrication **sont réalisées conformément à l'information donnée par le promoteur**> en application de l'article 9 [...]

R1.3 Le degré de sécurité est considéré comme satisfaisant lorsque <ces parties ont **été réalisées conformément aux prescriptions** d'une société de classification agréée.> [...]

40 Un deuxième sous-schéma (R2) se démarque assez nettement des emplois précédents par la présence systématique du circonstanciel <au niveau communautaire>. Le groupe verbal est typiquement modalisé ou explicitement évalué (notamment par l'adverbe « mieux »). La fonction discursive de cet ensemble paraît assez précise : il s'agit d'une déclaration sur la pertinence légale de la présente disposition qui doit être exposée sous forme de motivation dans les préambules. C'est l'une des « formules solennelles » mentionnées plus haut (section 2) :

R2.1 <Une évaluation scientifique de la question devrait **être réalisée au niveau communautaire**>, afin d'aboutir à une décision unique sur les points litigieux, et contraignante pour les États membres concernés [...]

R2.2 <L'objectif recherché peut donc **être mieux réalisé au niveau communautaire**> grâce à l'adoption de règles harmonisées concernant l'introduction de restrictions d'exploitation dans le cadre des règles de gestion [...]

R2.3 Étant donné que <cet objectif peut **être mieux réalisé au niveau communautaire**>, la Communauté peut arrêter des mesures conformément au principe de subsidiarité énoncé à l'article 5 du traité [...]

41 Enfin, le sous-schéma R3 comporte une conjonction subordonnante introduisant une proposition dite « contingente » (stipulant une conséquence ou un objectif plutôt que les conditions dans lesquelles s'accomplit le procès exprimé dans la proposition principale). Le groupe verbal de la proposition principale comporte systématiquement un modal (devoir). Sur le plan pragmatique, ce schéma correspond à une recommandation ou à une injonction technique portant sur un aspect de la sécurité ou sur des pratiques normatives :

R3.1 <L'humidification de l'échantillon doit **être réalisée de manière que**> l'action soit la même que celle de la pluie et de la rosée naturelles [...]

R3.2 <L'installation doit **être conçue et réalisée pour que**> les nuisances internes et externes [...] ne dépassent pas les valeurs limites [...]

R3.3 <une réduction sensible de ses effets négatifs devrait **être réalisé de telle sorte qu'**> une double imposition soit évitée [...]

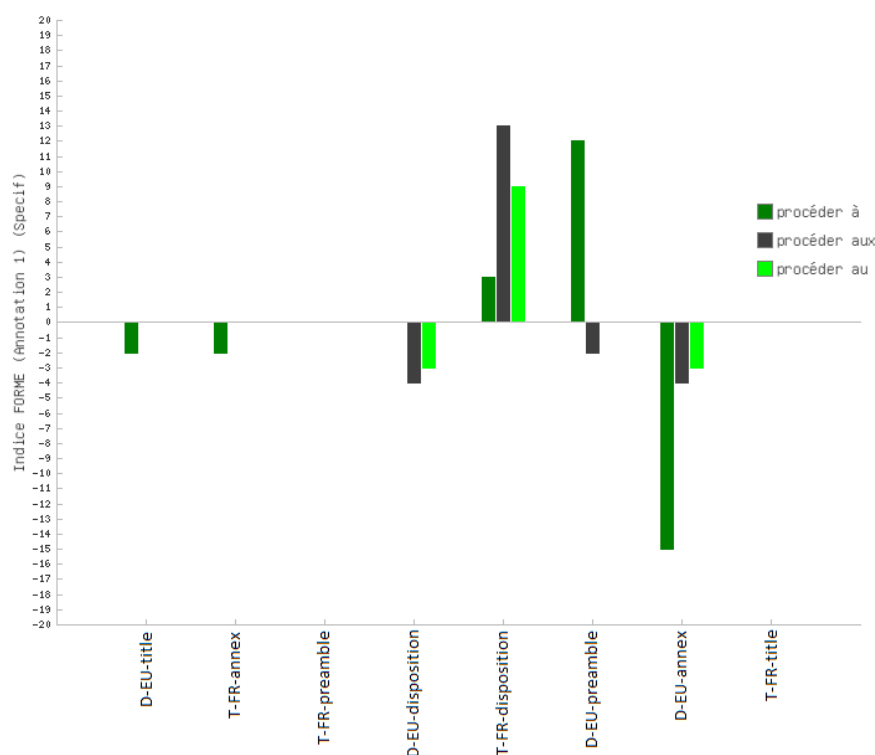
- 42 Il apparaît ainsi que les différents sous-schémas associés aux verbes « réaliser » et « effectuer » se distinguent en partie par les éléments qui les modifient (exprimant des rôles circonstanciels comme le moyen, le lieu, l'étendue, mais aussi parfois des participants agentifs). Mais, malgré des similitudes entre les deux V, les collocations associées à « réaliser » sont plus restreintes. On décèle aussi une certaine spécialisation rhétorique de ce V par rapport à « effectuer » : en particulier le schéma R2, dont la fonction communicative représente un emploi typique des Préambules, comme le souligne notamment le diagnostic de spécificité positive de ce V dans les Préambules du corpus D-EU (cf. Figure 4) et l'analyse des cooccurrents associés (cf. Figure 5).

4.3 Procéder

4.3.1 Poly-cooccurrences de « procéder »

- 43 Le calcul des poly-cooccurrences (Martinez 2012) permet de considérer les éléments en cooccurrence comme parties intégrantes d'un système dont les différents composants entretiennent une relation de dépendance simultanée. On peut préciser ces résultats en introduisant dans le calcul les partitions T-FR et D-EU (Titre, Préambule, Dispositif, Annexe), cf. Figure 7. Le calcul rapporte les associations lexicales dans chaque partie et permet de dresser une évolution de l'univers lexical du pôle au fil du texte. Par exemple, le verbe « procéder » dans les Préambules des directives est en cooccurrence avec les lemmes « il », « un », « convenir », « clarté », « souci », etc., comme le montre le Tableau 2 ci-après.

Figure 7. Segments répétés du verbe « procéder » suivi de « à|au|aux » dans les parties de T-FR et D-EU



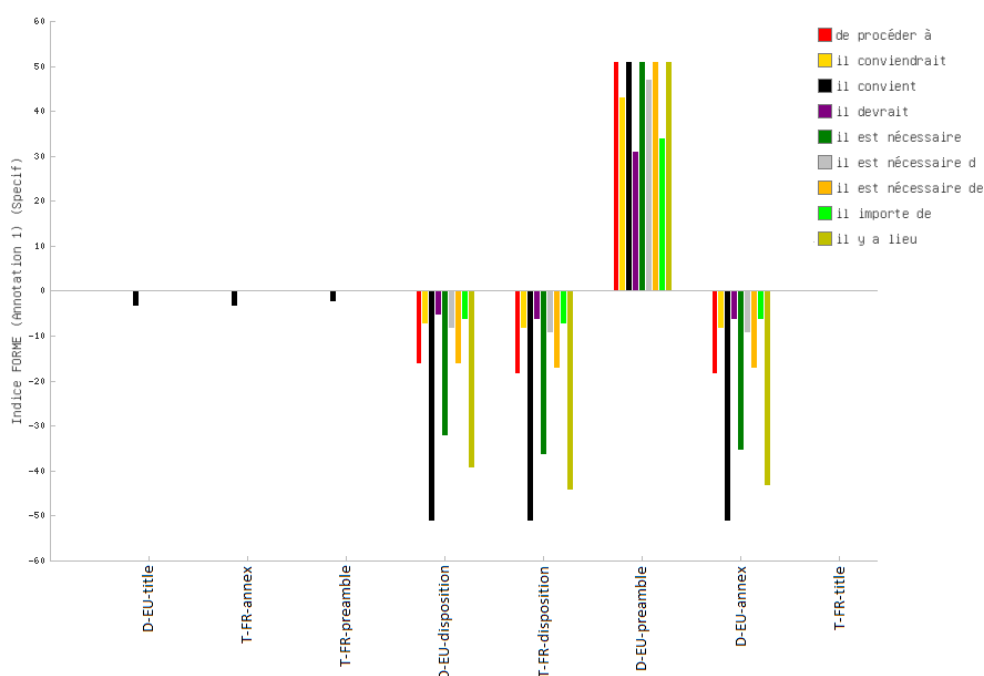
- 44 Le calcul des poly-cooccurrences livre une liste des attractions lexicales autour d'un pôle dans une unité contextuelle (ici phrase). On calcule d'abord pour le pôle « procéder » les cooccurrents spécifiques « de », « à », « il », « un », « convenir », etc. Dans leurs contextes communs, on calcule pour les pôles « procéder » + « de » les cooccurrents spécifiques « à » et « il ». Les pôles « procéder » + « de » + « à » ont pour cooccurrent spécifique « il », etc. Durant l'exploration, différents filtrages conditionnent l'épuisement des explorations contextuelles et réduisent le bruit dans les résultats pour privilégier l'information la plus spécifique : seuils maximaux de fréquence et de spécificité du cooccurrent, nombre minimal de contextes où se produit la cooccurrence, etc. À l'issue du calcul, on sélectionne les chemins originaux en écartant les chemins qui se contiennent (« procéder à » contenu dans « de + procéder + à ») ou qui se répètent.

Tableau 2. Cooccurrents (lemmes) du pôle « procéder » dans les Directives (partie D-EU-preamble)

Pôle	Cooc	F(cooc)	co-freq	Specif	contextes
procéder	de	43 280	658	31	172
procéder	à	15 853	294	26	162
procéder	il	9 168	155	12	134
procéder	un	9 902	189	18	132
procéder	convenir	2 050	85	28	85
procéder	clarté	121	65	**	65
procéder	souci	121	58	**	58
procéder	ledit	477	52	38	51
procéder	rationalité	49	45	**	45
procéder	codification	46	44	**	44
procéder	refonte	36	24	40	24
procéder	évaluation	458	29	15	23
procéder	modification	283	21	13	20

- 45 Les poly-cooccurrences peuvent être mises en rapport avec les segments répétés. Ainsi, les résultats nous amènent à considérer en contexte des attractions lexicales multiples où l'on observe une attraction stable entre le segment « de procéder à » et des distributions parallèles formées par les variantes « il convient » / « il conviendrait » / « il est nécessaire » / « il importe de », etc. utilisées dans les Préambules des directives (cf. Figure 8).

Figure 8. Segments répétés en relation de cooccurrence dans les parties de T-FR et D-EU



- 46 Muni des informations rapportées par l'exploration co-occurrentielle, on peut extraire les contextes où se réalisent les attractions constatées. Ces contextes correspondent aux sous-schémas LG commentés dans la section qui suit.

4.3.2 Schémas LG associés à « procéder »

- 47 Le V « procéder » se distingue des verbes « effectuer » et « réaliser » par le fait qu'il est typiquement employé à la voix active. La voix active n'implique pas pourtant l'expression d'un agent : dans de nombreux cas, le verbe est employé dans une construction impersonnelle. La fonction globale de ce type de construction est de mettre une construction impersonnelle ou modale en position thématique (= initiale) tout en mettant en position focale (= finale) le complément du V support (un N prédicatif exprimant, comme nous le verrons, un processus technique dont le cadre légal est défini dans le texte).
- 48 Le premier sous-schéma que nous observons (P1) comporte une structure impersonnelle au passif (veiller à ce qu'il soit procédé à...) ou une extraposition (il est nécessaire de..., il convient de... il y a lieu de) suivie par *V support + N prédicatif*. Ici, le procès exprimé par le N relève en général du domaine mental (examen, observation). Généralement, dans ces exemples, les auteurs du règlement expriment un jugement (formulé par une extraposition) sur le déroulement du processus (correspondant au N prédicatif) :
- P2.1 <Il y a lieu de **procéder** à une évaluation des différentes méthodes de recyclage> [...]
- P2.2 Si l'examen qualitatif met en évidence des signes d'altérations neuropathologiques, <il convient de **procéder** à un second examen de toutes les régions du système nerveux qui présentent ces altérations> [...]
- P2.3 Dans de nombreux cas, <il peut ne pas être nécessaire de **procéder** à une enquête en vue de déterminer la source de contamination [...]
- 49 Le deuxième sous-schéma (P2) correspond à une extension plus spécialisée du sous-schéma P1 : il s'agit d'une séquence relativement invariable introduite par l'extraposition

<il convient de>. Ce schéma a une fonction discursive très spécifique : les auteurs appellent à une reformulation de la réglementation existante. Ce schéma est semblable au sous-schéma R2, mais dans ce cas la structure du schéma LG comprend plusieurs constructions plus ou moins variables mais dont la présence est obligatoire (notamment le modifieur « dans un souci de clarté / pour des raisons de rationalité » etc.) :

P2.1 <A l'occasion de nouvelles modifications, il convient, dans un souci de clarté, de **procéder** à la refonte desdites directives> et, dans un but de simplification, de rassembler leurs dispositions en un texte unique [...]

P2.2 <À l'occasion de nouvelles modifications de ladite directive, il convient, pour des raisons de clarté, de **procéder** à une refonte des dispositions en question> [...]

P2.3 <Il convient dès lors, pour des raisons de rationalité et de clarté, de **procéder** à la codification desdites directives> en les regroupant en un texte unique [...]

- 50 Enfin, une troisième série d'exemples (sous-schéma P3) se distingue aussi des emplois précédents. Ici, nous retrouvons une proposition active prototypique, où le sujet correspond à l'agent du procès exprimé par le N prédicatif. Cette construction définit les compétences de l'agent en termes hiérarchiques (faire procéder d'office..., en lieu et place des professionnels mentionnés...). La responsabilité de l'agent est exprimée en termes de potentiel (présence fréquente du modal « pouvoir ») et peut également être dévolue à un autre agent (on note ici l'emploi d'une construction causative « faire procéder ») :

P3.1 <Le ministre peut faire procéder d'office, en lieu et place des professionnels mentionnés au premier alinéa et à leurs frais, à la réalisation de ce contrôle> [...]

P3.2 Article 8, <l'autorité compétente en matière de police peut faire procéder d'office à la suppression immédiate de cette publicité> [...]

P3.3 Elle [La Commission] peut également demander à <un État membre de la Communauté européenne de **procéder** ou de **faire procéder** à la notification d'actes ou de décisions afférents aux mêmes impôts> [...]

5. Visualisation de schémas LG

- 51 Les retours contextuels sont systématiquement nécessaires afin de cerner le fonctionnement communicatif des verbes analysés. Le fonctionnement des séquences textuelles qui rentrent dans la construction des schémas LG peut être analysé en contexte par la LTD : *Lecture Textométrique Différentielle* (Patin et al. 2016). Dans *Le Trameur*, les ensembles textuels (parties du corpus) peuvent être affichés simultanément à l'écran pour faciliter les comparaisons. La différenciation mobilise le calcul des *spécificités* : les éléments caractéristiques sélectionnés par seuillage dans chaque volet du corpus sont rendus « visibles » au fil des textes par un système de surlignage.
- 52 Ce type de visualisation est enrichi par les résultats issus de plusieurs analyses statistiques (segments répétés, poly-cooccurrences calculées sur annotations multiples : formes, lemmes, catégories morphosyntaxiques, etc.) pour garantir une prise en main « globale » de la différenciation à plusieurs niveaux d'analyse linguistique (Zimina & Fleury 2014). On peut orienter la LTD vers des passages textuels relevant d'une concentration de traits linguistiques saillants : nous utilisons les verbes « effectuer », « réaliser » et « procéder » comme points d'ancrage pour analyser les éléments caractéristiques de leurs voisinages contextuels dans les différentes parties du corpus.
- 53 La visualisation LTD sur la Figure 9 permet de contraster les contextes caractéristiques de « réaliser » dans les directives européennes et dans la législation nationale. L'appariement des contextes présentés sur la Figure 9 s'appuie sur la correspondance

entre la directive et sa transposition dans le texte législatif français. Dans la directive, on remarque que le verbe « réaliser » a plusieurs poly-cooccurents spécifiques (matérialisés en jaune au fil du texte). La séquence textuelle révélée a une fonction discursive précise dans les Préambules des directives : déclaration de la pertinence légale des dispositifs. Ces résultats reflètent les schémas LG analysés dans la section précédente (notamment le sous-schéma R2 pour le corpus D-EU) :

Figure 9. Lecture Textométrique Différentielle (LTD) de « réaliser » dans D-EU-preamble et T-FR-disposition

Directive européenne (Préambule)

(4) Dans un souci de clarté et de plus grande lisibilité, il convient d'intégrer dans la présente directive la disposition sur l'application de la directive 79/1072/CEE, qui figurait jusqu'à présent dans la directive 2006/112/CE du Conseil du 28 novembre 2006 relative au système commun de taxe sur la valeur ajoutée [4].

(5) Étant donné que les objectifs de la présente directive ne peuvent pas être réalisés de manière suffisante par les États membres et peuvent donc, en raison des dimensions de l'action, être mieux réalisés au niveau communautaire, la Communauté peut prendre des mesures, conformément au principe de subsidiarité consacré à l'article 5 du traité. Conformément au principe de proportionnalité tel qu'énoncé audit article, la présente directive n'excède pas ce qui est nécessaire pour atteindre ces objectifs.

Législation française (Dispositif)

intérieure de consommation sur les produits énergétiques [Dispositions déclarées non conformes à la Constitution par la décision du Conseil constitutionnel n° 2009-599 DC du 29 décembre 2009] ;

- « - les quotes-parts de résultat sur opérations faites en commun ;
- « - les charges financières afférentes aux immeubles d'exploitation ;
- « - les dotations aux amortissements d'exploitation ;
- « - les dotations aux provisions autres que les provisions techniques.

« VII. - Lorsque les plus-values de cession d'immobilisations ou de titres de nature à être comprises dans le chiffre d'affaires et la valeur ajoutée en application des I et III à VI sont réalisées l'année de création de l'entreprise, elles sont comprises dans le chiffre d'affaires et la valeur ajoutée retenus au titre de l'année suivante.

« Art. 1586 septies.-Le montant de la cotisation sur la valeur ajoutée des

6. Conclusion

- 54 Nous avons identifié et visualisé des différences d'emploi de trois verbes supports similaires dans deux corpus comparables (D-EU et T-FR). Ces spécificités ont été étudiées en mobilisant le concept de « schéma lexico-grammatical (LG) » correspondant à une unité de discours dont la définition formelle et fonctionnelle ne peut émerger que graduellement, lors de l'observation des contextes grammaticaux récurrents de tel ou tel élément lexical. L'observation des régularités contextuelles fait appel à la LTD qui matérialise les éléments constitutifs de schémas LG dans les différentes parties de corpus comparées au fil de la lecture. Ces éléments constitutifs (révélés par l'analyse textométrique) sont mémorisés par le gestionnaire de sélection dans *Le Trameur* afin d'être activés sur la trame textuelle à la demande de l'utilisateur (Patin *et al.* 2016)⁷. Ils sont rendus visibles sur la trame textuelle grâce à un système de surlignage.

- 55 Les différences entre schémas LG sont assez subtiles, mais les méthodes textométriques adoptées ici ont révélé un certain nombre de régularités :
- 56 a) le V « effectuer » est employé dans des schémas définissant les conditions légales dans lesquelles certaines activités techniques « sont effectuées » (majoritairement au passif) ; ces schémas se trouvent typiquement dans les Dispositifs et les Annexes du corpus D-EU ;
- 57 b) certains emplois du V « réaliser » sont caractéristiques des Préambules D-EU : comme le montre la Figure 9 (section 4.2.1), ces emplois correspondent à un schéma LG formulaïque (R2) exprimant une recommandation légale ; par contre, dans les autres sections des corpus D-EU et T-FR les schémas bâtis autour de « réaliser » correspondent à des définitions techniques (R1) ou à des injonctions techniques concernant des mesures de sécurité (R3) ;
- 58 c) le V « procéder » est sur-employé dans l'ensemble du corpus D-EU, que ce soit dans les Préambules (notamment le schéma formulaïque P1 exprimant une recommandation technique, et P2 exprimant une recommandation légale) ou dans les Dispositifs (le sous-schéma P3, exprimant une définition institutionnelle).
- 59 En somme, les régularités observées sont imputables à un principe général de la phraséologie : les collocations de chaque lexème – ses contextes grammaticaux habituels ainsi que ses co-occurrences lexicales typiques – se spécialisent, notamment dans les discours techniques et les types textuels spécifiques. De même, les lexèmes concurrents (des items partageant des traits sémantiques similaires) sont employés dans des contextes divergents, mais pas nécessairement de façon symétrique.
- 60 Dans le cadre de l'étude de la langue eurolectale des institutions européennes, le projet international *Observatoire sur l'eurolecte* envisage de mettre en regard les résultats issus des différentes langues d'étude et d'approfondir cette perspective contrastive. Il s'agira alors de faire émerger les divergences et convergences, non seulement entre des langues de familles différentes, mais aussi entre des langues issues d'une même langue matrice. Nous estimons que les méthodes textométriques adoptées ici permettront d'identifier ces différences de façon systématique, y compris dans le cas des distinctions très fines entre des formes lexicales dont le sémantisme est en principe très proche, et entre des corpus de textes aux fonctions discursives similaires.
-

BIBLIOGRAPHY

- Berber-Sardinha A. P. (1997). *Automatic identification of segments in written texts*. Thèse doctorale. English Department, University of Liverpool. Document en ligne : <http://www2.lael.pucsp.br/~tony/temp/publications/>.
- Biel L. (2014) *Lost in the Eurofog. The Textual Fit of Translated Law*. Frankfurt am Mein : Peter Lang.
- Ciostek A. (2014). « Eurolangue et sa productivité : quelques tendances », *Roczniki Humanistyczne*, vol. 62 (8), 65-77.

- Fleury S. et Zimina M. (2014). « Trameur : A Framework for Annotated Text Corpora Exploration », *Proceedings of COLING 2014 (25th International Conference on Computational Linguistics) : System Demonstrations*. Dublin, Ireland, 57-61. Document en ligne : <http://www.aclweb.org/anthology/C14-2013.pdf>.
- Gledhill C. (2000). *Collocations in Science Writing*. Tübingen : Gunter Narr.
- Gledhill C. (2011a). « A lexicogrammar approach to checking quality : Looking at one or two cases of comparative translation », in Depraetere, Ilse (éd.), *Perspectives on Translation Quality (Text, Translation, Computational Processing 9)*. Berlin : Mouton de Gruyter, 71-98.
- Gledhill C. (2011b). « The lexicogrammar approach to analysing phraseology and collocation in ESP texts », *Anglais de Spécialité* 59 : 5-23.
- Gledhill C. (2012). « The Discourse Function of Collocation in Research Article Introductions » in D. Biber et R. Reppen (éd.) *Benchmarks in language and linguistics (Volume I : Lexical Studies)*. London : Sage Publications, 23-45.
- Gledhill C. et Frath P. (2007). « Collocation, phrasème, dénomination : vers une théorie de la créativité phraséologique », *La Linguistique* 43/1 : 65-90.
- Gledhill C. et Todirascu A. (2008). « Collocations en contexte : extraction et analyse contrastive », *Texte et Corpus* 3 : 137-148.
- Goffin R. (1994). « L'Eurolecte : oui, jargon communautaire : non », *Meta*, Hommage à B. Quemada, 39/4 : 636-642.
- Goffin R. (2005). « Quels corpus et quelles approches pour une description contrastive de l'eurolecte ? » *Mots, termes et contextes. 7^{es} Journées scientifiques du réseau de chercheurs Lexicologie, Terminologie, Traduction*, 8-10 septembre 2005, Bruxelles, Belgique, cité par Thoiron Ph. dans : <http://perso.univ-lyon2.fr/~thoiron/JS%20LTT%202005/pdf/Goffin.pdf>.
- Goldberg A. (1995). *A construction grammar approach to argument structure*. Chicago : University of Chicago Press.
- Gontrand F. (1991). *Parlez-vous eurocrate ? Les 1 000 mots clés du Marché Unique*. Paris : Les Éditions d'Organisation.
- Guide pratique commun pour la rédaction des textes législatifs de l'Union européenne. (2014). Document en ligne : <http://eur-lex.europa.eu/content/techleg/FR-guide-de-redaction-legislative.pdf>.
- Halliday M.A.K. et Matthiessen C. (2014). *Introduction to Functional Grammar. 4th Edition*. London : Edward Arnold.
- Heiden S. (2006). « Un modèle de données pour la textométrie : contribution à une interopérabilité entre outils », in Viprey J.-M. et al. (éd.) *Actes des 8^{es} Journées internationales d'Analyse Statistique des Données Textuelles (JADT'06) « Archives, Bases, Corpus »*, 19-21 avril 2006, vol. 1. Besançon, Presses Universitaires de Franche-Comté, 747-487.
- Hoey M. (1991). *Patterns of lexis in text*. Oxford : Oxford University Press.
- Hunston S. et Francis G. (2000). *Pattern Grammar : a Corpus- Driven Approach to the Lexical Grammar of English*. Amsterdam/Philadelphia : John Benjamins.
- Khalifa R. (2015) « La phraséologie du discours juridique français. Étude de cas : les arrêts de la Cour de cassation » *Synergies Espagne* 8 : 49-64. Document en ligne : <http://gerflint.fr/Base/Espagne8/khalifa.pdf>.
- Lafon P. (1984). *Dépouillements et statistiques en lexicométrie*. Genève-Paris : Slatkine-Champion.

- Lebart L. et Salem A. (1994). *Statistique textuelle*. Paris : Dunod.
- Legallois D. et François J. (2006). *Autour des grammaires de constructions et de patterns*. Cahier du CRISCO 21, Université de Caen.
- Longrée D. et Mellet S. (2013). « Le motif : une unité englobante ? Étendre le champ de la phraséologie de la langue au discours », *Langages* 189 : 65-80.
- Maingueneau D. (2002). « Type de discours », in P. Charaudeau et D. Maingueneau (éd.). *Dictionnaire d'analyse du discours*. Paris : Seuil, p. 592.
- Martinez W. (2012). « Au delà de la cooccurrence binaire... Polyoccurrences et trames de cooccurrence », *Corpus* 11/1 : 191-218. Document en ligne : <http://www.aclweb.org/anthology/C14-2013.pdf>.
- Mayaffre D. (2007). « L'entrelacement lexical des textes, co-occurrences et lexicométrie », *Texte et corpus*, n° 3, août 2008, Actes des Journées de la linguistique de Corpus, 2007, Lorient, France, 91-102. Document en ligne : <http://docplayer.fr/8393891-Texte-et-corpus-n-3.html>.
- Office des publications de l'Union européenne. (2011). *Code de rédaction interinstitutionnel*. Luxembourg. Document en ligne : <http://publications.europa.eu/code/fr/fr-000500.htm>.
- Patin S., Zimina M. et Fleury S. (2016). « Lecture Textométrique Différentielle (LTD) de textes législatifs comparables de l'Union européenne », *Actes des 13^{es} Journées internationales d'Analyse statistique des Données Textuelles*, in D. Mayaffre et al. JADT 2016 - *Statistical Analysis of Textual Data*, vol. 2. Nice : Presses de FacImprimeur, 2016, 743-753. Document en ligne : <http://lexicometrica.univ-paris3.fr/jadt/jadt2016/01-ACTES/82326/82326.pdf>.
- Pawley A. et Syder F. H. (1983). « Two puzzles for linguistic theory : Native-like selection and native-like fluency » in J. C. Richards et R. W. Schmidt (éd.) *Language and communication*. New York : Longman, 191-226.
- Salem A. (1987). *Pratique des segments répétés. Essai de statistique textuelle*. Paris : Klincksieck.
- Schmid E. (1994). « Probabilistic Part-of-Speech Tagging Using Decision Tree », *Proceedings of International Conference on New Methods in Language Processing*, Manchester, UK, vol. 12, 44-49. Document en ligne : <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/tree-tagger1.pdf>.
- Sinclair J. McH. et Mauranen A. (2006). *Linear Unit Grammar : Integrating Speech and Writing*. Amsterdam : John Benjamins.
- Stefanowitsch A. et Gries S. (2003). « Collostructions : Investigating the interaction between words and constructions », *International Journal of Corpus Linguistics* 8(2): 209-43.
- Söze-Duval K. (2008.) « Pour une textométrie opérationnelle » Document en ligne : <http://www.cavi.univparis3.fr/-lexicometrica/jadt/ressources-textometriques/textes/-rti6provisoire.pdf>.
- Swales J. (1990). *Genre analysis : English in academic and research settings*. Cambridge : Cambridge University Press.
- Tutin A. et Grossmann F. (2002). « Collocations régulières et irrégulières : esquisse de typologie du phénomène collocatif », *Revue française de linguistique appliquée* 1(7): 7-25.
- Williams G. (2003a). « From meaning to words and back : Corpus linguistics and specialised lexicography », *Anglais de spécialité* 39-40 : 91-106.

Williams G. (2003b) « Les collocations et l'école contextualiste britannique » in F. Grossmann et A. Tutin (éd.) *Les collocations : analyse et traitement*, coll. Travaux de recherches en linguistique appliquée. Paris : Éditions De Werelt, 33-44.

Zimina M. et Fleury S. (2014). « Approche systémique de la résonance textuelle multilingue » *Actes des 12^{es} Journées internationales d'Analyse statistique des Données Textuelles*, in Née E. et al. (éd.) JADT 14, Paris, 717-728. Document en ligne : <http://lexicometrica.univ-paris3.fr/jadt/jadt2014/01-ACTES/59-JADT2014.pdf>.

Zimina M. et Fleury S. (2015). « Perspectives de l'architecture Trame/Cadre pour les alignements multilingues », *Nouvelles perspectives en sciences sociales : Revue internationale de systématique complexe et d'études relationnelles* 11(1), 325-353.

NOTES

1. Nous appliquons la différence entre « type » et « genre » de discours telle qu'elle est présentée dans Charaudeau & Maingueneau (2002 : 592) : le genre renverrait à un « secteur de production verbale d'une société » et le type « à un dispositif de communication particulier [...], à des modes fondamentaux de structuration qui se combinent dans les textes effectifs ».
2. Mayaffre (2012 : 92) « La cooccurrence est la co-présence ou présence simultanée de deux unités linguistiques (deux mots par exemple ou deux codes grammaticaux) au sein d'un même contexte linguistique (le paragraphe ou la phrase par exemple, ou encore une fenêtre arbitraire) ».
3. Plusieurs unités textuelles récurrentes (les *motifs*, par exemple, cf. Longrée & Mellet 2013) peuvent être utilisées pour identifier la composition d'un schéma LG. Cependant, sur le plan conceptuel, le schéma LG dépasse le cadre d'un marqueur formel que l'on parvient à identifier en corpus ; il est adossé à une fonction discursive globale que l'on peut lui associer dans un contexte communicationnel précis.
4. Ce projet dirigé par Laura Mori regroupe 11 juri-linguistes et linguistes dont Stéphane Patin. Une présentation détaillée du projet est disponible en ligne : <http://www.unint.eu/en/research/research-groups/39-higher-education/490-eurolect-observatory-interlingual-and-intralingual-analysis-of-legal-varieties-in-the-eu-setting.html>.
5. Comme nous l'avons indiqué plus haut, ces termes proviennent de la grammaire systémique fonctionnelle (Halliday & Matthiessen 2014).
6. *Le Trameur* (pour Windows et MacOSX) peut être téléchargé librement sur le site officiel <http://www.tal.univ-paris3.fr/trameur/>.
7. Les régularités identifiées dans les textes peuvent aussi alimenter des requêtes sous forme de patrons pré-écrits qui mémorisent certaines attractions lexico-grammaticales (Zimina & Fleury 2014).

ABSTRACTS

This article examines lexico-grammatical (LG) patterns associated with three support verbs in French (*effectuer*, *réaliser*, *procéder*) in two comparable corpora (a corpus of European directives

written in French, and a corpus of texts transposed into French national law). To this end, we use *Le Trameur*, a textometric tool designed for the systematic comparison of text data (and which measures: relative frequency in each corpus part, characteristic elements, repeated segments, co-occurrence, poly-cooccurrence, etc.). Quantitative methods such as these encourage the linguist to adopt a precise research protocol, as well as to revise his / her approach to the basic units of textual analysis during the analytical process. Finally, it is possible to display lexico-grammatical and semantic distinctions between lexemes, which would otherwise be unobservable out of context, thus allowing the researcher to identify constructions considered as characteristic features of 'Eurolectal' as opposed to 'national' technocratic French.

Nous examinons ici les schémas lexico-grammaticaux (LG) associés à trois verbes supports (« effectuer », « réaliser », « procéder ») dans deux corpus juridiques comparables (un corpus de directives européennes rédigées en français et leur transposition en France dans les textes de lois). A cette fin, nous utilisons *Le Trameur*, un logiciel de textométrie qui permet une comparaison systématique des données textuelles en mobilisant plusieurs méthodes statistiques exploratoires : analyse des fréquences par partie, spécificités, segments répétés, co-occurrences, poly-cooccurrences, etc. Le recours à ce type d'outils permet au linguiste d'élaborer des unités de base de son système analytique en suivant un protocole de recherche précis, qui tient compte des indices quantitatifs. Finalement, des distinctions lexico-grammaticales fines entre des lexèmes dont le sémantisme serait trop proche pour une comparaison hors contexte, sont visualisées de façon synthétique au fil du texte, tout en permettant de différencier certaines constructions spécifiques du français « eurolectal » et les éléments du discours technocratique « national ».

INDEX

Keywords: comparable corpora, language for specific purposes, lexico-grammatical patterns, support verb constructions, textometrics

Mots-clés: corpus comparables, constructions à verbes support, langues de spécialité, schémas lexico-grammaticaux, textométrie

AUTHORS

CHRISTOPHER GLEDHILL

CLILLAC-ARP EA 3967

Université Paris Diderot, Sorbonne Paris Cité

STÉPHANE PATIN

CLILLAC-ARP EA 3967

Université Paris Diderot, Sorbonne Paris Cité

MARIA ZIMINA

CLILLAC-ARP EA 3967

Université Paris Diderot, Sorbonne Paris Cité