

Faculté de philosophie, arts et lettres

Étude du vocabulaire de la presse généraliste dans les domaines des relations internationales et des conflits armés

Étude sur corpus

Auteur : Van Roey Camille
Promoteur(s) : Geneviève Maubille, Svetlana Aloushkova-Vogeleer
Lecteur(s) : Marina Riapolova
Année académique 2019-2020
Master en traduction à finalité spécialisée : terminologie et localisation

ÉTUDE SUR CORPUS DU VOCABULAIRE DE LA PRESSE GÉNÉRALISTE DANS LE DOMAINE DES RELATIONS INTERNATIONALES ET DES CONFLITS ARMÉS

Camille VAN ROEY
Année académique 2019-2020
Supervision : Geneviève MAUBILLE &
Svetlana ALOUSHKOVA-VOGELEER

Mémoire présenté en vue
de l'obtention du diplôme de
Master en Traduction

ÉTUDE SUR CORPUS DU VOCABULAIRE DE LA PRESSE GÉNÉRALISTE DANS LE DOMAINE DES RELATIONS INTERNATIONALES ET DES CONFLITS ARMÉS

Camille VAN ROEY
Année académique 2019-2020
Supervision : Geneviève MAUBILLE &
Svetlana ALOUSHKOVA-VOGELEER

Mémoire présenté en vue
de l'obtention du diplôme de
Master en Traduction

Nous souhaitons, en tout premier lieu, adresser nos plus sincères remerciements à notre promotrice, Madame Geneviève Maubille, qui nous a enseigné persévérance et rigueur. Ce travail n'aurait pu voir le jour sans ses nombreux conseils et le temps qu'elle nous a consacrés.

Nous tenions ensuite à remercier notre co-promotrice, Madame Svetlana Aloushkova-Vogeleer, dont les précieux conseils sur la langue et la culture russes nous ont été indispensables.

Enfin, nous remercions nos fidèles et précieuses amies de Marie-Haps – Liouba, Luna et Camille – dont le soutien sans faille a été immensément apprécié.

*Pour chaque langue que l'on parle, on vit une nouvelle vie.
Celui qui ne connaît qu'une seule langue ne vit qu'une seule fois.*

Proverbe tchèque

Table des matières

Table des matières	8
Préface	10
Introduction	11
1 Objectifs et destinataires du travail	12
2 Aspects théoriques	14
2.1 Définition	14
2.2 Historique	15
2.2.1 Les premiers corpus (un million de mots)	15
2.2.2 Les corpus de plus de 100 millions de mots	15
2.2.3 Les « méga-corpus » de plusieurs milliards de mots	16
2.3 Les différents types de corpus	17
2.3.1 Corpus de référence – corpus spécialisé	17
2.3.2 Corpus écrit – corpus oral	17
2.3.3 Corpus synchronique – corpus diachronique	17
2.3.4 Corpus fermé – corpus ouvert	18
2.3.5 Corpus d'apprenants	18
2.3.6 Corpus « officiel » – corpus « maison » (DIY corpus)	18
2.3.7 Corpus compilé manuellement – corpus compilé automatiquement	18
2.3.8 Corpus brut – corpus étiqueté	19
2.3.9 Corpus monolingue	19
2.3.10 Les corpus bilingues	20
2.3.11 Présentation du corpus de travail	21
2.4 L'utilisation des corpus	22
2.5 L'utilisation des corpus dans l'apprentissage du vocabulaire en langue étrangère	22
2.5.1 Approche directe – approche indirecte	23
3 Constitution des corpus	33
3.1 Critères de constitution	33
3.1.1 Taille du corpus	33
3.1.2 Extraits – textes entiers	33
3.1.3 Nombre de textes	33
3.1.4 Format	34
3.1.5 Sujet	34
3.1.6 Type de texte	34
3.1.7 Auteur	34
3.1.8 Langue	34
3.1.9 Date de publication	35
3.2 Présentation des journaux sélectionnés	35
3.2.1 Le Monde	35
3.2.2 Le Figaro	36
3.2.3 Le Vif/L'Express	36
3.2.4 Газета	36
3.2.5 Известия	37
3.2.6 Независимая	37
3.2.7 Новая Газета	37
3.2.8 Ведомости	38

3.2.9	Poids des journaux au sein du corpus	38
3.2.10	Évènements survenus pendant la période de compilation.....	41
3.3	La compilation du corpus	45
3.3.1	Procédé de compilation	45
3.3.2	Difficultés rencontrées	46
4	Sketch Engine	48
4.1	Présentation.....	48
4.1.1	Corpus disponibles sur le Sketch Engine	48
4.1.2	Utilisateurs et utilisations du <i>Sketch Engine</i>	49
4.2	Création et compilation de corpus dans le Sketch Engine	50
4.2.1	Création du corpus	50
4.2.2	Ajout de fichiers	51
4.2.3	Compilation du corpus	52
4.3	Interrogation de corpus dans le Sketch Engine	55
4.3.1	Outils d'extraction de la terminologie	55
4.3.2	Outils d'extraction de la combinatoire	59
5	Analyse des corpus.....	72
5.1	Extraction de la terminologie.....	72
5.1.1	Unités monolexicales.....	72
5.1.2	Unités polylexicales	82
5.2	Mise en regard des listes RU et FR	87
5.3	Comparaison et analyse de fréquence des termes simples.....	89
5.3.1	Déséquilibres.....	90
5.4	Cas particulier russe : mots-valises	94
5.5	Proportion des termes simples et complexes au sein du lexique.....	95
5.6	Erreurs et défauts du Sketch Engine	98
5.7	Extraction de la combinatoire.....	101
6	Élaboration des lexiques bilingues.....	104
6.1	Lexique débutant – lexique avancé.....	104
6.2	Présentation des lexiques.....	106
6.2.1	Présentation des collocations	107
6.3	Jeu de couleurs	110
6.4	Notes et remarques	111
	Conclusion	114
	Bibliographie	117

Préface

Ce mémoire consiste en l'élaboration de deux lexiques bilingues russe-français constitués à l'aide de corpus. Ces lexiques, débutant et avancé, s'adressent respectivement aux étudiants de bachelier et de master en traduction et interprétation.

Les thématiques choisies portent sur la politique, les relations internationales et les conflits armés. Au travers de ces lexiques, l'apprenant rencontrera du vocabulaire traitant de la vie politique, de l'armée, de la guerre, de l'espionnage ou encore de la diplomatie. Le choix des thématiques n'a pas été laissé au hasard : nous considérons qu'il s'agit de domaines de connaissance essentiels pour de futurs traducteurs et interprètes.

Ce travail comporte une partie théorique, ainsi qu'une partie pratique. Dans la partie théorique, nous expliquerons ce qu'est un corpus, présenterons les différents types de corpus et exposerons les caractéristiques et les critères de constitution des corpus compilés dans le cadre de ce travail. Dans la partie pratique, nous présenterons le *Sketch Engine*, outil utilisé pour interroger les corpus, et expliquerons le fonctionnement des différentes fonctionnalités utilisées tout au long de l'élaboration de ce travail. Dans cette partie, nous détaillerons également comment nous avons élaboré les lexiques (extraction de la terminologie et extraction de la combinatoire) et présenterons le produit final.

Bonne lecture !

Introduction

Ce mémoire consiste en l'élaboration de deux lexiques bilingues russe-français traitant des domaines des relations internationales et des conflits armés. Nous avons mis au point un lexique débutant et un lexique avancé, respectivement destinés aux étudiants de bachelier et de master de traduction et d'interprétation. Ces lexiques ont été constitués sur la base de deux corpus – l'un russe, l'autre français – compilés par nos soins et proposent une liste de termes, assortis pour la plupart de leur combinatoire, en russe ainsi que leurs équivalents en français.

Ces lexiques offriront donc des bases solides aux futurs interprètes et traducteurs tant sur le plan terminologique que sur le plan phraséologique. Il va sans dire que ces lexiques pourront être complétés par l'étudiant lui-même au cours de son parcours universitaire.

Dans un premier temps, nous expliquerons brièvement ce qu'est un corpus et nous nous pencherons sur les différentes façons dont les corpus peuvent être exploités.

Nous avons également dédié un chapitre sur l'utilisation des corpus dans l'apprentissage d'une langue étrangère. Dans ce chapitre, nous avancerons les avantages de cette approche et proposerons différents exercices à réaliser en salle de classe afin de tirer profit de toutes les facettes qu'offrent les corpus et le *Sketch Engine*.

Ensuite, nous présenterons le *Sketch Engine*, outil utilisé pour exploiter et interroger les corpus, puis nous aborderons les différentes étapes du processus d'élaboration des lexiques : choix des journaux, extraction des articles, compilation des corpus et fonctionnalités utilisées pour extraire la terminologie et la combinatoire. Enfin, nous dresserons le constat de différentes observations remarquées lors de l'élaboration de ce travail et présenterons les lexiques.

1 Objectifs et destinataires du travail

Ce travail a pour objectif de proposer un lexique, à la fois utile et fonctionnel, aux étudiants de traduction et d'interprétation, dont l'une des langues d'apprentissage est le russe. Nous joignons à ce travail deux lexiques (débutant et avancé) : le premier s'adresse aux étudiants de bachelier, tandis que le second s'adresse plutôt aux étudiants de master.

Ce travail comporte deux dimensions. D'une part, nous proposons une liste de termes bilingues et, d'autre part, nous avons assorti la plupart de ces termes de leur combinatoire. Les lexiques sont constitués d'une liste de termes relatifs à la politique, aux relations internationales et aux conflits armés et sont classés en différents sous-domaines.

« [La combinatoire] est au cœur de toute production linguistique : sa maîtrise dénote une richesse d'expression, et, de ce fait, elle constitue un contenu d'apprentissage essentiel pour la maîtrise de la langue » (Tremblay & Ancil, 2016 : 5).

Nous tenions à ce que la combinatoire fasse partie intégrante de ce travail. En effet, que ce soit dans le cadre de l'apprentissage d'une langue étrangère ou de l'enrichissement de sa langue maternelle, nombreux sont les experts de la langue à mettre l'accent sur l'importance de la maîtrise de la combinatoire. Citons par exemple Di Vito (2013) qui déclare que *« [l']enseignement des collocations consiste non seulement à fournir aux étudiants un échantillon de possibilités lexicales pour accroître leur vocabulaire, mais surtout à leur faire découvrir les associations qui existent entre les mots »* ou James Carl (1998 : 152) qui souligne que *« adherence to the collocational conventions of a foreign language contributes greatly to one's idiomaticity and nativelikeness »*.

Idéalement, ces lexiques accompagneraient l'apprenant au fil de ses études et seraient alimentés du nouveau vocabulaire rencontré tout au long de son parcours universitaire. Certaines informations pourront également être sujettes à une actualisation, nous pensons par exemple aux dirigeants de pays et aux capitales.

Ces informations devront donc faire l'objet de mises à jour au fil des années par l'enseignant ou par l'étudiant.

Idéalement, les corpus que nous avons constitués et qui nous ont permis d'élaborer les lexiques devraient être mis à la disposition de l'enseignant et des étudiants en vue d'une exploitation en salle de classe. Cela n'est envisageable que si les étudiants sont formés aux principales fonctionnalités des outils d'interrogation de corpus, tels que le *Sketch Engine*.

2 Aspects théoriques

Dans la section qui suit, nous définirons la notion de corpus, exposerons les types de corpus et aborderons les différentes utilisations possibles des corpus.

2.1 Définition

Il existe différentes définitions de la notion de « corpus ». Du côté des dictionnaires de langue générale, le *Larousse* le définit comme un « *[e]nsemble fini d'énoncés écrits ou enregistrés, constitué en vue de leur analyse linguistique* » ; le *Cambridge Dictionary* comme « *a collection of **written or spoken** material stored on a **computer** and used to find out **how language is used*** » et le *Trésor de la Langue Française* comme un « *[e]nsemble de textes établi selon un principe de documentation exhaustive, un **critère thématique ou exemplaire** en vue de leur étude linguistique* ».

Ces trois définitions contiennent toutes la dimension de l'analyse linguistique. Le *Larousse* et le *Cambridge Dictionary* apportent tous deux une précision sur le format, qui peut être écrit ou oral, mais seul le *Cambridge Dictionary* précise que les corpus se présentent sous format numérique. Le *Trésor de la Langue Française* ajoute, quant à lui, la notion de critère de sélection des textes.

Du côté des ouvrages spécialisés, Bowker et Pearson (2002) nous fournissent la définition suivante : « *A corpus can be described as a large collection of **authentic** texts that have been gathered in electronic form according to a **specific set of criteria*** » (2002 : 9). Sager définit un corpus comme « *a **representative** body of texts of a **subject field** (1990 : 130) assembled [...] according to previously **established criteria** of representativeness, completeness and relevance (1990 : 155)* ».

Contrairement aux définitions des dictionnaires généraux, ces définitions mettent davantage l'accent sur les notions de représentativité des textes et de sélection de critères de constitution spécifiques.

Si nous regroupons les différents éléments de ces définitions, nous pouvons proposer la définition suivante : « *Un corpus est un ensemble de textes*

authentiques, oraux ou écrits, représentatifs d'une langue et regroupés sous format numérique selon une série de critères spécifiques en vue d'une étude linguistique ».

2.2 Historique

Dans ce chapitre, nous retracerons brièvement la création et l'évolution des premiers corpus jusqu'à ceux que nous utilisons aujourd'hui.

2.2.1 Les premiers corpus (un million de mots)

Le *Brown Corpus* est souvent considéré comme le premier corpus. Compilé entre 1963 et 1964 à l'Université de Brown aux États-Unis, le *Brown Corpus* est un corpus de référence représentatif de l'anglais américain. Il est constitué d'échantillons de 2 000 mots issus de 500 textes différents publiés en 1961 et relevant de 15 genres textuels (Loock, 2016 et Varieng, 2014).

Compilé entre 1970 et 1978, le *LOB* a été compilé par les universités de Lancaster, Oslo et Bergen selon les mêmes critères que le *Brown*, à ceci près qu'il porte sur l'anglais britannique (Loock, 2016 et Varieng, 2014).

Alors que le *Brown Corpus* et le *LOB* sont composés de textes écrits, le *London-Lund Corpus*, compilé entre 1959 et 1990 par l'University College de Londres et l'Université de Lund, est un corpus d'anglais britannique parlé. Il est composé d'échantillons de 5 000 mots issus de 100 textes oraux, pour un total de 500 000 mots (Chateau-Smith, 2013 et Varieng, 2014).

Comme le souligne Loock (2016 : 29), « *Ces corpus de première génération ont permis la réalisation de nombreux travaux et ont posé les jalons méthodologiques de la compilation de corpus pour les décennies qui ont suivi.* »

2.2.2 Les corpus de plus de 100 millions de mots

Parmi ces corpus, citons le *British National Corpus (BNC)*, qui, avec ses 100 millions de mots, constitue, encore aujourd'hui, le corpus de référence pour l'anglais britannique. Compilé entre 1991 et 1994, il est composé à 90 % de textes écrits et à 10 % de textes oraux, issus les uns et les autres de la seconde moitié du

XX^e siècle. Il regroupe 4 049 textes et échantillons ne dépassant pas 45 000 mots et appartenant à différents genres textuels (fiction, articles de presse, ouvrages académiques...) (Loock, 2016 et Varieng, 2015).

Nous retrouvons également le *COCA* (*Corpus of Contemporary American English*), compilé depuis 1990 par la Brigham Young University, et qui représente le corpus de référence actuel de l'anglais américain (Varieng, 2015). Disponible en ligne depuis 2008, il contient environ un milliard de mots (English Corpora, 2019) et est alimenté en moyenne de 20 millions de mots chaque année (Varieng, 2015).

En France, nous retrouvons *Frantext*, un corpus de 285 millions de mots compilé par l'ATILF (Analyse et Traitement Informatique de la Langue Française). Disponible en ligne depuis 1998, ce corpus comprend principalement des textes philosophiques et littéraires datant du IX^e au XXI^e siècle et est composé à 10 % de textes scientifiques et techniques (Loock, 2016).

2.2.3 Les « méga-corpus » de plusieurs milliards de mots

Les « méga-corpus » se distinguent des précédents corpus par leur taille et leur méthode de compilation et se présentent comme des corpus de référence (Loock, 2016 : 72).

Contrairement aux précédents corpus, où la compilation est effectuée manuellement et selon des critères précis, les « méga-corpus » sont compilés de manière automatique par rapatriement de pages web. Loock souligne qu'il est essentiel d'exploiter ces corpus avec précaution compte tenu du manque d'informations et de contrôle sur les textes constitutifs du corpus (auteur, authenticité et qualité des textes, types de textes, degré de spécialité, etc.). Il ne convient donc pas de placer ces « méga-corpus » au même rang que le *COCA* ou le *BNC* qui font « *l'objet d'un contrôle éditorial strict* » (Loock, 2016 : 31).

La seconde distinction porte sur la taille des « méga-corpus » qui peuvent accueillir plusieurs milliards de mots. Citons par exemple le *Google Books Corpus*, constitué de quelque 500 milliards de mots, ou encore le *Wikipedia Corpus*, constitué de 4,4 millions d'articles issus de l'encyclopédie en ligne Wikipédia, pour un total de 2 milliards de mots (Loock, 2016 : 31).

2.3 Les différents types de corpus

Dans la catégorisation qui suit, nous présentons succinctement les différents types de corpus.

2.3.1 Corpus de référence – corpus spécialisé

Représentatif d'une langue dans son ensemble, le **corpus de référence** est utilisé pour observer le comportement de cette même langue. Pour garantir cette représentativité, le corpus de référence est composé de documents écrits et oraux relevant de genres textuels différents.

À l'inverse, un **corpus spécialisé** est consacré à un aspect particulier d'une langue : genre textuel précis ; domaine de spécialité ; groupe de locuteurs, etc. Un corpus spécialisé ne permet donc pas de faire des observations générales sur une langue (Bowker et Pearson, 2002 : 11-12).

2.3.2 Corpus écrit – corpus oral

Un **corpus écrit** contient uniquement des textes provenant de sources écrites, tandis qu'un **corpus oral** est constitué de retranscriptions de documents oraux (conversations, discours, allocutions, conférences, journal parlé...). À noter qu'un corpus peut à la fois réunir des documents écrits et des documents oraux (Bowker et Pearson, 2002 : 12).

2.3.3 Corpus synchronique – corpus diachronique

Un **corpus synchronique** contient des documents sélectionnés sur une période de temps déterminée. À l'inverse, un **corpus diachronique** rassemble des textes issus de différentes périodes et permet donc d'analyser l'évolution d'une langue dans le temps (Bowker et Pearson, 2002 : 12).

2.3.4 Corpus fermé – corpus ouvert

Une fois compilé, un **corpus fermé** n'est plus alimenté. Un **corpus ouvert**, aussi appelé corpus de suivi, peut, quant à lui, accueillir de nouveaux textes (Bowker et Pearson, 2002 : 12-13).

2.3.5 Corpus d'apprenants

Un **corpus d'apprenants** est constitué de textes produits par des apprenants d'une langue étrangère ou de traductions réalisées par des apprenants en traduction. En général, ce type de corpus est utilisé par les enseignants pour repérer les erreurs typiques commises par les apprenants en vue d'améliorer l'apprentissage (Loock, 2016 : 91).

2.3.6 Corpus « officiel » – corpus « maison » (DIY corpus)

Un **corpus officiel** est un corpus compilé dans le cadre, le plus souvent, d'un projet de recherche et mis à la disposition, parfois gratuitement, du grand public. Un **corpus « maison »** est compilé généralement par un seul individu pour répondre à un besoin spécifique ; c'est le cas par exemple d'un traducteur qui compilerait un corpus dans le cadre d'un projet de traduction (Loock, 2016 : 78).

2.3.7 Corpus compilé manuellement – corpus compilé automatiquement

La compilation d'un corpus peut être effectuée manuellement ou automatiquement. Manuellement, chaque texte est sélectionné selon des critères précis par l'auteur du corpus, ce qui octroie davantage de contrôle sur son contenu. Aujourd'hui, des logiciels permettent de compiler rapidement et automatiquement du contenu disponible sur le Web. Comme le précise Loock, la compilation automatique présente des avantages et des inconvénients :

« Les avantages sont donc la rapidité, mais aussi la possibilité de récupérer un corpus de taille bien plus importante que dans le cadre d'une compilation manuelle, ce qui in fine permet de diminuer la prise de risques du fait d'un nombre accru de sources. Du côté des inconvénients, il est indéniable que le contrôle des

textes inclus dans le corpus est moindre que pour une compilation manuelle même si des vérifications sont toujours possibles »

(2016 : 82).

2.3.8 Corpus brut – corpus étiqueté

Un **corpus étiqueté** est un corpus dans lequel à chaque mot est associée une catégorie grammaticale (nom, verbe, préposition, etc.). Cette opération, appelée étiquetage, est réalisée le plus souvent automatiquement à l'aide d'un logiciel spécifique appelé étiqueteur (L'Homme, 2004). Un **corpus brut** est un corpus contenant uniquement du texte sans aucune information linguistique.

2.3.9 Corpus monolingue

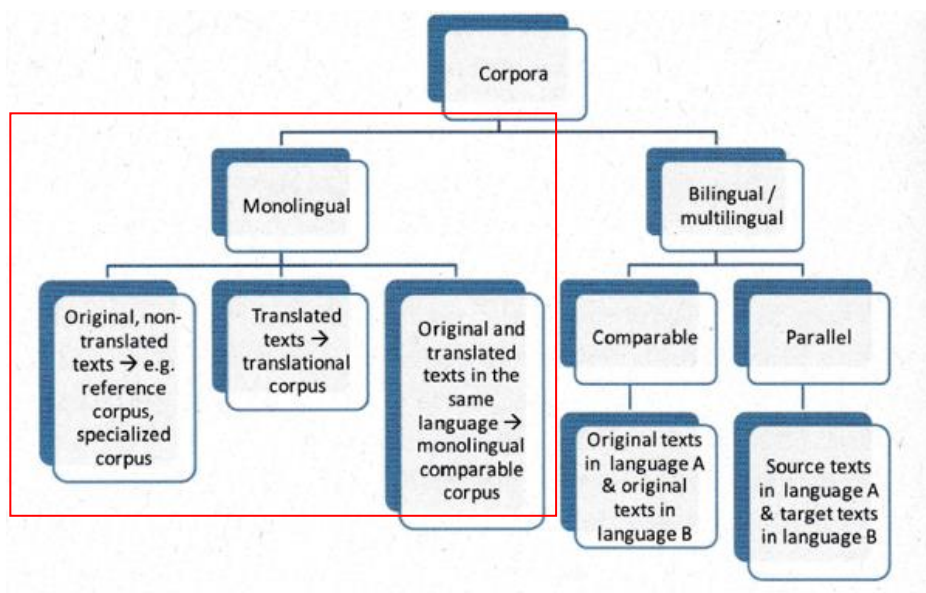


Figure 1. Types of corpora (Lefer, 2015)

Un corpus peut être monolingue, bilingue ou multilingue. Parmi les corpus monolingues, nous retrouvons les corpus constitués de textes originaux, les corpus constitués de traductions et les corpus monolingues comparables (Figure 1).

Dans les **corpus monolingues constitués de textes originaux**, nous retrouvons les corpus de référence et les corpus spécialisés qui ont respectivement pour objectif de représenter l'utilisation d'une langue et de représenter un domaine de

spécialité. Ces corpus comprennent donc uniquement des textes originaux rédigés par des natifs de la langue (Loock, 2016).

Dans la seconde sous-catégorie, nous retrouvons les **corpus monolingues de langue traduite** (*translational corpus*). Ces corpus sont constitués de traductions en langue cible et sont fréquemment utilisés en traductologie dans le but d'analyser la langue traduite (Loock, 2016 : 74-75).

Enfin, nous retrouvons dans la troisième sous-catégorie les corpus monolingues constitués de deux sous-corpus : les **corpus monolingues comparables** (*monolingual comparable corpus*). Le premier sous-corpus est constitué de textes originaux d'une langue A, et le second est constitué de textes traduits de cette même langue A. Ces corpus permettent d'analyser et de comparer la langue originale et la langue traduite (Loock, 2016).

2.3.10 Les corpus bilingues

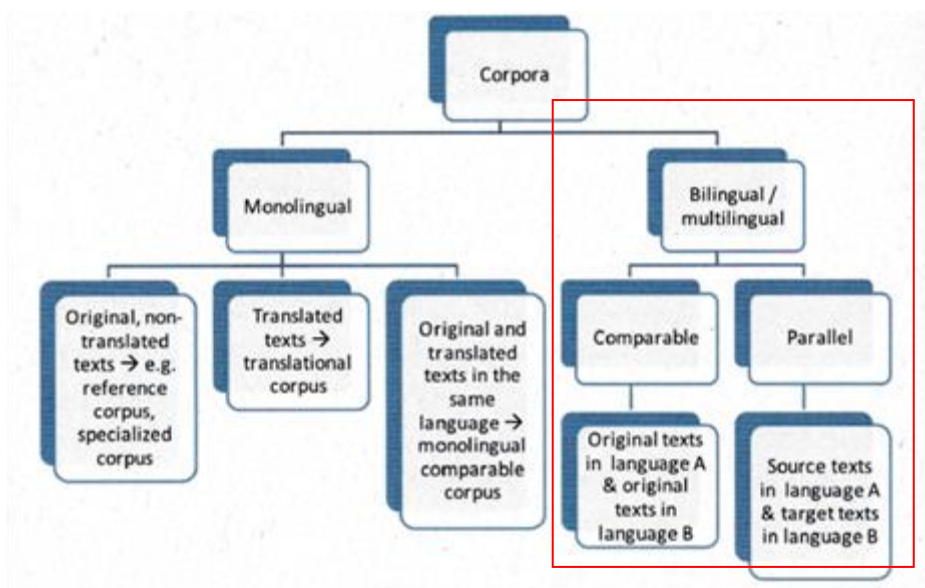


Figure 2. Types of corpora (Lefer, 2015)

Les corpus bilingues se composent de deux sous-corpus : l'un dans une langue A, et le second dans une langue B. Les corpus bilingues peuvent être parallèles ou comparables (Figure 2).

Les **corpus parallèles** sont composés de deux sous-corpus reprenant des textes en langue source ainsi que leur traduction en langue cible (Loock, 2016 : 76). Le site [Linguee](#) ou les mémoires de traduction sont des exemples de corpus parallèles.

Les corpus parallèles peuvent également être monodirectionnels ou bidirectionnels. Les corpus parallèles monodirectionnels contiennent une langue source et sa traduction (EN > FR), tandis que les corpus bidirectionnels contiennent les deux sens de traduction (FR > EN et EN >FR) (Lefer, à paraître).

Les **corpus comparables** sont, quant à eux, constitués de textes originaux en langue A et de textes originaux en langue B. Ces textes sont sélectionnés selon des critères identiques (registre, type de texte, domaine, sujet, public cible, date de publication, etc.) (Lefer, à paraître). Les corpus comparables peuvent pallier l'absence de corpus parallèles dans certaines combinaisons linguistiques et représentent plus fidèlement les caractéristiques linguistiques et culturelles de chaque langue (Goeuriot, 2009 : 13).

2.3.11 Présentation du corpus de travail

Les corpus compilés dans le cadre de ce travail sont :

- multilingues et comparables : les corpus sont constitués de textes originaux en langue russe et de textes originaux en langue française, sélectionnés selon les mêmes critères de constitution.
- spécialisés : les corpus ont été constitués selon des critères spécifiques (thématique, domaine, genre textuel, période). Toutefois, le contenu en lui-même des textes n'est pas spécialisé étant donné qu'il s'agit d'articles issus de la presse généraliste.
- écrits : seuls les articles écrits ont été compilés, aucune retranscription de discours n'a été sélectionnée.
- synchroniques : les corpus ont été compilés sur une période déterminée : du 1^{er} septembre au 31 octobre pour le corpus français et du 1^{er} septembre au 2 décembre pour le corpus russe.

- fermés : nous proposons les corpus en tant que corpus fermés. Toutefois, il pourrait être envisagé que les corpus soient alimentés ultérieurement par un mémorant ou par un apprenant selon les mêmes critères de constitution.
- étiquetés : les corpus ont été étiquetés lors de la compilation par le *Sketch Engine*. Le corpus francophone a été étiqueté par l'étiqueteur [Freeling](#), et le corpus russe par l'étiqueteur [Multext-East V4](#).
- maison : les corpus ont été compilés par nos soins afin de répondre à un besoin précis, en l'occurrence, l'élaboration de lexiques russe-français traitant de la politique, des relations internationales et des conflits armés.
- compilés manuellement : chaque article a été sélectionné manuellement et ce, selon les critères définis.

2.4 L'utilisation des corpus

Les corpus sont utilisés dans différents domaines : lexicographie, traductologie et interprétologie, terminologie, traduction et interprétation, linguistique, sociolinguistique, etc. Comme ce travail porte sur l'utilisation des corpus dans l'apprentissage du vocabulaire en langue étrangère, nous proposons de nous focaliser uniquement sur cette utilisation des corpus dans la section qui suit.

2.5 L'utilisation des corpus dans l'apprentissage du vocabulaire en langue étrangère

Bien que les corpus soient depuis longtemps utilisés par les linguistes et les experts de la langue, ils ont également beaucoup à offrir aux apprenants de langues étrangères. Comme le souligne Chambers (2010) :

« L'intégration des corpus dans l'enseignement des langues vise non seulement l'utilisation de ressources linguistiques à la fois authentiques et pertinentes, mais passe aussi par la mise au point d'une nouvelle approche pédagogique. »

Dans ce chapitre, nous expliquerons en quoi consistent l'approche directe et l'approche indirecte, proposerons des idées d'exercices, en approche directe et

indirecte, à réaliser en salle de classe et présenterons les avantages qu'offre l'utilisation de corpus dans l'apprentissage d'une langue étrangère.

Les exercices proposés dans ce chapitre peuvent être réalisés dans le *Sketch Engine* qui propose des corpus « prêts à l'emploi », les *SkELL* (*Sketch Engine for Language Learning*). L'objectif de ces corpus est de fournir du contenu utile et facile à exploiter dans le cadre de l'apprentissage d'une langue et sont pour le moment disponibles en anglais, allemand, italien, russe, tchèque et estonien. Le *Sketch Engine* propose, par ailleurs, différents types de corpus disponibles en plus de 90 langues que nous citons dans de plus amples détails au point [4.1.1 Corpus disponibles sur le Sketch Engine](#). Enfin, l'enseignant peut compiler ses propres corpus en fonction des besoins du cours, ou utiliser les corpus que nous avons compilés dans le cadre de ce travail.

2.5.1 Approche directe – approche indirecte

Chambers (2010) distingue deux manières d'utiliser les corpus en salle de classe : l'approche directe, également appelée *Data-Driven-Learning* par Johns (1997 : 101), et l'approche indirecte.

2.5.1.1 Approche directe

Dans le cas de l'**approche directe**, l'étudiant consulte et interroge lui-même des corpus à l'aide d'un concordancier, ce qui lui octroie un rôle actif et autonome dans son apprentissage (Chambers, 2010). Corino et Onesti (2019) soulignent que cette approche active « *leads to better metalinguistic competencies [...] and activate[s] HOTS (higher order thinking skills) processes that will probably result into a longer-lasting knowledge and better language skills* ».

Dans le cadre de l'approche directe, il revient donc à l'étudiant d'interroger les corpus, mais à l'enseignant de guider l'apprenant et de lui soumettre des exercices à réaliser. Gavioli (2001 : 108) souligne d'ailleurs que les activités doivent impérativement être guidées par l'enseignant, en l'absence de quoi l'étudiant risque de se sentir désorienté ou découragé.

Ci-dessous, nous présentons un exercice en approche directe proposé par Di Vito (2013).

2.5.1.1.1 Propositions d'exercices en approche directe

Di Vito (2013) propose un exercice ¹ ciblé pour systématiser l'emploi de « beaucoup », « beaucoup de » et « très ». Boulton et Landure (2016) soulignent d'ailleurs qu'il est important de se concentrer sur un nombre restreint d'erreurs considérées comme « graves » et de les corriger de manière active afin d'enclencher les processus cognitifs de mémorisation.

Cet exercice peut être adapté aux difficultés propres à différentes langues. Pour l'anglais, l'emploi de « *much* » et « *many* » ou de « *fewer* » et « *less* » par exemple. En russe, nous pourrions comparer l'emploi de « *мал* » et « *мало* » ou encore de « *очень* » et « *много* ».

L'exercice élaboré par Di Vito consiste à observer les occurrences des mots « beaucoup » et « très » à l'aide de la série de questions suivante :

- Quels sont les mots qui constituent le contexte gauche et droit de « très » et « beaucoup » ?
- Quelle est la catégorie syntaxique de ces mots ?
- Y a-t-il des constructions récurrentes ?
- À quelle catégorie grammaticale correspond le co-texte à gauche de « beaucoup » ?
- Les mots qui suivent la construction « beaucoup de » sont-ils au pluriel, au singulier... ?

Les étudiants doivent ensuite trier les co-occurents en deux colonnes, l'une pour « beaucoup » et l'autre pour « très », afin d'identifier les colligations et de découvrir les règles d'emploi.

¹ Cet exercice est destiné à des étudiants italiens étudiant la langue française.

La structure de cet exercice peut être adaptée à différentes langues et permet à l'apprenant de tirer des conclusions et d'assimiler activement les structures grammaticales.

En langue russe, cet exercice peut être adapté pour analyser et comparer différentes règles d'emploi, comme celles de « *так* » et « *такой* » qui représentent fréquemment une difficulté pour les apprenants de cette langue. À première vue, nous constatons (Figure 3) que « *так* » est suivi d'adverbes (*так легко, так много*), de verbes (*так кричать, так мыслить*), de constructions sujet-verbe (*так Макрон говорил, так мы и проведем выборы*), tandis que « *такой* » (Figure 4) est suivi de substantifs (*такое вообще государство, такой вопрос, такое впечатление, такое время*).

Di Vito (2013) a constaté que l'étude des concordances peut « *aider les étudiants à améliorer leur compétence en expression écrite : ils apprennent non seulement à enrichir leur lexique mais surtout à tenir compte des associations, des contraintes d'emploi et du critère de fréquence entre des mots ou expressions presque équivalents en français, et à éviter des calques* ».

1351	☐ doc#0 иция американцев -- например, непонятно почему США	так	легко	уступают Башару Асаду нефтяные месторождения ,
1352	☐ doc#0 носить неуверенность>>, -- жалуются Гладких. </s><s> Он	так	легко	позволяет себе критиковать работу коллег, потому
1353	☐ doc#0 признали свою ошибку и теперь покинут регион. </s><s>	Так	ли это, разбиралась <<Газета.Ru> >. </s><s> Секретарь Вь	
1354	☐ doc#0 рал, а Трамп не умеет говорить правду. <<Вот почему он	так	лихорадочно говорит наглую ложь, рассказывает о теори	
1355	☐ doc#0 гволав обеспечению мира и безопасности в Ливане>>, --	так	Макрон говорил	Шираке в телеобращении к нации по
1356	☐ doc#0 гволав обеспечению мира и безопасности в Ливане>>, --	так	Макрон говорил о	Шираке в телеобращении к нации по
1357	☐ doc#0 эссийские самолеты, а не ваши [европейские]>>. </s><s>	Так	Махатхир прокомментировал намерение Евросоюза отк	
1358	☐ doc#0 епились и оборонные связи России и Кубы -- прошло не	так	много	времени и с визита на остров секретаря российски
1359	☐ doc#0 эффектов после отравления <<Новичком> >, <<Есть еще	так	много	вопросов без ответа, и именно поэтому он попрос
1360	☐ doc#0 у. </s><s> У нас, к сожалению, сейчас в мире осталось не	так	много	политиков -- мировых лидеров, и эти решения, ко
1361	☐ doc#0 ид войск. </s><s> Участников протеста, впрочем, было не	так	много	-- всего несколько сотен человек. </s><s> На следу
1362	☐ doc#0 ет: <<Отлично. </s><s> Ни одна другая страна не сделала	так	много	для Украины]], как мы. </s><s> Но знаете что? </s>
1363	☐ doc#0 три звонка от Зеленского, и не думаю, что он ответил на	так	много	звонков от Петра Порошенко>>, -- отметил Тейлор
1364	☐ doc#0 три звонка от Зеленского, и не думаю, что он ответил на	так	много	звонков от Петра Порошенко>>, -- отметил Тейлор
1365	☐ doc#0 >><s> Однако на реализацию своих планов у Макрона не	так	много	времени -- в 2022 году во Франции должны пройти
1366	☐ doc#0 о объективен и санкции на самом деле себя изжили, их	так	много,	что все уже забыли, в связи с чем и какими событ
1367	☐ doc#0 ку сторонникам Brexit, но времени для этого осталось не	так	много.	</s><s> Полномочия Юнкера на посту главы Евро
1368	☐ doc#0 .> Времени у оппонентов американского лидера тоже не	так	много.	</s><s> Демпартия рассчитывает завершить про
1369	☐ doc#0 а конфликта в Донбассе. </s><s> Вариантов, впрочем, не	так	много и,	судя по действиям и заявлениям самого Зеленс
1370	☐ doc#0 ным способом. </s><s> Однако времени у президента не	так	много,	если националисты так и не услышат Зеленского,
1371	☐ doc#0 и выходе социалистов из коалиции с АСУМ у них есть не	так	много	вариантов, с кем сформировать новое правительс
1372	☐ doc#0 тва назвали уже 83% опрошенных. </s><s> Празднику не	так	много	лет -- официально отмечаться он стал лишь в 2005
1373	☐ doc#0 же произойдет и пост президента займет Пенс, у него не	так	много	шансов противостоять сильному кандидату от дел
1374	☐ doc#0 ошению к Турции, которая в настоящее время содержит	так	много	членов ИГ в тюрьме и в то же время контролирует
1375	☐ doc#0 ил стал российским навсегда. <<Таких историй с картами	так	много,	что можно говорить о тенденции: европейцы прс
1376	☐ doc#0 ышинства Вариантов для выхода из кризиса остается не	так	много.	</s><s> Одним из них является формирование пр
1377	☐ doc#0 ию Die Welt Горбачев подчеркнул, что Россия сделала <<	так	много	для того, чтобы положить конец <<холодной войн
1378	☐ doc#0 да называлась Мьянма) с рекой заблудших душ. </s><s>	Так	много	британских чиновников и просто авантюристов н.
1379	☐ doc#0 обстоятельство с тех пор сильно изменились, появилось	так	много	новой информации, которой у нас не было в 2016-
1380	☐ doc#0 жий вопрос не сходит с повестки дня, однако по нему не	так	много	вариантов из-за того, что исчерпаны возможности
1381	☐ doc#0 мплицитне Трампу в информационном пространстве и	так	много.	</s><s> Вот и постарались ее притянуть>>, - счита
1382	☐ doc#0 и реальных хозяйственных достижений у губернатора не	так	много,	но всем в крае понравилось, как Фургал разобрал
1383	☐ doc#0 графом Гани с позиции силы. </s><s> Потому, что только	так	можно	получить максимальную выгоду от будущего ком
1384	☐ doc#0 з как у Ленина, помните было: <<Грabb награбленное!>>	Так	мы и живем в Латвии>>,. -- добавил экс-глава правительс	
1385	☐ doc#0 акону. </s><s> Как определили по конституции и закону,	так	мы и проведем выборы. </s><s> Они для нашего народа,	
1386	☐ doc#0 ия за преступление, которое я не совершал, начинаеш	так	мыслить. </s><s> Я думаю, если бы Россия в лице право	

Figure 3. Aperçu des occurrences de *mak*

186	❏ ① doc#0 с улицы власть, они очень плохо себе представляли, что	такое	вообще государство, что такое государственная машина
187	❏ ① doc#0 его разговора с президентом России. </s><s> Для них это	такой	вопрос политический, а для меня человеческий. </s><s>
188	❏ ① doc#0 а Марианна Уварова. <<У вас есть офшорные счета?>> --	такое	вопрос предложила Ульяна Гродно. </s><s> Был также пс
189	❏ ① doc#0 упали земли, конечно, люди скажут <<нет> >. </s><s> Но	такой	вопрос и ставить не будут. </s><s> А вот если вы постави
190	❏ ① doc#0 тежности Крыма. <<Мы неоднократно говорили, в целом	такого	вопроса не существует и существовать не может: [Крым]
191	❏ ① doc#0 ания роли Украины, но возражать против рассмотрения	такого	вопроса не стал. </s><s> При этом он отметил, что при ре
192	❏ ① doc#0 иет конфронтации с Пекином, а ищет взаимодействия по	таким	вопросам, как экономика и контроль над вооружениями,
193	❏ ① doc#0 повека может способствовать активизации контактов по	таким	вопросам, как Договор о сокращении стратегических нас
194	❏ ① doc#0 чия. </s><s> Сложности, главным образом, возникают по	таким	вопросам, как преемственность руководства, материалы
195	❏ ① doc#0 онной политике, скептицизму в отношении Евросоюза и	таким	вопросам, как аборт и права ЛГБТ>>. </s><s> А потому <<
196	❏ ① doc#0 астенчивым способом. </s><s> Происходит политизация	таких	вопросов, которые должны оставаться вне политики. </s>
197	❏ ① doc#0 ует между еврейской страной и США. </s><s> В решении	таких	вопросов между Россией и Израилем власти стран опире
198	❏ ① doc#0 айте тот закон, о котором я говорил, у вас будет меньше	таких	вопросов. </s><s> В законе не просто прямой призыв, та
199	❏ ① doc#0 зл крайне удивлен, что глава контрразведки занимается	таким	вопросом. </s><s> Президент Чехии в ответ на предостер
200	❏ ① doc#0 в какой-то мере затрагивающего вторую поправку. <<По	такому	вопросу требуется солидарная позиция большинства Ко
201	❏ ① doc#0 й президент лидер нации и все-таки пока он определяет	такие	вопросы наиболее важные для страны, поэтому корректн
202	❏ ① doc#0 уоте Крамп-Карренбауэр предложение не дает ответа на	такие	вопросы, как кто будет вести борьбу со сторонниками ИГ
203	❏ ① doc#0 ттагона. </s><s> В этих исследованиях были, в том числе,	такие	вопросы, которые пытались установить отношение насе
204	❏ ① doc#0 ановлении отмечается, что журналисты имеют право на	такие	вопросы и оценочные суждения по закону <<О СМИ>>. </s>
205	❏ ① doc#0 нии почитали за живого бога, сошедшего на землю, и на	таком	восприятии было воспитано не одно поколение. -- После
206	❏ ① doc#0 това возвращаться в страну, которая поступила со мной	таким	вот образом. </s><s> И вообще всем россиянам я бы пор
207	❏ ① doc#0 на руках до воды не представилось возможным. </s><s>	Такое	вот украинское Аральское море>>, -- написал запорожски
208	❏ ① doc#0 совершенно разные у них были подходы. </s><s> У меня	такое	впечатление, что стороны ведут речь не о Brexit, а о том,
209	❏ ① doc#0 тавы американского внешнеполитического ведомства. --	Такое	впечатление, что то ли дух покойного Маккейна в него в
210	❏ ① doc#0 збраны Хабек и Бербок. </s><s> Тогда их победа не была	такой	впечатляющей, как год спустя. </s><s> Но за истекшее вр
211	❏ ① doc#0 ы. </s><s> Потому что у нас не прописаны критерии, что	такое	вредоносный контент. </s><s> Это сложная и опасная ве
212	❏ ① doc#0 проведении своей политики. </s><s> По иронии судьбы,	такие	времена бывали, и, несмотря на <<холодную войну>>, Ге
213	❏ ① doc#0 годна демократия переживает мрачный день, но даже в	такие	времена поражение не должно одержать верх. </s><s>
214	❏ ① doc#0 омневаются, стоит ли вообще проводить мероприятие в	такое	время. </s><s> Бывший глава НАТО по планированию Фа
215	❏ ① doc#0 час Виктор содержится в тюрьме строго режима. </s><s>	Таких	всего несколько в США. </s><s> Впрочем, какого-то желан

Figure 4. Aperçu des occurrences de *макой*

Dans le cadre de ce travail, nous avons exploité le contenu des corpus compilés dans le but de constituer des lexiques bilingues contenant des termes et, pour la plupart, leur combinatoire. À une plus petite échelle, le même principe peut être appliqué en salle de classe. L'enseignant peut demander aux étudiants de chercher un terme dans *Concordance* (les fonctionnalités du *Sketch Engine* sont détaillées au point [4.3 Interrogation de corpus dans le Sketch Engine](#)) et d'analyser, avec la fonctionnalité *Sort* ou *Collocations*, avec quels verbes ou adjectifs ce terme est le plus fréquemment utilisé. Comme le souligne Rao (2018), la maîtrise des collocations faire partie intégrante de l'apprentissage d'une langue et permet aux apprenants de « *[i]mprove language use, [u]se native like expressions, [e]xpress their ideas with greater power, [i]mprove writing sophistication [and] [f]urther develop their language proficiency* ».

Dans l'exemple ci-dessous (Figure 5), le terme recherché est *правительство* (FR *gouvernement*). Grâce à la fonctionnalité *Collocations*, nous obtenons les collocatifs, sous forme lemmatisée, suivants :

- *глава + правительство (141 occ.)* ;
- *сформировывать + правительство (58 occ.)* ;
- *формирование + правительство 53 occ.)* ;
- *создание + правительство (31 occ.)* ;
- *отставка + правительство (30 occ.)*.

Une fois les lignes de concordance analysées, cette liste permet de dégager les collocations suivantes :

- *глава правительства (FR chef du gouvernement)* ;
- *сформировать правительство (FR former le gouvernement)* ;
- *формирование правительства (FR la formation du gouvernement)* ;
- *создание правительства (FR la formation du gouvernement)* ;
- *отставка правительства (FR la démission du gouvernement)*.

	Lemma	Cooccurrences ?	Candidates ?	T-score	MI	↓ LogDice
1	☐ глава	141	2,515	11.60	5.43	10.14 ...
2	☐ сформировывать	58	137	7.59	8.35	10.11 ...
3	☐ формирование	53	364	7.22	6.81	9.80 ...
4	☐ новый	84	1,887	8.90	5.10	9.63 ...
5	☐ британский	37	522	5.97	5.77	9.17 ...
6	☐ создание	31	399	5.47	5.90	9.00 ...
7	☐ отставка	30	406	5.38	5.83	8.95 ...

Figure 5. Aperçu des collocations de *правительство*

CONCORDANCE RI_RU

simple правительство 58 filter сформировывать -5,-1 58 (47.75 per million) X

Details Left context KWIC Right context

1	☐ doc#0	читать одобрение большинства и сформировать новое правительство -- у них есть все шансы остановить <<жесткий Brexit>>
2	☐ doc#0	партии Курца еще есть возможности сформировать правительство с социал-демократами, что тоже логично -- создать кс
3	☐ doc#0	и, так как ни одна из партий не может сформировать правительство страны. </s><s> Между политическими силами немал
4	☐ doc#0	нейших партий сформировать широкое либеральное правительство . </s><s> То, что препятствует созданию такого правит
5	☐ doc#0	уации, когда ни одна партия не может сформировать правительство с опорой на союзников, именно <<Наш дом Израиль>
6	☐ doc#0	ли 9 мест в парламенте. </s><s> Чтобы сформировать правительство , кандидат на пост премьер-министра должен заручит
7	☐ doc#0	ая впервые за почти 50 лет не смогла сформировать правительство большинства в регионе. </s><s> Сама Меркель заявил
8	☐ doc#0	зра Израиля, без встречи невозможно сформировать правительство единства. </s><s> На выборах в парламент партия Га
9	☐ doc#0	ых партийных лидеров не может быть сформировано правительство единства>>, -- написал Нетаньяху в твиттере. </s><s> I
10	☐ doc#0	> По его словам, Израилу необходимо сформировать правительство единства, которое будет учитывать интересы все гра

Figure 6. Aperçu des 10 premières lignes de concordance de *сформировывать* + *правительство*

Enfin, remarquons également la difficulté que représentent les déclinaisons en russe. À nouveau, les corpus peuvent contribuer à l'élaboration d'exercices où l'étudiant analyse quels sont les cas utilisés après certains verbes ou prépositions. Après avoir analysé les lignes de concordance et tiré ses propres conclusions, l'apprenant aura activement assimilé l'information.

Ce type d'exercices pourrait être mis au point pour les verbes tels que *требовать* (FR *exiger*). Ce verbe représente une difficulté pour les apprenants francophones de la langue russe. En effet, FR *exiger* est un verbe transitif direct et est donc suivi par un complément d'objet direct (*exiger la démission, exiger des explications*), ce qui équivaldrait à un accusatif en russe. Or, RU *требовать* requiert un génitif (*требовать создания, требовать смены режима, требовать борьбы, требовать помощи*) (Figure 7), et non un accusatif.

122	☐	① doc#0	одоходе турецкое руководство действительно вправе	требовать	больших прав в сфере ядерной энергетики. 05.09.201	
123	☐	① doc#0	т добиваться отмены итогов выборов 28 сентября, и	требовать	создания временного правительства, в котором докт	
124	☐	① doc#0	ионной логики. 18.11.2019 21:18:00 НАТО продолжает	требовать	от союзников тратить больше на оборону США стрем	
125	☐	① doc#0	лиллионам простых алжирцев, которые продолжают	требовать	смены режима. </s><s> Правда, что конкретно будет ,	
126	☐	① doc#0	ления. </s><s> Разумеется, бизнес-сообщество будет	требовать	наладить отношения, в первую очередь в США, ибо е	
127	☐	① doc#0	ния, в первую очередь в США, ибо в Китае у властей	требовать	что-либо опасно>>, - сказал Карнеев. </s><s> Похоже	
128	☐	① doc#0	ссор Филатова. </s><s> В частности, Кремль не будет	требовать	борьбы с коррупцией, повышения прозрачности, дел	
129	☐	① doc#0	эбирать свои команды, а затем от имени президента	требовать	помощи и послушания. </s><s> Работа вне правител	
130	☐	① doc#0	им делам, критику лидеров ЕС и экс-посла США и т.д.,	требуем	обнародования стенограмм всех телефонных разгов	
131	☐	① doc#0	устошена бездействием, голодом и бедностью, и мы	требуем	ответов>>, -- говорил ранее один из организаторов п	

Figure 7. Aperçu des occurrences de *требовать*

2.5.1.1.2 Avantages de l'utilisation de corpus en approche directe

Que ce soit par rapport à l'acquisition du lexique (Cobb, 1997 : 303) ou par rapport à la grammaire (Gaskell et Cobb, 2004 : 304), la langue étrangère est souvent trop peu présente dans la vie quotidienne des apprenants. Selon Chambers (2010), l'utilisation des corpus permet de fournir aux étudiants des exemples issus de textes authentiques rédigés par des natifs, ce qui offre aux étudiants un apprentissage multicontextuel (*multicontextual learning*) (Cobb, 1997 : 303).

Selon Corino et Onesti, il est indispensable de disposer de contexte lors de l'apprentissage d'une langue. Ackerley les rejoint et affirme que « *[a] phrase-focused approach to teaching and learning may lead to more fluent, native-like, or expert production* » (2017 : 197). Grâce à l'affichage *KWIC* (*keyword in context*) des lignes de concordance, l'étudiant visualise une série d'informations relatives au terme recherché : expressions idiomatiques, collocations, expressions figées, fréquence... (Corino et Onesti, 2019).

Di Vito (2013) souligne également l'importance de l'affichage *KWIC* et explique que « *la signification et l'emploi d'un mot peuvent être étudiés à partir de l'observation de son contexte linguistique, le co-texte, qui met en évidence les liens syntaxiques et sémantiques caractérisant le mot-clé et ses co-occurents* ».

Di Vito (2013) attire toutefois notre attention sur le fait que l'apprentissage par corpus « *ne peut pas couvrir tous les aspects de l'apprentissage, ni remplacer d'autres pratiques d'enseignement* ».

2.5.1.2 Approche indirecte

Dans le cas de **l'approche indirecte**, l'apprenant n'interroge pas lui-même de corpus mais réalise des exercices élaborés grâce au contenu de corpus. Bien que cette approche puisse être considérée comme moins fructueuse d'un point de vue du développement de l'autonomie, elle permet toutefois d'introduire l'étudiant à l'utilisation des corpus (Chambers, 2010). Chambers ajoute également que l'approche indirecte peut être plus facilement mise en place dans les cas où il n'est pas possible de mettre à la disposition des étudiants le matériel informatique et les logiciels nécessaires à l'interrogation de corpus. McEnery et Wilson (1997 : 6) soulignent par ailleurs que *« l'enseignant peut préparer des exercices qui encouragent une forme constructiviste d'apprentissage »*.

Les exercices proposés ci-dessous seraient donc préparés au préalable par l'enseignant et ne requièrent pas d'interrogation de corpus de la part de l'apprenant.

2.5.1.2.1 Propositions d'exercices en approche indirecte

Boulton et Landure (2016) soulignent que l'utilisation des prépositions constitue fréquemment une source de difficultés pour les étudiants. Nous pourrions dès lors élaborer des exercices présentés sous la forme de textes à trous où l'étudiant doit choisir la préposition correcte. Notons par exemple l'utilisation de la préposition « pour » en russe qui peut être exprimée de plusieurs manières : « за », « для » ou « на ». Une fois l'aspect théorique expliqué par l'enseignant, l'étudiant pourra mettre en pratique ce qu'il a appris et développer activement sa maîtrise des différentes utilisations.

Ce même type d'exercice pourrait être mis au point pour les termes interrogatifs « зачем » et « почему » (ex. 1), pour le choix des aspects perfectif et imperfectif (ex. 2) ou encore pour le choix des déclinaisons (ex. 3).

Les exercices² proposés ci-dessous ont été élaborés à l'aide de la fonctionnalité GDEX (*Good dictionary examples*) accessible depuis Concordance (Figure 8). La fonctionnalité GDEX « *automatically identif[ies] sentences which are easy to*

² Exercices élaborés à l'aide du corpus ruSkell 1.6.

understand and illustrative enough to serve as Good Dictionary EXamples or as sentences suitable for teaching. The sentences are evaluated for length, advanced vocabulary, sufficient context, pronouns pointing outside of the sentence and other criteria » (Sketch Engine).

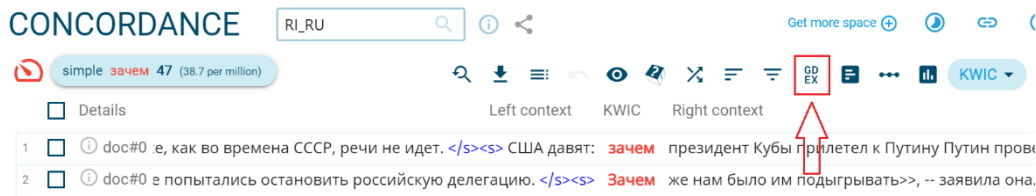


Figure 8. Interface GDEX

Ex. 1 : Remplacez les pointillés par le terme interrogatif adéquat : *зачем* ou *почему*.

- (1) вообще ехать домой чтобы заглянуть в интернет? [зачем]
- (2) А угрожать, если тебя не боятся? [зачем]
- (3) вообще говорить о том, что прошло? [зачем]
- (4) Вы то им нужны? [зачем]
- (5) мне нужна физика и химия, я же ведь собираюсь стать историком! [зачем]
- (6) тогда он не пишет заявление в прокуратуру? [почему]
- (7) Вы ушли из бизнеса во власть? [почему]
- (8) он не опубликовал статью о своей концепции энергии? [почему]
- (9) Сэм ничего не рассказывал об этом? [почему]
- (10) И это ты решила, что ты лучше нее? [почему]

Ex. 2 : Remplacez les pointillés par l'aspect adéquat (perfectif ou imperfectif) du verbe *рассказывать* – *рассказать*.

- (1) Фильм о любви двух молодых людей. [рассказывает]
- (2) Хочу о своем отдыхе в Абхазии. [рассказать]
- (3) Ты не понимаешь, почему я упорно о себе. [рассказываю]
- (4) Безусловно, мастер ей об устройстве и организации Ордена. [рассказывал]

- (5) Пришло время, как моя жизнь. [рассказать]
- (6) Очень хочется о книге, но не выходит. [рассказать]
- (7) Конечно, абсолютно всё невозможно, но самые яркие моменты этого мероприятия мы постарались вам описать.
[рассказать]
- (8) Ему все хотелось говорить, говорить, [рассказать]
- (9) Здравствуйте, мне сказали, что здесь можно свои сны.
[рассказывать]
- (10) Хочу вам как вылечить аскаридоз народными средствами. [рассказать]

Ex. 3 : Remplacez les pointillés par la déclinaison adéquate.

- (1) Это говорит о профессиональн... уровн... коллектива всей компании. [-ом, -е]
- (2) Могу сказать только о хими... и математик.... [-и, -е]
- (3) Тема о войн... двух различных миров была всегда актуальна. [-е]
- (4) Недавно начала читать книг... об истории государства Российского. [-у]
- (5) Эта ситуация требует срочн... разбирательств..., поверьте мне.
[-ого, -а]
- (6) Таковы вопросы, которые требуют сегодня срочн... ответ....
[-ого, -а]
- (7) Купить полный комплект документов для получени... кредит....
[-я, -а]
- (8) Сегодня карты можно купить в люб... районе Москв.... [-ом, -ы]
- (9) Марина, журналист, 28 лет: Несколько лет назад купила машин....
[-у]
- (10) За пять месяц... россияне уже купили один миллион машин.
[-ев]

3 Constitution des corpus

Dans ce chapitre, nous exposerons les critères appliqués pour la sélection des textes constitutifs des corpus, présenterons les journaux retenus et expliquerons le processus de compilation.

3.1 Critères de constitution

Bowker et Pearson (2002) énumèrent neuf critères de constitution de corpus, que nous avons retenus dans le cadre de ce travail et que nous exposons ci-après.

3.1.1 Taille du corpus

Il n'existe pas de taille idéale à respecter, le volume du corpus dépend des besoins de la recherche (Bowker et Pearson, 2002 : 45-48). Toutefois, Williams (1999) recommande un seuil de 500 000 mots, taille « *généralement acceptée pour les corpus spécialisés* ».

Dans le cadre de ce travail, nous avons décidé de nous aligner aux critères de précédents mémoires et avons fixé le seuil à un million de mots pour chaque corpus. Le corpus français est plus exactement composé de 1 308 446 mots, tandis que le corpus russe compte 982 520 mots.

3.1.2 Extraits – textes entiers

Les corpus sont constitués d'articles de presse sélectionnés dans leur intégralité. Comme le soulignent Bowker et Pearson (2002 : 49), il convient de sélectionner les textes dans leur intégralité dans le cas de corpus spécialisés, l'information terminologique pouvant se trouver dans n'importe quelle partie du texte.

3.1.3 Nombre de textes

Les corpus sont constitués de plusieurs milliers d'articles rédigés par une multitude de journalistes afin de garantir la représentativité de la langue et d'éviter les répétitions d'une certaine manière de dire d'un auteur (Bowker et Pearson, 2002 : 49). Ce critère est extrêmement important selon Bowker et Pearson (2002) qui

affirment que « *there is a big difference between a 20,000-word corpus that is made up of twenty different texts written by twenty different authors and a 20,000-word corpus that is made up of just two long texts written by the same author* ».

3.1.4 Format

Les corpus sont constitués uniquement d'articles issus de la presse généraliste écrite. Les articles traduits et les transcriptions de vidéos ont été exclus lors de la constitution des corpus.

3.1.5 Sujet

Chaque article constitutif des corpus est issu des rubriques « relations internationales » et « conflits armés » des journaux [Le Monde](#), [Le Vif/L'Express](#) et [Le Figaro](#) pour le corpus francophone, et des journaux [Газета](#) (Gazeta), [Известия](#) (Izvestia), [Независимая](#) (Nezavissimaïa), [Новая Газета](#) (Novaïa Gazeta) et [Ведомости](#) (Vedomosti) pour le corpus russe.

3.1.6 Type de texte

Les textes sélectionnés sont issus de la presse généraliste. Nous retrouvons des articles sur l'actualité, des comptes rendus, des reportages, des enquêtes et des éditoriaux.

3.1.7 Auteur

Les auteurs sont des journalistes et des envoyés spéciaux des journaux mentionnés ci-dessus.

3.1.8 Langue

Les articles ont été rédigés par des locuteurs natifs de la langue. Comme expliqué dans la section [3.1.4 Format](#), les articles traduits n'ont pas été sélectionnés lors de la constitution des corpus.

Les journaux sélectionnés pour le corpus russe sont représentatifs du russe de Russie. Le corpus francophone contient, pour sa part, du français de France et de

Belgique. Notons cependant que le poids du journal belge *Le Vif/L'Express* est minime, le corpus est, de ce fait, majoritairement représentatif du français de France (voir le point [3.2.9 Poids des journaux au sein du corpus](#)).

3.1.9 Date de publication

Initialement, la période d'extraction des articles a été fixée à deux mois pour les deux langues, soit du 1^{er} septembre 2019 au 31 octobre 2019. Cependant, au terme de ces deux mois, nous avons remarqué que, si le seuil du million avait bien été atteint en français, ce n'était pas le cas pour le russe. Nous avons donc prolongé la période d'extraction des articles russes jusqu'au 02 décembre 2019, date où le seuil du million a été franchi.

Nous pouvons expliquer ce déséquilibre par la taille des articles. En effet, nous avons remarqué une nette différence entre la longueur des articles extraits des journaux francophones et ceux extraits de la presse russe. Il a donc été nécessaire de compiler sur une plus longue période pour atteindre un équilibre entre les deux langues.

3.2 Présentation des journaux sélectionnés

Le corpus français est composé d'articles issus des journaux [Le Monde](#), [Le Figaro](#) et [Le Vif/L'Express](#). Le corpus russe est constitué d'articles issus des journaux [Газета](#) (Gazeta), [Известия](#) (Izvestia), [Независимая](#) (Nezavissimaïa), [Новая Газета](#) (Novaïa Gazeta) et [Ведомости](#) (Vedomosti).

3.2.1 Le Monde

[Le Monde](#) est un quotidien français créé en 1944 par Hubert-Beuve-Méry à la demande du général de Gaulle qui souhaitait fonder un journal de référence disponible à l'internationale (Gralon, 2008).

Aujourd'hui, *Le Monde* est considéré comme le quotidien de référence de la presse francophone et est accessible dans plus de 120 pays. Il est disponible sous format papier et sous format électronique et propose de nombreux suppléments hebdomadaires tels que « Le Monde Économie » ou « Le Monde Radio TV »

(Gralon, 2008). Selon l'Alliance pour les Chiffres de la Presse et des Médias (ACPM), le tirage papier s'élève à 188 123 exemplaires par jour pour septembre 2019.

Le Monde est un journal qui se veut indépendant, que ce soit vis-à-vis de la politique ou des grandes puissances financières (L'Histoire, 2001) et appartient au groupe de presse *Groupe Le Monde* (Le Monde, 2019). Son orientation politique est de gauche (Euro Topics).

3.2.2 Le Figaro

Fondé en 1826 par Maurice Alhoy et Étienne Arago, [Le Figaro](#) est à l'époque un journal « *littéraire satirique et non-politique* » (Laroche-Signorile, 2016).

Le Figaro devient un journal quotidien en 1866 sous la direction d'Hyppolyte de Villemessant et fait aujourd'hui partie du groupe de presse *Groupe Figaro*, filiale du *Groupe Dassault* (Groupe Dassault, 2016). Il est le plus ancien journal français encore en activité et est le premier quotidien conservateur de France (Euro Topics). *Le Figaro* est disponible sous format papier et sous format numérique. Selon l'ACPM, le tirage papier de septembre 2019 s'élève à 211 470 exemplaires par jour.

3.2.3 Le Vif/L'Express

Fondé en 1983, [Le Vif/L'Express](#) est un journal hebdomadaire belge francophone qui fusionne en 1986 avec le journal français *L'Express* (Courrier International). Ce journal appartient au groupe de presse *Roularta Media* (Roularta Media) et est de centre gauche (Euro Topics). *Le Vif/L'Express* est disponible tant au format papier qu'au format électronique. Selon le Centre d'information sur les médias (CIM), le tirage papier moyen pour l'année 2019 s'élève à 61 665 exemplaires.

3.2.4 Газета

Fondé en 1999 par l'homme d'affaires Anton Nosik, [Gazeta](#) est un quotidien russe disponible uniquement au format électronique et dont l'accès est gratuit. Il figure régulièrement dans le classement des sites internet russes les plus consultés et est

de tendance libérale (Courrier International). *Gazeta* appartient au groupe de presse russe *Rambler Group* (Rambler Group).

3.2.5 Известия

Créé pendant la révolution russe de 1917, [Izvestia](#) est aujourd'hui un des quotidiens les plus influents de Russie (Courrier International). Il est disponible au format papier ainsi qu'au format électronique dont l'accès est gratuit. *Izvestia* appartient au *National Media Group* (NMG) et est de tendance pro-gouvernementale ; il représente le « *porte-voix des positions du Kremlin* » (Euro Topics). Le tirage papier de ce quotidien s'élève à 263 600 exemplaires (Courrier International).

3.2.6 Независимая

[Nezavissimaïa](#) est un quotidien russe né en 1990, un an avant la chute de l'URSS. Il constitue le journal russe le plus tourné vers la politique étrangère, et plus particulièrement vers l'actualité des pays de la CEI (Courrier International). *Nezavissimaïa* est disponible sous format papier et sous format électronique en libre accès et est de tendance social-démocrate (Courrier International). Ce quotidien propose divers suppléments traitant de littérature, de religion, des affaires internationales ou encore de science (Euro Topics). Depuis 2005, *Nezavissimaïa* est la propriété de l'homme d'affaires et journaliste Konstantin Remchukov. Selon le Courrier International, le tirage papier de ce journal s'élève à 42 000 exemplaires.

3.2.7 Новая Газета

[Novaïa Gazeta](#) est un journal russe bihebdomadaire fondé en 1993 par d'anciens journalistes de la *Komsomolskaïa Pravda* dont l'objectif était de fonder un nouveau quotidien indépendant et influent (Courrier International). Ce journal est considéré comme l'un des journaux les plus ouvertement critiques envers le gouvernement et la société russe (Courrier International). Il est disponible sous format papier et sous format électronique en libre accès et est détenu par l'ex-président soviétique Mikhaïl Gorbatchev et l'ancien député de la Douma Alexandre Lebedev (The Guardian, 2015). Son tirage papier s'élève à 530 000 exemplaires (Courrier International).

3.2.8 Ведомости

[Vedomosti](#) est un quotidien russe fondé en 1999 par le groupe de presse russe *Sanoma Independent Media* en partenariat avec le *Financial Times* et *The Wall Street Journal* (Courrier International). Il est revendu en 2015 au propriétaire du *Moscow Times*, Demian Koudryavtsev (Euro Topics). *Vedomosti* constitue le quotidien de référence pour les domaines du monde des affaires et de la finance. Le journal est disponible sous format papier et sous format électronique en libre accès et est d'orientation libérale (Courrier International). Le tirage papier de *Vedomosti* s'élève à 56 500 exemplaires (Courrier International).

3.2.9 Poids des journaux au sein du corpus

3.2.9.1 Corpus francophone

Le corpus francophone est constitué des journaux *Le Monde*, *Le Figaro* et *Le Vif/L'Express*. *Le Figaro* détient 50 % du volume total, suivi de près par *Le Monde*. Nous constatons dès lors un déséquilibre au sein du corpus, *Le Vif/L'Express* détenant à peine 3 % du volume total (Figure 9). Si le corpus était à refaire, nous supprimerions vraisemblablement *Le Vif/L'Express* dont l'apport est négligeable. Ces 3 % peuvent être expliqués par le fait que *Le Vif/L'Express* est un hebdomadaire, tandis que *Le Monde* et *Le Figaro* sont des quotidiens.

D'autre part, le journal *Le Soir* avait été sélectionné lors du choix des journaux afin de disposer de deux journaux représentatifs du français de Belgique et de deux journaux représentatifs du français de France. Cependant, les archives du journal *Le Soir* n'étaient pas disponibles sur le site, et nous n'avons dès lors pas pu extraire les articles de ce journal.

Journaux	Poids
Le Monde	613 922 mots
Le Vif	36 178 mots
Le Figaro	658 346 mots
Total	1 308 446 mots

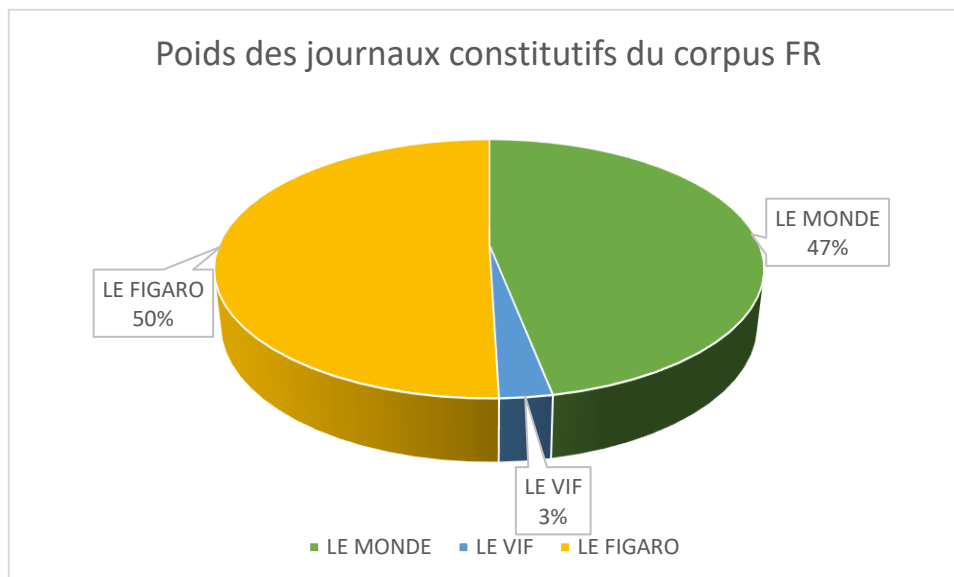


Figure 9. Poids des journaux constitutifs du corpus FR

3.2.9.2 Corpus russe

Nous constatons également un déséquilibre au sein du corpus russe, composé à près de 50 % du journal *Gazeta* (Figure 10). Cette prédominance peut être expliquée par le fait que deux rubriques du journal relevaient des relations internationales (« *Внешняя политика* » et « *Политика в мире* ») et ont donc été compilées.

Nous pouvons également proposer une explication pour les journaux *Novaïa Gazeta* et *Vedomosti* dont le poids au sein du corpus est plus faible.

Concernant le journal *Novaïa Gazeta*, l'explication relève de la périodicité du journal, qui est publié deux fois par semaine, contrairement aux autres journaux qui sont, eux, publiés quotidiennement.

Quant au journal *Vedomosti*, nous avons constaté une nette infériorité du nombre d'articles publiés quotidiennement, ce qui peut être expliqué par le fait que le domaine d'expertise du journal relève plutôt de l'économie et du monde des affaires.

Journaux	Poids
Gazeta	450 172 mots
Izvestia	163 207 mots
Nezavissimaïa	243 120 mots
Novaïa Gazeta	79 710 mots
Vedomosti	46 311 mots
Total	982 520 mots

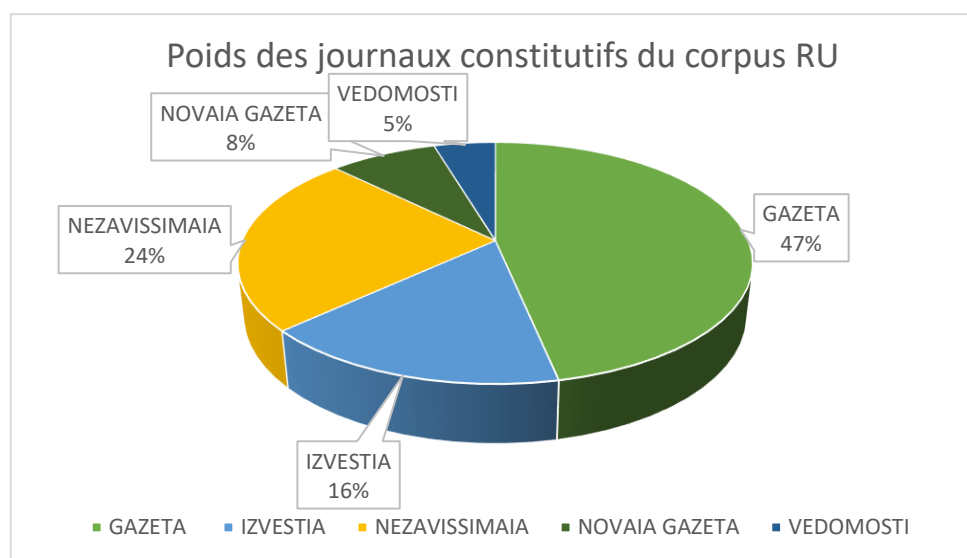


Figure 10. Poids des journaux constitutifs du corpus RU

3.2.10 Évènements survenus pendant la période de compilation

Comme nous l'avons souligné au point [3.1.9 Date de publication](#), les articles constitutifs du corpus français ont été publiés entre le 1^e septembre 2019 et le 31 octobre 2019, et ceux du corpus russe ont été publiés entre le 1^e septembre 2019 et le 02 décembre 2019. Durant cette période, nombre d'événements ont fait la une de la section internationale de la presse. Ci-dessous, nous énumérons ces principaux événements, en distinguant les événements traités à la fois par la presse russe et francophone de ceux traités uniquement par la presse russe ou par la presse francophone.

- États-Unis : procédure de destitution du président Donald Trump ;
- États-Unis : élections présidentielles ;
- Royaume-Uni : Brexit ;
- Mort du chef de Daech Abou Bakr al-Baghdadi ;
- Guerre en Syrie ;
- Discussion à Genève entre le régime de Damas et l'opposition ;
- Crise des migrants ;
- États-Unis : la Chambre des représentants reconnaît le génocide arménien ;
- Espagne : manifestations à Barcelone pour l'indépendance de la Catalogne ;
- Espagne : déménagement de la sépulture du général Franco ;
- Allemagne : montée en puissance du parti d'extrême-droite « AfD » ;
- Italie : victoire de Matteo Salvini en Ombrie ;
- Irak : contestation du pouvoir ;
- Israël : Benyamin Nétanyahou renonce à tenter de former un gouvernement ;
- Japon : intronisation de l'empereur Naruhito ;
- Canada : victoire du parti de Justin Trudeau aux élections parlementaires ;
- Chili : grèves et manifestations contre les inégalités socio-économiques ;

- Bolivie : crise post-électorale et suspicions de fraude électorale à la suite de la victoire d'Evo Morales au premier tour ;
- Liban : démission du premier ministre ;
- Liban : soulèvement populaire contre la hausse des taxes et la corruption de la classe politique ;
- Chine : éviction de l'opposant Joshua Wong aux élections locales de Pékin ;
- Argentine : victoire d'Alberto Fernandez au premier tour de l'élection présidentielle ;
- Tunisie : élections présidentielles ;
- Corée du Nord : test d'un système de lancement de « missiles multiples » ;
- Russie : Poutine supervise des tirs de missiles stratégiques à travers le pays ;
- France : décès de Jacques Chirac ;
- Irlande : Légalisation de l'avortement et du mariage homosexuel ;
- Incendies dévastateurs en Californie et en Australie ;
- Proposition d'aide de l'ONU face à la marée noire au Brésil ;
- Contestation de l'hypothèse de suicide de Jeffrey Epstein ;
- Russie : un militaire tue huit soldats sur une base de Sibérie ;
- Chine : Sophia Huang Xueqin, instigatrice du mouvement « Me too », arrêtée ;
- Royaume-Uni : 39 corps découverts dans un camion arrivé de Belgique.

Sujets mentionnés uniquement par la presse russe :

- Ukraine : guerre du Donbass ;
- Ukraine : Kiev souhaite négocier la récupération de la Crimée ;
- Ukraine : ouverture d'une enquête contre l'ancien procureur général Iouri Loutsenko ;

- Géorgie : démission du premier ministre ;
- Kosovo : élections parlementaires remportées par l'opposition ;
- Tatarstan : la Cour suprême se déclare en faveur du retour de la peine de mort ;
- La Biélorussie refuse le projet de création d'une confédération avec la Russie.

Sujets mentionnés uniquement par la presse francophone :

- Belgique : Sophie Wilmès nommée première ministre par intérim ;
- Guinée-Bissau : le président José Mario Vaz limoge avec effet immédiat le gouvernement en place ;
- Algérie : élections présidentielles ;
- Algérie : manifestations contre le pouvoir.

Des graphiques circulaires (Figures 11 et 12) ont été réalisés afin de comparer les régions du monde les plus mentionnées par la presse francophone d'une part, et par la presse russe d'autre part. Le graphique FR contient les articles publiés par le journal *Le Monde*, tandis que le graphique RU contient les articles publiés par le journal *Gazeta*. Ces deux journaux constituent, l'un comme l'autre, légèrement moins de 50 % de leur corpus respectif et ont donc été sélectionnés pour représenter un échantillon du contenu des journaux de la presse francophone et de la presse russe.

Du côté de la presse francophone (Figure 11), nous pouvons constater que, sur les 800 articles extraits du journal *Le Monde*, la moitié de ces articles traitent de l'Europe et du Moyen-Orient. La troisième région du monde la plus mentionnée est l'Amérique du Nord (14 %), suivie de près par l'Afrique (11 %). Le dernier quart du graphique reprend l'Asie (8 %), l'Europe de l'Est (7 %) et l'Amérique du Sud (7 %). La proportion des articles traitant de l'Amérique centrale, des Balkans, de l'Australie et du Caucase est quasiment insignifiante (environ 1 % chacun).

Du côté de la presse russe (Figure 12), nous constatons d'emblée la prédominance de l'Europe de l'Est. En effet, sur les 843 articles extraits du journal *Gazeta*, plus de la moitié (54 %) traite des pays d'Europe de l'Est. L'Amérique du Nord vient en seconde place avec 22 % et le dernier quart du graphique reprend l'Europe (8 %), le Moyen-Orient (6 %), l'Asie (4 %), le Caucase (2 %), l'Amérique du Sud (2 %) et enfin l'Amérique centrale et les Balkans (1 % chacun). L'Afrique représente moins d'1 % (un article sur 843) et l'Australie ne figure pas dans les données du graphique russe parce qu'aucun article du journal *Gazeta* ne traitait de cette région du monde.

Les 10 pays les plus mentionnés par le journal *Le Monde* sont : les **États-Unis** (109 articles), le **Royaume-Uni** (91 articles), la **Syrie** (49 articles), la **Chine** (36 articles), la **Russie** (31 articles), la **France** (29 articles), **Israël** (26 articles), **l'Espagne** (23 articles), **l'Iran** (21 articles) et **l'Irak** (18 articles).

Les 10 pays les plus mentionnés par le journal *Gazeta* sont : **l'Ukraine** (234 articles), les **États-Unis** (183 articles), la **Russie** (173 articles), le **Royaume-Uni** (29 articles), la **Biélorussie** (28 articles), la **Syrie** (20 articles), la **Géorgie** (17 articles), **l'Allemagne** (14 articles), la **Chine** (14 articles) et **Israël** (13 articles).

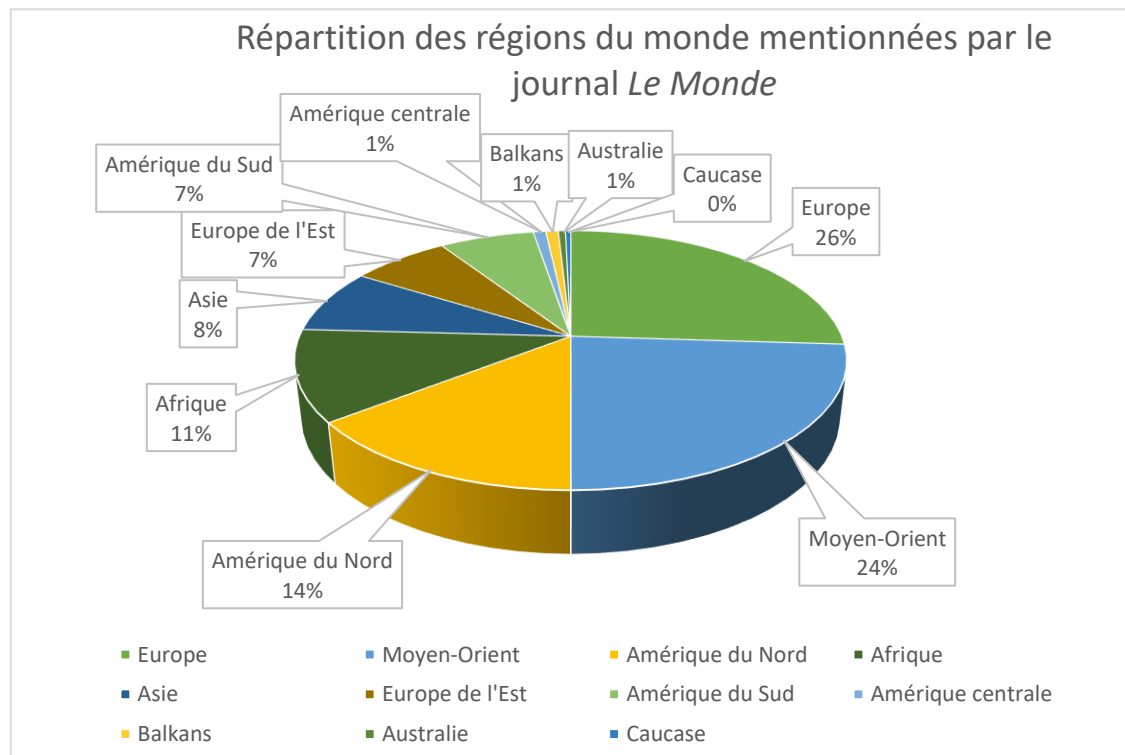


Figure 11. Répartition des régions du monde mentionnées par le journal *Le Monde*

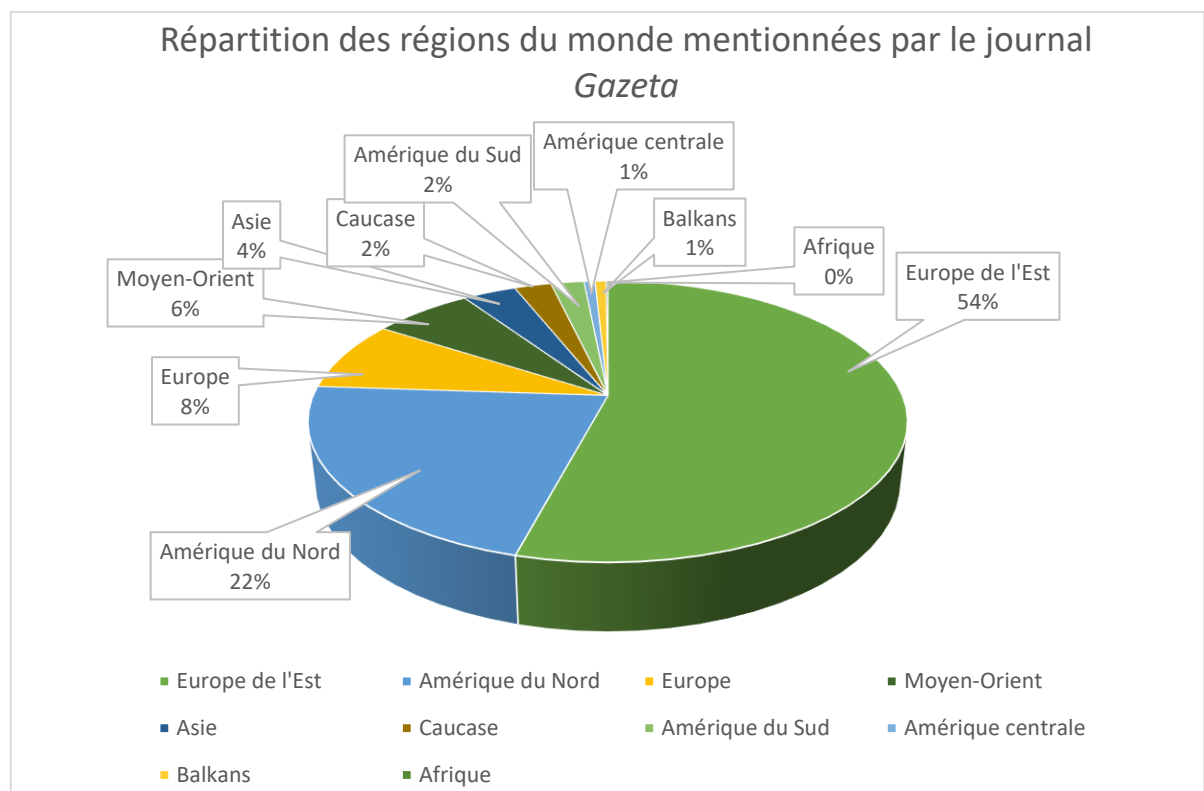


Figure 12. Répartition des régions du monde mentionnées par le journal *Gazeta*

3.3 La compilation du corpus

Selon le Professeur Michael Thomas (2009), la compilation d'un corpus est « *[t]he process of collecting samples of language according to predefined criteria, such as medium, register, genre, etc., and putting them together either in a single file or a file collection to serve as data for concordance queries* ».

Nous expliquons dans ce chapitre comment nous avons compilé les corpus et quelles difficultés ont été rencontrées lors de ce processus.

3.3.1 Procédé de compilation

Nous avons procédé méthodiquement lors de l'extraction des textes des différents journaux. Les textes sont issus des rubriques de la politique internationale et ont été compilés manuellement. Nous avons créé un document *Word* par journal afin de garder une trace de la provenance des articles. Lors de la compilation, nous avons extrait un à un les articles de la période du 01/09/2019 au 31/10/2019 pour le français, et du 01/09/2019 au 02/12/2019 pour le russe.

Afin d'accélérer et de faciliter le processus d'extraction, nous avons mis au point une macro sur *Word*. Cette macro a été configurée pour automatiquement sélectionner l'entièreté du texte, supprimer toutes les mises en forme (comme les liens hypertextes), nettoyer le fichier et insérer un saut de page. Nous avons cependant dû supprimer manuellement toutes les images et leur légende ainsi que tout texte étranger à l'article (textes de publicité, liens vers articles externes...).

L'objectif était donc de disposer de documents nettoyés, contenant uniquement du texte brut et dépourvus d'informations impropres à l'article. Cette étape du travail s'est révélée extrêmement chronophage, et nous avons rencontré quelques difficultés lors de l'extraction des articles.

3.3.2 Difficultés rencontrées

3.3.2.1 Difficultés rencontrées avec les journaux francophones

Nous avons rencontré deux principales difficultés lors de l'extraction des articles de certains journaux francophones.

- Accès : l'accès aux journaux francophones a posé un problème car il ne nous était pas possible de consulter les articles gratuitement. Nous avons toutefois pu accéder aux articles grâce aux comptes fournis par l'Université. Cependant, certains abonnements ont été modifiés au cours de la période de compilation et nous avons notamment perdu l'accès au journal *Le Monde*. Ne souhaitant pas perdre cette source ni nos avancées dans le processus de compilation, nous avons pris la décision de souscrire un abonnement payant.
- Archives : certains journaux ne mettaient pas leurs archives à la disposition des utilisateurs. Il nous était donc impossible de consulter les articles plus anciens qu'un certain nombre de jours. C'est pour cette raison que le journal *Le Soir*, initialement sélectionné dans notre liste, a dû être mis de côté. Il aurait été intéressant de disposer de ce quotidien belge dans le corpus, mais l'impossibilité d'accéder aux archives nous en a empêché.

3.3.2.2 Difficultés rencontrées avec les journaux russes

Tous les journaux russes sélectionnés étaient par chance disponibles gratuitement à travers le monde. Le principal problème rencontré avec certains journaux russes ne relevait pas de l'accès, mais de la praticité des sites.

Certains journaux, notamment *Gazeta* et *Vedomosti*, présentaient leurs articles sous la forme de liste. Cette liste chargeait les articles au fur et à mesure que nous descendions le curseur vers le bas de la page. Malheureusement, il est arrivé à de nombreuses reprises que la page du site s'actualise et nous ramène au début de la liste. Le temps perdu à redescendre les centaines d'articles et à retrouver l'article où nous nous trouvions a été considérable et a représenté un réel frein dans l'avancée de la compilation du corpus russe.

4 Sketch Engine

4.1 Présentation

Créé en 2004 par Adam Kilgarriff, le *Sketch Engine* est un système de gestion de corpus disponible en ligne (www.sketchengine.eu). Non seulement il héberge quelque 500 corpus dans plus de 90 langues, mais il offre aussi à l'utilisateur la possibilité de compiler et d'interroger ses propres corpus. Le *Sketch Engine* est un outil payant, toutefois, les institutions académiques européennes bénéficient d'une subvention de l'Union européenne et y ont accès gratuitement jusqu'au 1^{er} avril 2022.

4.1.1 Corpus disponibles sur le Sketch Engine

Les corpus hébergés par le *Sketch Engine* sont multiples (Figure 14) : corpus de référence (*British National Corpus*, *Brown*) ; corpus spécialisés (*Brexit corpus*, *Medical Web Corpus*) ; corpus parallèles (*DGT*, *Europarl*) ; corpus Web (la suite des corpus *TenTen*, disponibles dans plus de 40 langues) ; corpus oraux (*Open American National Corpus*) ; corpus d'apprenants (*Arabic Learner Corpus*), etc. Bien que les corpus de langue anglaise soient les plus nombreux, le *Sketch Engine* propose des corpus dans plus de 90 langues (Figure 13), de l'afrikaans au yoruba, en passant par l'islandais, le népalais ou encore le maltais.

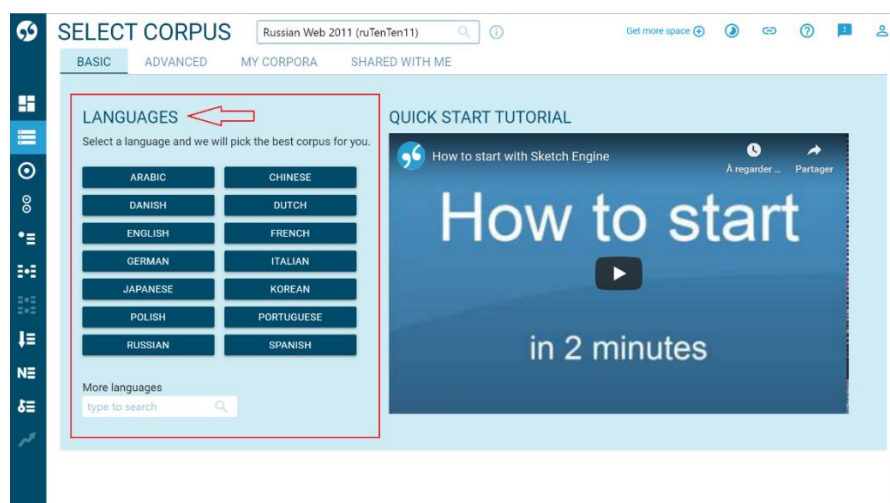


Figure 13. Recherche par langue

SELECT CORPUS Russian Web 2011 (ruTenTen11) Get more space

BASIC ADVANCED MY CORPORA SHARED WITH ME

Corpus or language... Only with word sketches Specialized x Corpus language...

191 corpora 42 languages

Language	Name	Words
Afrikaans	CHILDES Afrikaans Corpus	26,020
Arabic	Quran annotated corpus [unvowelled Arabic]	128,243
Arabic	Quran annotated corpus [unvowelled Latin]	99,268
Arabic	Quran annotated corpus [vowelled Arabic]	128,241
Arabic	Quran annotated corpus [vowelled Latin]	97,970
Arabic	KSUCCA (Classical Arabic)	46,705,577
Bulgarian	EUROPARL7, Bulgarian	9,215,233
Bulgarian	DGT, Bulgarian	25,912,721
Bulgarian	EUR-Lex Bulgarian 2/2016	329,071,554
Bulgarian	EUR-Lex judgments Bulgarian 12/2016	17,071,495
Catalan	CHILDES Catalan Corpus	209,525
Chinese Simplified	Chinese GigaWord 2 Corpus: Mainland, simplified	205,031,379
Chinese Simplified	Guangwai - Lancaster Chinese Learner Corpus	1,289,060
Chinese Traditional	Chinese GigaWord 2 Corpus: Taiwan, traditional	382,600,557
Croatian	CHILDES Croatian Corpus	300,832
Croatian	DGT, Croatian	3,968,608
Croatian	EUR-Lex Croatian 2/2016	109,138,184
Croatian	EUR-Lex judgments Croatian 12/2016	5,613,468
Czech	EUROPARL7, Czech	13,013,774
Czech	DGT, Czech	43,621,933
Czech	EUR-Lex Czech 2/2016	350,230,088
Czech	EUR-Lex judgments Czech 12/2016	18,226,505
Czech	CzechParl 2012 (v2 with lemmos)	37,184,025
Danish	CHILDES Danish Corpus	285,231
Danish	EUROPARL7, Danish	48,343,860

CORPUS CATEGORY

- ALL 567
- RECENT 0
- MY CORPORA 2
- SHARED WITH ME 46
- FEATURED 16
- GENERAL PURPOSE 307
- WEB 252
- NON-WEB 315
- PARALLEL 142
- SPOKEN 52
- SPECIALIZED 191**
- DIACHRONIC 129
- MULTIMEDIA 1
- LEARNER 4
- ERROR-ANNOTATED 2
- GRAMMAR DEVELOPMENT 0

Figure 14. Recherche par catégorie

4.1.2 Utilisateurs et utilisations du *Sketch Engine*

Kilgarrif *et al.* (2014) identifient six utilisateurs et utilisations du *Sketch Engine* : lexicographie, universités, apprentissage des langues, traducteurs, terminologues et entreprises d'industrie langagière.

Les premiers utilisateurs du *Sketch Engine* étaient et restent les lexicographes. Encore aujourd'hui, de nombreux dictionnaires tels que le *Harper Collins* ou le *Cambridge University Press* utilisent régulièrement le *Sketch Engine* pour mettre au jour de la nouvelle terminologie et suivre l'évolution du sens de certains termes.

Nous retrouvons également cet outil dans les universités. Il est utilisé dans les départements de linguistique pour l'enseignement et la recherche, mais aussi dans l'enseignement de la traduction et dans l'analyse du discours.

En traduction, les corpus permettent d'identifier la terminologie et la phraséologie d'un domaine et de s'en approprier les manières de dire spécifiques et correctes.

En terminologie, l'utilisation consistante d'un même terme pour chaque langue est primordiale. Les corpus permettent d'extraire la traduction d'un terme spécifique en différentes langues et de mettre en place des répertoires terminologiques pour différents domaines de spécialité.

Enfin, nous pouvons également citer les développeurs d'application qui se servent du matériel brut fourni par les corpus du *Sketch Engine* afin de développer des fonctionnalités liées à la reconnaissance vocale, à la correction orthographique ou encore à la saisie intuitive.

4.2 Création et compilation de corpus dans le Sketch Engine

L'utilisateur peut créer et compiler des corpus unilingues ou multilingues. Dans le cas des corpus unilingues, ceux-ci peuvent être créés sur la base de textes sélectionnés automatiquement sur le Web (*corpus compilé automatiquement*) ou sur la base de textes sélectionnés par l'utilisateur selon des critères précis (*corpus compilé manuellement*). Comme dans le cadre de ce travail nous avons compilé manuellement les corpus, seule cette méthode de création et de compilation sera expliquée dans les lignes qui suivent.

4.2.1 Création du corpus

Lorsque l'utilisateur s'est identifié, la page « *Dashboard* » s'affiche. En haut à droite de cette page, il suffit de cliquer sur le bouton « *NEW CORPUS* » pour créer un nouveau corpus (Figure 15).

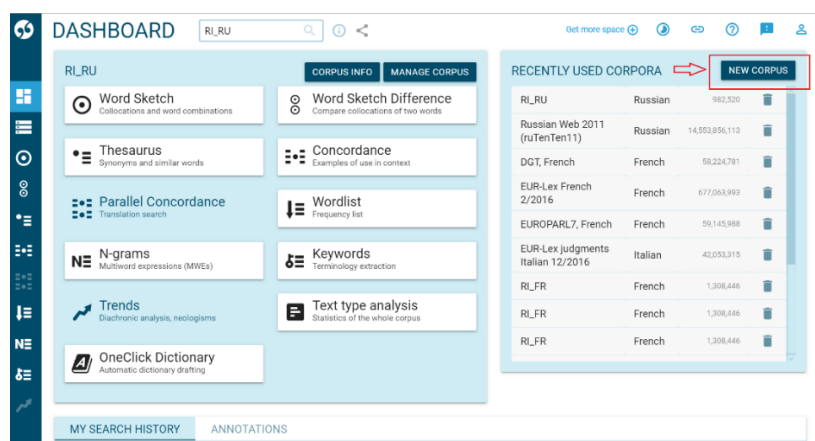


Figure 15. Créer un nouveau corpus

Une fois sur la page de création du corpus, l'utilisateur est invité à nommer son corpus, à choisir entre corpus unilingue et corpus multilingue et, dans le cas d'un corpus unilingue, à sélectionner la langue du corpus (Figure 16). Cette dernière étape est essentielle puisqu'au choix de la langue est associé un étiqueteur. Enfin, l'utilisateur peut, s'il le souhaite, inclure une brève description du corpus.

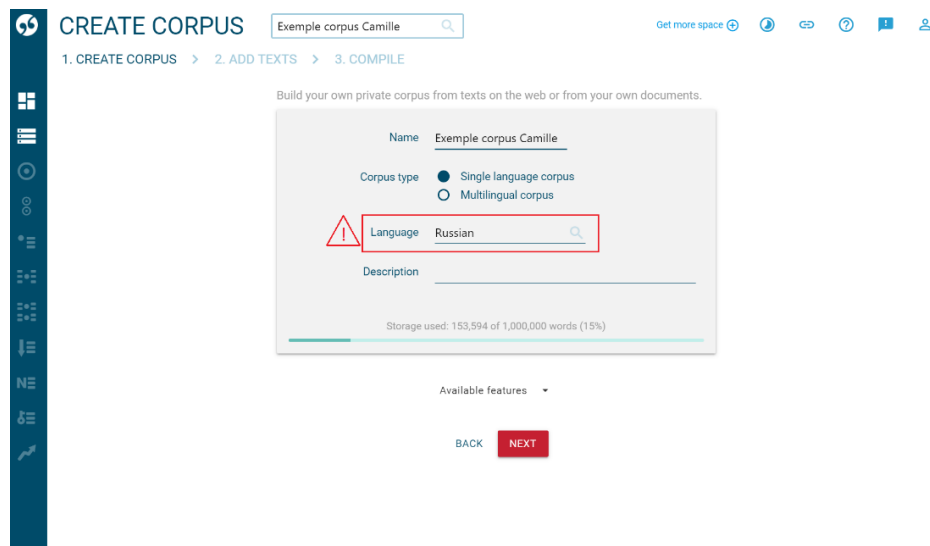


Figure 16. Caractéristiques du corpus

4.2.2 Ajout de fichiers

L'étape suivante consiste à ajouter les textes constitutifs du corpus. Le *Sketch Engine* offre le choix entre « *Find texts on the web* » ou « *I have my own texts* ». Comme nous l'avons souligné précédemment, nous avons privilégié l'option « *I have my own texts* » dans le cadre de ce travail.

Le *Sketch Engine* accepte un large éventail de formats de fichiers (.csv, .doc, .docx, .htm, .html, .ods, .pdf, .tar.bz2, .tar.gz, .tei, .tgz, .tmx, .txt, .vert, .xlf, .xliff, .xml, .zip). Si les fichiers sont nombreux, il est conseillé de les sauvegarder dans un ou plusieurs dossiers compressés pour en faciliter l'importation dans le *Sketch Engine*.

En fonction du nombre et du poids des fichiers, le téléchargement sera plus ou moins long (Figure 17). Une fois le téléchargement terminé, il est possible d'afficher

chaque fichier en cliquant sur « *upload* » : le format du fichier ainsi que le nombre de mots qu'il contient apparaîtront (Figure 18).

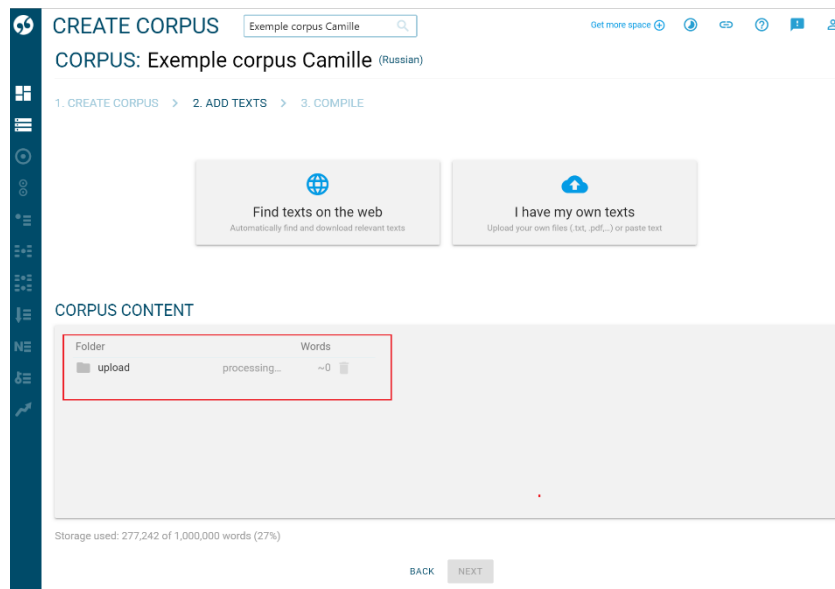


Figure 17. Téléchargement des fichiers

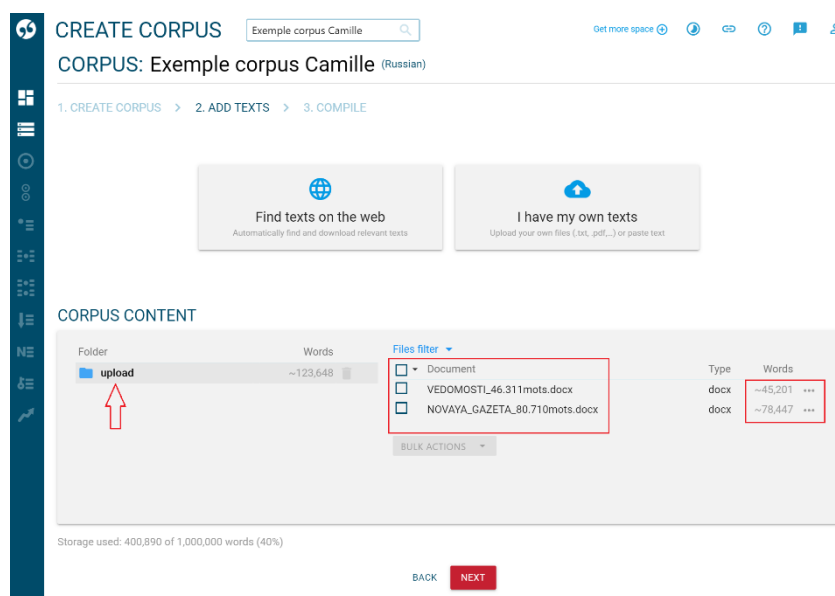


Figure 18. Fichiers téléchargés

4.2.3 Compilation du corpus

La dernière étape consiste à compiler le corpus en cliquant sur le bouton « *compile* » (Figure 19). À nouveau, le temps de compilation varie en fonction du volume du corpus.

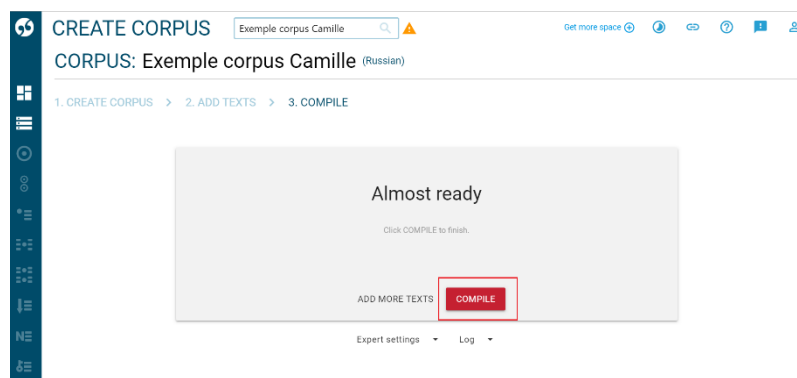


Figure 19. Démarrer la compilation

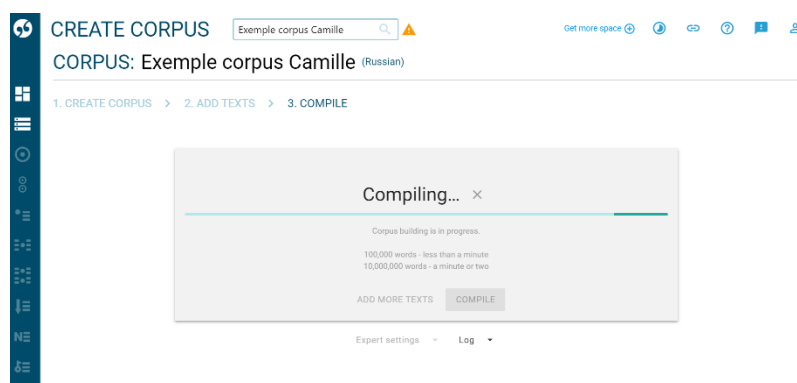


Figure 20. Compilation

Dès la compilation terminée, l'utilisateur peut afficher le « *Dashboard* » de son corpus ou explorer son corpus en cliquant sur « *Extract keywords and Terms* » ou sur « *Corpus details and Statistics* » (Figure 21).

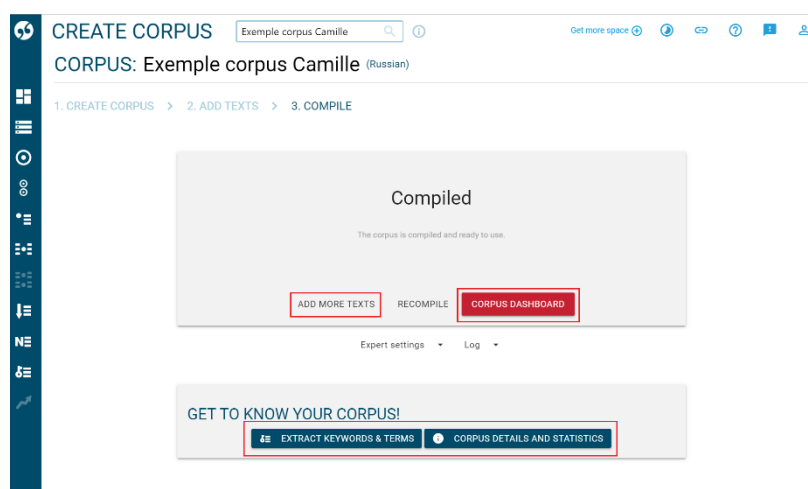


Figure 21. Corpus compilé

L'interface « *Corpus details and Statistics* » (Figure 22) apporte différentes informations sur les caractéristiques du corpus : auteur, date de création, étiqueteur, langue, nombre de tokens, de mots et de phrases.

Un token est « *the smallest unit that a corpus consists of. A token normally refers to: a word form, punctuation, digit, abbreviations, product names or anything else between spaces* » (Sketch Engine).

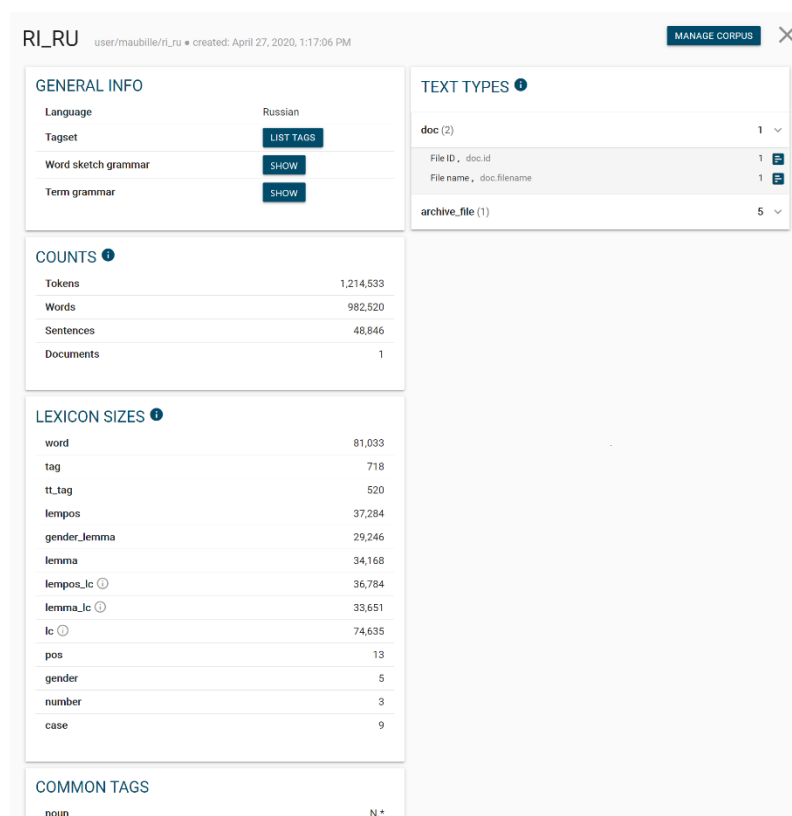


Figure 22. Corpus details and statistics

L'interface « *Manage corpus* » est le centre de contrôle du corpus (Figure 23). C'est depuis cette page qu'il est possible notamment d'alimenter le corpus, de le partager, de créer un ou plusieurs sous-corpus ou encore de supprimer le corpus.

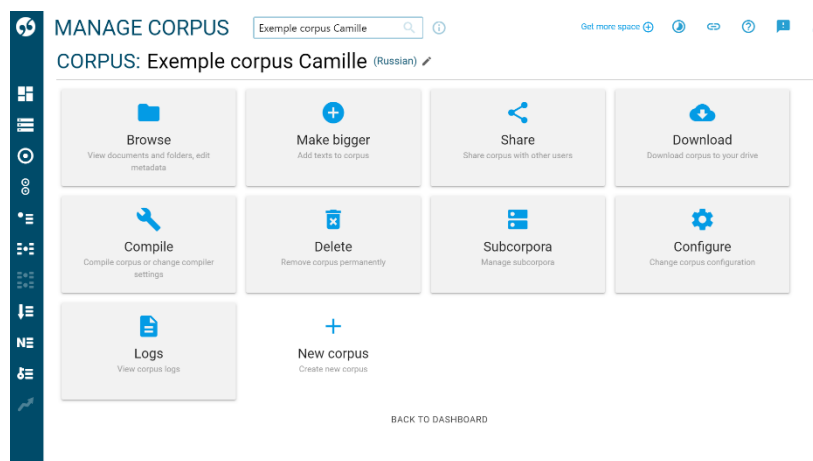


Figure 23. Interface *Manage corpus*

4.3 Interrogation de corpus dans le Sketch Engine

4.3.1 Outils d'extraction de la terminologie

Dans ce chapitre, nous détaillerons et expliquerons les différentes fonctionnalités du *Sketch Engine* utilisées dans le cadre de ce travail, à savoir d'une part les fonctionnalités *Wordlist*, *N-grams*, qui permettent d'extraire la terminologie, et, d'autre part, les fonctionnalités *Concordance* et *Collocations*, qui permettent de dégager la combinatoire.

4.3.1.1 N-Grams

Les unités polylexicales générées par la fonctionnalité *N-grams* sont composées de deux à six mots. L'utilisateur a le choix entre différentes possibilités : générer dans une seule liste toutes les unités polylexicales ; générer certains types d'unités polylexicales (à la fois les bigrammes et les trigrammes) ou générer un seul type d'unités polylexicales (uniquement les quadrigrammes) (Figure 24).

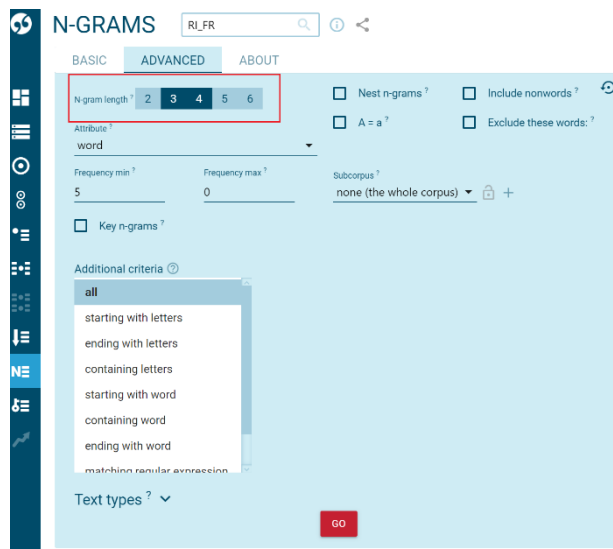


Figure 24. Paramètres de longueur

Dans « *Attribute* », il convient de sélectionner « *word* », puisque « *lemma* » ramènerait chaque mot de l'unité polylexicale à sa forme canonique, ce qui donnerait des unités dépourvues de sens telles que « droit de le homme » ou « état de urgence ». Les listes d'unités polylexicales ont été traitées par ordre décroissant, soit des 6-grammes aux bigrammes, selon les fréquences minimales suivantes :

- 6 et 5-grammes : 10 occurrences
- Quadrigrammes, trigrammes et bigrammes : 25 occurrences

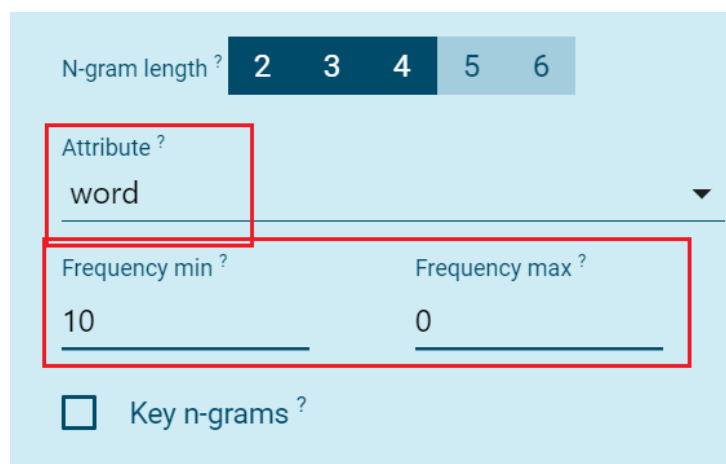


Figure 25. Choix de l'attribut et du nombre d'occurrences

4.3.1.2 Wordlist

La fonctionnalité *Wordlist* génère la liste des unités monolexicales.

Contrairement aux *N-grams*, il faut sélectionner « *lemmas* » dans « *Attribute* » (Figure 26) qui, comme son nom l'indique, ramène chaque unité monolexicale à sa forme canonique.

La fréquence minimale a été fixée à 50 occurrences.

find ?

- words
- lemmas**
- nouns
- verbs
- adjectives
- adverbs
- pronouns
- conjunctions

all

- starting with
- ending with
- containing
- matching regex
- from this list:

Exclude these words:

Include nonwords ?

A = a ?

Frequency min ? 0 Frequency max ? 0

result format

- Simple list ?**
- Display as ?

Subcorpus ?

none (the whole corpus) ▼ 🔒 +

Text types ? ▼

GO

Figure 26. Interface *Wordlist*

Lemma	Absolute Frequency ?	Lemma	Absolute Frequency ?
1 premier	2,909 ***	18 général	661 ***
2 nouveau	2,235 ***	19 jeune	655 ***
3 américain	2,017 ***	20 étranger	650 ***
4 autre	1,626 ***	21 international	647 ***
5 politique	1,619 ***	22 public	609 ***
6 européen	1,573 ***	23 seul	606 ***
7 dernier	1,573 ***	24 nombreux	577 ***
8 grand	1,332 ***	25 social	555 ***
9 ancien	1,181 ***	26 iranien	531 ***
10 même	918 ***	27 russe	528 ***
11 britannique	913 ***	28 économique	504 ***
12 militaire	902 ***	29 blanc	492 ***
13 français	808 ***	30 démocrate	472 ***
14 syrien	788 ***	31 principal	456 ***
15 national	772 ***	32 présidentiel	452 ***
16 turc	710 ***	33 prochain	451 ***

Figure 27. Exemple d'extraction d'adjectifs

La fonctionnalité *Wordlist* permet également de générer la liste de résultats sous la forme « *lemposes* » (Figure 28). Ce paramètre extrait une liste sous forme lemmatisée et apporte en outre une précision sur la classe grammaticale de chaque mot. Dans le cas d'ambiguïté grammaticale, comme pour les mots « pouvoir » ou « savoir », ce paramètre permet de lever le doute quant à la classe grammaticale. Nous constatons donc en Figure 29 que « pouvoir » (ligne 27) est le verbe, et non le substantif. La légende explicative des différentes abréviations utilisées peut être retrouvée dans l'interface « *Corpus details and statistics* » dans la section « *Lempos suffixes* » (Figure 30).

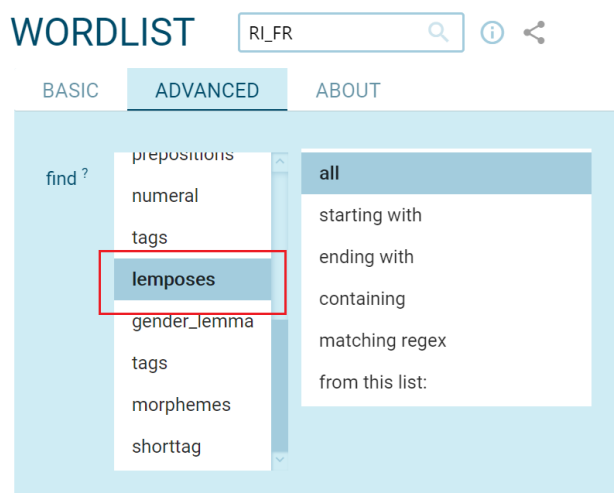


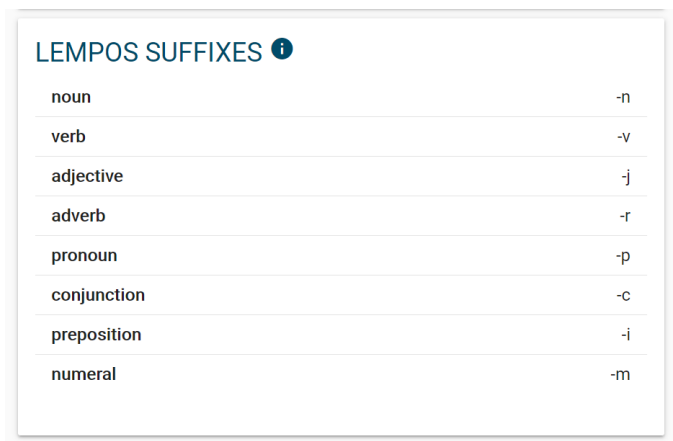
Figure 28. Interface *Wordlist* - attribut *Lempos*

WORDLIST RI_FR

lempos (28,521 items | 1,308,444 total frequency)

Lempos	Absolute Frequency ?	Lempos	Absolute Frequency ?
1 le-x	129,626 ***	18 par-i	8,802 ***
2 de-i	128,038 ***	19 sur-i	7,496 ***
3 à-i	39,208 ***	20 pas-r	6,236 ***
4 avoir-v	34,485 ***	21 plus-r	5,306 ***
5 un-x	31,585 ***	22 avec-i	5,004 ***
6 être-v	26,569 ***	23 faire-v	4,713 ***
7 et-c	21,730 ***	24 ce-p	4,263 ***
8 en-i	17,146 ***	25 mais-c	3,835 ***
9 son-x	12,676 ***	26 président-n	3,803 ***
10 dans-i	11,855 ***	27 pouvoir-v	3,284 ***
11 que-c	11,477 ***	28 leur-x	3,247 ***

Figure 29. Aperçu des résultats en *Lempos*



noun	-n
verb	-v
adjective	-j
adverb	-r
pronoun	-p
conjunction	-c
preposition	-i
numeral	-m

Figure 30. Légende des abréviations *Lempos suffixes*

4.3.2 Outils d'extraction de la combinatoire

Après avoir dégagé les termes simples et complexes à la suite de l'analyse des listes d'unités monolexicales et polylexicales, nous avons généré la combinatoire de chacun de ces termes. Pour ce faire, nous avons utilisé l'option « *Collocations* », accessible via la fonctionnalité « *Concordance* » et, dans une moindre mesure, la fonctionnalité « *Word Sketch* » et les options « *Sort* » et « *Filter* » dans « *Concordance* ».

4.3.2.1 Collocations

Accessible depuis l'interface « *Concordance* », l'option « *Collocations* » génère l'environnement linguistique immédiat d'un terme et permet donc d'identifier la combinatoire de ce même terme.

Au préalable, il faut introduire dans « *Concordance* » le terme recherché, en veillant bien, dans le cas d'un terme simple, à sélectionner « *lemma* » dans « *Query type* » (Figure 31) ainsi que la catégorie lexicale dans « *Part of Speech* » (PoS).

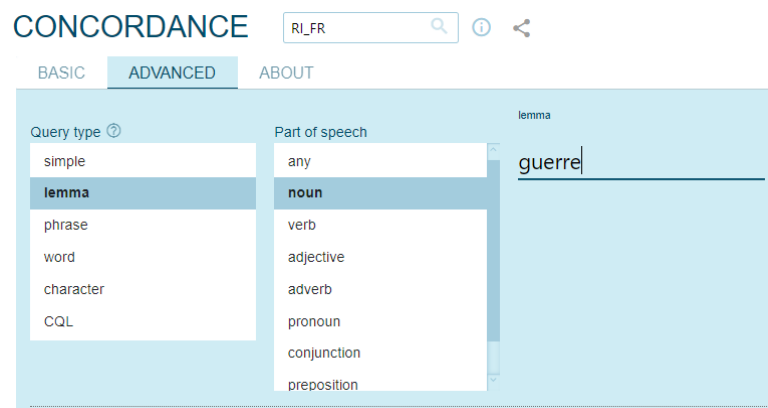


Figure 31. Interface *Concordance*

Comme on peut le voir à la Figure 32, l'option « *Collocations* », symbolisée par le pictogramme  est accessible via l'écran d'affichage des occurrences de la requête introduite (« guerre » dans notre exemple).



Left context	KWIC	Right context
raison à la fin de la	guerre	de Sécession, Richard Nixon et Bill Clinton au XXe siècle, Trump fait face semaine après semai
maru au cours des	guerres	passées, des sinistres ou des catastrophes naturelles. </s><s> Dernièrement, le château de Ku
ent aux lois de la	guerre	dans les 24 heures suivant sa mort>>, a-t-il ajouté. </s><s> Le général McKenzie a prévenu qu
uis le début de la	guerre	civile, la production a chuté de 90 % et le régime est presque entièrement dépendant des impor
pour <<crime de	guerre	>> et <<assassinats en relation avec une entreprise terroriste>>. </s><s> Il aurait été un cadre
de d'une nouvelle	guerre	civile semble peu probable, juge la journaliste: <<l'armée, et c'est historique, a protégé les mani
ois en huit ans de	guerre	, le gouvernement syrien et l'opposition ont accepté de s'asseoir à une table pour discuter de la
endant la Première	Guerre	mondiale par les troupes de l'Empire ottoman, alors allié à Allemagne et à l'Autriche-Hongrie. </
riques sur fond de	guerre	civile et de famine ayant fait des centaines de milliers de morts dans les deux camps. </s><s> <

Figure 32. Résultats de la requête « guerre »

Dans les paramètres avancés de l'option « *Collocations* », il faut sélectionner « *lemma* » dans « *Attribute* », définir la « fenêtre » de requête (jusqu'à 5 tokens³ à droite et/ou à gauche du terme) et modifier éventuellement les fréquences minimales (« *Minimum frequency in corpus* » et « *Minimum frequency in given range* »).

³ Token: the smallest unit that a corpus consists of. A token normally refers to: a word form, punctuation, digit, abbreviations, product names or anything else between spaces (Sketch Engine).

Dans le cadre de ce travail, nous avons conservé les fréquences définies par défaut du *Sketch Engine* : minimum 3 occurrences du collocatif en combinaison avec la base, et minimum 5 occurrences du collocatif dans tout le corpus.

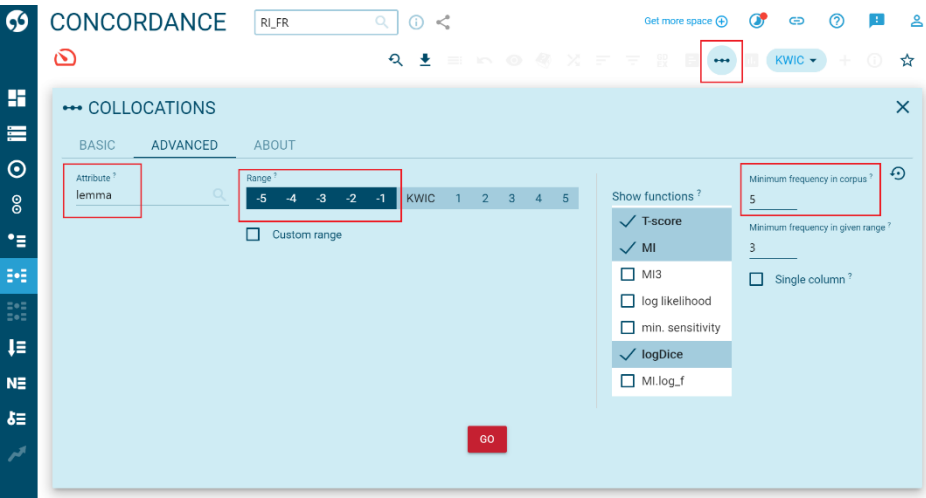


Figure 33. Interface *Collocations*

Collocations [CHANGE CRITERIA](#) [BACK TO CONCORDANCE](#)

	Lemma	Cooccurrences [?]	Candidates [?]	T-score	MI	↓ LogDice	
1	☐ civil	98	555	9.87	8.45	11.24	***
2	☐ mondial	54	270	7.33	8.63	10.73	***
3	☐ froid	33	64	5.74	10.00	10.34	***
4	☐ commercial	25	152	4.98	8.35	9.80	***
5	☐ contre	55	2,633	7.24	5.37	9.05	***
6	☐ 1975-1990	10	10	3.16	10.95	8.72	***
7	☐ irak	14	415	3.69	6.07	8.60	***
8	☐ entre	29	2,280	5.17	4.66	8.28	***
9	☐ fin	17	1,110	3.99	4.93	8.21	***
10	☐ susceptible	7	90	2.63	7.27	8.06	***
11	☐ syrien	14	1,009	3.61	4.78	8.01	***
12	☐ malouines	6	6	2.45	10.95	7.99	***
13	☐ ennemi	7	150	2.62	6.53	7.96	***
14	☐ commettre	7	160	2.62	6.44	7.95	***
15	☐ afghanistan	7	180	2.61	6.27	7.92	***
16	☐ indépendance	7	181	2.61	6.26	7.92	***

Figure 34. Résultats des collocations à droite de la base « guerre »

Collocations

CHANGE CRITERIA

BACK TO CONCORDANCE

	Lemma	Cooccurrences ?	Candidates ?	T-score	MI	↓ LogDice
1	☐ civil	98	Only this			
2	☐ mondial	54	Whole concordance but this			
3	☐ froid	33	64	5.74	10.00	10.34 ...
4	☐ commercial	25	152	4.98	8.35	9.80 ...

Figure 35. Accès à la mise en contexte

simple guerre 98 > filter civil 1,5 98 (64.2 per million) X

Left context KWIC Right context

1	☐ ⓘ doc#0 intérieure et d'en exporter le reste. </s><s> Depuis le début de la	guerre civile	, la production a chuté de 90 % et le régime est presque ent
2	☐ ⓘ doc#0 manifestants pacifistes, ces derniers jours, le risque d'une nouvelle	guerre civile	semble peu probable, juge la journaliste: <[l'armée, et c'est
3	☐ ⓘ doc#0 nocide> <s> >, évoquant des massacres réciproques sur fond de	guerre civile	et de famine ayant fait des centaines de milliers de morts c
4	☐ ⓘ doc#0 asrallah, avait aussi mis en garde vendredi contre le risque de <<	guerre civile	>> au Liban. </s><s> Trump annonce que le successeur <<i
5	☐ ⓘ doc#0 services de base, comme l'électricité 24 h/24, depuis la fin de la	guerre civile	en 1990. </s><s> Pour apaiser la colère populaire, Hassan
6	☐ ⓘ doc#0 itte contre la corruption tout en agitant le spectre d'un retour à la	guerre civile	. </s><s> Selon lui, des manifestants demanderaient la car
7	☐ ⓘ doc#0 tête quelques années plus tard. </s><s> Profitant du chaos de la	guerre civile	, ses combattants s'installent en Syrie en 2013 avant une o
8	☐ ⓘ doc#0 at d'une classe politique quasiment inchangée depuis la fin de la	guerre civile	(1975-1990). </s><s> Le président du Parlement par exem
9	☐ ⓘ doc#0 capable de répondre aux besoins des Libanais depuis la fin de la	guerre civile	en 1990. </s><s> D'autres annonces, dont la baisse des sai
10	☐ ⓘ doc#0 at d'une classe politique quasiment inchangée depuis la fin de la	guerre civile	(1975-1990). </s><s> Dans un communiqué, l'armée a dépl
11	☐ ⓘ doc#0 iforme à la fois la sépulture de plus de 33.000 combattants de la	guerre civile	- cont, estime-t-on, un tiers de républicains enterrés contre
12	☐ ⓘ doc#0 universel aux soins, ne sont pas assurés 30 ans après la fin de la	guerre civile	(1975-1990). </s><s> Commission européenne: Macron pr
13	☐ ⓘ doc#0 sés de jeunes, femmes et hommes>>. </s><s> Depuis la fin de la	guerre civile	en 1990, les infrastructures du pays sont restées en déliqu
14	☐ ⓘ doc#0 i saint, adoptés par référendum et qui, en 1998, mettent fin à une	guerre civile	ayant fait plus de 3500 morts en trente ans, le parti unionis
15	☐ ⓘ doc#0 s> La classe politique, elle, n'a quasiment pas changée depuis la	guerre civile	(1975-1990). </s><s> SOS Méditerranée cherche de nouve
16	☐ ⓘ doc#0 >s> Aucun membre de l'élite qui dirige le pays depuis la fin de la	guerre civile	en 1990 n'échappe à leurs insultes et à leur rage, cristallisé
17	☐ ⓘ doc#0 part indiqué les secours. </s><s> <<Je n'ai jamais vu ça, c'est la	guerre civile	>> déplore Andrés, sorti sur le pas de sa porte. </s><s> Il a,
18	☐ ⓘ doc#0 bre 2019 La scène résume parfaitement le nouvel épisode de la	guerre civile	syrienne. </s><s> Sur une route au nord-est du pays file un
19	☐ ⓘ doc#0 it retiré son armée et ses services de renseignement alors que la	guerre civile	syrienne commençait. </s><s> Ce mois d'octobre 2019 ma
20	☐ ⓘ doc#0 entielle et les législatives. </s><s> Cependant, l'ex-rébellion de la	guerre civile	(1975-1992) devenue le principal parti d'opposition, la Ren

Figure 36. Affichage en contexte de la recherche « guerre » + « civil »

4.3.2.2 Concordance

La fonctionnalité *Concordance* permet de rechercher un terme et de visualiser toutes ses occurrences en contexte. L'outil permet de choisir entre un affichage de type *KWIC* (*keyword in context*) ou un affichage en « sentence ».

CONCORDANCE

simple politique 2,360 (1,546,14 per million)

Left context KWIC Right context

1	☐ doc#0	et avait conseillé plusieurs personnalités conservatrices sur la	politique	étrangère. </s><s> Malgré l'absence de percées dans les négoc
2	☐ doc#0	çait comme le succès numéro un du milliardaire républicain en	politique	étrangère, patine. </s><s> Steve Biegun, qui n'a pu rencontrer ei
3	☐ doc#0	semble pas avoir été jusqu'à présent capable de proposer une	politique	autre que celle de battre Trump Les autres personnalités plus c
4	☐ doc#0	semble pas avoir été jusqu'à présent capable de proposer une	politique	autre que celle de battre Trump. </s><s> Les propositions d'auc
5	☐ doc#0	omnique Reynié, le directeur de la Fondation pour l'innovation	politique	, observe: <<De la même façon qu'il y a eu un cycle d'expansion
6	☐ doc#0	l ces partis, l'un après l'autre, se sont installés dans le paysage	politique	, il semble bien qu'ils soient depuis peu entrés dans une phase
7	☐ doc#0	constance témoignent d'un effort pour s'adapter à un paysage	politique	de plus en plus fragmenté, analyse Susi Dennison, chercheuse
8	☐ doc#0	pas, explique Marc Lazar, professeur d'histoire et de sociologie	politique	et spécialiste de l'Italie. </s><s> C'est un dilemme pour les part
9	☐ doc#0	nce des mouvements populistes: les causes sociales liées aux	politiques	d'austérité, au chômage, au sentiment d'insécurité économique
10	☐ doc#0	e, ainsi que leur style, leur langage et leur volonté de faire de la	politique	autrement les rassemblent. </s><s> Mais ce sont avant tout de
11	☐ doc#0	à formuler des promesses bien plus illusoires que le reste des	politiques	, comme l'ont montré l'Ukip ou les promoteurs les plus fervents
12	☐ doc#0	en dans le débat de l'entre-deux-tours, ou avec la grosse erreur	politique	de Salvini cet été, note Marc Lazar. </s><s> La grande question
13	☐ doc#0	cablantes, sur sa façon très particulière d'utiliser à son profit la	politique	étrangère américaine. </s><s> Visiblement nerveux, Trump ne s
14	☐ doc#0	er avec l'enquête parlementaire; il transforme le débat en sujet	politique	, multipliant les messages sur les réseaux sociaux et les chaîne
15	☐ doc#0	ous ses apparences de procédure légale, une affaire avant tout	politique	. </s><s> L'un des indicateurs importants est la perception de l'i
16	☐ doc#0	bat sur le terrain qui lui réussit le mieux: celui de l'affrontement	politique	, et la mise en scène de Trump face au < </s> <marcage> <s> :

Figure 37. Affichage KWIC

CONCORDANCE

simple politique 2,360 (1,546,14 per million)

Left context sentence Right context

1	☐ doc#0	<s> Mais ce républicain avait déjà fait partie de l'équipe de sécurité nationale du président George W. Bush au début des années 2000 et		
2	☐ doc#0	<s> Car le rapprochement avec la Corée du Nord, qui s'annonçait comme le succès numéro un du milliardaire républicain en	politique	
3	☐ doc#0	<s> Le Parti démocrate ne semble pas avoir été jusqu'à présent capable de proposer une	politique	autre que celle de battre Trump Les autres personnalités plus centristes comme Beto O'Rourke, ancien représentant du Texas, ou Kamala Harris, sénateur de Californie, n'ont pas transformé en soutien massif la curiosité médiatique qu'ils ont suscitée. </s>
4	☐ doc#0	<s> Enfin, le Parti démocrate ne semble pas avoir été jusqu'à présent capable de proposer une	politique	autre que celle de battre Trump. </s>
5	☐ doc#0	<s> Sans trancher, Dominique Reynié, le directeur de la Fondation pour l'innovation	politique	, observe: <<De la même façon qu'il y a eu un cycle d'expansion lors duquel ces partis, l'un après l'autre, se sont installés dans le paysage politique, il semble bien qu'ils soient depuis peu entrés dans une phase de stagnation.>> Quand le < </s>
6	☐ doc#0	<s> Sans trancher, Dominique Reynié, le directeur de la Fondation pour l'innovation	politique	, observe: <<De la même façon qu'il y a eu un cycle d'expansion lors duquel ces partis, l'un après l'autre, se sont installés dans le paysage politique, il semble bien qu'ils soient depuis peu entrés dans une phase de stagnation.>> Quand le < </s>
7	☐ doc#0	<s> <<Ces alliances de circonstance témoignent d'un effort pour s'adapter à un paysage	politique	de plus en plus fragmenté, analyse Susi Dennison, chercheuse au centre de réflexion European Council on Foreign Relations, mais elles ne peuvent faire reculer les mouvements populistes que de façon temporaire. </s>
8	☐ doc#0	<s> Leurs marges de progression sont celles de la crise >><<Il y a une contamination par le style des populistes, qu'ils soient au pouvoir ou pas, explique Marc Lazar, professeur d'histoire et de sociologie	politique	et spécialiste de l'Italie. </s>

Figure 38. Affichage en sentence

Lors de l'entrée de la requête, nous pouvons choisir entre différents critères de recherche.

La recherche en « simple » extrait toutes les occurrences du mot recherché, peu importe sa nature grammaticale ou ses inflexions. Si nous cherchons « politique » nous retrouvons autant le substantif que l'adjectif.

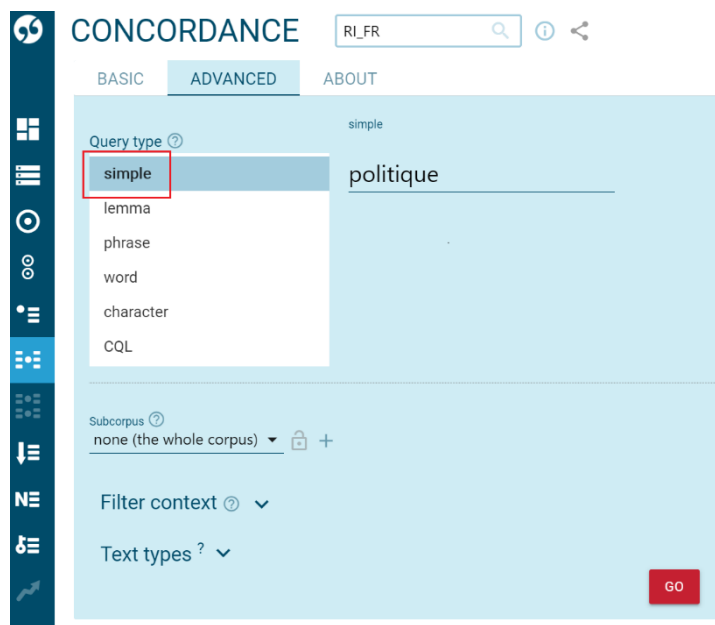


Figure 39. Recherche de concordance en « simple »

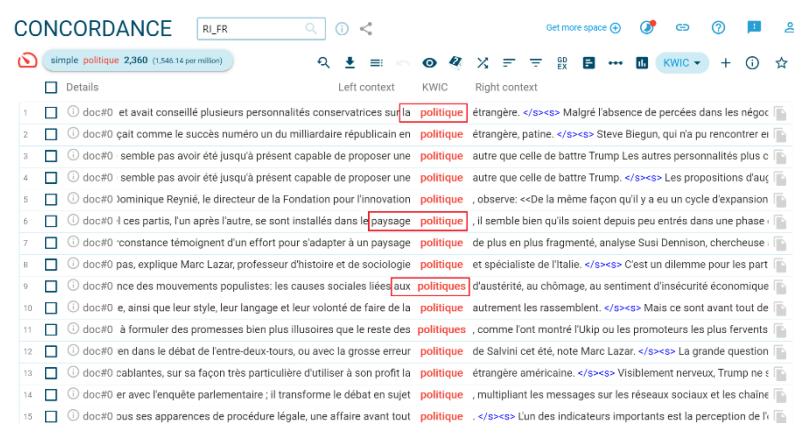


Figure 40. Résultats d'une recherche en « simple »

Effectuer ses recherches en « simple » est très utile pour le russe car ce paramètre permet d'obtenir toutes les occurrences d'un terme, peu importe la forme sous laquelle il est décliné.

	Details	Left context	KWIC	Right context
1	doc#0	эслиса к независимой позиции Кубы, к ее суверенной	политике	. </s><s> Мы рады дальнейшему укреплению кубинской
2	doc#0	исключением стала Гавана. <<В рамках линии внешней	политике	президента эти действия сдерживают получение доход
3	doc#0	не очередного транша кредита. </s><s> Американская	политике	в отношении Украины не имеет никакого смысла, да к
4	doc#0	льтернативы для Германии>> (AfD) в поддержку новой	политике	в отношении России. </s><s> Законопроект предлагал
5	doc#0	Меркель. </s><s> По его словам, в Тюрингии немецкие	политике	ждут развития диалога европейских стран с Россией, та
6	doc#0	осударства неоднократно говорили об оккупационной	политике	соседней страны. </s><s> Стоит отметить, что согласен
7	doc#0	достоинством людей>>, -- приводит слова украинского	политике	<<Интерфакс-Украина> >, </s><s> Вместе с тем глава па
8	doc#0	юдцы: объединяются в группы, чтобы продолжать <...>	политике	непризнания аннексии Крыма! </s><s> Как можно приз
9	doc#0	'ИА <<Новости> >, <<Запад оправдывает свою военную	политике	надуманной необходимостью сдерживать якобы агрес
10	doc#0	ичество в развитии вооруженных сил, согласованная	политике	в области обороны и взаимное использование военно
11	doc#0	же в планы властей входит объединение таможенной	политике	и создание единого регулятора нефтегазовой отрасли
12	doc#0	и суверенитета этого государства. </s><s> Касаемо же	политике	американской стороны в Сирии, Песков заметил, что я
13	doc#0	тава комиссии Совета Федерации по информационной	политике	Алексей Пушков в ответ на это указал, что Крым <<поте
14	doc#0	оссии. <<Это типичное вмешательство во внутреннюю	политике	России, свидетельствующее о гегемонии Запада>>, -- з
15	doc#0	думать, что мы вот таким образом выстраиваем свою	политике	на основе того, что вы сейчас сказали. </s><s> Я не гов
16	doc#0	тель комиссии Совета Федерации по информационной	политике	Алексей Пушков со своей стороны указал на странное г
17	doc#0	ем, что происходит в мировой экономике и в мировой	политике	<...> Наш крупнейший партнер в Европе -- Германия, а
18	doc#0	ведывательные службы США обслуживают нынешнюю	политике	этой страны. </s><s> Но они пока не представили ника
19	doc#0	бъединяются в группы, чтобы продолжать российскую	политике	-- политику непризнания аннексии Крыма! </s><s> Как
20	doc#0	я в группы, чтобы продолжать российскую политику --	политике	непризнания аннексии Крыма! </s><s> Как можно приз
21	doc#0	тель комитета Верховной рады по вопросам внешней	политике	и межпарламентского сотрудничества Богдан Яременк
22	doc#0	ий разговор с Трампом 31 июля. </s><s> В ходе беседы	политике	обсуждали в основном тему с лесными пожарами, буш
23	doc#0	ция в 6-м округе Парижа, съехались многие известные	политике	, в том числе -- президент России Владимир Путин. </s>
24	doc#0	де Голяя, Ширак как активист пришел во французскую	политике	в 1965 году. </s><s> Его первым правительственным по
25	doc#0	>, <<Это визовая война>>; МИД рассказал об ответе на	политике	США МИД обвинил США в развязывании визовой войн

Figure 41. Résultats d'une recherche RU en « simple »

Nous avons principalement utilisé la fonctionnalité « *Concordance* » pour chercher manuellement la présence de collocations précises dans le corpus, par exemple dans le cas où une collocation serait apparue dans une langue mais pas dans l'autre. Dans ces cas de figure, nous avons cherché manuellement si la collocation était présente sous le seuil des 3 occurrences.

Par exemple, nous n'avons pas d'équivalent en français pour la collocation « *высказывать – высказать свою кандидатуру* » (FR *présenter sa candidature*). Nous avons donc entré « candidature » dans « *Concordance* » et cherché le verbe « présenter » (Figure 42). Comme nous le constatons dans la Figure 43, la collocation « présenter sa candidature » est bien présente dans le corpus, mais sous le seuil des 3 occurrences.

CONCORDANCE RI_FR

BASIC ADVANCED ABOUT

Query type ①
 simple
lemma
 phrase
 word
 character
 CQL

Part of speech
 any
noun
 verb
 adjective
 adverb
 pronoun
 conjunction
 preposition

lemma
 candidature

Subcorpus ①
 none (the whole corpus) +

Filter context ① ^

☐ Do not filter
☒ Lemma context
☐ Part-of-speech context

Only keep lines with

all of présenter within 7 Tokens

left

Text types ② v

GO

Figure 42. Recherche d'une collocation spécifique via *Concordance*

CONCORDANCE RI_FR

Get more space +

lemma candidature 2 > context [lemma="présenter"] (-7..1) 2 (1.31 per million) x

Details Left context KWIC Right context

1	doc#0	nt d'amender la Constitution pour se présenter à sa quatrième candidature . </s><s> Il s'est également vu reprocher sa gestion des incend
2	doc#0	une a également annoncé aux médias qu'il allait présenter sa candidature . << Je ne crois pas à ces élections dans les conditions actuelle

Figure 43. Résultats de la recherche de collocations via *Concordance*

4.3.2.3 Sort

Accessible depuis l'interface *Concordance*, l'option « *Sort* » permet de trier alphabétiquement la liste de résultats obtenue en fonction des mots se trouvant à gauche ou à droite du terme recherché. Dans l'exemple ci-dessous (Figure 45), nous avons trié alphabétiquement par le premier mot à gauche du terme.

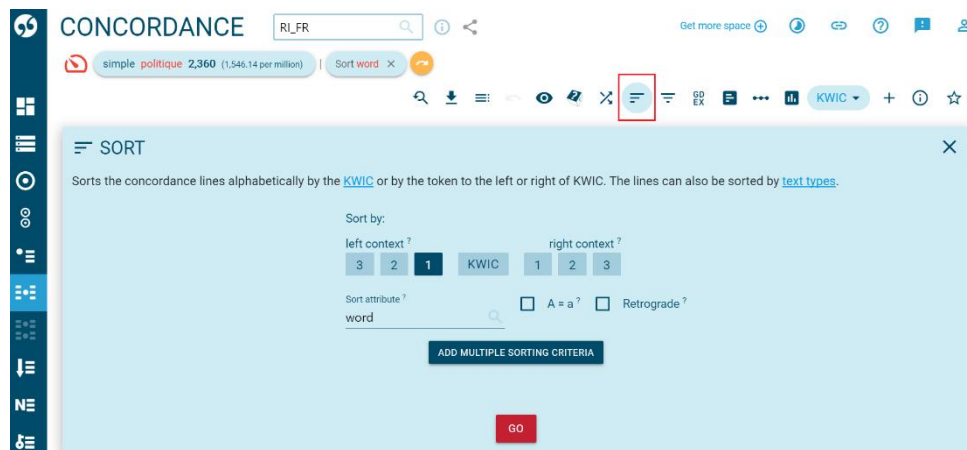


Figure 44. Interface sort

154	☐	doc#0	nénager la susceptibilité de Donald Trump et son propre avenir	politique	.	</s><s>	A 70 ans, dépourvu de la moindre ambition électorale
155	☐	doc#0	ins parlementaires de 2021 et contribueront à façonner l'avenir	politique	de la Russie, alors que le président Vladimir Poutine entame sa		
156	☐	doc#0	</s><s> Ainsi s'achève, en moins d'une semaine, une aventure	politique	inédite de sept ans. </s><s>	Sommet de l'UE: Emmanuel Macr	
157	☐	doc#0	issus de la loi>>, a déclaré Nancy Pelosi. </s><s> Une aventure	politique	risquée à laquelle s'étaient ralliés de plus en plus d'élus modéré		
158	☐	doc#0	ur les accords de paix avec la guérilla des Farc que le balancier	politique	colombien semble se réorienter. </s><s>	Alvaro Uribe a admis l	
159	☐	doc#0	e subordonner la sécurité nationale à des objectifs basement	politiques	</s><s>	Il dénonce sans relâche la corruption. <<La simple idé	
160	☐	doc#0	iment que les démocrates veulent les priver d'une vraie bataille	politique	électorale en déclenchant l'< </s><s> >>? </s><s>	Ce sera l'un ou	
161	☐	doc#0	comme l'appela John Kerry, se fit donc sur fond d'après bataille	politique	, et de compromis arraché aux forceps par les Américains. </s>		
162	☐	doc#0	finalement, peu désireuse de se retrouver mêlée à une bataille	politique	américaine. </s><s>	Le 1er septembre, le chargé d'affaires des	
163	☐	doc#0	our la paix, il faut d'abord s'unir face à eux et gagner la bataille	politique	</s><s>	On continuera à se battre le temps qu'il faudra. </s><!	
164	☐	doc#0	n de plus de 3 millions de réfugiés. </s><s>	politique	hors norme, leur renvoi présente en outre l'avantage de modifie		
165	☐	doc#0	rendum n'ayant jamais réuni de majorité. </s><s>	politique	persiste, la perspective d'élections anticipées reviendra en forc		
166	☐	doc#0	l de [sa] famille et l'intérêt national>>. </s><s>	politique	complet entre le gouvernement et le parlement, Boris Johnson		
167	☐	doc#0	iban. </s><s>	politique	avaient été suivies d'une poussée de fièvre entre Israël et le He:		
168	☐	doc#0	de mécontentement des électeurs de gauche après le blocage	politique	auquel ils ont assisté, souligne Pablo Simon, professeur de scie		
169	☐	doc#0	précédent scrutin, en 2014, à une nouvelle situation de blocage	politique	, de quoi éloigner encore davantage une population lasse de di		
170	☐	doc#0	it meilleures lorsque les dirigeants n'étaient pas du même bord	politique	. </s><s>	Le < </s> <bulldozer> <s> > avait su tirer tout le bénéf	
171	☐	doc#0	ieux placés pour battre ceux du pouvoir, quel que soit leur bord	politique	. </s><s>	Dans les urnes, ce vote s'est traduit dimanche 8 sept	
172	☐	doc#0	étudiants, se souvient-elle, n'arrivaient pas à identifier son bord	politique	, car il critiquait tous les partis au pouvoir quelle que soit leur cr		
173	☐	doc#0	li, il s'est dit certain d'obtenir un << fort soutien des deux bords	politiques	>>, remontés par l'inaction de l'exécutif. </s><s>	La proposition	
174	☐	doc#0	it pas une majorité constitutionnelle, qui lui ouvre un boulevard	politique	à la hongroise. </s><s>	Derrière le PIS, la Coalition civile (KO), c	
175	☐	doc#0	e pour réclamer le changement>>. </s><s>	politique	ouvre une ère d'incertitude. </s><s>	Le gouvernement qu'il dirig	
176	☐	doc#0	Stefan Jagsch à se faire élire. << Celui qui n'a pas de boussole	politique	et morale et prend une décision électorale aussi irresponsable l		
177	☐	doc#0	lations extérieures au Conseil démocratique syrien, la branche	politique	des FDS. </s><s>	Une interprétation a minima qui contredit une	
178	☐	doc#0	upplétifs syriens, jamais avarés en atrocités, le PYP, la branche	politique	du proto-Etat kurde, a consenti au retour des pro-Damas. </s><!		
179	☐	doc#0	sime et la crise économique s'est engouffrée dans les brèches	politiques	héritées de l'ère Chirac. </s><s>	Après avoir accompagné l'enfa	
180	☐	doc#0	lace significative aux élus des territoires >>. </s><s>	politique	, plus large, sera lui renouvelé après les municipales, a-t-il préc		
181	☐	doc#0	à l'égard du Parti et de Xi, qui règne sans partage sur le Bureau	politique	depuis un Congrès triomphal, en 2017. </s><s>	L'application a l	
182	☐	doc#0	qui sont prêts à se servir de l'Ukraine pour atteindre leurs buts	politiques	étroits. >> </s><s>	Donald Trump : << La procédure d'impeachn	
183	☐	doc#0	ntéressé ou obligé à une quelconque feuille de route par calcul	politique	>>, promet-il en développant son programme: changement de r		
184	☐	doc#0	e fois accusé les démocrates d'agiter ce spectre par pur calcul	politique	. <<Ils ne savent pas comment m'arrêter>> avant l'élection prési		

Figure 45. Résultats triés à gauche

4.3.2.4 Filter

L'option « *Filter* » permet de filtrer la liste de résultats en fonction d'une certaine combinaison de mots. Nous pouvons par exemple demander de conserver

uniquement les lignes contenant le terme « paysage » afin d'analyser comment ce terme se comporte en combinaison avec « politique ». À l'inverse, il est également possible de demander uniquement les combinaisons ne contenant pas un certain terme.

Dans l'exemple ci-dessous (Figure 46), nous demandons d'obtenir uniquement les résultats contenant le mot « paysage » dans le contexte gauche de « politique ».

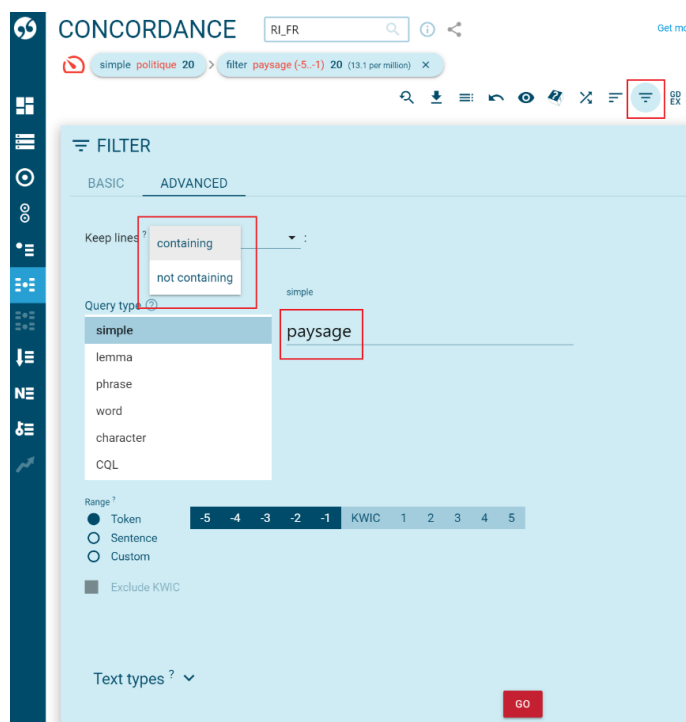


Figure 46. Interface *filter*

	Details	Left context	KWIC	Right context
1	☐	doc#0 et ces partis, l'un après l'autre, se sont installés dans le	paysage politique	il semble bien qu'ils soient depuis peu entrés dans une phase d
2	☐	doc#0 l'ironie témoignant d'un effort pour s'adapter à un	paysage politique	de plus en plus fragmenté, analyse Susi Dennison, chercheuse a
3	☐	doc#0 un journaliste espagnol qui lui demandait de parler du	paysage politique	argentin, Juan Perón avait répondu : « Regardez en Argentine, il
4	☐	doc#0 s>< Depuis cette date, il a complètement disparu du	paysage politique	</s><s> Figure de ce qu'on appelait le courant « barbe FLN », l
5	☐	doc#0 s>< Il se faisait fort de recomposer complètement le	paysage politique	de l'hémicycle ; son groupe Renew n'est qu'un groupe libéral élar
6	☐	doc#0 Etoiles (M5), qui pourrait transformer durablement le	paysage politique	</s><s> Le texte, qui avait l'appui des principaux partis, proposi
7	☐	doc#0 gardes rouges. </s><s> Ces derniers disparaissent du	paysage politique	et Mao reprend la tête du pays. </s><s> Il y reste jusqu'à sa mo
8	☐	doc#0 ssi évidente que lors de la présidentielle de 2014. <Le	paysage politique	tunisien est aujourd'hui très fragmenté. </s><s> Les candidats it
9	☐	doc#0 rangs dispersés>>, observe Sophie Bessis. </s><s> Le	paysage politique	de 2019 s'éparpille désormais en plusieurs pôles : islamistes, laïc
10	☐	doc#0 ment absorber l'extrême droite et ses électeurs dans le	paysage politique	? </s><s> La question se pose depuis vingt ans en France avec l
11	☐	doc#0 Depuis 2013, le parti s'implante de plus en plus dans le	paysage politique	allemand, et aujourd'hui ses résultats fragilisent la coalition de li
12	☐	doc#0 l'échelle de la Suisse - connue pour la stabilité de son	paysage politique	, c'est un tremblement de terre. </s><s> Les formations écologi
13	☐	doc#0 démocrates. </s><s> Si les écologistes ont bousculé le	paysage politique	il semble toutefois difficile « de renverser cette formule >>, pré
14	☐	doc#0 utés au Parlement. </s><s> Ces résultats dessinent un	paysage politique	radicalement différent de celui en place ces quatre dernières an
15	☐	doc#0 té sur le Brexit, leur formation perde du terrain dans un	paysage politique	ultra-polarisé avec, d'un côté, un parti conservateur devenu le <
16	☐	doc#0 ire mixte. </s><s> Cette échéance devrait redessiner le	paysage politique	au vu du rejet essuyé par les partis traditionnels. </s><s> La for
17	☐	doc#0 te historique de la formation la plus proeuropéenne du	paysage politique	italien, le Parti démocrate (centre gauche). </s><s> Par la suite, l
18	☐	doc#0 nédit qu'annonciateur d'une restructuration radicale du	paysage politique	tunisien. « Sombre tableau, écrit la chroniqueuse Synda Tajine,
19	☐	doc#0 s Le double scrutin de lundi a confirmé l'éclatement du	paysage politique	</s><s> Alors que les travaillistes enregistrent une défaite hist
20	☐	doc#0 juant à une fragmentation de plus en plus marquée du	paysage politique	</s><s> Les élections norvégiennes ne font pas exception. </s>

Figure 47. Résultats filtrés avec le terme « paysage »

4.3.2.5 Word Sketch

L'outil « *Word Sketch* » est la fonctionnalité qui a donné son nom au *Sketch Engine* (Kilgarriff et al., 2015). Le *Word Sketch* est un résumé d'une page de l'environnement d'un terme. Cet outil est très utile pour analyser le comportement grammatical et combinatoire d'un terme. Les résultats sont triés en différentes colonnes : nous retrouvons par exemple les colonnes verbe + objet, sujet + verbe, substantif + modificateur, etc.

Lors de la recherche, il convient d'entrer sa requête sous forme lemmatisée. Ensuite, nous avons le choix entre laisser le *Sketch Engine* déterminer automatiquement la nature grammaticale du terme, ou sélectionner nous-même la classe grammaticale.

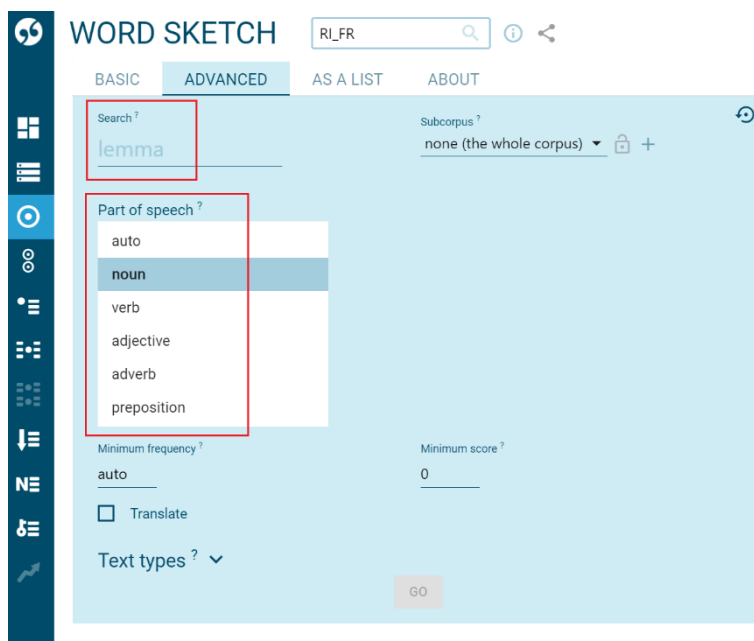


Figure 48. Interface du *Word Sketch*

Pour la requête ci-dessous (Figures 49 et 50), nous avons introduit le terme « politique » et précisé que nous souhaitons la forme substantive, et non adjectivale.

WORD SKETCH RI_FR

politique as noun 740x

verbs with "politique" as object	verbs with "politique" as subject	modifiers of "politique"	pronominal possessors of "politique"	usage patterns
mener mener une politique	être la politique est	étranger de la politique étrangère	son de sa politique	with_definite_article with_indefinite_article
définir l'excentricité qui définissaient la politique britannique ont disparu		migratoire la politique migratoire	leur leur politique	
infléchir infléchir sa politique		intérieur la politique intérieure		
durcir durci sa politique		industriel la politique industrielle, le		
promouvoir promouvoir une politique		britannique la vie politique britannique		
renforcer renforcer la politique		rival enquêter sur son rival politique Joe Biden		
défendre défend une politique		responsable responsable politique		
devenir l'avortement est devenue une actrice politique de poids et		général son discours de politique générale		
être est une politique		extérieur politique extérieure		
		italien de la politique italienne		
		prisonnier comme un « prisonnier politique		
		agricole la politique agricole commune		

Figure 49. Résultats du *Word Sketch* pour le substantif « politique »

prepositions preceding noun/nouns after preposition	prepositional phrases	"politique" and/or ...	nouns modified by noun "politique"
sur sur la politique	"politique" de	marché la politique industrielle, le marché unique numérique	vie la vie politique britannique
contre contre la politique	"politique" à	prise domaine de la politique et de la prise de décision	classe de la classe politique
en en politique	"politique" en	défense politique industrielle, la défense	scène de la scène politique
avec avec la politique	"politique" dans		
de de la politique	"politique" avec		
par de la justice par le politique			
dans dans la politique			
entre entre le politique et			
pour haut représentant pour la politique étrangère			
à à la politique			

Figure 50. Résultats du *Word Sketch* pour le substantif « politique »

Grâce à cet outil, nous apercevons en quelques secondes que l'on peut « mener une politique », « renforcer une politique », « infléchir sa politique » ou encore qu'il existe la « politique étrangère », la « politique migratoire », ...

Cette fonctionnalité offre des résultats moins complets que les listes fournies par « Collocations », néanmoins, il s'agit de l'outil le plus facile d'utilisation et il convient

aux étudiants et aux traducteurs qui ont besoin d'un aperçu rapide et clair du comportement d'un terme.

5 Analyse des corpus

5.1 Extraction de la terminologie

Comme nous l'avons expliqué au point [4.3.1 Outils d'extraction de la terminologie](#), les fonctionnalités utilisées pour extraire les unités polylexicales et les unités monolexicales sont respectivement *N-Grams* et *Wordlist*.

Nous avons procédé de la manière suivante :

- Extraction, par ordre décroissant, des 6-grammes, 5-grammes, quadrigrammes, trigrammes, bigrammes et des unités monolexicales selon les seuils définis aux points [4.3.1.1](#) et [4.3.1.2](#) ;
- Sauvegarde des listes au format *Excel* : un fichier par liste ;
- Extraction, à partir de ces listes, des termes simples et des termes complexes.

5.1.1 Unités monolexicales

Dans ce sous-chapitre, nous détaillons le processus de tri des unités monolexicales et expliquons les raisons pour lesquelles certaines unités ont été écartées. Les mots grammaticaux, les mots relevant de la langue générale et les mots relevant d'un autre domaine de spécialité ont notamment été rejetés. Nous expliquons également comment nous avons levé le doute dans le cas de mots présentant une ambiguïté grammaticale ou sémantique.

5.1.1.1 Mot lexical – mot grammatical

Lors de l'analyse des unités monolexicales, nous avons d'abord distingué les mots lexicaux grammaticaux des mots grammaticaux, également appelés respectivement mots pleins et mots vides (Bertin, 2009). Les mots lexicaux sont les noms, les adjectifs qualificatifs, les verbes et les adverbes (Cordial). Les mots grammaticaux sont les articles, les adjectifs non qualificatifs, les conjonctions (coordination et subordination) et les prépositions (Cordial). Comme l'illustrent les Figures 51 et 52, les mots grammaticaux figurent parmi les unités monolexicales les

plus fréquentes, tant en français qu'en russe. En effet, les mots grammaticaux sont très fréquents dans les textes (L'Homme, 2004 : 149). Cependant, utilisés seuls, ils ne sont pas retenus par le terminologue, car ils ne constituent pas des unités significatives. En revanche, ils entrent, pour certains d'entre eux, dans la composition de termes complexes qui, eux, retiennent toute l'attention du terminologue.

En français, les 20 mots grammaticaux du corpus les plus fréquents sont : *le, de, à, avoir, un, être, et, en, que, ce, son, dans, pour, se, qui, ne, il, par, sur* et *pas*. En russe, les 20 premiers mots grammaticaux sont : *в, и, на, что, с, не, быть, он, по, тот, который, это, этот, о, к, как, они, для, из* et *свой*.

WORDLIST   

lemma (2,398 items | 1,165,117 total frequency)

Lemma	Absolute Frequency ?	Lemma	Absolute Frequency ?
1 le	132,485 ***	11 son	12,682 ***
2 de	128,530 ***	12 dans	11,855 ***
3 à	39,208 ***	13 pour	11,133 ***
4 avoir	34,494 ***	14 se	10,059 ***
5 un	32,400 ***	15 qui	9,892 ***
6 être	26,623 ***	16 ne	9,412 ***
7 et	21,730 ***	17 il	9,082 ***
8 en	18,668 ***	18 par	8,802 ***
9 que	14,877 ***	19 sur	7,496 ***
10 ce	13,653 ***	20 pas	6,338 ***

Figure 51. Mots grammaticaux les plus fréquents du corpus FR

WORDLIST   

lemma (2,426 items | 833,674 total frequency)

Lemma	Absolute Frequency ?	Lemma	Absolute Frequency ?
1 в	44,082 ***	11 который	6,956 ***
2 и	24,925 ***	12 это	6,699 ***
3 на	17,085 ***	13 год	5,600 ***
4 что	15,166 ***	14 этот	5,482 ***
5 с	12,446 ***	15	5,438 ***
6 не	12,221 ***	16 о	5,430 ***
7 быть	10,156 ***	17 президент	5,104 ***
8 он	9,573 ***	18 к	5,024 ***
9 по	9,238 ***	19 сша	5,000 ***
10 тот	7,334 ***	20 как	4,991 ***

Figure 52. Mots grammaticaux les plus fréquents du corpus RU

Une fois les mots grammaticaux écartés, nous nous sommes penchée sur les mots lexicaux. Parmi ceux-ci, nous avons d'emblée rejeté les adjectifs et les adverbes, qui entrent pour la plupart dans la composition de termes complexes, pour ne retenir que les verbes et les substantifs. Dès lors, de nombreux adjectifs sont présents dans la composition de termes complexes. Par exemple, l'adjectif « international » / « международный » entre dans la composition des bigrammes *relations internationales* / *международные отношения*, *scène internationale* / *международная арена* ou encore *coalition internationale* / *международная коалиция*. L'adjectif « humanitaire » / « гуманитарный » se retrouve dans les bigrammes *crise humanitaire* / *гуманитарный кризис* et *mission humanitaire* / *гуманитарная миссия*. Citons encore, par exemple, « diplomatique » / « дипломатический » qui se retrouve dans *relations diplomatiques* / *дипломатические отношения*, *valise diplomatique* / *дипломатическая почта* ou encore *immunité diplomatique* / *дипломатический иммунитет*.

5.1.1.2 Mots relevant de la langue générale

Après avoir écarté les adverbes et les adjectifs, nous avons exclu les mots lexicaux relevant de la langue générale. Les 20 mots lexicaux les plus fréquents dans le corpus FR, hors adjectifs et adverbes, sont : *faire*, *pouvoir*, *président*, *pays*, *octobre*, *an*, *politique*, *ministre*, *septembre*, *gouvernement*, *devoir*, *parti*, *accord*, *chef*, *force*, *élection*, *jour*, *aller*, *vouloir* et *mettre*. En russe, les 20 premiers mots lexicaux sont : *год*, *президент*, *США*, *Россия*, *страна*, *Украина*, *мочь*, *заявить*, *глава*, *лидер*, *время*, *отношение*, *партия*, *слово*, *сторона*, *дело*, *вопрос*, *власть*, *человек* et *статья*.

Dans cette liste, les verbes *faire*, *pouvoir*, *devoir*, *aller*, *vouloir* et *mettre* n'ont pas été retenus, puisqu'il ne s'agit pas de termes propres aux domaines des relations internationales et de la politique. Toutefois, certains de ces verbes constituent des collocs qui, en combinaison avec différents termes, forment par exemple les collocations suivantes : ***mettre*** en accusation, ***faire*** une déclaration ou encore ***aller*** en prison.

Il en va de même pour le russe, où les verbes *мочь*, *заявить* et *статья*, qui relèvent de la langue générale, n'ont donc pas été conservés en tant que tels. Cependant,

certaines verbes entrent eux-aussi dans la combinatoire de différents termes du lexique et sont donc présents sous la forme de collocations telles que **стать целью атак, стать мишенью, объявлять – объявить войну, взять штурмом, взять в руки оружие.**

Tandis que *président, pays, politique, ministre, gouvernement, parti, accord* et *élection* ont été retenus car il s'agit de termes relevant des domaines à l'étude, *octobre, an, septembre* et *jour*, qui appartiennent à la langue générale, ont été supprimés.

Le nombre très élevé d'occurrences de « septembre » (2 288 occurrences) et « octobre » (2 687 occurrences) s'explique par le fait que les articles constitutifs du corpus FR ont été publiés en septembre ou en octobre (voir point [3.1.9 Date de publication](#)).

CONCORDANCE RI_FR

simple **septembre** 2,288 (1,496.97 per million)

Details Left context KWIC Right context

1	doc#0	amaïne dernière, en Ombrie, en est un exemple. </s><s> Début septembre , la présidente du nouvel exécutif européen, Ursula von der Ley
2	doc#0	lement par <<l'immigration multipliée par trois depuis le début septembre et la réouverture des ports>>, même si la menace pour les terri
3	doc#0	>><s> Après un léger accès de faiblesse dans les sondages en septembre , elle remonte depuis quelques semaines. </s><s> Selon l'enqu
4	doc#0	its>>, a ajouté Nancy Pelosi sur un ton solennel. </s><s> Le 24 septembre , Nancy Pelosi avait décidé d'engager son parti sur la voie périll
5	doc#0	chef des démocrates au Congrès Nancy Pelosi. </s><s> Le 24 septembre , elle avait décidé d'engager son parti sur la voie périlleuse de l'
6	doc#0	</s><s> John Bolton, limogé par le président républicain le 10 septembre dernier, pourrait être auditionné la semaine prochaine, a précis
7	doc#0	chef des démocrates au Congrès Nancy Pelosi. </s><s> Le 24 septembre , elle avait décidé d'engager son parti sur la voie périlleuse de l'
8	doc#0	ukrainien Volodymyr Zelensky au coeur de l'été. </s><s> Le 24 septembre , elle avait décidé d'engager son parti sur la voie périlleuse de l'
9	doc#0	Wong de s'être rendu au Congrès américain à Washington en septembre , pour appuyer un projet de loi soutenant la cause démocratiqu
10	doc#0	alvanisant l'opposition démocrate. </s><s> Le retrait formel en septembre du projet de loi controversé qui aurait placé les Hongkongais à

Figure 53. Mise en contexte des occurrences de « septembre »

Pour le russe, les termes *президент, страна, лидер, партия, сторона* et *власть*, qui relèvent des domaines d'étude, ont été conservés. Cependant, les termes relevant de la langue générale tels que *год, время, слово, вопрос* et *человек* ont été écartés.

Après une analyse attentive des lignes de concordance et du contexte, le nombre d'occurrences, étonnamment élevé, des mots *время* (2 322 occ.) et *слово* (2 047 occ.) peut être facilement expliqué. En effet, *время* (FR *temps*) entre dans dans la composition de locutions temporelles fréquemment utilisées (liste non-exhaustive) :

- *в ближайшее время* (113 occ) : *bientôt, sous peu, prochainement* ;

- *во время* (568 occ.) : *pendant, durant* ;
- *долгое время* (75 occ.) : *longtemps* ;
- *в то же время* (252 occ.) : *en même temps* ;
- *за время* (106 occ.) : *pendant, durant* ;
- *в настоящее время* (190 occ.) : *actuellement*.

Il est intéressant de remarquer que, là où le russe utilise principalement des locutions adverbiales, le français privilégie l'utilisation d'adverbes.

Au regard de *слово* (FR *mot*), on le retrouve également dans la composition de locutions fréquemment utilisées (liste non-exhaustive) :

- *по [кого-то] словам* (1 310 occ.) : *selon* ;
- *к слову* (234 occ.) : *d'ailleurs, au fait*.

456	☐	① doc#0 « выборам в Белоруссии. </s><s> Теперь они, по его	словам	, приезжают без заранее готового решения, что <<в	
457	☐	① doc#0 президента России Дмитрий Песков. </s><s> По его	словам	, непонятно, как его будут согласовывать, если Киев	
458	☐	① doc#0 анизации процесса интеграции. </s><s> Это, по его	словам	, ведет к экономическим потерям его государства. <	
459	☐	① doc#0 цую президента Трампа книгу и продолжает, по его	словам	, работать в администрации США. </s><s> Вы работ	
460	☐	① doc#0 эрил генсек блока Йенс Столтенберг. </s><s> По его	словам	, страны-участницы Североатлантического альянса	
461	☐	① doc#0 разил надежду на диалог с Путиным. </s><s> По его	словам	, они с российским президентом могут найти взаим	
462	☐	① doc#0 знить политику в отношении Крыма. </s><s> По его	словам	, необходимо налаживать <<человеческие>> отнош	
463	☐	① doc#0 иностранных дел РФ Сергей Лавров. </s><s> По его	словам	, возможная встреча президентов Путина и Зеленск	
464	☐	① doc#0 дующей поставки соседним странам. </s><s> По его	словам	, Белоруссия может покупать нефть у Ирана, США, С	
465	☐	① doc#0 зние дипломатического иммунитета. </s><s> По его	словам	, дипломаты не совершали никаких нарушений. <<F	
466	☐	① doc#0 дать контроль над Крымом Украине. </s><s> По его	словам	, в дипломатическом ведомстве отнеслись к подобн	

Figure 54. Aperçu des occurrences de la locution *по кого-то словам*

5.1.1.3 Mots simples entrant dans la composition de termes complexes

Remarquons que, en russe comme en français, *chef* et *глава* n'ont pas été conservés en tant que termes simples. Cependant, ils entrent dans la composition de termes complexes tels que *chef d'État* / *глава государства*, *chef du gouvernement* / *глава правительства* ou encore *chef de l'exécutif* / *глава исполнительной власти*.

Il en va de même pour *organisation* / *организация*, *opération* / *операция* ou encore *communauté* / *сообщество* qui n'ont pas été conservés en tant que termes simples, mais qui apparaissent toutefois dans *organisation terroriste* / *террористическая*

организация, opération militaire / военная операция et communauté internationale / международное сообщество, мировое сообщество.

5.1.1.4 Mots relevant d'un autre domaine de spécialité

Tout mot ne relevant pas des domaines de la politique, des relations internationales et des conflits armés a été écarté. Parmi les mots lexicaux supprimés, un grand nombre d'entre eux relevait, entre autres, des domaines de l'économie, de l'environnement ou encore de la communication. Nous avons donc supprimé des termes tels que *presse (490 occ.) / пресса (173 occ.), journaliste (394 occ.) / журналист (544 occ.), interview (77 occ.) / интервью (289 occ.) ; banque (101 occ.) / банк (94 occ.), économie (256 occ.) / экономика (376 occ.), grève (93 occ.) / забастовка (40 occ.), financement (69 occ.) / финансирование (128 occ.), commerce (110 occ.) / торговля (161 occ.), croissance (86 occ.) / рост (171 occ.), libre-échange (35 occ.) / свободная торговля (37 occ.) ; climat (143 occ.) / климат (58 occ.), environnement (115 occ.) / окружающая среда (42 occ.).*

5.1.1.5 Cas particulier : le Brexit

Trier la liste des termes définitifs n'a pas toujours été une tâche aisée, et il nous est arrivé de changer plusieurs fois d'avis au cours de l'élaboration du lexique. Tel fut notamment le cas avec les termes découlant du Brexit, évènement au cœur de l'actualité internationale lors de la période de compilation des articles. Initialement, il avait été décidé de proposer une rubrique « Brexit », dans la partie « Politique », reprenant le vocabulaire propre à la sortie du Royaume-Uni de l'Union européenne dont *report du Brexit / перенос Брекзит, Brexit sans accord / Брекзит без сделки, reporter le Brexit / переносить – перенести Брекзит.*

Toutefois, nous avons changé d'avis en cours de route et avons finalement décidé de ne pas inclure les entrées relatives au Brexit, ce vocabulaire étant trop lié à l'actualité et à un évènement précis.

5.1.1.6 Ambiguïté grammaticale

Il est intéressant de noter l'ambiguïté grammaticale des termes « politique » et « pouvoir » qui apparaissent respectivement 2 359 fois et 3 864 fois. Le nombre d'occurrences reprend à la fois les formes nominale et adjectivale de « politique » et les formes nominale et verbale de « pouvoir ». Afin de connaître le nombre d'occurrences réel de ces deux mots en tant que substantifs, nous pouvons utiliser la fonctionnalité « *Concordance* » et rechercher le terme en précisant la catégorie grammaticale (PoS).

Dès lors, nous remarquons (Figure 56) que « politique » en tant que substantif apparaît 740 fois sur les 2 359 occurrences.

Notons toutefois l'erreur d'étiquetage en lignes 14 et 15 de la Figure 56, où « politique » est adjectif et ne devrait dès lors pas apparaître dans les résultats de la requête. Après avoir analysé le contexte, nous pouvons émettre l'hypothèse que l'étiqueteur a commis cette erreur à cause de la juxtaposition des adjectifs « allemande » et « britannique » à droite de « politique ».

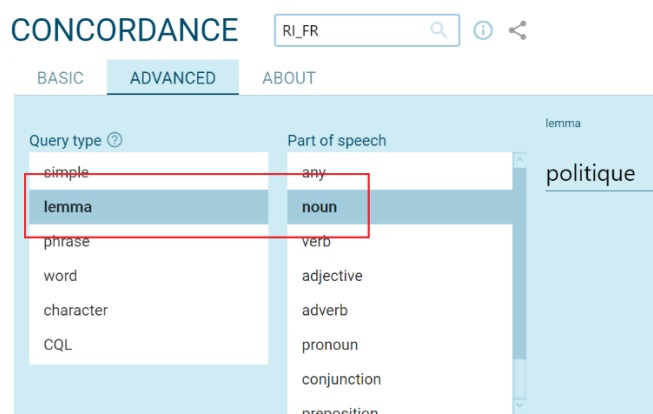


Figure 55. Interface *Concordance*

	Details	Left context	KWIC	Right context
1	☐ ① doc#0	et avait conseillé plusieurs personnalités conservatrices sur la	politique	étrangère. </s><s> Malgré l'absence de percées dans les négoc
2	☐ ① doc#0	çait comme le succès numéro un du milliardaire républicain en	politique	étrangère, patine. </s><s> Steve Biegun, qui n'a pu rencontrer ei
3	☐ ① doc#0	semble pas avoir été jusqu'à présent capable de proposer une	politique	autre que celle de battre Trump Les autres personnalités plus c
4	☐ ① doc#0	semble pas avoir été jusqu'à présent capable de proposer une	politique	autre que celle de battre Trump. </s><s> Les propositions d'auç
5	☐ ① doc#0	e, ainsi que leur style, leur langage et leur volonté de faire de la	politique	autrement les rassemblent. </s><s> Mais ce sont avant tout de
6	☐ ① doc#0	cablantes, sur sa façon très particulière d'utiliser à son profit la	politique	étrangère américaine. </s><s> Visiblement nerveux, Trump ne s
7	☐ ① doc#0	ous ses apparences de procédure légale, une affaire avant tout	politique	. </s><s> L'un des indicateurs importants est la perception de l'i
8	☐ ① doc#0	choqué par l'irruption des réseaux parallèles de Trump dans la	politique	étrangère de son pays, le colonel bardé de médailles, lui-même
9	☐ ① doc#0	is la salle sécurisée où était auditionnée une responsable de la	politique	ukrainienne au Pentagone. </s><s> Pendant plusieurs heures, li
10	☐ ① doc#0	le président et les parlementaires démocrates pour parler de la	politique	syrienne, après le retrait décidé par Trump du Rojava, l'ambianc
11	☐ ① doc#0	itter, ne se préoccupant guère du suivi ni de la cohérence de sa	politique	, sans équipes ni entourage, qui navigue à vue, en fonction de s
12	☐ ① doc#0	la marge du fonctionnement institutionnel. </s><s> Avec cette	politique	de cordon sanitaire, ils poussent la droite < </s> <illibérale> <s>
13	☐ ① doc#0	à des circonstances historiques particulières et repose sur une	politique	spécifique, ne peut pas être transféré à d'autres pays. </s><s> L
14	☐ ① doc#0	it parvenu en l'espace de quelques scrutins à déstabiliser la vie	politique	allemande, qui était structurée depuis des décennies autour de
15	☐ ① doc#0	Live résumé en vidéo ce feuilleton du Brexit qui a dominé la vie	politique	britannique et européenne depuis plus de trois ans. </s><s> Po
16	☐ ① doc#0	l'amille a des racines roumaines. </s><s> Il s'engage très tôt en	politique	, dès l'université, où il étudie les sciences politiques. </s><s> Cr
17	☐ ① doc#0	différente> <s> > dans un pays de tradition conservatrice, où la	politique	s'exerce en famille. </s><s> En témoignent ces adversaires dar
18	☐ ① doc#0	jour financer ses études. </s><s> Elle s'est très tôt engagée en	politique	avec le mouvement étudiant de la <<septima papelera>> (septi
19	☐ ① doc#0	it de travail est <<adhésion à la protestation sociale contre les	politiques	économiques et sociales qui nous affectent en tant que travail
20	☐ ① doc#0	rêt de travail est l'adhésion à la protestation sociale contre les	politiques	économiques et sociales qui nous affectent en tant que travail

Figure 56. Résultats de la recherche du terme « politique » sous forme nominale

5.1.1.7 Ambiguïté sémantique

Dans certains cas, l'ambiguïté ne relevait pas de la classe grammaticale, mais du sens du mot. Dans ces cas de figure, nous avons procédé à une analyse attentive du contexte de chaque occurrence afin de lever toute ambiguïté. Citons par exemple le cas de « mobilisation » (202 occ.). Initialement conservé dans le chapitre « guerre » dans le sens de « *[m]ise sur le pied de guerre d'une partie ou de l'ensemble des forces armées d'un pays.* » (TLFi), il s'est avéré, après analyse du contexte, qu'aucune des occurrences ne revêtait cette signification et l'entrée « mobilisation » a, de ce fait, été supprimée du lexique.

5.1.1.8 Noms de pays

Les noms de pays ont été écartés des listes. La liste des pays en annexe du lexique a été mise au point sur la base d'une liste issue d'un mémoire précédent et a été étoffée selon les besoins de ce travail. Nous avons notamment choisi d'ajouter les capitales en raison des difficultés que peut engendrer la translittération du russe vers le français et du français vers le russe. Les pays membres de la CEI ont également été ajoutés à la liste (FR *Biélorussie, Ukraine, Moldova, Arménie, Azerbaïdjan, Géorgie, Turkménistan, Kazakhstan, Ouzbékistan, Tadjikistan et Kirghizstan* ; RU

Беларусь, Украина, Молдова, Армения, Азербайджан, Грузия, Туркменистан, Казахстан, Узбекистан, Таджикистан, Кыргызстан).

Pour le français, les noms de pays ont été vérifiés et ajoutés selon la graphie en usage reprise dans le *Code de rédaction interinstitutionnel*. Les notes et remarques d'usage qui figurent en bas de page sont également issues du *Code de rédaction interinstitutionnel* (Figure 57).

Pour le russe, les noms de pays proviennent de la base de données terminologique *UNTerm*. Les notes et remarques d'usage proviennent elles aussi d'*UNTerm*, le *Code de rédaction interinstitutionnel* ne reprenant pas la langue russe parmi ses langues proposées.

Étant donnée la dimension pédagogique du lexique, qui s'adresse à des apprenants en traduction et en interprétation, nous avons jugé utile de fournir des cartes représentant chaque région du monde reprise dans le lexique. Des connaissances approfondies en géographie étant indispensables dans le cadre des relations internationales, nous avons ajouté des chiffres correspondant aux pays de la liste sur chaque carte (Figure 58). Les régions du monde proposées dans l'annexe sont : *l'Europe, le Moyen-Orient, l'Asie, l'Amérique du Nord, l'Amérique centrale, l'Amérique du Sud, et l'Afrique.*

Des sigles représentant l'UE, l'OTAN, l'ONU et la CEI (Figure 59) ont également été ajoutés à côté de chaque pays membre.

Annexe : Pays, capitales et gentils

16	Дания	Копенгаген	датчанин, датчанка	le Danemark	Copenhague	Danois, e
17	Германия	Берлин	немец, немка	l'Allemagne	Berlin	Allemand, e
18	Польша	Варшава	поляк, полька	la Pologne	Varsovie	Polonais, e
19	Чехия	Прага	чех, чешка	la République tchèque	Prague	Tchèque
20	Словакия	Братислава	словак, словачка	la Slovaquie	Bratislava	Slovaque
21	Австрия	Вена	австриец, австрийка	l'Autriche	Vienne	Autrichien, ne
22	Венгрия	Будапешт	венгр, венгерка	la Hongrie	Budapest	Hongrois, e
23	Словения	Любляна	словенец, словенка	la Slovénie	Ljubljana	Slovène
24	Хорватия	Загреб	хорват, хорватка	la Croatie	Zagreb	Croate
25	Босния и Герцеговина	Сараево	босниец, боснийка	la Bosnie-Herzégovine	Sarajevo	de Bosnie-Herzégovine ²
26	Черногория	Подгорица	черногорец, черногорка	le Monténégro	Podgorica	Monténégrin, e
27	Албания	Тирана	албанец, албанка	l'Albanie	Tirana	Albanais, e

¹ Ne pas confondre avec « finnois », qui est l'adjectif relatif à la langue (Code de rédaction interinstitutionnel).² Ne pas utiliser « Bosnien » ni « Bosniaque » (Code de rédaction interinstitutionnel).

Figure 57. Aperçu liste et remarques

L'Europe – Европа

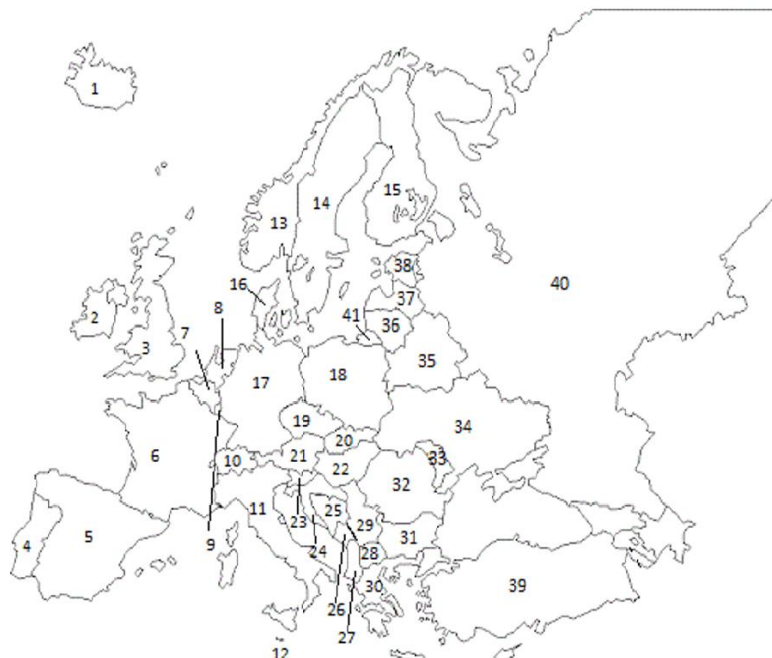


Figure 58. Aperçu carte

Légende :

-  : pays membres de l'UE en 2020
-  : pays membres de l'OTAN en 2020
-  : pays membres de l'ONU en 2020
-  : pays membres de la CEI en 2020

Figure 59. Légende des sigles utilisés

5.1.2 Unités polylexicales

Comme indiqué au point [4.3.1 Outils d'extraction de la terminologie](#), les unités polylexicales ont été extraites à l'aide de la fonctionnalité *N-Grams* du *Sketch Engine*.

Parmi les unités polylexicales générées par la fonctionnalité *N-Grams*, nous avons extrait principalement des termes complexes, mais aussi de la combinatoire, comme l'illustre la Figure 60 (ligne 253 « sortir de l'impasse »).

N-GRAMS    Get more space     

4-grams  (Items: 2,067, total frequency: 41,292)     

Word	↓ Count ?	Word	↓ Count ?
1 droits de l' homme	208 ...	251 Il n' est pas	32 ...
2 des droits de l'	198 ...	252 Afrique de l' Ouest	32 ...
3 forces de l' ordre	184 ...	253 sortir de l' impasse	31 ...
4 nord de la Syrie	174 ...	254 n' ont pas été	31 ...
5 de l' Union européenne	173 ...	255 le retour d' une	31 ...
6 la Chambre des communes	170 ...	256 le lanceur d' alerte	31 ...
7 la Chambre des représentants	169 ...	257 la création d' une	31 ...
8 nord-est de la Syrie	163 ...	258 la chaîne de télévision	31 ...
9 chef de l' Etat	155 ...	259 l' est de la	31 ...

Figure 60. Extrait de la liste des 4-grammes

Proportionnellement, la fonctionnalité *N-Grams* a permis de mettre au jour 3 % (Figure 61) des collocations FR présentes dans le lexique de niveau avancé (36 avec *N-Grams* et 1 026 avec *Collocations*).

Nous obtenons exactement la même proportion pour les collocations russes (Figure 62) dont 31 ont été mises au jour avec la fonctionnalité *N-Grams* et 1 117 avec la fonctionnalité *Collocations*.

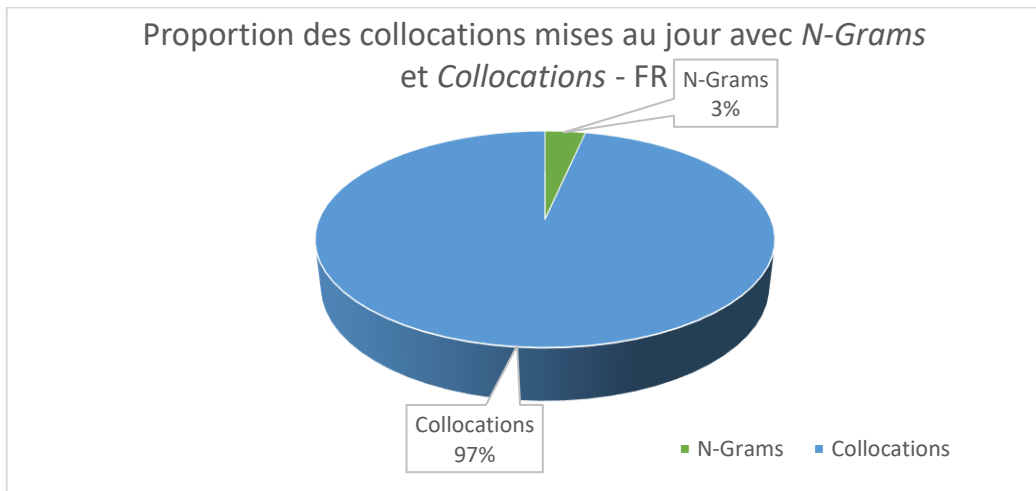


Figure 61. Proportion des collocations mises au jour avec *N-Grams* et *Collocations* - FR

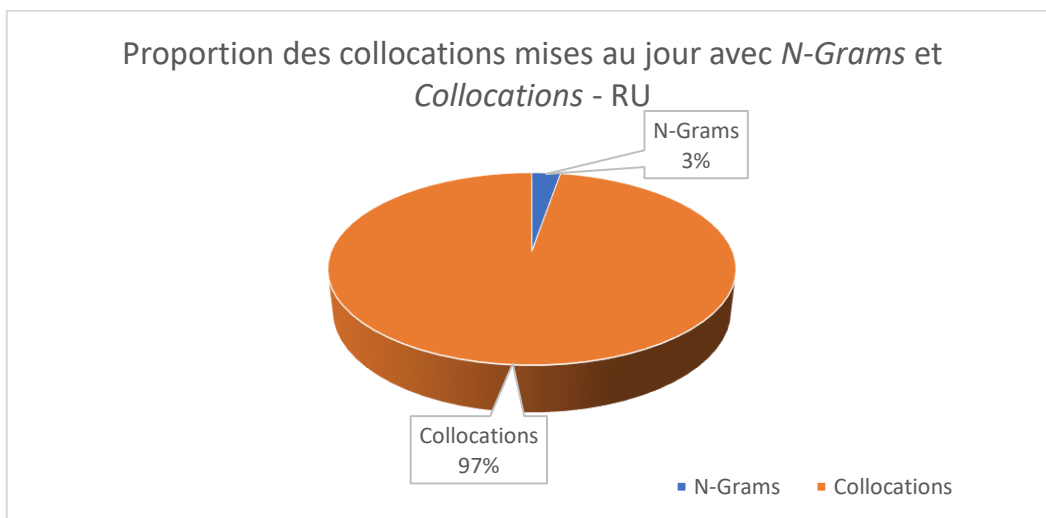


Figure 62. Proportion des collocations mises au jour avec *N-Grams* et *Collocations* – RU

Nous avons procédé à un tri des listes par ordre décroissant, c'est-à-dire des 6-grammes, 5-grammes, quadrigrammes, trigrammes, bigrammes aux unités monolexicales. Cette manière de procéder a permis de mettre au jour les termes complexes et de rejeter les mots simples entrant dans la composition de termes complexes, comme expliqué au point [5.1.1.3 Mots simples entrant dans la composition de termes complexes](#).

5.1.2.1 Premier nettoyage

Un premier tri a d'abord été effectué afin d'éliminer le bruit et les unités dépourvues de sens. En effet, la fonctionnalité *N-Grams* extrait des ensembles de mots purement en fonction de leur fréquence. C'est donc à l'utilisateur de déterminer quels résultats sont pertinents, ou non, et d'ainsi dégager une liste de termes complexes. Comme nous le constatons en Figure 63, sur les 50 premiers trigrammes extraits, seul *Chambre des représentants* (ligne 46) constitue un terme complexe composé de trois mots.

3-grams (Items: 1,815, total frequency: 99,260)

Word	↓ Count ?	Word	↓ Count ?	Word	↓ Count ?
1 n' est pas	675 ...	18 la Maison Blanche	250 ...	35 de l' État	212 ...
2 n' a pas	549 ...	19 président de la	243 ...	36 dans le nord	211 ...
3 le premier ministre	488 ...	20 n' ont pas	234 ...	37 de la Commission	209 ...
4 de la Syrie	450 ...	21 de l' ONU	233 ...	38 qu' il a	207 ...
5 il y a	438 ...	22 lors d' une	231 ...	39 nord de la	207 ...
6 chef de l'	380 ...	23 l' accord de	231 ...	40 à la tête	199 ...
7 la Chambre des	376 ...	24 de l' homme	228 ...	41 dans un communiqué	197 ...
8 l' Union européenne	344 ...	25 Donald Trump a	227 ...	42 dans le cadre	197 ...
9 de l' Etat	341 ...	26 dans la région	226 ...	43 la première fois	192 ...
10 de Donald Trump	318 ...	27 Le premier ministre	226 ...	44 forces de l'	191 ...
11 de l' UE	302 ...	28 de l' opposition	225 ...	45 qu' il n'	190 ...
12 le chef de	299 ...	29 et de la	224 ...	46 Chambre des représentants	190 ...
13 de l' Union	271 ...	30 de l' accord	222 ...	47 von der Leyen	189 ...
14 de l' armée	270 ...	31 de l' ordre	221 ...	48 la fin de	188 ...
15 l' un des	267 ...	32 le président américain	219 ...	49 de la ville	188 ...
16 droits de l'	261 ...	33 de la Chambre	216 ...	50 de la frontière	188 ...
17 Il y a	259 ...	34 des droits de	214 ...		

Figure 63. Aperçu des 50 premiers trigrammes

5.1.2.2 Mots complexes relevant d'un autre domaine de spécialité

Lors du second tri, les mots complexes relevant d'un autre domaine de spécialité, notamment des domaines de l'économie et de la communication, ont été rejetés de la liste. Nous avons rejeté des termes tels que *FR homme d'affaires* (101 occ.) et *chaîne de télévision* (59 occ.).

5.1.2.3 Forme tronquée de termes complexes

Les formes tronquées de termes complexes ont été rejetées de la liste. C'est le cas notamment des trigrammes FR *gouvernement d'union* (55 occ) et *président du Conseil* (90 occ.) qui sont en réalité les formes tronquées des termes complexes *gouvernement d'union nationale* et *président du Conseil européen*.

5.1.2.4 Noms propres

Parmi les unités polylexicales, les noms de journaux (*Washington Post*, *New York Times*, *Financial Times*, *Los Angeles Times*, *South China Morning Post*), d'agences de presse (*AFP / Agence France-Presse*, *Reuters*, *PIA Новосту*) et de partis politiques (*AfD – DE*, *Front national – FR*, *Labour – UK*, *PS – BE*, *Слуга народа – UA...*) ont été rejetés de la liste.

5.1.2.5 Bigrammes

5.1.2.5.1 Noms propres

En raison des difficultés que peut poser la translittération du russe vers le français et du français vers le russe, les noms et prénoms ont été conservés. Nous nous sommes toutefois limitée aux noms et prénoms des chefs d'État et des premiers ministres en exercice au moment de la rédaction de ce travail. Dès lors, les noms de dirigeants devront faire l'objet d'une mise à jour au fil des années si les lexiques sont utilisés en salle de classe.

Le tableau inséré ci-dessous reprend les 10 dirigeants les plus mentionnés par les presses russe et francophone. Si Donald Trump figure en première place, tant dans la presse russe que dans la presse francophone, nous constatons une divergence dès la seconde place. En effet, les 2^e et 3^e places sont occupées par Boris Johnson et Emmanuel Macron pour le français, mais sont occupées par Volodymyr Zelensky et Vladimir Poutine pour le russe. Ces résultats concordent avec les Figures présentées au point [3.2.10 Événements survenus pendant la période de compilation](#), où l'Europe occidentale prédominait dans l'échantillon du corpus francophone, tandis que l'Europe de l'Est prédominait dans l'échantillon du corpus russe.

Remarquons que le président de la Biélorussie, Alexandre Loukachenko, est la 8^e figure politique (311 occ.) la plus mentionnée dans la presse russe mais est absente du corpus francophone.

	Presse francophone		Presse russe	
	Nom du dirigeant	Occurrences	Nom du dirigeant	Occurrences
1	Donald Trump	2 239	Дональд Трамп	3 534
2	Boris Johnson	1 205	Владимир Зеленский	2 103
3	Emmanuel Macron	528	Владимир Путин	1 517
4	Recep Tayyip Erdogan	370	Реджеп Тайип Эрдоган	454
5	Benjamin Netanyahu	358	Биньямин Нетаньяху	395
6	Vladimir Poutine	333	Борис Джонсон	384
7	Volodymyr Zelensky	258	Эммануэль Макрон	345
8	Angela Merkel	159	Александр Лукашенко	311
9	Jair Bolsonaro	114	Ангела Меркель	229
10	Justin Trudeau	111	Николас Мадуро	98

5.1.2.5.2 *Bigrammes relevant de la langue générale*

Les bigrammes relevant de la langue générale ont été rejetés de la liste. Nous avons dès lors rejeté des bigrammes tels que *réseaux sociaux* (185 occ.) / *социальные сети* (48 occ.) et *conversation téléphonique* (58 occ.) – *appel téléphonique* (41 occ.) / *телефонный разговор* (256 occ.).

5.1.2.5.3 *Bigrammes relevant d'un autre domaine de spécialité*

De nombreux bigrammes relevaient d'autres domaines de spécialité, notamment des domaines de l'économie et de l'écologie, et ont donc été supprimés. Nous

avons rejeté des bigrammes tels que *crise économique* (46 occ.) / *экономический кризис* (18 occ.), *свободная торговля* (37 occ.) (FR *libre-échange*), *marché intérieur* (51 occ.) / *внутренний рынок* (9 occ.), *grève générale* (33 occ.) / *всеобщая забастовка* (5 occ.) ; *réchauffement climatique* (25 occ.) / *глобальное потепление* (11 occ.).

5.2 Mise en regard des listes RU et FR

Une fois les listes de termes simples et complexes obtenues pour le français, puis pour le russe, nous les avons mises en regard afin de trouver des équivalents dans chaque langue. Cette étape a été cruciale étant donné que, en cas d'incompatibilité des listes francophone et russe, l'élaboration du lexique aurait échoué.

À la suite de la mise en regard des listes francophone et russe, 79 % des termes (Figure 64) ont trouvé d'emblée un équivalent. Au sein des 21 % de termes n'ayant pas trouvé d'équivalent (Figure 65), 97 % d'entre eux ont trouvé un équivalent sous les seuils définis, et seulement 3 % des termes n'ont pas trouvé d'équivalent (FR *pays fondateur*, *balistique* et *table ronde*). Ce taux de pertes très faible nous permet d'affirmer, d'une part, que les corpus ont été correctement compilés et étaient équivalents et, d'autre part, que les seuils ont été judicieusement définis, près de 80 % des entrées ayant trouvé un équivalent d'entrée de jeu.

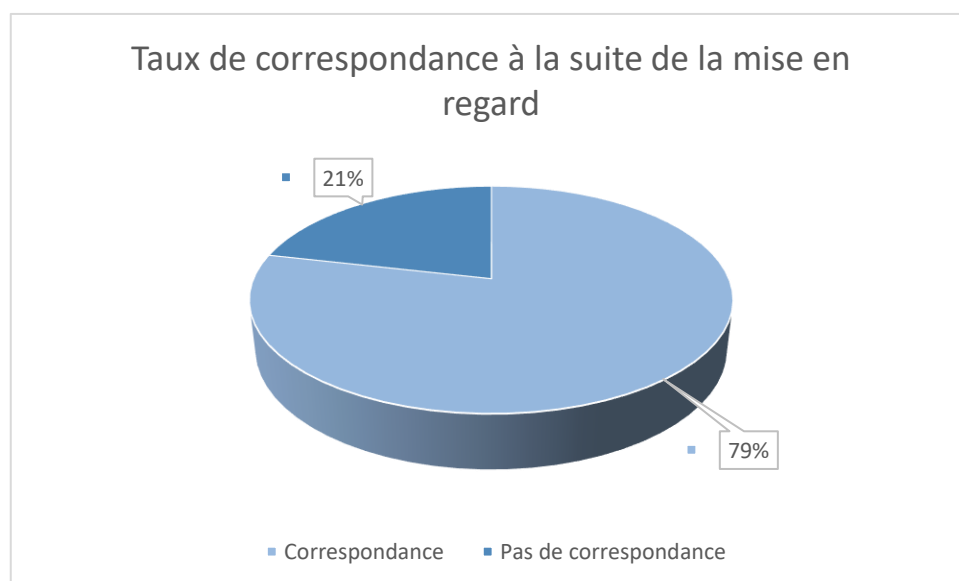


Figure 64. Taux de correspondance à la suite de la mise en regard

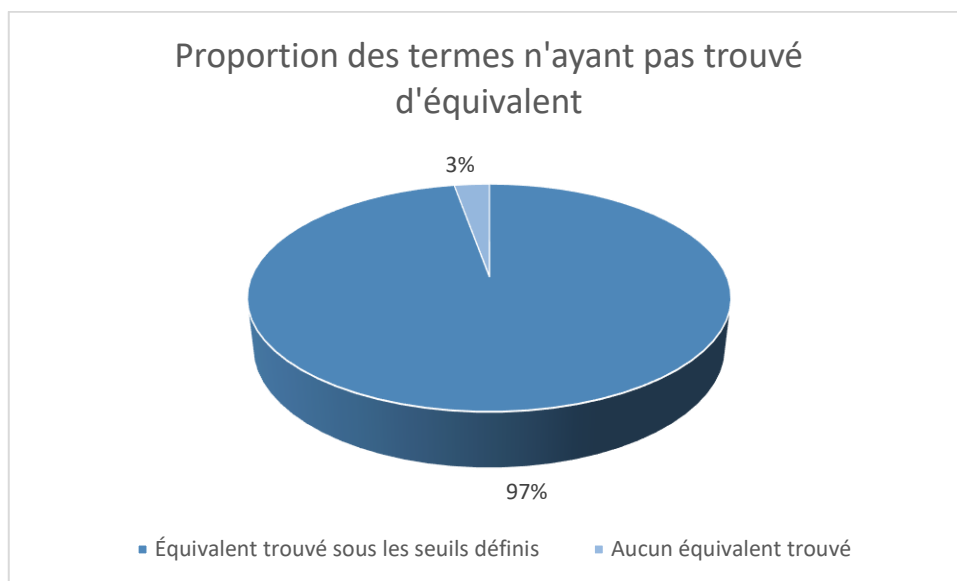


Figure 65. Proportion des termes n'ayant pas trouvé d'équivalent

Sur l'entièreté du lexique, 87 % des entrées russes (Figure 66) ont été mises au jour selon les seuils définis et 13 % des entrées russes ont donc été trouvées sous ces seuils. En ce qui concerne le français (Figure 67), 93 % des entrées ont été mises au jour selon les seuils définis et 7 % des entrées ont été trouvées sous ces seuils.

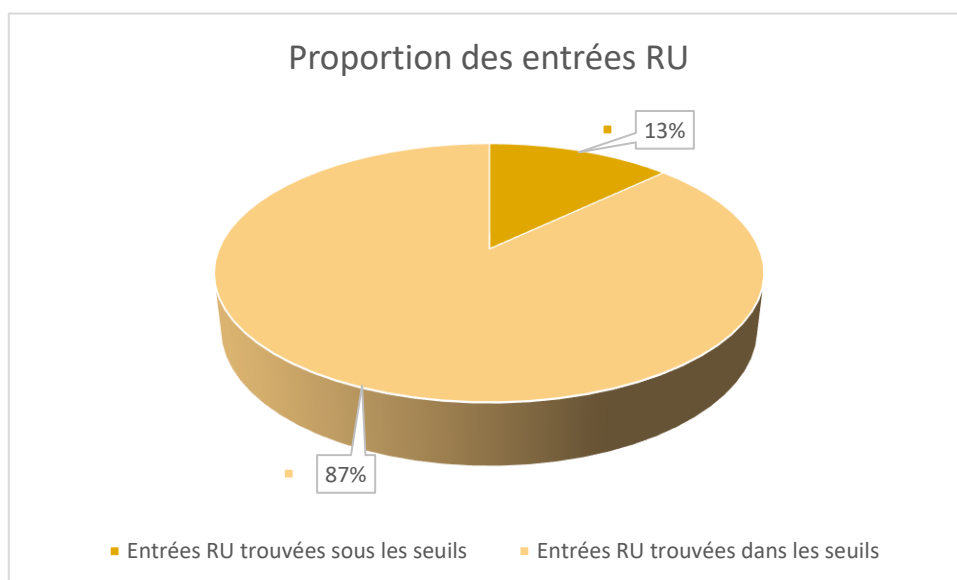


Figure 66. Proportion des entrées RU

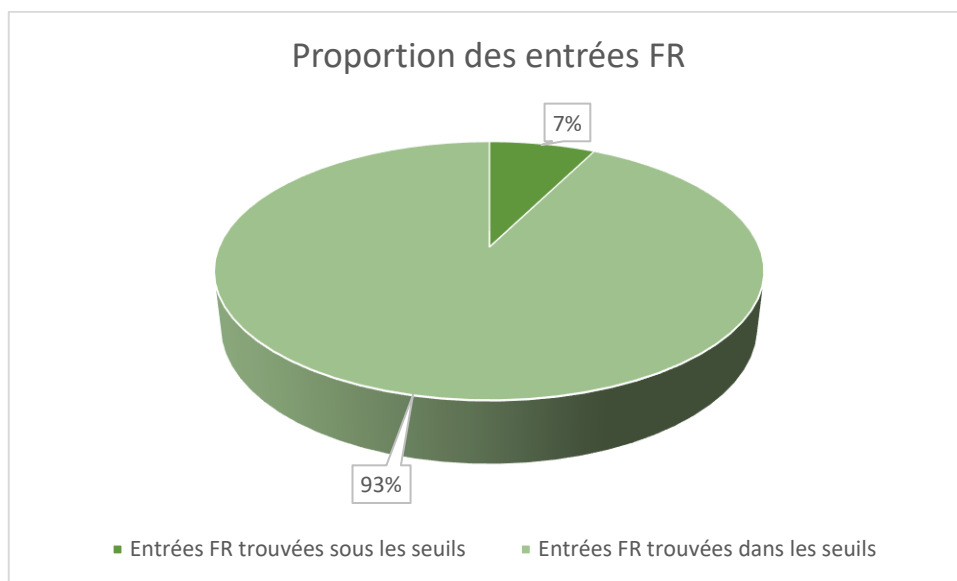


Figure 67. Proportion des entrées FR

Nous pouvons également noter que le lexique contient 10 % de termes russes en plus que de termes français. Cela peut être expliqué par le fait que, dans plusieurs cas, une entrée FR correspond à plus d'une entrée RU (Figures 68 et 69).

вступление в должность президента	l'investiture présidentielle
дебаты, дискуссия, обсуждение, спор	un débat
<i>бурный спор</i>	<i>un débat houleux</i>

Figure 68. Aperçu de l'entrée *débat*

мандат, срок, пребывание на посту	le mandat
<i>президентский мандат / президентский срок</i>	<i>le mandat présidentiel</i>

Figure 69. Aperçu de l'entrée *mandat*

5.3 Comparaison et analyse de fréquence des termes simples

Ci-dessous, le tableau inséré reprend les 10 termes simples les plus fréquents en russe et en français. Les cases de même couleur représentent les termes russes et français correspondants. Nous pouvons constater que 60 % de ces 10 premiers

termes se recoupent, et que les deux premières places sont occupées, pour le russe comme pour le français, par les termes *président* / *президент* et *pays* / *страна*.

	Corpus FR		Corpus RU	
	Termes simples les plus fréquents	Occurrences	Termes simples les plus fréquents	Occurrences
1	président	3 803	президент	5 104
2	pays	2 827	страна	4 294
3	ministre	2 333	лидер	2 346
4	gouvernement	2 253	партия	2 064
5	parti	2 079	власть	1 902
6	accord	1 893	правительство	1 581
7	élection	1 521	представитель	1 491
8	Parlement	1 158	выборы	1 454
9	État	1 053	государство	1 454
10	sécurité	974	политика	1 381

5.3.1 Déséquilibres

Nous constatons un taux d'occurrences très élevé pour les termes russes *лидер* (2 346 occ.) (FR *leader*) et *представитель* (1 491 occ.) (FR *représentant*). *Leader* et *représentant* n'apparaissant respectivement que 301 et 786 fois en français, nous

avons décidé d'analyser leurs occurrences en contexte afin d'éventuellement fournir une explication à ce déséquilibre.

Concernant *leader*, qui peut faire référence à un dirigeant de pays ou à un leader politique, en russe comme en français, nous avons constaté que ce terme est très fréquemment utilisé en russe en tant que synonyme de *président*. Cette utilisation, bien que présente dans le corpus francophone, est nettement moins utilisée. Dans le corpus russe, le terme *лидер* est principalement utilisé pour faire référence aux présidents des États-Unis (270 occ.), de l'Ukraine (242 occ.) et de la Russie (228 occ.) (Figure 70).

CONCORDANCE RI_RU [Get more space](#)

simple лидер 2,346 (1,931.61 per million)

Collocations [CHANGE CRITERIA](#) [BACK TO CONCORDANCE](#)

	Lemma	Cooccurrences ?	Candidates ?	T-score	MI	↓ LogDice	
1	☐ американский	270	2,221	16.17	5.98	10.92	***
2	☐ украинский	242	1,969	15.31	5.99	10.84	***
3	☐ российский	228	2,720	14.75	5.44	10.53	***
4	☐ встреча	97	1,133	9.63	5.47	9.84	***

Figure 70. Aperçu des principaux collocatifs de *лидер*

Au regard du terme *представитель*, nous avons relevé deux éléments pouvant expliquer ce déséquilibre.

D'une part, le terme *представитель* entre dans la composition du terme complexe *Палата представителей* (FR *Chambre des représentants*), qui compte pour 319 occurrences (Figure 71) dans le corpus russe contre 199 occurrences dans le corpus francophone.

CONCORDANCE RI_RU [Get more space](#)

simple представитель 1,491 (1,227.63 per million)

Collocations [CHANGE CRITERIA](#) [BACK TO CONCORDANCE](#)

	Lemma	↓ Cooccurrences ?	Candidates ?	T-score	MI	LogDice
1	.	468	56,668	18.42	2.75	8.04 ***
2	,	446	83,187	16.28	2.13	7.43 ***
3	палата	319	710	17.81	8.52	12.21 ***

Figure 71. Aperçu du nombre d'occurrences du collocatif *палата*

D'autre part, nous retrouvons *представитель* dans le trigramme *официальный представитель МИД* (Figure 72) et sous les formes complètes *официальный представитель Министерства иностранных дел* et *официальный представитель внешнеполитического ведомства*. Littéralement, ces termes se traduiraient par *représentant officiel du ministère des Affaires étrangères*, or, en français, on parlera de *porte-parole du ministère*, et non de *représentant* (Figure 73). Nous retrouvons le même cas de figure pour *представитель Кремля* (35 occ.) (Figure 74) qui, en français, sera désigné par *porte-parole du Kremlin* (Figure 75), et non *représentant du Kremlin*.

75	☐	doc#0	игланий. </s><s> Об этом сообщила официальный представитель МИД России Мария Захарова, слова которой приводи
76	☐	doc#0	оролества и его экономике. </s> <s> Официальный представитель МИД России Мария Захарова также оценила приняты
77	☐	doc#0	«Рос»», заявила в четверг, 3 октября, официальный представитель МИД России Мария Захарова. </s><s> По ее словам, и
78	☐	doc#0	«> Демократическую партию, заявила официальный представитель МИД России Мария Захарова. </s><s> По ее словам, у
79	☐	doc#0	ийской делегации. 18:34 Как отмечала официальный представитель МИД, отказ в визах россиянам станет «центральной
80	☐	doc#0	ребят капитально сломалось>>. 17:22 Официальный представитель МИД России Мария Захарова назвала действия амер
81	☐	doc#0	цию стенограммы уже отреагировала официальный представитель МИД России Мария Захарова. <<Эта такая работа у д
82	☐	doc#0	говора Трампа и Зеленского оценила официальный представитель МИД РФ Мария Захарова. <<Прочитала стенограмму
83	☐	doc#0	эти утверждения неверными. </s> <s> Официальный представитель МИД РФ Мария Захарова заверила, что все документ
84	☐	doc#0	. </s><s> В то же время, как сообщила официальный представитель МИД РФ Мария Захарова, документы подавались все
85	☐	doc#0	роки. </s><s> При этом, как отметила официальный представитель МИД РФ Мария Захарова, все документы подавались
86	☐	doc#0	эфакс слова Рябкова. 13:50 По словам официального представителя МИД, отказ в визах россиянам станет «центральной
87	☐	doc#0	эти утверждения неверными. </s> <s> Официальный представитель МИД РФ Мария Захарова заверила, что все документ
88	☐	doc#0	юведено не было. </s><s> Кроме того, официальный представитель МИД РФ Мария Захарова заявила, что в России нико
89	☐	doc#0	ержке вылета атташе США 03.11.2019 Официальный представитель МИД РФ Мария Захарова назвала «фейком» > и <<п
90	☐	doc#0	</s><s> В середине прошлого месяца официальный представитель МИД России Мария Захарова оценила принятие пар.

Figure 72. Aperçu des occurrences de *официальный представитель*

CONCORDANCE RI_FR

simple porte-parole 30 > filter ministère 1, 5 30 (19.65 per million) X | Sort word X

Left context KWIC Right context

1	doc#0 d'«une rumeur politique politiquement motivée», a déclaré le	porte-parole	du ministère des Affaires étrangères chinois, Hua Chunying, au
2	doc#0 dans la province du Jiangsu (est), a déclaré devant la presse le	porte-parole	du ministère chinois des Affaires étrangères, Geng Shuang. </s><s>
3	doc#0 de sécurité, ont été tuées et 27 ont été blessées», a déclaré le	porte-parole	du ministère de l'Intérieur Nasrat Rahimi. </s><s> L'explosion a
4	doc#0 iardi n'est pas le père de famille, a indiqué la police. </s><s> Un	porte-parole	du ministère autrichien des Affaires étrangères a affirmé qu'il s
5	doc#0 la position de l'adversaire. </s><s> Maria Zakharova, l'influente	porte-parole	du ministère russe des Affaires étrangères (MID), a posté tout i
6	doc#0 l est destiné à promouvoir «la paix dans le monde», selon un	porte-parole	du ministère de la Défense. </s><s> Face à «l'unilatéralisme»
7	doc#0 tribunal ne peut le permettre», conclut Dolly Gee. </s><s> Un	porte-parole	du ministère de la Justice a déclaré que l'administration était <
8	doc#0 ontestablement parrainée par l'Iran», a déclaré de son côté le	porte-parole	du ministère de la Défense saoudien, Turki al-Maliki, au cours d
9	doc#0 que nous avons vu de mémoire récente». </s><s> Mercredi, le	porte-parole	du ministère chinois des Affaires étrangères, Geng Shuang, a d
10	doc#0 mier checkpoint donnant accès au meeting électoral, selon un	porte-parole	du ministère de l'Intérieur. </s><s> D'après une porte-parole du
11	doc#0 zations américaines. </s><s> Soufflant le chaud et le froid, le	porte-parole	du ministère iranien des Affaires étrangères a déclaré lundi que
12	doc#0 sées <s> > et < </s> < incompréhensibles <s> >, par la voix du	porte-parole	du ministère des Affaires étrangères, Abbas Moussavi, qui a lai
13	doc#0 ont assuré que la situation était «sous contrôle». </s><s> Le	porte-parole	du ministère de l'Intérieur, Nasrat Rahimi, a précisé que l'armée
14	doc#0 </s><s> La Russie dit n'avoir << pas d'informations fiables >> Le	porte-parole	du ministère russe de la défense, Igor Konashenkov, a quant à l

Figure 73. Aperçu des occurrences de *porte-parole* + *ministère*

CONCORDANCE RI_RU

simple представитель 35 > filter кремль 1, 5 35 (28.82 per million) X

Left context KWIC Right context

1	doc#0 неожиданном», -- приводит РИА «Новости» слова	представителя	Кремля. </s><s> Накануне спецпредставитель америк.
2	doc#0 осьбе сирийского руководства, передает НСН. </s><s>	Представитель	Кремля также добавил, что финальной целью являе
3	doc#0 ище и, тем самым, заставив воевать с турками. </s><s>	Представитель	Кремля считает, что если курдские формирования с вс
4	doc#0 го формата», ни на любых других площадках. </s><s>	Представитель	Кремля указал, что полуостров вошел в состав РФ в 20
5	doc#0 з будут какие-то сигналы, будем обсуждать», -- указал	представитель	Кремля. </s><s> Более того - на официальном сайте п
6	doc#0 одолжению шагов в рамках интеграции», -- указал	представитель	Кремля. </s><s> Вместе с тем конгрессмены от Демок
7	doc#0 советских стран и воссоздать СССР к 2024 году. </s><s>	Представитель	Кремля Дмитрий Песков опроверг доводы бывшего пр
8	doc#0 й Песков. «Ведется обсуждение этой темы», - заявил	представитель	Кремля. </s><s> В Киеве, несмотря на отказ от «форм
9	doc#0 жет: [Крым] это один из регионов России», -- заявил	представитель	Кремля. </s><s> Перед этим президент Украины Влади
10	doc#0 его минского ««комплекса мер». </s><s> В этой связи	представитель	Кремля Дмитрий Песков заявил о налаживании диало
11	doc#0 одолжению шагов в рамках интеграции», -- пояснил	представитель	Кремля. </s><s> В исходном сообщении СМИ говорили
12	doc#0 ского лидера Дмитрий Песков, передает ТАСС. </s><s>	Представитель	Кремля отметил, что ограничительные меры -- это «с
13	doc#0 ны освобожденных людей непринципиален - пояснил	представитель	Кремля Дмитрий Песков. «Путин свою часть пути про
14	doc#0 Мы... приветствуем состоявшийся обмен», -- добавил	представитель	Кремля. </s><s> По его словам, обмен также дает шан
15	doc#0 о телефону от своих зарубежных коллег», -- сообщил	представитель	Кремля. </s><s> При этом, по словам Пескова, нередк

Figure 74. Aperçu des occurrences de *представитель Кремля*

CONCORDANCE RI_FR

simple porte-parole 10 > filter kremlin 1, 5 10 (6.55 per million) X | Sort word X

Left context KWIC Right context

1	doc#0 as le processus de règlement politique en Syrie», a souligné le	porte-parole	du Kremlin. </s><s> Néanmoins, selon le ministre des Affaires
2	doc#0 s et d'identifier les causes de l'accident, a déclaré à la presse le	porte-parole	du Kremlin, Dmitri Peskov. </s><s> Des images télévisées mor
3	doc#0 s mardi sur les réseaux sociaux. </s><s> Sur la vidéo, on voit le	porte-parole	du Kremlin, Dmitri Peskov, s'adresser aux étudiants en journali
4	doc#0 as plus grandes chaînes d'information américaines. </s><s> Le	porte-parole	du Kremlin, Dmitri Peskov, a confirmé que l'espion avait bien ti
5	doc#0 sont pressés de minimiser le rôle de Smolenkov. </s><s> Le	porte-parole	du Kremlin, Dmitri Peskov, a confirmé que l'espion avait bien ti
6	doc#0 abandonné les Kurdes et les ont de facto trahis >>, a résumé le	porte-parole	du Kremlin, mercredi 23 octobre. </s><s> De son côté, Assad é
7	doc#0 terrogé mardi sur la mort d'un soldat russe au Mozambique, le	porte-parole	du Kremlin, Dmitri Peskov, a nié tout déploiement de militaires
8	doc#0 igne qui aidera à l'apaiser >>, a déclaré lundi aux journalistes le	porte-parole	du Kremlin, Dmitri Peskov, disant espérer que «< l'Arabie saou
9	doc#0 officielle, dont la date sera annoncée prochainement, a ajouté le	porte-parole	du Kremlin. </s><s> << L'attaque d'hier par des drones contre c
10	doc#0 n contact avec le président >> Poutine, a déclaré à la presse le	porte-parole	du Kremlin, Dmitri Peskov, en qualifiant les informations diffus

Figure 75. Aperçu des occurrences de *porte-parole du Kremlin*

5.4 Cas particulier russe : mots-valises

Au cours de l'analyse des listes, nous avons remarqué la présence de nombreux mots-valises en russe, phénomène qui n'a pas été constaté en français, et avons décidé de nous pencher sur le sujet. Ci-dessous, le tableau inséré reprend les mots-valises, leur forme complète ainsi que le nombre d'occurrences de l'un et de l'autre. Ce tableau nous a permis d'analyser sous quelle forme ces termes étaient le plus fréquemment utilisés et il en a été dégagé que, dans 84 % des cas, le mot-valise constitue la forme privilégiée. Soulignons néanmoins que cette analyse vaut pour les cas observés et listés ci-dessous, nous ne prétendons pas en déduire une règle générale pour la langue russe.

Forme complète	Occurrences	Mot-valise	Occurrences
национальная безопасность	176	нацбезопасность	65
Министерство обороны	56	Минобороны	100
Министерство юстиции	32	Минюст	66
проект закона	8	законопроект	289
Государственный департамент	12	Госдепартамент	128
Государственный департамент	12	Госдеп	37
Государственный секретарь	5	Госсекретарь	100
Государственная дума	14	Госдума	146

разведывательные службы	19	разведслужбы	26
специальные службы	3	спецслужбы	194
Генеральная Ассамблея	74	Генассамблея	7
Генеральный секретарь	57	генсек	118
Европейский союз	74	Евросоюз	161
Европейская комиссия	11	Еврокомиссия	9
Европейский парламент	9	Европарламент	41
специальное назначение	4	спецназ	50
Генеральный штаб	4	Генштаб	35
террористический акт	9	теракт	61

5.5 Proportion des termes simples et complexes au sein du lexique

Nous avons analysé et calculé la proportion des termes simples et des termes complexes pour chaque langue du lexique. En russe, 67 % des entrées sont des termes simples (Figure 76) et, parmi les 33 % de termes complexes (Figure 77), trois-quarts de ces termes sont des bigrammes. Le dernier quart est, quant à lui, constitué à 17 % de trigrammes, à 4 % de quadrigrammes, à 3 % de 5-grammes et seulement à 1 % de 6-grammes.

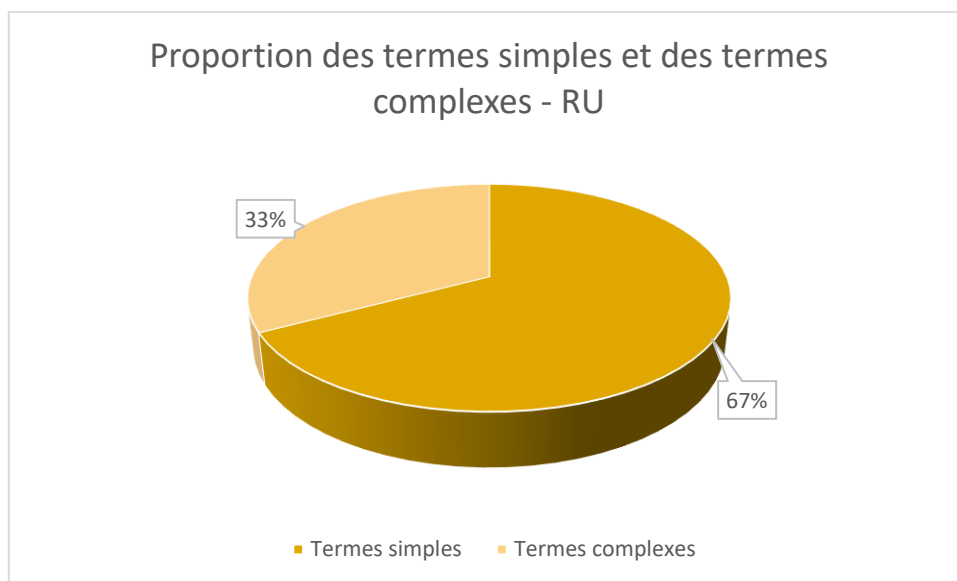


Figure 76. Proportion des termes simples et des termes complexes - RU

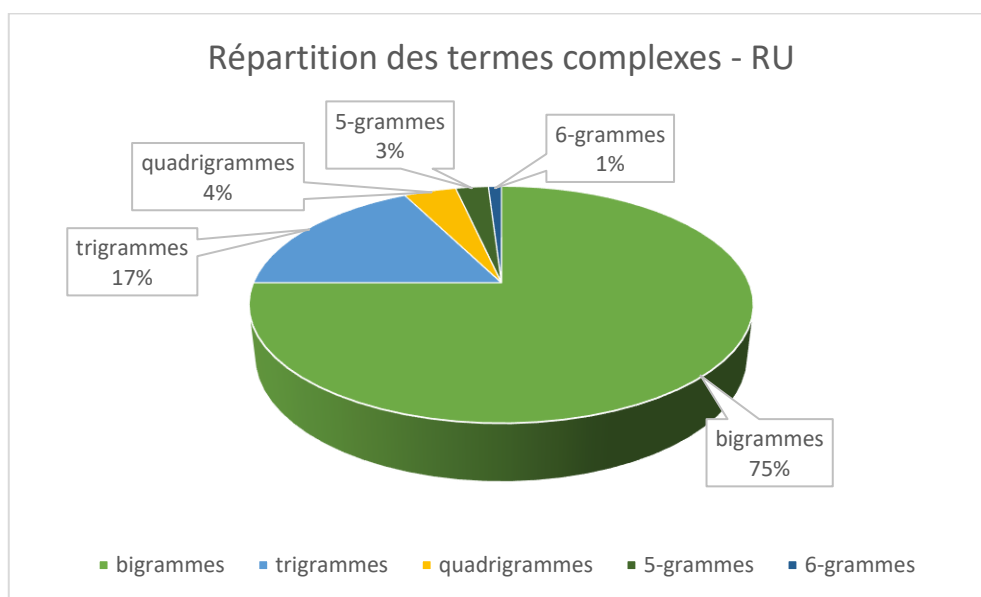


Figure 77. Répartition des termes complexes – RU

En ce qui concerne le français (Figure 78), nous obtenons exactement la même proportion entre termes simples et termes complexes. Cependant, la répartition des termes complexes (Figure 79) présente plusieurs différences par rapport au russe. En effet, les bigrammes sont moins présents (54 % contre 75 % pour le russe), tandis que les trigrammes sont presque deux fois plus nombreux en français qu'en russe (31 % contre 17 % en russe). À la suite de cette constatation, nous avons

décidé d'analyser les bigrammes et trigrammes en russe et en français afin de proposer une explication quant à cette différence.

L'explication réside dans le fait que les déterminants définis (FR *le, la, les*) et indéfinis (FR *un, une, des*) n'existent pas en russe. De nombreux trigrammes en français équivalent dès lors à un bigramme (base + génitif) en russe, et inversement, dont voici quelques exemples : *Conseil des ministres* / кабинет министров, *membre du Parlement* / член парламента, *procédure de destitution* / процедура импичмента, *séparation des pouvoirs* / разделение властей, *projet de loi* / проект закона, *motion de censure* / вотум недоверия ou encore *fonctionnaire de police* / сотрудник полиции.

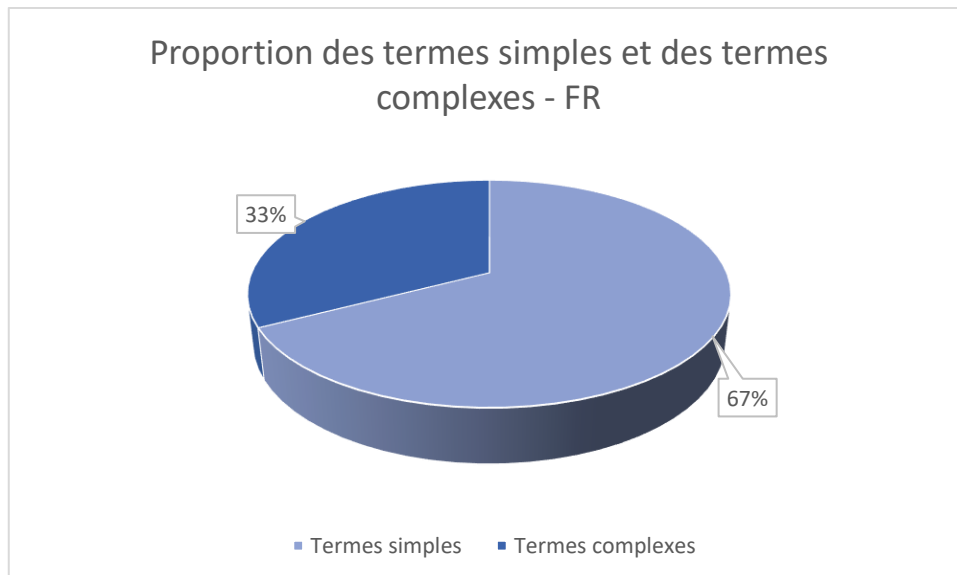


Figure 78. Proportion des termes simples et des termes complexes - FR

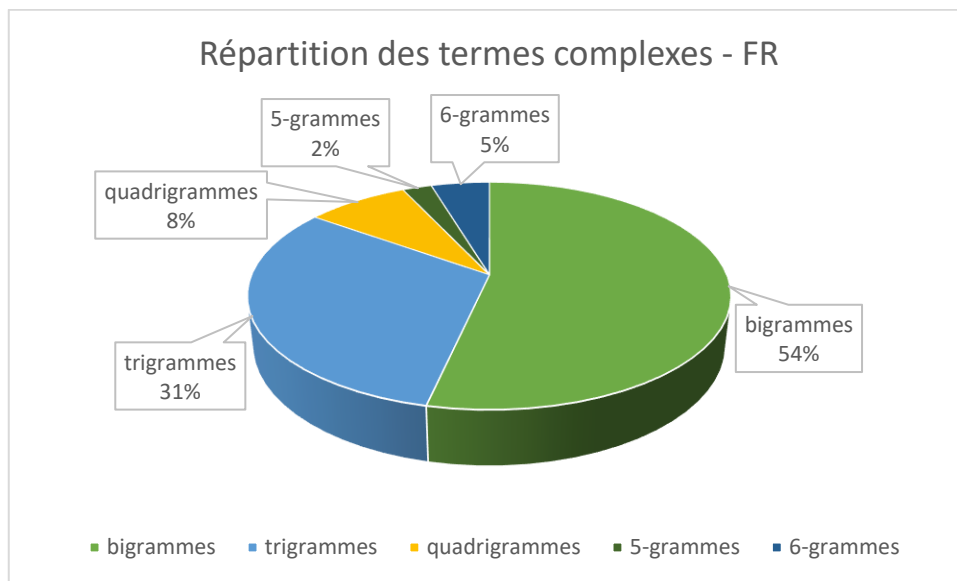


Figure 79. Répartition des termes complexes – FR

5.6 Erreurs et défauts du Sketch Engine

Quelques erreurs d'étiquetage et de lemmatisation ont été repérées lors du nettoyage et du tri des listes. Nous avons listé ci-dessous les erreurs repérées dans le corpus russe et dans le corpus français.

Erreurs repérées dans le corpus russe :

- Figure 80 : lemmatisation de *mpu* (FR *trois*) en *тереть* (FR *perdre*) ;
- Figure 81 : mise au nominatif de sigles et d'acronymes :
 - ПАСЕ (FR APCE) → *пас* (FR *une passe*) ;
 - ИГ (FR l'État islamique) → *уго* (FR *le joug*) ;
- Mise au nominatif de noms propres : *Ангела* (Angela [Merkel]) → *ангел* (FR *ange*) ;
- L'étiqueteur ne fait pas la différence entre *власть* (FR *le pouvoir*) et *власти* (FR *les autorités*).

Line	Left context	KWIC	Right context
1	doc#0 онбасса обходится в \$22 за тонну, то есть, как минимум, в	три	раза дешевле. </s><s> При этом украинские власти подня
2	doc#0 еферендума, так как ее лишили ряда прав. </s><s> Спустя	три	года Москва объявила, что замораживает выплату членск
3	doc#0 еского ведомства. </s><s> Ранее стало известно, что сразу	три	комитета палаты представителей США, подконтрольных ч
4	doc#0 гам своего визита в Россию, у Москвы и Вашингтона есть	три	сферы для успешного сотрудничества. <<Во-первых, наши
5	doc#0 ><s> Затем Захарова обратила внимание на то, что спустя	три	месяца, 10 сентября, президент США Дональд Трамп уволи
6	doc#0 на следующий день. </s><s> Джон Болтон был третьим за	три	года советником Трампа по нацбезопасности. </s><s> В ср
7	doc#0 и переговоры в «нормандском формате» в ближайшие	три	недели, заявил глава украинского МИД Вадим Пристайко.
8	doc#0 зoenных из тех регионов, которые были согласованы еще	три	года назад, а затем и на всем протяжении линии соприкос
9	doc#0 сельского хозяйства», -- подчеркнул политик. </s><s> За	три	дня до этого Комитет постоянных представителей стран Е
10	doc#0 </s><s> С конца июля по начало августа в Москве прошли	три	несогласованные демонстрации. </s><s> Причиной для ак
11	doc#0 з полном объеме», -- отметили в комитете. </s><s> Сразу	три	комитета палаты представителей в рамках процедуры им
12	doc#0 мен заключенными и удерживаемыми лицами, диалог по	трём	группам людей находится на решающей стадии. </s><s> П
13	doc#0 н состоит из нескольких групп людей. </s><s> Скажем так,	три	группы людей, которые мы хотим обменять», -- сказал ук
14	doc#0 лента майской инаугурации Зеленского у него состоялась	три	телефонных разговора с Владимиром Путиным. </s><s> Гр
15	doc#0 м в Австрии. <<Во времена Советского Союза Крым почти	три	десятка лет находился в составе РСФСР, после этого -- 37 л
16	doc#0 ингтон безразличен. </s><s> На африканском рынке есть	три	ключевых российских актора -- Роснефть, Росатом и Росо
17	doc#0 ующего года будут доставлены четыре вертолета Ми-35 и	три	Ми-17 (по контракту с Россией). </s><s> Ранее Белград уже
18	doc#0 лационный фон вблизи места взрыва превышает норму в	три	раза. </s><s> Кроме того, в день взрыва в Норвегии сообщ

Figure 80. Aperçu de l'erreur de lemmatisation de *mpu* en *тереть*

Line	Left context	KWIC	Right context
1	doc#0 борьбе против террористов <<Исламского государства>> (	ИГ	, организация запрещена в России). <<Мы и пришли в Сир
2	doc#0 и как террористическая) и <<Исламским государством>> (	ИГ	, террористическая организация запрещена в России). </s>
3	doc#0 шила атаковать регионального партнера в борьбе против	ИГ	. </s><s> Такие действия президент Франции назвал груб
4	doc#0 акже боевиков группировки <<Исламское государство>> (	ИГ	, запрещена в России) на севере Сирии в начале октября.
5	doc#0 ь-Багдади Появились подробности уничтожения главаря	ИГ	аль-Багдади 27.10.2019 США сообщили об уничтожении л
6	doc#0 агдади 27.10.2019 США сообщили об уничтожении лидера	ИГ	Абу Бакра аль-Багдади. </s><s> О том, как <<Аль-Каида>> >
7	doc#0 зористической организации <<Исламское государство>> (	ИГ	, запрещена в России) Абу Бакра аль-Багдади. </s><s> О то
8	doc#0 трупы людей. </s><s> В поиске местонахождения главаря	ИГ	участвовала разведка Ирака. </s><s> По утверждению Баг,
9	doc#0 срой> > в Сирии. </s><s> В итоге бои развернулись между	ИГ	и официальным отделением <<Аль-Каиды>> в Сирии -- <<
10	doc#0 ды, Балкарии и Карачая заявили о единогласной присяге	ИГ	. </s><s> Гибель аль-Багдади вне сомнения скажется на бо
11	doc#0 юридическую организацию <<Исламское государство>> (	ИГ	, запрещена в России). </s><s> Что же касается президент
12	doc#0 литика как раз и привела к росту численности и влиянию	ИГ	, <<Принудительно>> >: Украина требует от России новую г

Figure 81. Aperçu de l'erreur de lemmatisation de *ИГ* en *узо*

Erreurs repérées dans le corpus français :

- Lemmatisation erronée de noms propres :
 - o Figure 82 : lemmatisation de *Gaza* en *gazer*. Sur les 57 occurrences, une seule (ligne 4) provient effectivement d'un participe passé du verbe *gazer* ;
 - o Figure 83 : lemmatisation de *Cuba* en *cuber* ;
- Figure 84 : lemmatisation de *[nous] sommes* et du substantif *sommes* (en lignes 1 et 21) en *sommer* ;
- Figure 85 : *pourparlers* mis au singulier ;
- Lemmatisation de certaines occurrences de *ouvrant* et *ouvrent* en *ouvrer*.

CONCORDANCE RI_FR

simple **gazer** 57 (37.34 per million)

Details Left context KWIC Right context

- 1 ☐ doc#0 ntre Palestiniens et l'armée israélienne ont éclaté à la frontière de **Gaza** , faisant 59 morts palestiniens. </s><s> Si elle n'a pas mis le feu a
- 2 ☐ doc#0 e de groupes tels que le Hamas, dissimulé parmi la population de **Gaza** , Tsahal assure prévenir les civils avant une frappe pour les convoi
- 3 ☐ doc#0 exprimé leurs doutes sur le nombre d'inscrits dans la province de **Gaza** (sud), où les ONG locales ont indiqué avoir relevé la présence de 3
- 4 ☐ doc#0 rg des parties civiles. </s><s> Cette survivante dont la mère a été **gazée** en novembre 1944, raconte avoir vu des gardes frapper des enf
- 5 ☐ doc#0 's a appris à gérer l'affrontement avec le Hamas dans la bande de **Gaza** , au sud du pays, la menace iranienne, au nord, change de nature e
- 6 ☐ doc#0 r les Israéliens au début de la deuxième intifada, et s'était rendu à **Gaza** , qui n'était pas encore sous le contrôle du Hamas. </s><s> Une au
- 7 ☐ doc#0 itaire israélienne contemporaine, des guerres du Liban à celles de **Gaza** , en passant par les soulèvements palestiniens. </s><s> Elle éclai
- 8 ☐ doc#0 utation d'être un dur à cuire. </s><s> Lors de la dernière guerre à **Gaza** en 2014, il était à la manœuvre. </s><s> Dans une de ses premièr
- 9 ☐ doc#0 ionné en novembre 2018 à la suite d'un accord de cessez-le-feu à **Gaza** , qu'il avait qualifié de <<capitulation face au terrorisme>>. </s><s>
- 10 ☐ doc#0 les Palestiniens. </s><s> Il dénonce la situation dans la bande de **Gaza** et accuse le gouvernement actuel d'avoir perdu sa capacité de dis
- 11 ☐ doc#0 ienne islamio-nationaliste qui règne sans partage sur la bande de **Gaza** . </s><s> Avigdor Lieberman voulait aller au clash. </s><s> Il reco
- 12 ☐ doc#0 Nétanyahou écourte un meeting après un tir de roquettes depuis **Gaza** Des images montrant le premier ministre israélien poussé hors de
- 13 ☐ doc#0 s><s> Elles annonçaient, dans ce secteur frontalier avec la bande **Gaza** , des tirs de roquettes palestiniennes. </s><s> La chaîne publique

Figure 82. Aperçu de l'erreur de lemmatisation de *Gaza* en *gazer*

CONCORDANCE RI_FR

simple **cuber** 24 (15.72 per million)

Details Left context KWIC Right context

- 1 ☐ doc#0 bie saoudite, le Venezuela et a gâché une relation privilégiée avec **Cuba** , très influente dans la Caraïbe. </s><s> Du jamais-vu dans l'histoi
- 2 ☐ doc#0 ef du Parti populaire du Canada (droite), Maxime Bernier. </s><s> **Cuba** , en besoin cruel d'alliés, accueille le premier ministre russe 3 octo
- 3 ☐ doc#0 re russe Dmitri Medvedev entame jeudi une visite de deux jours à **Cuba** au moment où l'île, en pénurie de carburant en raison des sanctor
- 4 ☐ doc#0 pas seuls >> après le don par Pékin de 100 millions de dollars à **Cuba** . </s><s> Ne pas être seul, c'est l'obsession du gouvernement cub
- 5 ☐ doc#0 sportant du pétrole du Venezuela, principal fournisseur de brut de **Cuba** . </s><s> Ces dernières semaines, des files d'attente de plusieurs
- 6 ☐ doc#0 <conjoncturelle> <s> selon le gouvernement, force littéralement **Cuba** à vivre au ralenti. </s><s> De quoi peser sur sa croissance alors q
- 7 ☐ doc#0 rdiction depuis juin, par Washington, des croisières américaines à **Cuba** . </s><s> La solidarité de la Russie, dont le chef de la diplomatie S
- 8 ☐ doc#0 <<Alors que s'intensifient les mesures de blocus américain contre **Cuba** , toute aide en provenance de grandes puissances mondiales est i
- 9 ☐ doc#0 e politologue Carlos Alzugaray. </s><s> Mais <<sur le long terme, **Cuba** devra résoudre (ses) déséquilibres économiques et faire progress
- 10 ☐ doc#0 'acebook. <<Elle oublie que le Chili n'est pas devenu un (nouveau) **Cuba** grâce à ceux qui ont eu le courage de renverser la gauche en 1973
- 11 ☐ doc#0 > Un retrait dû en partie à la maladie de sa fille, Florencia, traitée à **Cuba** , où l'ex-présidente s'est rendue six fois en sept mois. </s><s> Un i
- 12 ☐ doc#0 'sition >>, a-t-il ajouté. </s><s> Washington suspend les vols vers **Cuba** , sauf La Havane Neuf aéroports internationaux sont concernés pi
- 13 ☐ doc#0 ont concernés par cette mesure, notamment celui de Santiago de **Cuba** , la deuxième ville de l'île. </s><s> 26 octobre 2019 Le ministère d

Figure 83. Aperçu de l'erreur de lemmatisation de *Cuba* en *cuber*

CONCORDANCE RI_FR

simple **sommer** 276 (100.82 per million)

Details Left context KWIC Right context

- 1 ☐ doc#0 'r sa campagne, facteur crucial dans une élection qui coûte des **sommes** colossales. </s><s> Ses principaux concurrents, Elizabeth Warr
- 2 ☐ doc#0 'urgence c'est aussi de fournir des médicaments. </s><s> Nous **sommes** dans un foyer épidémique de fièvre jaune et de paludisme qui rit
- 3 ☐ doc#0 > Lorsqu'il s'agit de nous opposer aux partis traditionnels, nous **sommes** bien sûr disposés à bâtir des coalitions ad hoc. </s><s> Mais le
- 4 ☐ doc#0 de son ego, de trop en faire et de malmenier les ministres en les **sommant** avec des préavis très courts de venir s'expliquer devant les dépu
- 5 ☐ doc#0 'un premier ministre sunnite. </s><s> Outre Saad Hariri, <<nous **sommes** face à une communauté sunnite pauvre en premier ministrables
- 6 ☐ doc#0 'l'incendie <<Kincade Fire>>, plus de 180 000 personnes ont été **sommées** de quitter leur domicile. </s><s> L'État d'urgence a été décrété le
- 7 ☐ doc#0 'incipe des règles du jeu équitables? </s><s> L'accord que nous **sommes** prêts à discuter c'est zéro tarif, zéro quota, zéro dumping. </s><s>
- 8 ☐ doc#0 'péronistes >>?, a alors demandé le journaliste. <<Ah, mais nous **sommes** tous péronistes >> Mais une autre question se pose : le péronis
- 9 ☐ doc#0 'arti, ils sont venus soutenir leur chef, Hassan Nasrallah. <<Nous **sommes** à ton service, Nasrallah>>, scandaient-ils alors que la nuit tomb
- 10 ☐ doc#0 'mble pour nous diviser>> <<Il peut dire ce qu'il veut, mais nous **sommes** toujours là, il n'est pas notre chef>>, abonde Christine Kharrat, 21
- 11 ☐ doc#0 'youth, les manifestants sont protégés par leur nombre, ici nous **sommes** au maximum 2000 personnes et tout le monde se connaît>>, co
- 12 ☐ doc#0 'bats, entre puissance étatique et non étatique. </s><s> <<Nous **sommes** face à une rupture systémique>> dans la nature des guerres, est
- 13 ☐ doc#0 'l'incendie <<Kincade Fire>>, plus de 180 000 personnes ont été **sommées** de quitter leur domicile. </s><s> L'État d'urgence a été décrété le
- 14 ☐ doc#0 'sécurité. </s><s> Déjà, quelque 180.000 personnes avaient été **sommées** d'évacuer à cause de l'incendie <<Kincade Fire>>, au nord de Sar
- 15 ☐ doc#0 'a déclaré le chef de l'Etat dans un message à la nation. <<Nous **sommes** dans une nouvelle réalité. </s><s> Le Chili est différent de celui
- 16 ☐ doc#0 'radicalisation potentielle de la crise libanaise. </s><s> Nous ne **sommes** pas à l'abri d'une réaction de certains groupes qui sont mis en c
- 17 ☐ doc#0 > Il y a là une confluence de facteurs qui nous a menés où nous **sommes** . </s><s> Et tandis que nos problèmes viennent de l'intérieur, no
- 18 ☐ doc#0 'ats-Unis n'ont plus qu'une <<stratégie de sortie>>. </s><s> Nous **sommes** à court terme, tactiques, réactifs. </s><s> Les Russes et les Iran
- 19 ☐ doc#0 's aussi pour adresser un ultime adieu au généralissime: <<Nous **sommes** là pour rendre un dernier hommage au meilleur gouvernant d'Es
- 20 ☐ doc#0 s><s> <<Dans notre région, il n'y a pas de traitres. </s><s> Nous **sommes** juste des gens qui ont voulu porter les armes pour défendre nos
- 21 ☐ doc#0 'dicaments que s'échangent les familles séparées ou encore des **sommes** d'argent pour des héritages ou des ventes de terrain>>, raconte l

Figure 84. Erreurs de lemmatisation de *sommes* en *sommer*

828	spécialiste
829	pourparler
830	défi

Figure 85. Forme lemmatisée erronée de *pourparlers*

5.7 Extraction de la combinatoire

Une fois les listes de termes simples et complexes obtenues, nous nous sommes attelée à la deuxième étape de l'élaboration des lexiques : l'extraction de la combinatoire.

Comme expliqué au point [4.3.2 Outils d'extraction de la combinatoire](#), la combinatoire a été extraite à l'aide de la fonctionnalité *Collocations* du *Sketch Engine* et, en moindre mesure, des fonctionnalités *Concordance* et *Word Sketch*.

Nous avons procédé de la manière suivante :

- Recherche du terme dans *Concordance* afin d'accéder à la fonctionnalité *Collocations* ;
- Extraction à gauche, puis à droite, des listes de potentiels collocatifs ;
- Analyse des listes extraites, puis analyse en contexte des potentiels collocatifs ;
- Ajout des collocations retenues dans le lexique sous le terme correspondant ;
- Répétition du même procédé pour la seconde langue ;
- Mise en regard des collocations obtenues pour chaque langue.

Ce travail reposant sur une approche par corpus, nous avons décidé de conserver uniquement les collocations présentes dans les deux langues, et donc de ne pas

ajouter de traduction pour celles n'ayant pas trouvé d'équivalent. À la suite de la mise en regard des collocations obtenues pour chaque langue, un quart des collocations⁴ (Figure 86) n'a pas trouvé d'équivalent. Nous tenions dès lors à souligner que davantage de collocations sont présentes au sein des corpus mais n'ont pas été conservées, faute d'équivalents.

Nous avons, entre autres, perdu des collocations telles que RU *повлиять на политику, предоставлять гражданство, лишать – лишить гражданства, отказаться от гражданства, признавать – признать суверенитет, подготовить законопроект, вносить – внести законопроект, проводить – провести законопроект, привести к власти, термоядерная бомба, водородная бомба, закладывать – заложить бомба* ; FR *proroger le Parlement, lever l'état d'urgence, expédier les affaires courantes, pourvoir un siège [au Parlement], fermeture des frontières, capitale économique, capitale constitutionnelle, capitale administrative, annexer un territoire, bande de territoire, être déchu de sa nationalité, infléchir sa politique*.

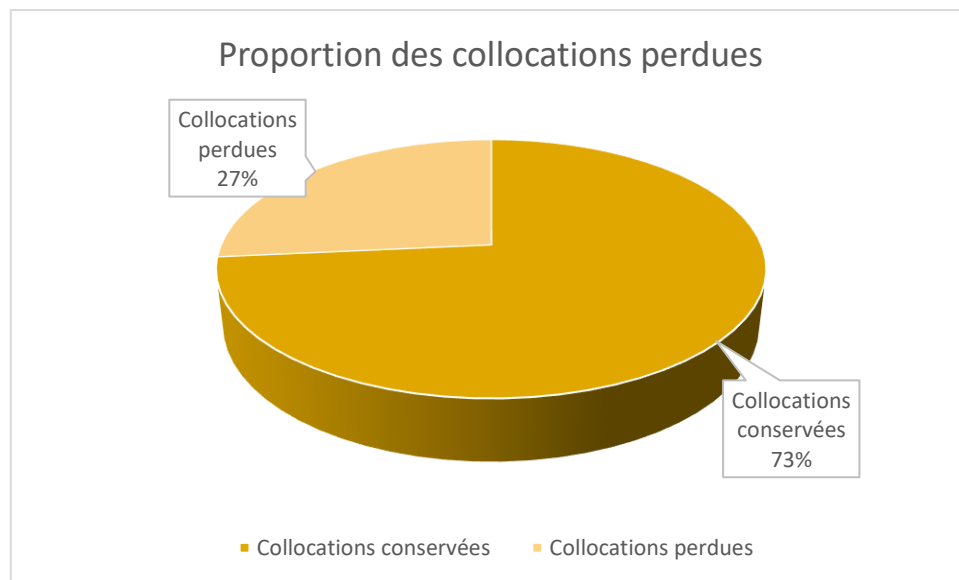


Figure 86. Proportion des collocations perdues

Afin de limiter le volume des lexiques et de ne pas surcharger l'apprenant d'informations, nous avons également décidé d'exclure la combinatoire de la

⁴ Proportion calculée par rapport au lexique « Avancé ».

combinatoire, comme *представлять интересы страны* / *préserver les intérêts d'un pays*, *запрещать – запретить въезд на территорию* / *interdire l'entrée sur le territoire*, *устанавливать – установить контроль на территории* / *prendre le contrôle d'un territoire*, *вернуть контроль над территорией* / *reprendre le contrôle d'un territoire*.

Il convient donc de souligner que les corpus permettent de mettre au jour davantage de collocations que celles présentes dans le lexique.

6 Élaboration des lexiques bilingues

Dans ce chapitre, nous présentons nos lexiques, mise en page et jeu de couleurs, et expliquons les raisons pour lesquelles nous avons décidé de proposer un lexique débutant et un lexique avancé.

6.1 *Lexique débutant – lexique avancé*

Deux lexiques distincts sont joints en annexe de ce travail. Le lexique débutant est destiné aux étudiants de 2^e et de 3^e années de bachelier, tandis que le lexique avancé vise plutôt les étudiants de 1^e et de 2^e années de master. Initialement, il avait été prévu de proposer un lexique unique destiné aux étudiants de bachelier. Cependant, au fil de l'élaboration du lexique, nous nous sommes aperçue que le contenu serait trop volumineux, et parfois trop compliqué, pour des étudiants de bachelier. Nous avons dès lors décidé d'emprunter un chemin différent et, au lieu de supprimer du contenu, de proposer une seconde version plus avancée.

Les lexiques débutant et avancé reprennent, l'un comme l'autre, les mêmes entrées. La différence entre les deux lexiques réside dans le nombre de collocations proposées pour chaque niveau. Dans un premier temps, comme expliqué au point [5.7 Extraction de la combinatoire](#), nous avons supprimé les collocations ne trouvant pas d'équivalent, ainsi que la combinatoire de la combinatoire. Une fois ce premier tri effectué, nous avons décidé de compter le nombre d'occurrences de chaque collocation présente dans le lexique et, à partir de là, déterminer un seuil d'occurrences minimum pour le lexique débutant. Ce seuil a été fixé à 10 occurrences minimum. Le lexique débutant comprend, donc, toute collocation présente plus de 10 fois dans un des deux corpus, tandis que le lexique avancé reprend toutes les collocations : collocations du lexique débutant et collocations présentes moins de 10 fois dans les deux corpus. Le lexique avancé contient dès lors 120 % de collocations en plus que le lexique de niveau débutant, soit un peu plus du double (Figure 87).

Dans le cas de collocations apparaissant une seule fois dans le corpus, nous avons consulté le chapitre « La politique » du dictionnaire *Palabras* afin de déterminer si,

oui ou non, il convenait de les conserver. Tel fut notamment le cas pour FR *mettre aux voix*, *État souverain*, *ratifier un traité* qui ne sont apparus qu'une fois dans le corpus mais qui, étant présents dans le dictionnaire *Palabras*, ont été conservés.

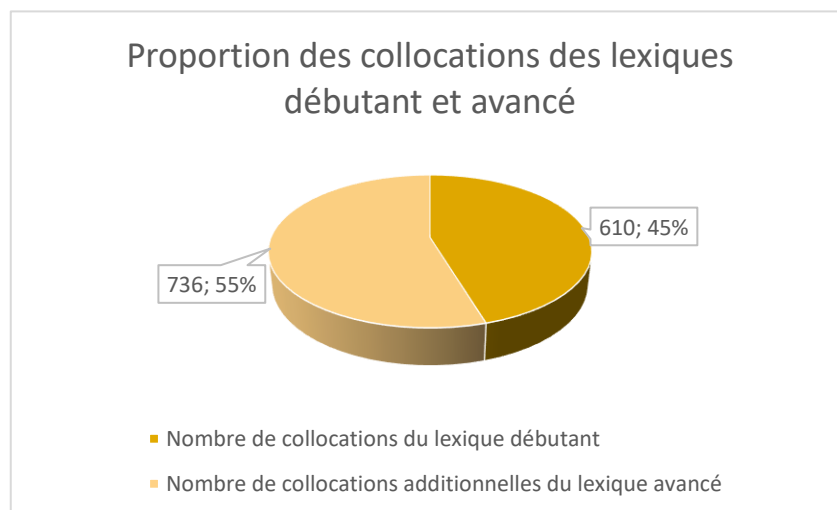


Figure 87. Proportion des collocations des lexiques débutants et avancés

напряженность	les tensions
<i>усиление напряженности</i>	<i>la montée des tensions</i>
<i>снизить напряженность (снижение напряженности, разрядка напряженности)</i>	<i>apaiser les tensions</i>

Figure 88. Aperçu de l'entrée *tensions* - lexique débutant

напряженность	les tensions
<i>усиление напряженности</i>	<i>la montée des tensions, un regain des tensions</i>
<i>эскалация напряженности</i>	<i>l'escalade des tensions</i>
<i>создавать напряженность</i>	<i>provoquer des tensions / créer des tensions</i>
<i>усиливать напряженность</i>	<i>attiser les tensions, exacerber les tensions</i>
<i>снизить напряженность (снижение напряженности, разрядка напряженности), ослабить напряженность</i>	<i>faire baisser les tensions, apaiser les tensions, désamorcer les tensions</i>

Figure 89. Aperçu de l'entrée *tensions* - lexique avancé

6.2 Présentation des lexiques

Les lexiques débutant et avancé contiennent tous deux les mêmes rubriques et sont constitués de trois parties principales (« La politique », « Les relations internationales » et « Les conflits armés »), chaque partie étant divisée en différentes rubriques et sous-rubriques.

La première partie contient les rubriques suivantes :

- La politique ;
- La Constitution ;
- Le pouvoir ;
- L'ordre public ;
- Les élections ;
- Termes propres aux États-Unis, à la Russie, à l'Ukraine, à la France, à l'Allemagne et au Royaume-Uni ;
- Chefs d'État et figures politiques importantes.

La seconde partie contient les rubriques suivantes :

- La politique internationale ;
- La diplomatie ;
- Les ONG ;
- L'espionnage ;
- La crise des réfugiés ;
- L'ONU ;
- L'Union européenne.

La troisième partie contient les rubriques suivantes :

- L'armée ;
- La guerre ;
- Le terrorisme ;
- Les armes ;
- Les crimes et délits.

6.2.1 Présentation des collocations

Afin d'être facilement distinguées, les collocations sont placées en décalage de l'entrée et sont au format italique. Elles sont rangées, en fonction du russe, dans l'ordre suivant : substantif + substantif, adjectif + substantif et verbe + substantif. Nous précisons « en fonction du russe » car les classes grammaticales ne sont pas toujours équivalentes dans les deux langues. Nous avons donc fait le choix de classer les collocations en fonction de la langue source (Figure 90).

договор	un traité
1 <i>подписание договора</i>	<i>la signature d'un traité</i>
2 <i>мирный договор</i>	<i>un traité de paix</i>
3 <i>заключать – заключить договор</i> <i>подписывать – подписать договор</i>	<i>conclure un traité</i> <i>signer un traité</i>

Figure 90. Aperçu de l'entrée *traité* du lexique débutant

Comme nous venons de le mentionner, les classes grammaticales des collocations ne correspondent pas toujours en russe et en français. C'est notamment le cas pour les collocations substantif + substantif et verbe + substantif. Pour le français, nous constatons que les collocations verbe + substantif sont les plus fréquentes (Figure 91) et représentent plus de la moitié des collocations (54 %). En russe, les collocations verbe + substantif sont également les plus fréquentes mais ne représentent cependant que 47 % des collocations (Figure 92). En effet, la proportion de collocations substantif + substantif (32 %) et adjectif + substantif

(21 %) est légèrement plus élevée qu'en français (respectivement 29 % et 17 % pour le français).

Au cours de l'élaboration des lexiques, nous avons constaté que cette différence était due aux substantifs russes en *-ehue* qui équivalent très souvent à un infinitif en français (Figures 93-100).

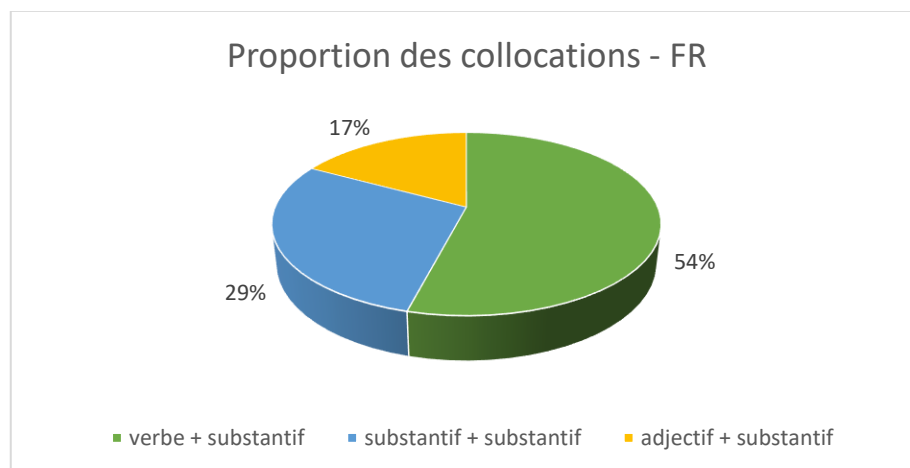


Figure 91. Proportion des collocations - FR

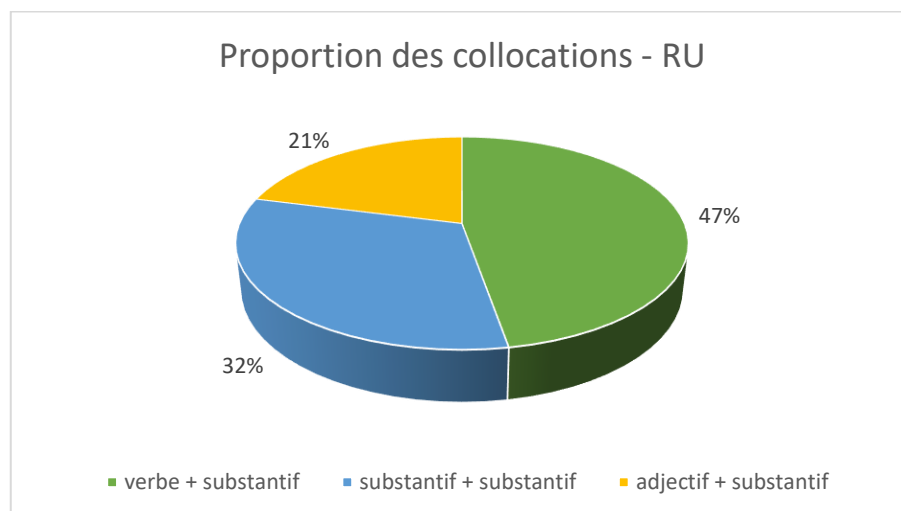


Figure 92. Proportion des collocations – RU

<i>вступление в армию</i>	<i>intégrer l'armée / rejoindre l'armée, s'enrôler dans l'armée</i>
---------------------------	---

Figure 93. Aperçu de l'entrée *armée*

<i>достигать – достичь / достигнуть</i> <i>компромисса (достижение компромисса)</i>	<i>arriver à un compromis</i>
---	--------------------------------------

Figure 94. Aperçu de l'entrée *compromis*

гражданин, выходец <i>возвращение граждан</i>	un ressortissant <i>rapatrier ses ressortissants</i>
---	--

Figure 95. Aperçu de l'entrée *ressortissant*

дипломатические отношения <i>установление дипломатических отношений</i>	les relations diplomatiques <i>établir des relations diplomatiques</i>
---	--

Figure 96. Aperçu de l'entrée *relations diplomatiques*

<i>вводить – ввести санкции против</i> <i>(введение санкций), накладывать – наложить</i> <i>санкции</i>	<i>imposer des sanctions</i>
--	-------------------------------------

Figure 97. Aperçu de l'entrée *sanctions*

<i>победить своего конкурента, поражать</i> <i>противника (поражение противника)</i>	<i>battre son adversaire</i>
--	-------------------------------------

Figure 98. Aperçu de l'entrée *adversaire*

кандидатура <i>выставлять – выставить свою</i> <i>кандидатуру на (выдвижение своей кандидатуры)</i>	la candidature <i>présenter sa candidature</i>
--	--

Figure 99. Aperçu de l'entrée *candidature*

<i>провозглашение республики</i>	<i>proclamer une république</i>
---	--

Figure 100. Aperçu de l'entrée *république*

6.3 Jeu de couleurs

Le lexique débutant contient les collocations apparaissant minimum 10 fois dans un des deux corpus. Les collocations en bleu représentent dès lors les collocations apparues moins de 10 fois dans une des deux langues (Figure 101).

Pour le lexique avancé, qui contient à la fois les collocations du lexique débutant et les collocations apparues moins de 10 fois, la couleur mauve a été utilisée pour distinguer les collocations additionnelles de celles déjà présentes dans le lexique débutant (Figure 102).

диалог	le dialogue
возобновление диалога	<i>la reprise du dialogue</i>
конструктивный диалог	<i>un dialogue constructif</i>
вести диалог, налаживать – наладить	<i>engager le dialogue, nouer le dialogue</i>
диалог (налаживание диалога)	
начинать – начать диалог	<i>ouvrir le dialogue</i>
<i>возобновлять – возобновить диалог</i>	<i>renouer le dialogue</i>
<i>призывать – призвать к диалогу</i>	<i>appeler au dialogue</i>

Figure 101. Aperçu de l'entrée *dialogue* – lexique débutant

уход, вывод	le retrait
полный вывод / <i>полный уход</i>	<i>un retrait total</i>
<i>поспешный вывод</i>	<i>un retrait précipité</i>
объявить об уходе / объявить о выводе	<i>annoncer le retrait</i>
<i>требовать ухода / требовать вывода</i>	<i>ordonner le retrait</i>
<i>начинать – начать вывод</i>	<i>entamer le retrait</i>

Figure 102. Aperçu de l'entrée *retrait* – lexique avancé

La traduction de trois entrées (Figure 103) qui ne figuraient pas dans le corpus français a été ajoutée parce que nous avons jugé qu'il était important pour les étudiants de connaître ces termes relatifs à la Russie. Ces trois entrées ont, elles aussi, été mises en couleur et sont *Российский совет по международным делам (РСМД) / Conseil russe pour les affaires internationales (RSMD)*, *Комитет*

Государственной Думы по международным делам / commission des Affaires étrangères de la Douma et Евразийский экономический союз (ЕАЭС) / Union économique eurasiatique (UEEA).

Российский совет по международным делам (РСМД)	le Conseil russe pour les affaires internationales (RSMD)
---	--



Partie 1 : La politique

 <i>Le RSMD a un statut informel : il est soutenu par l'État mais n'est pas un organisme officiel de la Fédération de Russie.</i>	
Комитет Государственной Думы по международным делам	la commission des Affaires étrangères de la Douma
Совет Безопасности Российской Федерации	le Conseil de sécurité nationale de la Fédération de Russie
Советский Союз	l'Union soviétique
Содружество Независимых Государств (СНГ)	la Communauté des États indépendants (CEI)
Организация Договора о коллективной безопасности (ОДКБ)	l'Organisation du traité de sécurité collective (OTSC)
Евразийский экономический союз (ЕАЭС)	l'Union économique eurasiatique (UEEA)

Figure 103. Aperçu des entrées ajoutées pour le français

6.4 Notes et remarques

Les lexiques sont assortis de quelques notes et remarques. Nous retrouvons les capsules précédées du logo « attention ! » (Figure 104) qui servent à attirer l'attention de l'apprenant sur une remarque d'usage (Figure 105), de vocabulaire (Figure 106) ou de grammaire (Figure 107).

статс-секретарь	un secrétaire d'État  « le <i>secrétaire d'État</i> » (<i>Государственный секретарь / Госсекретарь</i>) est l'équivalent du ministre des Affaires étrangères aux États-Unis.
-----------------	---

Figure 104. Aperçu de la capsule « attention ! » de *secrétaire d'État*

Организация Объединённых Наций, ООН	L'Organisation des Nations unies, l'ONU  La graphie en usage à l'ONU est Organisation des Nations <u>Unies</u> .
-------------------------------------	--

Figure 105. Aperçu de la capsule « attention » - remarque d'usage

представитель  À ne pas confondre avec « <i>председатель</i> » qui signifie « <i>président</i> » (d'un comité, d'une organisation).	un représentant
---	-----------------

Figure 106. Aperçu de la capsule « attention » - remarque de vocabulaire

кому-то гроз <u>ит</u> тюрьма, грозить до X лет тюрем <u>ы</u>  Construction dative : datif + v. + sujet nominatif	risquer la prison, risquer X années de prison, encourir X années de prison
---	---

Figure 107. Aperçu de la capsule « attention » - remarque de grammaire

La note (*adj. subs.*) (Figure 108) est présente à côté de chaque adjectif substantivé afin d'attirer l'attention de l'apprenant sur ces substantifs particuliers qui se déclinent comme des adjectifs.

задержанный (<i>adj. subs.</i>)	un détenu
-----------------------------------	-----------

Figure 108. Aperçu de l'entrée *задержанный*

Enfin, nous avons mis en rouge les terminaisons d'adjectifs mous (Figure 109) et des adjectifs masculins en « -ой » (Figure 1).

внешн я я политика	la politique étrangère
внутренн я я политика	la politique intérieure

Figure 109. Aperçu de la mise en forme des adjectifs mous

свидетель	le témoin
ключев ой свидетель	un témoin clé

Figure 110. Aperçu de la mise en forme des adjectifs masculins en « -ой »

Conclusion

L'objectif de ce travail était de déterminer ce que pouvait apporter le contenu des corpus dans l'élaboration d'un lexique bilingue russe-français. Après avoir défini nos domaines, nous avons procédé à la constitution des corpus, à l'extraction et au tri des listes d'unités monolexicales et polylexicales, à la mise en regard des listes de termes obtenues et enfin à l'élaboration des lexiques bilingues. Nous avons pu constater, grâce au taux très faible de termes perdus, que les corpus avaient été correctement constitués et ce, selon les mêmes critères, et que les seuils d'occurrences avaient été correctement choisis.

Cette approche par corpus nous aura permis de mettre au point, non pas un, mais deux lexiques bilingues et d'assortir la plupart des termes obtenus de leur combinatoire. Il va sans dire que les lexiques pourront être complétés ou réagencés selon les besoins de l'enseignant ou de l'apprenant. Nous invitons également les potentiels utilisateurs de ces lexiques à actualiser les informations nécessitant une mise à jour, comme les noms de dirigeants, les noms de capitales ainsi que les noms des pays membres de l'OTAN, de l'UE, de l'ONU et de la CEI.

Les corpus pourraient, eux-aussi, être exploités en salle de classe. Outre les exercices proposés dans le chapitre « Utilisation des corpus en salle de classe », les corpus pourraient être utilisés lors d'exercices de traduction afin de dégager davantage de combinatoire. Comme nous l'avons souligné, certaines collocations n'ont pas trouvé d'équivalent lors de la mise en regard, et ont donc été supprimées, mais sont toutefois présentes dans les différents corpus. Il en va de même pour la combinatoire de la combinatoire, délibérément supprimée des lexiques, qui pourrait tout à fait être exploitée en salle de classe lors d'exercices de traduction.

Au regard des corpus, ils sont pour le moment fermés. Cependant, nous pourrions envisager de les rendre ouverts dans le cas où un mémorant souhaiterait s'en servir et les alimenter pour un futur mémoire, ou si un professeur les utilise en salle de classe et désire accroître leur volume ou ajouter du contenu plus actuel. Alimenter les corpus permettrait, en outre, de dégager davantage de termes et de

combinatoire. Nous soulignons néanmoins que les corpus devront impérativement être alimentés selon les mêmes critères de constitution et avec des fichiers nettoyés.

Bibliographie

Ouvrages de référence

Bowker, L., Pearson, J., (2002). *Working with Specialized Language. A Practical Guide to Using Corpora*. London : Routledge.

James, C., (1998). *Errors in language learning and use: Exploring error analysis*. London-New York : Longman.

L'Homme, M.-C., (2004). *La terminologie : principes et techniques*. Montréal : Les Presses de l'Université de Montréal.

Loock, R. (2016). *La traductologie de corpus*. Villeneuve d'Ascq : Presses Universitaires du Septentrion.

SAGER, Juan C. (1990). *A Practical Course in Terminology Processing*. Amsterdam-Philadelphia : John Benjamins Publishing. Retrieved from https://books.google.be/books?id=Be4nBVIfj0wC&printsec=copyright&redir_esc=y#v=onepage&q&f=false

Thomas, M. (2009). *Handbook of Research on Web 2.0 and Second Language Learning*. s.l. : IGI Global. Retrieved from <http://doi:10.4018/978-1-60566-190-2>

Articles

Ackerley, K. (2017). « Effects of corpus-based instruction on phraseology in learner English », in *Language Learning & Technology* [En ligne], Vol. 21, pp. 195–216. Retrieved from https://www.researchgate.net/publication/322615473_Effects_of_corpus-based_instruction_on_phraseology_in_learner_English

Ancil, D. & Tremblay, O. (2016). « Les collocations : des combinaisons de mots privilégiées », in *Correspondance* [En ligne], Vol. 21-3. Retrieved from <https://correspo.ccdmd.qc.ca/wp-content/uploads/2018/09/correspondance-la->

[lecture-dans-tous-ses-etats-les-collocations-des-combinaisons-de-mots-privilegiees-.pdf](#)

Bertin, A. (2009). « Grammaire et dictionnaire : le parti pris des mots », in *LINX* [En ligne], Vol. 61, pp. 71-85. Retrieved from <https://journals.openedition.org/linx/1332> ; DOI : <https://doi.org/10.4000/linx.1332>

Boulton, A. & Landure, C. (2016). « Using corpora in language teaching, learning and use », in *Recherches et pratiques pédagogiques en langues de spécialité* [En ligne], Vol. 35-1. Retrieved from <https://journals.openedition.org/apliut/5433>

Chambers, A. (2010). « L'apprentissage de l'écriture en langue seconde à l'aide d'un corpus spécialisé », in *Revue française de linguistique appliquée* [En ligne], Vol. 15, pp. 9-20. Retrieved from <https://www.cairn.info/revue-francaise-de-linguistique-appliquee-2010-2-page-9.htm>

Chateau-Smith, C. (2013). « Corpus Linguistics – Une modeste contribution à l'histoire des sciences », in *Texto !* [En ligne], Vol. 18-2. Retrieved from http://www.revue-texto.net/docannexe/file/3239/carmela_texto_corpus_linguistics.pdf

Cobb, T. (1997). « Is there any measurable learning from hands on concordancing? », in *System*, Vol. 25-3, pp. 301-315.

Corino E. & Onesti C. (2019). « Data-driven learning: A scaffolding methodology for CLIL and LSP teaching and learning », in *Frontiers in Education* [En ligne], Vol. 4-7. Retrieved from <https://www.frontiersin.org/articles/10.3389/feduc.2019.00007/full#h15>

Di Vito, S. (2013). « L'utilisation des corpus dans l'analyse linguistique et dans l'apprentissage du FLE », in *LINX* [En ligne], Vol. 68-69, pp. 159-176. Retrieved from <http://journals.openedition.org/linx/1519> ; DOI : <https://doi.org/10.4000/linx.1519>

Gaskell, D. & Cobb, T. (2004). « Can learners use concordance feedback for writing errors? », in *System*, Vol. 32-3, pp. 301-319.

Gavioli, L. (2001), « The learner as a researcher: Introducing corpus concordancing in the classroom », in Aston, G. (éd.), *Learning with Corpora*. Bologna : Clueb, pp. 108-137.

Johns, T. (1997). « Contexts: The background, development, and trialling of a concordance-based CALL program », in Wichmann, A. (éd.), *Teaching and Language Corpora*. London-New York : Longman, pp. 100-115.

Kilgariff, A. (éd.) (2015). « Corpora and Language Learning with the Sketch Engine and SKELL », in *Revue française de linguistique appliquée* [En ligne], Vol. 20, pp. 61-80. Retrieved from <https://doi.org/10.3917/rfla.201.0061>

Kilgariff, A. (éd.) (2014). « The Sketch Engine: Ten Years On », in *Lexicography* [En ligne], Vol. 1, pp. 7-36. Retrieved from https://www.researchgate.net/publication/271848017_The_Sketch_Engine_Ten_Years_On/references

Lefer, M-A. (à paraître). « Parallel Corpora », pp. 257-282.

McEnery, T. & Wilson, A. (1997). « Teaching and language corpora », in *ReCALL* [En ligne], Vol. 9-1, pp. 5-14. Retrieved from https://www.academia.edu/22660666/McEnery_T_and_Wilson_A_1997_Teaching_and_Language_Corpora_TALC?auto=download

McEnery, T. & Xiao, R. (2010). « What corpora can offer in language teaching and learning », in Hinkel, E. (éd.), *Handbook of Research in Second Language Teaching and Learning* [En ligne], Vol. 2, pp. 364-380. London : Routledge. Retrieved from https://www.lancaster.ac.uk/fass/projects/corpus/ZJU/xpapers/McEnery_Xiao_teaching.PDF

Rao, CS. (2018). « The Importance of Collocations in Teaching of Vocabulary », in *Journal for Research Scholars and Professionals of English Language Teaching* [En ligne], Vol. 2. Retrieved from

https://www.academia.edu/36773173/The_Importance_of_Collolocations_in_Teaching_of_Vocabulary?auto=download&email_work_card=download-paper

Schaeffer-Lacroix, E. (2015). « Analyse de trois systèmes de gestion de corpus pour l'enseignement-apprentissage des langues étrangères », in *Alsic* [En ligne], Vol. 18-1. Retrieved from <https://journals.openedition.org/alsic/2852>; DOI : <https://doi.org/10.4000/alsic.2852>

Williams G. (1999). « Looking in before looking out: Internal selection criteria in a corpus of plant biology », in Kiefer, F., Kiss, G. & Pajzs, J. (1999). *Papers in Computational Lexicography - Complex '99*, pp. 195-204.

Thèse de doctorat

Goeuriot, L. (2009). « Découverte et caractérisation des corpus comparables spécialisés », Interface homme-machine [cs.HC]. Université de Nantes. Retrieved from <https://tel.archives-ouvertes.fr/tel-00474405/document>

Documents électroniques

Annuaire culturel Gralon. *Journal Le Monde : Présentation et Histoire*, [En ligne], adresse URL : <https://www.gralon.net/articles/news-et-media/magazines/article-le-journal-le-monde---presentation-et-histoire-2045.htm> (page web consultée le 15/09/2020)

Centre d'information sur les médias. *Rapports de marque 2019*, [En ligne], adresse URL : <https://www.cim.be/nl/pers/brand-reports> (page web consultée le 18/09/2020)

Courrier International. *Le Vif-L'Express*, [En ligne], adresse URL : <https://www.courrierinternational.com/notule-source/le-vif-l-express> (page web consultée le 15/09/2020)

Courrier International. *Gazeta.ru*, [En ligne], adresse URL : <https://www.courrierinternational.com/notule-source/gazeta-ru> (page web consultée le 15/09/2020)

Courrier International. *Izvestia*, [En ligne], adresse URL :

<https://www.courrierinternational.com/notule-source/izvestia> (page web consultée le 15/09/2020)

Courrier International. *Nezavissimaïa Gazeta*, [En ligne], adresse URL :

<https://www.courrierinternational.com/notule-source/nezavissimaia-gazeta> (page web consultée le 15/09/2020)

Courrier International. *Novaïa Gazeta*, [En ligne], adresse URL :

<https://www.courrierinternational.com/notule-source/novaia-gazeta> (page web consultée le 20/09/2020)

Courrier International. *Vedomosti*, [En ligne], adresse URL :

<https://www.courrierinternational.com/notule-source/vedomosti> (page web consultée le 20/09/2020)

English-Corpora. *Corpora overview*, [En ligne], adresse URL : <https://www.english-corpora.org/corpora.asp> (page web consultée le 06/10/2020)

Euro Topics. *Le Figaro*, [En ligne], adresse URL :

<https://www.eurotopics.net/fr/148679/le-figaro> (page web consultée le 15/09/2020)

Euro Topics. *Le Monde*, [En ligne], adresse URL :

<https://www.eurotopics.net/fr/148683/le-monde> (page web consultée le 15/09/2020)

Euro Topics. *Le Vif/L'Express*, [En ligne], adresse URL :

<https://www.eurotopics.net/fr/148692/le-vif-l-express> (page web consultée le 15/09/2020)

Euro Topics. *Izvestia*, [En ligne], adresse URL :

<https://www.eurotopics.net/fr/179544/izvestia> (page web consultée le 15/09/2020)

Euro Topics. *Nezavissimaïa Gazeta*, [En ligne], adresse URL :

<https://www.eurotopics.net/fr/180109/nezavissimaia-gazeta> (page web consultée le 15/09/2020)

Euro Topics. *Novaia Gazeta*, [En ligne], adresse URL :

<https://www.eurotopics.net/fr/155260/novaia-gazeta> (page web consultée le 20/09/2020)

Euro Topics. *Vedomosti*, [En ligne], adresse URL :

<https://www.eurotopics.net/fr/170578/vedomosti> (page web consultée le 15/09/2020)

Frantext. *Bienvenue dans Frantext*, [En ligne], adresse URL :

<https://www.frantext.fr/> (page web consultée le 06/10/2020)

Groupe Dassault. *Groupe Figaro*, [En ligne], adresse URL :

<http://www.dassault.fr/filiale.php?docid=2410> (page web consultée le 06/10/2020)

L'Alliance pour les Chiffres de la Presse et des Médias. *Fiche support presse*

« *Le Figaro* », [En ligne], adresse URL : <https://www.acpm.fr/Support/le-figaro> (page web consultée le 15/09/2020)

L'Alliance pour les Chiffres de la Presse et des Médias. *Fiche support presse*

« *Le Monde* », [En ligne], adresse URL : <https://www.acpm.fr/Support/le-monde> (page web consultée le 15/09/2020)

L'Histoire. *Le Journal Le Monde*, [En ligne], adresse URL :

<https://www.lhistoire.fr/le-journal-%C2%AB-le-monde-%C2%BB> (page web consultée le 15/09/2020)

Le Figaro. Laroche-Signorile, V. (2019). *Il y a 150 ans Le Figaro devenait quotidien*, [En ligne], adresse URL :

<https://www.lefigaro.fr/histoire/archives/2016/11/15/26010-20161115ARTFIG00272-il-y-a-150-ans-le-figaro-devenait-quotidien.php> (page web consultée le 15/09/2020)

Le Monde. *Comment l'actionnariat du Groupe Le Monde est organisé*, [En ligne],

adresse URL : https://www.lemonde.fr/actualite-medias/article/2019/09/10/comment-l-actionnariat-du-groupe-le-monde-est-organise_5508557_3236.html (page web consultée le 06/10/2020)

National Media Group. *MIC "Izvestia" is the first private multimedia information center in Russia*, [En ligne], adresse URL : http://nm-g.com/en/actives/sections.php?ID_BLOCK=35&ID_SECTION=145 (page web consultée le 16/10/2020)

Rambler Group. *Gazeta.ru*, [En ligne], adresse URL : <https://ramblergroup.com/assets/media/gazeta> (page web consultée le 15/10/2020)

Roularta Media. *Le Vif/L'Express*, [En ligne], adresse URL : <https://www.roulartamedia.be/fr/marques/magazines/le-viflexpress> (page web consultée le 15/10/2020)

Sketch Engine. *GDEX – Good Dictionary Examples*, [En ligne], adresse URL : <https://www.sketchengine.eu/guide/gdex/> (dernière consultation le 01/12/2020)

Sketch Engine. *Glossary « token »*, [En ligne], adresse URL : <https://www.sketchengine.eu/guide/glossary/> (page web consultée le 30/10/2020)

Sketch Engine. *Word Sketch – collocations and word combinations*, [En ligne], adresse URL : <https://www.sketchengine.eu/guide/word-sketch-collocations-and-word-combinations/> (page web consultée le 24/09/2020).

The Guardian. Greenslade. R. (2015). *Alexander Lebedev gives up bankrolling Russia's Novaya Gazeta*, [En ligne], adresse URL : <https://www.theguardian.com/media/greenslade/2015/mar/20/alexander-lebedev-gives-up-bankrolling-russias-novaya-gazeta> (page web consultée le 16/10/2020)

Varieng. *British National Corpus (BNC)*, [En ligne], adresse URL : <http://www.helsinki.fi/varieng/CoRD/corpora/BNC/> (page web consultée le 05/10/2020)

Varieng. *The Corpus of Contemporary American English (COCA)*, [En ligne], adresse URL : <http://www.helsinki.fi/varieng/CoRD/corpora/COCA/> (page web consultée le 05/10/2020)

Varieng. *The Lancaster-Oslo/Bergen Corpus (LOB Corpus)*, [En ligne],
adresse URL : <http://www.helsinki.fi/varieng/CoRD/corpora/LOB/> (page web
consultée le 05/10/2020)

Varieng. *The London-Lund Corpus of Spoken English (LLC)*, [En ligne],
adresse URL : <http://www.helsinki.fi/varieng/CoRD/corpora/LLC/index.html> (page
web consultée le 05/10/2020)

Varieng. *The Standard Corpus of Present-Day Edited American English (the
Brown Corpus)*, [En ligne], adresse URL :
<http://www.helsinki.fi/varieng/CoRD/corpora/BROWN/> (page web consultée le
05/10/2020)

Dictionnaire papier

Gusdorf, F. (éd.) (2003). *Palabras – le vocabulaire espagnol*. Lonrai : Roto
Normandie impression s.a.s.

Dictionnaires électroniques

Cambridge Dictionaries. *Définition de « corpus »*, [En ligne], adresse URL :
<https://dictionary.cambridge.org/fr/dictionnaire/anglais/corpus> (page web consultée
le 12/09/2020)

Cordial. *Définition de « mot lexical » et de « mot grammatical »*, [En ligne],
adresse URL : <https://www.cordial.fr/grammaire/manuels/lexical.htm> (page web
consultée le 09/11/2020)

Dictionnaire Larousse. *Définition de « corpus »*, [En ligne], adresse URL :
<https://www.larousse.fr/dictionnaires/francais/corpus/19410> (page web consultée
le 12/09/2020)

Le Trésor de la Langue Française informatisé. *Définition de « corpus »*, [En ligne],
adresse URL :
[http://atilf.atilf.fr/dendien/scripts/tlfiv5/visusel.exe?11;s=4261509765;r=1;nat=;sol=](http://atilf.atilf.fr/dendien/scripts/tlfiv5/visusel.exe?11;s=4261509765;r=1;nat=;sol=0;)
[0](http://atilf.atilf.fr/dendien/scripts/tlfiv5/visusel.exe?11;s=4261509765;r=1;nat=;sol=0;); (page web consultée le 12/09/2020)

Le Trésor de la Langue Française informatisé. *Définition de « mobilisation »*,

[En ligne], adresse URL :

<http://stella.atilf.fr/Dendien/scripts/tlfiv5/advanced.exe?8;s=835641630>; (page web consultée le 30/11/2020)

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
Faculté de philosophie, arts et lettres

Place Blaise Pascal, 1 bte L3.03.11, 1348 Louvain-la-Neuve, Belgique | www.uclouvain.be/fial