


**Intégration des outils de
traitement de corpus en contexte
professionnel : étude de cas dans
la traduction juridique et
financière.**



#TranslatingEurope Forum 2017

mardi 7 Novembre 2017,
Commission européenne, Bruxelles

Julie Remfort

Directrice adjointe du centre de traduction du Ministère de l'Économie et des Finances

Sandrine Peraldi

Professeur assistant de Linguistique, School of Languages, Cultures and Linguistics, UCD

julie.remfort@ssi.gouv.fr

sandrine.peraldi@ucd.ie



1.

La linguistique de corpus

appliquée à la traduction spécialisée

Linguistique de corpus

- ▷ Étude et observation des faits de langue fondées sur un corpus
- ▷ Démarche empirique, inductive et quantitative.

« 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. »

(Bowker, Pearson 2002 : 9)

Apports et applications

En terminologie

Appropriation facilitée de domaines de spécialité

Extraction des informations terminologiques, sémantiques, techniques, linguistiques et grammaticales

Identification des concepts clés du domaine, concurrents éventuels, collocations, relations sémantiques et associatives, etc.

En traduction

Mise en lumière de solutions à des problèmes de traduction

Détection des normes et des stratégies de traduction

Étude des similitudes et des différences d'usage entre les langues

Mise en avant des éléments d'ordre stylistique.

(Aston 1999; Bowker 1998, 2001; Kübler 2011; Laviosa 1998, Loock 2016),



Les origines du projet

Questionnements théoriques



- ▷ Exploitation quasi exclusive des corpus parallèles en traductologie de corpus.
- ▷ La linguistique de corpus sert fondamentalement la formation des traducteurs.

Questionnements pragmatiques



- ▷ Comment convaincre les traducteurs de la pertinence de ces outils ?
- ▷ Comment insérer ces outils dans l'environnement de travail du traducteur ?
- ▷ Quelles fonctionnalités exploiter et à quelles étapes du processus de traduction ?

Le contexte du MINEFI

Diversité des traductions

Traduction dans des domaines très variés (économie, finances, fiscalité, consommation, recouvrement fiscal etc.) avec des traducteurs polyvalents

Haute technicité

Forte composante juridique dans les traductions.

Contraintes de temps/qualité

Des délais contraints et absence de juriste-linguiste au sein du centre de traduction pour valider les hypothèses de traduction/terminologies.

Les objectifs du projet

Évaluer l'apport de Sketch Engine pendant l'activité traduisante :



Terme

Accéder à, comprendre un terme en langue source via l'extraction d'un contexte définitoire.



Hypothèse de traduction

Trouver un ou des équivalent(s) potentiels en langue cible sur la base de cooccurrences



Équivalent

Valider les équivalents trouvés à l'aide du corpus.



Collocation

Identifier la phraséologie associée à un terme.

2.

La méthodologie de recherche

Périmètre du projet

Experimentation :

- ▷ Constitution de **corpus comparables** sur les contrats de licences de marque
- ▷ Utilisation de ces corpus dans un contexte de traduction par les apprentis du MINEFI
- ▷ Expérimentation de l'outil avec un groupe de traducteurs du centre
- ▷ Retour d'expérience sur l'utilisation des outils

Livrables :

- ▷ Guide d'utilisation de Sketch Engine dans un contexte de traduction
- ▷ Compte-rendu d'expérience et préconisations

Constitution du corpus: critères linguistiques

Domaine d'application : contrats de licences de marque

- ▷ Objectif
- ▷ Représentativité
- ▷ Taille du corpus
- ▷ Cadre de communication
- ▷ Langue
- ▷ Date de publication
- ▷ Parcours et renommée de l'auteur
- ▷ Analyse du genre

(Pearson 1998, Bowker 2001)

Constitution du corpus : une triple approche

(Candel 2001, Peraldi 2011)

 **312 858 mots**
Corpus explicatif en anglais

 **608 332 mots**
Corpus explicatif en français

 **93 206 mots**
Corpus phraséologique

Démarche intellectuelle

- A. Comprendre la notion dans la langue source
- B. Faire une ou des hypothèse(s) de traduction
- C. Vérifier la validité d'un équivalent potentiel et son sens
- D. Choisir entre différentes hypothèses
- E. Explorer les collocations d'un terme et choisir la bonne phraséologie

3. L'expérimentation

Comprendre la notion dans la langue source

Exemple : « certification mark »

Requête **certification, mark** 168 (449.27 par million) 

Page de 9 | [Dernier](#)

[file2970995](#) applicant 3.3.2 Content of regulations 3.4 **Certification mark** regulations 3.4.1 Status and competence

[file2970995](#) certification, collective and ordinary trade marks A **certification mark** is defined in Section 50(1) of the Act

[file2970995](#) defined in Section 50(1) of the Act as: “A **certification mark** is a mark indicating that the goods and

[file2970995](#) characteristics.” The main feature of a **certification mark** is that it is used not by the proprietor

[file2970995](#) of examination peculiar to collective and **certification marks** . Secondly comes an examination of the regulations

[file2970995](#) 2 gives the distinguishing function of a **certification mark** as: “...distinguishing goods or services

[file2970995](#) these goods) but would be acceptable as a **certification mark** as it is capable of performing that particular

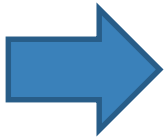
[< précédent](#) applications for certification and collective marks. 1.2 Definitions of, and comparisons between, certification, collective and ordinary trade marks A certification mark is defined in Section 50(1) of the Act as: “A **certification mark** is a mark indicating that the goods and services in connection with which it is used are certified by the proprietor of the mark in respect of origin, material, mode of manufacture of goods or performance of services, quality, accuracy or other characteristics.” The main feature of a certification mark is that it is used not by the proprietor of the mark but instead by his authorised users for the purpose of guaranteeing to the relevant public that goods or services possess a particular characteristic. The proprietor’s mark certifies the presence of [suivant >](#)

Faire une ou des hypothèse(s) de traduction

Exemple : « Limited License »

Equivalent possible :

- licence limitée,
- licence non exclusive,
- licence restreinte ?



Utilisation de la fonctionnalité **Word Sketch**



Faire une ou des hypothèse(s) de traduction


licence (noun)
Corpus explicatif FR fréquence = 2,228 (2,794.83 par million)

modifier			 sujet_de			 pp_de	
	<u>187</u>	0.08		<u>1,011</u>	0.45		<u>301</u>
exclusif	<u>82</u>	12.27	être †	<u>303</u>	10.38	marque †	<u>214</u>
		<i>une licence exclusive</i>			<i>licence est</i>		<i>de licence de marque</i>
obligatoire	<u>11</u>	10.62	pouvoir †	<u>113</u>	10.13	brevet	<u>32</u>
		<i>licence obligatoire</i>			<i>de licence peut</i>		<i>la licence de brevet</i>
mixte	<u>6</u>	9.94	devoir	<u>57</u>	9.85	exploitation	<u>16</u>
simple	<u>6</u>	9.87			<i>de licence doit</i>		<i>une concession de licence d' exploitation e</i>
gratuit	<u>5</u>	9.60	concéder	<u>27</u>	9.45	fabrication	<u>5</u>
partiel	<u>5</u>	9.44			<i>la licence concédée</i>	droit	<u>7</u>
principal	<u>5</u>	9.32	avoir	<u>51</u>	9.19	durée	<u>3</u>
réciproque	<u>4</u>	9.31			<i>de licence a</i>	technologie	<u>3</u>
classique	<u>4</u>	9.30	octroyer	<u>17</u>	9.07	usage	<u>3</u>
opposable	<u>3</u>	8.88			<i>pour lesquels la licence est octroyée , le territoire</i>		
temporaire	<u>3</u>	8.86	limiter	<u>18</u>	9.03		
relatif	<u>3</u>	7.71	interdire	<u>16</u>	8.83		
"	<u>5</u>	7.20			<i>du contrat de licence interdisant , pour des</i>		
			concerner	<u>18</u>	8.74		
					<i>licence concerne</i>		

Vérifier la validité d'un équivalent potentiel et son sens

Vérification de l'hypothèse : « licence partielle »

licence partielle    [Corpus explicatif FR](#)

Requête **licence, partielle** 4 (5.02 par million) 

[file2953756](#) sur une partie des produits ou services (**licence partielle**), ceux-ci doivent également être énumérés

[file2882297](#) le certificat d'enregistrement. En cas de **licence partielle** , il est fondamental d'énumérer très précisément

[file2917747](#) les produits et services couverts par une **licence partielle** ne sont publiés. Il est possible d'accéder

[file2917747](#) Il n'est pas possible d'enregistrer une **licence partielle** limitée à certains produits auxquels le

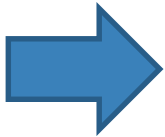


< [précédent](#) alléger la rédaction des clauses. Les parties doivent, par ailleurs, énumérer les marques données en licence en indiquant leurs coordonnées exactes, et lorsque la licence ne porte que sur une partie des produits ou services (**licence partielle**), ceux-ci doivent également être énumérés (sur l'énumération des marques, voir noF2.320-5). II.320-40 - Objet du contrat ou concession de licence a) La licence d'une marque peut être totale ou partielle La licence est totale lorsque le Licencié est autorisé à utiliser la marque pour tous les produits (ou services) qu'elle désigne. La licence est partielle lorsqu'elle est limitée à une partie des produits (ou services) couverts par la marque ou à un produit bien déterminé. Mais, [suivant](#) >

Choisir entre différentes hypothèses

Choisir un équivalent pour « trade name » :

- nom commercial
- désignation commerciale ?



Utilisation de la fonctionnalité **Sketch Diff** au sein du corpus français

Choisir entre différentes hypothèses

nom/désignation

Corpus explicatif FR fréqs = 1,041 | 96

nom 6.0 4.0 2.0 0 -2.0 -4.0 -6.0 désignation

et_ou	282	8	1.80	0.60
enseigne	<u>64</u>	0	12.4	--
dénomination	<u>47</u>	0	11.8	--
adresse	<u>17</u>	0	10.8	--
mot	<u>10</u>	0	9.9	--
marque	<u>34</u>	0	9.7	--
nom	<u>10</u>	0	9.2	--
image	<u>6</u>	0	9.1	--
prénom	<u>4</u>	0	8.8	--
logo	<u>4</u>	0	8.7	--
signe	<u>6</u>	0	8.6	--
pseudonyme	<u>3</u>	0	8.4	--
raison	<u>3</u>	0	8.3	--
droit	<u>5</u>	0	8.1	--
forme	<u>3</u>	0	8.0	--

sujet_de	93	1	0.70	0.10
choisir	<u>6</u>	0	9.9	--
tirer	<u>3</u>	0	9.7	--
connaître	<u>3</u>	0	8.8	--
marquer	<u>3</u>	0	8.7	--
utiliser	<u>3</u>	0	8.3	--
être	<u>29</u>	0	7.2	--
avoir	<u>8</u>	0	7.1	--
pouvoir	<u>6</u>	0	6.4	--

modifier	422	55	2.10	3.10
patronymique	<u>110</u>	0	12.7	--
commercial	<u>198</u>	0	12.7	--
géographique	<u>60</u>	0	11.2	--
générique	<u>6</u>	0	8.6	--
commun	<u>6</u>	0	8.3	--
noble	<u>3</u>	0	7.8	--
célèbre	<u>3</u>	0	7.8	--
notoire	<u>3</u>	0	7.4	--
''	0	<u>3</u>	--	7.3
nécessaire	0	<u>18</u>	--	11.7
usuel	0	<u>26</u>	--	12.7

pp_du	37	12	0.60	2.30
créateur	<u>4</u>	0	11.3	--
déposant	<u>5</u>	0	11.0	--
fabricant	<u>3</u>	0	10.6	--
titulaire	<u>4</u>	0	8.5	--
produit	0	<u>6</u>	--	7.0

objet_de	178	38	1.50	3.70
utiliser	<u>22</u>	0	10.5	--
choisir	<u>12</u>	0	10.3	--
comporter	<u>9</u>	0	10.1	--
employer	<u>6</u>	0	9.8	--

Explorer et exploiter la phraséologie

contrat (*noun*)
Corpus phraséologique FR fréquence = 1,074 (9,766.03 par million)

<u>modifier</u>	<u>16</u>	<u>0.01</u>	<u>sujet_de</u>	<u>376</u>	<u>0.35</u>	<u>et_ou</u>	<u>106</u>	<u>0.10</u>	<u>pp_de</u>	<u>105</u>	<u>0.10</u>
opposable	<u>3</u>	12.26	être +	<u>135</u>	12.10	tort	<u>8</u>	11.15	licence	<u>76</u>	13.02
valide	<u>2</u>	11.67	résilier	<u>18</u>	10.50	résiliation	<u>10</u>	11.14	<i>contrat de licence</i>		
"	<u>4</u>	10.98	conclure	<u>17</u>	10.38	<i>terme du présent contrat ou de la résiliation de la licence</i>			sous-licence	<u>17</u>	12.04
autre	<u>2</u>	10.95	<i>présent contrat est conclu</i>			partie	<u>10</u>	11.01	<i>contrats de sous-licence .</i>		
			prendre	<u>16</u>	10.28	terme	<u>8</u>	10.84	sous-traitance	<u>2</u>	9.26
			<i>Contrat prend</i>			concessionnaire	<u>6</u>	10.68	mariage	<u>2</u>	9.26
			pouvoir	<u>18</u>	10.15	redevance	<u>4</u>	10.11	achat	<u>2</u>	9.17
			<i>Contrat peut être</i>			concedant	<u>4</u>	9.92			
			régir	<u>12</u>	9.97	an	<u>3</u>	9.81			
			avoir	<u>10</u>	9.39	moi mois	<u>3</u>	9.78			
			soumettre	<u>8</u>	9.36	destruction	<u>2</u>	9.24			
			emporter	<u>6</u>	9.00	concession	<u>2</u>	9.21			
			savoir	<u>6</u>	8.97	changement	<u>2</u>	9.20			
			conférer	<u>5</u>	8.73	annexe	<u>2</u>	9.20			
			si	<u>5</u>	8.73	autorisation	<u>2</u>	9.16			
			constituer	<u>5</u>	8.71	>	<u>2</u>	9.16			
			résoudre	<u>4</u>	8.43	tiers	<u>2</u>	9.12			
			annuler	<u>4</u>	8.42	renouvellement	<u>2</u>	9.10			
			établir	<u>4</u>	8.42	usage	<u>2</u>	8.98			
									<u>objet_de</u>	<u>99</u>	<u>0.09</u>
									résilier	<u>22</u>	12.49
									<i>peut résilier le contrat</i>		
									present	<u>15</u>	12.07
									<i>. Le present contrat</i>		
									conclure	<u>16</u>	11.95
									signer	<u>12</u>	11.74
									rendre	<u>6</u>	10.68
									durer	<u>4</u>	10.24
									céder	<u>3</u>	9.81

4.

Résultats et perspectives

Conclusions

Atouts:

- ▷ Outil prometteur
- ▷ Bien accueilli par les traducteurs professionnels pour explorer les collocations et la phraséologie, qui sécurise la recherche terminologique dans le domaine
- ▷ Traductions beaucoup plus idiomatiques

Inconvénients:

- ▷ Temps de mise place opérationnelle et de prise en main
- ▷ Ne peut pas être utilisé seul pour formuler des hypothèses de traduction

Perspectives



Marqueurs définitoires

Rechercher les marqueurs de contexte définitoire à fournir aux traducteurs pour extraire ces derniers plus rapidement

.



Volume

Vérifier l'apport de cet outil avec des corpus plus volumineux.
Enjeux : accessibilité des ressources électroniques



Environnement

Réfléchir à l'intégration de Sketch Engine dans la palette d'outil du traducteur et optimiser son utilisation au regard des autres outils.



Révision

Réfléchir à l'utilisation des outils de linguistique de corpus dans le processus de révision

.



Terminologie

Possible évolution du métier de terminologue vers les sciences de l'information.

5. Bibliographie

- ASTON, Guy (1999) "Corpus use and learning to translate", *Textus*, 12, 289-313.
- BEEBY Allison, Rodriguez Inés Patricia and Sánchez-Gijón Pilar (eds). (2009). *Corpus Use and Translating*. Amsterdam: Benjamins
- Biel, Łucja (2010): *Corpus-Based Studies of Legal Language for Translation Purposes: Methodological and Practical Potential*. In: Carmen Heine and Jan Engberg, eds. *Reconceptualizing LSP. Online proceedings of the XVII European LSP Symposium 2009*. (Aarhus, Denmark, 17-21 August 2009).
- BOWKER, Lynne (2001). 'Towards a Methodology for a Corpus-Based Approach to Translation Evaluation'. *Meta*, XLVI, 2
- BOWKER, Lynne (1998): *Using Specialized Monolingual Native-Language Corpora as a Translation Resource : A Pilot Study*. In : *Meta : journal des traducteurs / Meta: Translators' Journal*, vol. 43, n° 4, 1998, pp. 631-651.
- BOWKER, Lynne et Jennifer PEARSON (2002) : *Working with Specialized Languages*. In : *A Practical Guide to Using Corpora*. London, Routledge.
- CANDEL, Danielle, MARCHAUDON Pierrette & Violette TOLÉDANO (2001). « Le groupe nominal dans l'activité définitoire des scientifiques », "Langue et discours spécialisés", *Revue française de linguistique appliquée* VI/2, p. 17-28.
- CONDAMINES, Anne (2003) : *Sémantique et corpus spécialisés : constitution de bases de connaissances terminologiques*. Habilitation à diriger des recherches, Université de Toulouse 2.

- ENGUEHARD, Chantal (2005) : Un banc de test pour la reconnaissance de termes en corpus. In : La linguistique de corpus, Williams, Geoffrey (ed.), Rennes, Presses Universitaires de Rennes, 2005, 420 pages, p. 273
- KÜBLER, Natalie & Aston, Guy. 2010. "Using Corpora in Translation". in M. Mc Carthy & A. O'Keefe (eds) The Routledge Handbook of Corpus Linguistics, London : Routledge.
- KÜBLER, Natalie (2011). 'Working with different corpora in translation teaching'. In Ana Frankenberg-Garcia, Lynne Flowerdew, and Guy Aston (eds) New Trends in Corpora and Language Learning. London : Continuum.
- KÜBLER, Natalie (2014). «Traduction pragmatique, linguistique de corpus, traducteur : un ménage à trois explosif ?», Tralogy [En ligne], Tralogy II, Session 3 - Machine and Human Translation: Finding the Fit? / TA et Biotraduction, mis à jour le : 02/06/2014
- LAVIOSA Sara (1998), The Corpus-based Approach: A New Paradigm in. Translation Studies. Volume 43, numéro 4.
- LOOCK, Rudy (2016). La traductologie de corpus. Villeneuve-d'Ascq : Presses Universitaires Du Septentrion, 261 pages
- PEARSON, Jennifer (1998) : Terms in Context. Amsterdam : John Benjamins Publishing Company. (Studies in corpus linguistics ; 1).
- PERALDI, Sandrine (2011) : Indétermination terminologique et multidimensionnalité dans le domaine de la chimie organique. Thèse sous la direction de John Humbley, Université Paris-Diderot.

Merci de votre
attention!

Des questions?

Contacts :

julie.remfort@ssi.gouv.fr
sandrine.peraldi@ucd.ie

Sources

- ▷ Modèle PowerPoint par [SlidesCarnival](#)
- ▷ Captures d'écran de SketchEngine (outil conçu par [Lexical Computing](#))
- ▷ Photos par [Unsplash](#)