



**HAL**  
open science

# Quels corpus pour l'analyse contrastive? L'exemple des constructions verbo-nominales de sentiment en français et en russe

Elena Melnikova, Iva Novakova, Olivier Kraif

## ► To cite this version:

Elena Melnikova, Iva Novakova, Olivier Kraif. Quels corpus pour l'analyse contrastive? L'exemple des constructions verbo-nominales de sentiment en français et en russe. 6e Journées de la Linguistique de Corpus, 2009, Lorient, France. pp.197-208. hal-01073699

**HAL Id: hal-01073699**

**<https://hal.science/hal-01073699v1>**

Submitted on 11 Mar 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# QUELS CORPUS POUR L'ANALYSE CONTRASTIVE ? L'EXEMPLE DES CONSTRUCTIONS VERBO-NOMINALES DE SENTIMENT EN FRANÇAIS ET EN RUSSE

Elena Melnikova<sup>1,2</sup>, Iva Novakova<sup>2</sup> et Olivier Kraif<sup>2</sup>

1. Département des langues romanes – Université d'état d'Astrakhan (Russie)

2. LIDILEM – Université de Grenoble (Stendhal-Grenoble 3)

## RÉSUMÉ

Les corpus comparables et parallèles constituent une base de données solide pour une analyse contrastive fiable. Nous nous référons aux deux types de corpus pour étudier les constructions verbo-nominales de sentiment et leurs fréquences en français et en russe. Le corpus comparable nous a servi à calculer les fréquences des noms de sentiment afin de définir la sélection des noms pour notre étude. Grâce à lui, nous avons pu également analyser les fréquences des verbes par rapport aux constructions verbo-nominales. Le corpus parallèle (ayant constitué notre corpus de contrôle) nous a permis d'établir les équivalents fonctionnels des constructions françaises en russe (par ex. les constructions impersonnelles, les verbes de sentiments, les constructions verbo-nominales de sentiment) et inversement. Ces deux types de corpus sont complémentaires en linguistique contrastive.

## 1 LA PROBLÉMATIQUE ET LES OBJECTIFS

L'analyse contrastive qui a pour but de décrire les similitudes et les différences entre les langues nécessite la constitution de corpus bilingues de différents types, à savoir des corpus *comparables* et des corpus *parallèles* (i.e. des corpus de traduction) (Williams 2005, Degand 2005, Celle 2006, Dalbera 2003, Lewis 2005). Cet article a pour objectif d'explicitier l'apport de ces deux types de corpus pour l'étude de la fréquence des constructions verbo-nominales de sentiment (CVN\_sent) en français et en russe. Nous défendons ici l'idée que le corpus comparable et le corpus parallèle peuvent servir de base à une analyse contrastive qui vise à découvrir, à comparer et à mettre en évidence l'emploi des CVN\_sent dans les deux langues.

Chaque type de corpus a ses avantages et ses inconvénients. Les corpus comparables, composés de textes originaux dans deux ou plusieurs langues qui respectent les mêmes critères de genre, de registre, de public visé, d'époque, ne permettent pas d'établir des équivalences de façon simple. Quant aux corpus de traductions (parallèles), qui offrent un accès direct aux équivalences, on leur reproche souvent de contenir des traces de la langue source et, du fait de ces artefacts traductionnels, de ne pas être totalement fiables. Ces deux types de corpus présentent donc des caractéristiques complémentaires<sup>26</sup>.

Selon nos premières observations et en nous référant à la recherche de G. Gak (1983, [2006]), nous avançons l'hypothèse que les CVN\_sent sont moins fréquentes en russe que les verbes distributionnels de sentiment en tant qu'équivalents de la structure *Verbe+N\_sent* en français. Le russe dispose d'un aspect morphologique encodé dans le verbe (*zlit'sja* -> être en

<sup>26</sup> Cf. aussi à ce sujet, Degand (2005).

*colère et razozlit'sja -> tomber en colère*) Le français, au contraire, aurait tendance à utiliser plus de CVN, du fait de l'absence d'un aspect verbal morphologique.

Nous allons vérifier notre hypothèse dans les corpus comparables et parallèles.

En premier lieu, nous étudierons le statut des deux types de corpus pour l'analyse contrastive. En deuxième lieu, nous essaierons de clarifier l'apport des deux types de corpus pour la sélection des N\_sent et aussi pour l'analyse de la fréquence des équivalents fonctionnels des CVN\_sent français en russe.

## 2 LA CONSTITUTION DES CORPUS ET LA MÉTHODOLOGIE

Pour le corpus comparable (Tableau 1) nous avons utilisé les bases de données Frantext (<http://atilf.atilf.fr/frantext.htm>) et Ruscorpora (<http://www.ruscorpora.ru/>): 2029 textes d'auteurs contemporains entre 1950 et 2007 (environ 60 millions de mots), qui ont fourni 22 942 occurrences de CVN\_sent en français et en russe.

	Français	Russe
Bases de données	Frantext	Ruscorpora
Genre	Romans, récits de voyage, essais, correspondance, mémoires	Romans, récits, essais, nouvelles, correspondance
Epoque	1950 – à nos jours	1950 – à nos jours
Nombre de mots	29 670 087	29 618 550
Nombre d'occurrences (Vsup + N_sent et N_sent+Vsup) retenues de l'échantillon	6 514 (sur 12 772)	4 853 (sur 10 170)

Tableau 1 : Corpus comparable

Pour constituer le corpus parallèle (Tableau 2) et aligner les textes, nous avons utilisé le logiciel Alinea (cf. Kraif 2006, et <http://w3.u-grenoble3.fr/kraif/>). Les textes parallèles (français-russes) sont des textes d'auteurs du XIX<sup>e</sup> et XX<sup>e</sup> siècles (libres de droits), traités dans leur version originale et leur traduction. Nous avons réuni un corpus d'environ 10 millions de mots (84 textes) qui nous a fourni 1500 occurrences de CVN\_sent dans les deux langues :

	Français - Russe		Russe - Français	
Genre	Romans, récits		Romans, récits	
Epoque	1830 - 1930		1830 - 1930	
Nombre de textes	27 + 27 traductions		15 + 15 traductions	
Nombre de mots	2 007 048	2 619 016	3 674 812	2 717 174
Nombre d'occurrences (Vsup+N_sent) retenues de l'échantillon	750		750	

Tableau 2 : Corpus parallèle

L'alignement des textes russes dans *Alinea* présente certains aspects techniques spécifiques comme l'utilisation de listes de transfert (pour amorcer l'extraction des points d'ancrage) ou la translittération des textes en cyrilliques (pour identifier les noms propres et les mots apparentés).

La constitution du corpus parallèle nécessite la collecte des textes en version originale et des textes traduits et ceci dans les deux langues. Ces textes sont soumis à l'alignement (alinéa par alinéa) et sont analysés, par la suite, à l'aide de l'extraction. L'exploitation des textes parallèles (original et traduit) s'effectue en plusieurs étapes : le prétraitement avant l'alignement, l'alignement, l'élaboration des requêtes et l'analyse des résultats. Le prétraitement consiste d'abord à translittérer les textes russes. Nous utilisons ici le système le plus répandu de translittération des caractères cyrilliques russes (ISO/R 9 : 1968), repris par GOST (1983). Ce changement des cyrilliques en lettres latines nous facilite la reconnaissance des chaînes apparentées (cognats ou emprunts) par le logiciel. Ensuite, si les textes téléchargés contiennent des irrégularités liées aux conversions de format (présence de macros, présence de retour-chariots à la fin de chaque ligne, etc.), il faut les éliminer en procédant à un nettoyage (qui permet de restaurer des lignes coupées, etc.). Après avoir étiqueté les textes prétraités à l'aide du logiciel Treetagger, on lance le logiciel Alinea, qui crée le projet des textes alignés. L'étape suivante est l'extraction des points d'ancrage. Ce sont les "transfuges" : des équivalents traductionnels identiques qui sont souvent nombreux comme les nombres, les noms propres, les sigles, les emprunts, les mots apparentés etc. L'étape finale est l'extraction de l'alignement complet (Figure 1).

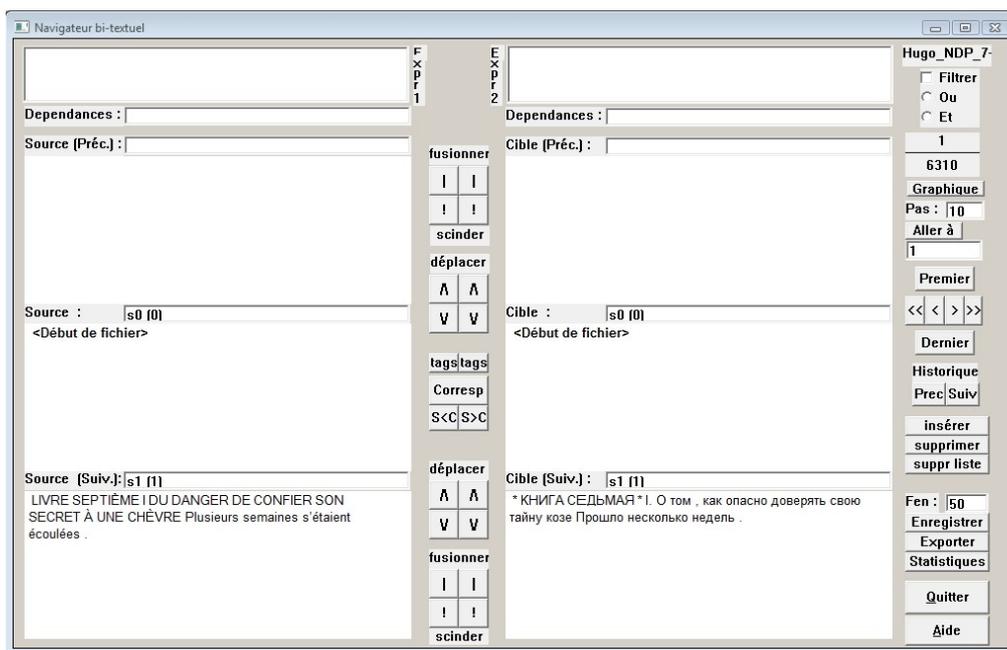


Figure 1

Le projet des textes alignés est visualisé sous forme des deux colonnes de textes français – russes, répartis en alinéas où chaque binôme (les petites fenêtres à gauche et à droite) représente une partie minimale (alinéa ou phrase) du texte original et sa traduction. Toutefois l'alignement peut donner des binômes irréguliers, ce qui demande leur ajustement (le regroupement pour mettre les segments en correspondance). Une fois l'alignement effectué, une requête peut être formulée autour d'un nom de sentiment, selon une formule intégrant des

expressions régulières, du type « </^colère.\*> ». Toutes les occurrences trouvées peuvent être exportées en un seul fichier (Figure 2).

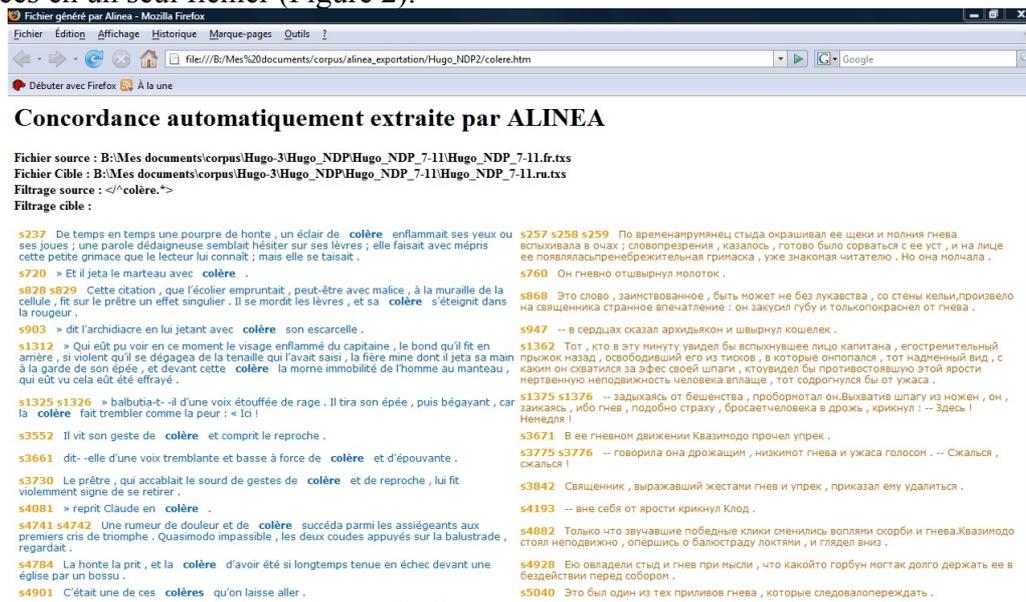


Figure 2

L'identification du patron syntaxique  $V + N_{sent}$ , qui nous a servi pour l'extraction des CVN<sub>sent</sub> dans les textes alignés, exige une analyse minutieuse des résultats recueillis.

Le corpus de traduction, collecté grâce à Alinea, présente des avantages, ainsi que des inconvénients qui apparaissent au cours du prétraitement des textes, de l'alignement et de l'analyse des résultats. Le premier grand inconvénient est le grand nombre de manipulations de prétraitement des textes qui impose un travail à long terme. La seconde difficulté est due au fait que la majorité des textes traduits sont soumis aux droits d'auteur. La conséquence est que nous avons été obligés d'utiliser des textes du XIX<sup>e</sup> et du début de XX<sup>e</sup>, libres de droits. Toutefois, malgré ces difficultés, Alinea permet de mener la recherche à l'aide de différentes formules de requête, d'obtenir des résultats immédiats à travers un grand nombre de textes, de rechercher et d'extraire des constructions syntaxiquement pertinentes.

Dans la collecte des données, nous avons été aussi confrontés à d'autres problèmes comme le fait que les textes dans les deux types de corpus relèvent d'époques différentes (à cause des problèmes de droits d'auteurs) et aussi qu'ils ne sont pas de taille comparable (pour la même raison).

Conformément à notre objectif, nous allons étudier l'apport de chaque type de corpus. Nous procéderons à des analyses quantitatives des données bilingues. Dans un premier temps, nous étudierons la fréquence des N<sub>sent</sub> sélectionnés dans les deux langues (corpus comparable). Dans un deuxième temps, nous calculerons le taux de fréquences des CVN<sub>sent</sub> françaises par rapport aux verbes, aux CVN et aux autres constructions en russe (corpus parallèle).

### 3 L'APPORT DES DEUX CORPUS POUR L'ANALYSE CONTRASTIVE

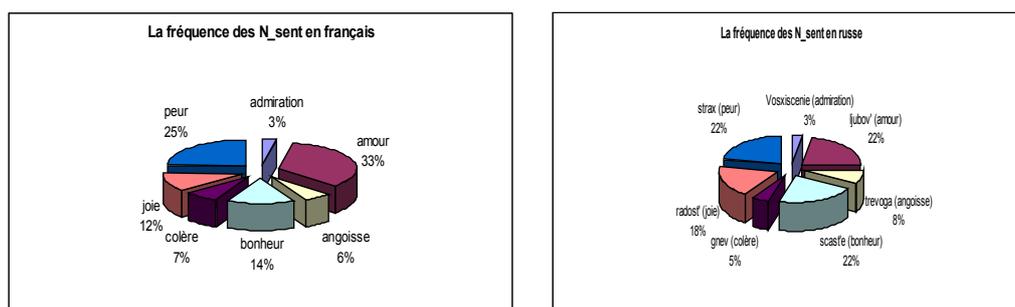
#### 3.1 La sélection des N<sub>sent</sub> et leurs fréquences dans les CVN (corpus comparable)

Nous avons extrait les occurrences des sept N\_sent sélectionnés selon le critère de fréquence en français et en russe (*admiration / vosxiščenie, amour/ ljubov', angoisse / trevoga, bonheur / sčast'ie, colère / gnev, joie / radost', peur / strax*)<sup>27</sup> à l'aide du patron syntaxique *V+N\_sent* dans les corpus comparables monolingues Frantext et Ruscorpora (Tableau 3) :

N_sent français	Nombre d'occ.	N_sent russe	Nombre d'occ.
admiration	619	Vosxiščenie	767
amour	6 803	ljubov'	5 958
angoisse	1 163	trevoga	2 174
bonheur	2 815	sčast'e	5 821
colère	1 468	gnev	1 245
joie	2 324	radost'	4 929
peur	4 937	strax	6 063

Tableau 3 : La fréquence de sept N\_sent (Frantext & Ruscorpora)

Le pourcentage des fréquences absolues pour *admiration* (3%) / *vosxiščenie* (3%), *angoisse* (6%) / *trevoga* (8%) et *colère* (7%) / *gnev* (5%) est presque le même dans les deux langues. Les autres noms manifestent un léger décalage dans leur fréquence, ce qui est probablement dû aux spécificités des deux corpus. Ainsi *sčast'e* (*bonheur*) est plus fréquent dans le russe (22%) que *bonheur* dans le corpus français (14%), tandis que *ljubov'* (*amour*) est moins fréquent (22%) qu'*amour* (33%) (Figures 3 et 4).



Figures 3 et 4

Les fréquences des CVN russes par rapport aux CVN en français dans le corpus comparable montrent que le russe utilise plus de verbes que de CVN. En français, inversement c'est les CVN qui sont plus fréquentes que les verbes pour la plupart des N\_sent étudiés (Figures 5 et 6).

<sup>27</sup> Tous les exemples russes sont translittérés en lettres latines (ISO/R 9 : 1968, repris par GOST 1983) pour faciliter la lecture aux non-russophones.

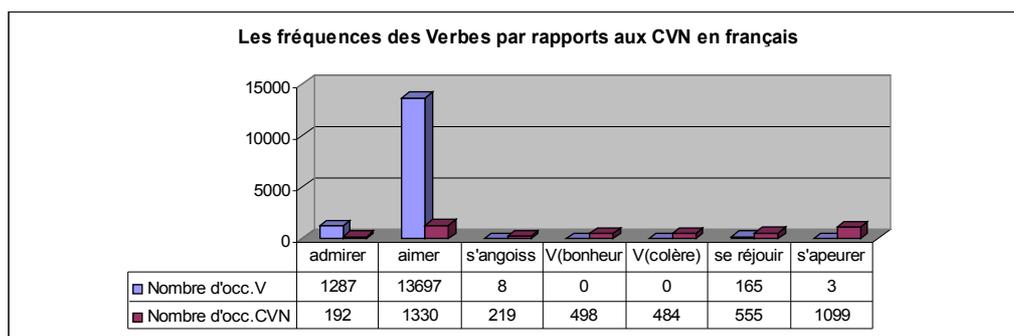


Figure 5

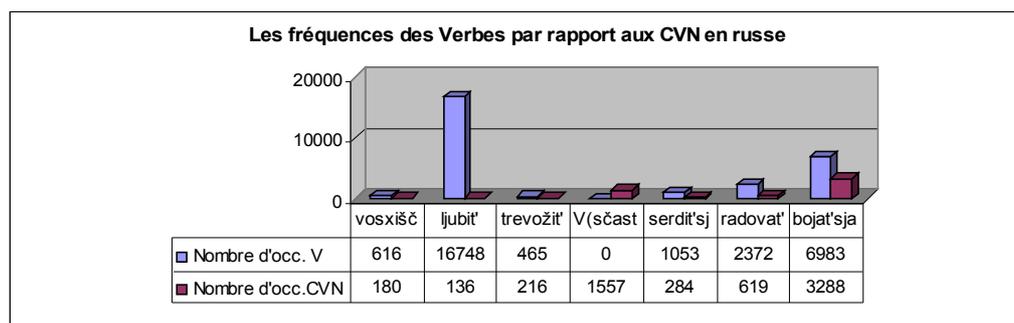


Figure 6<sup>28</sup>

En français, sur les sept N\_sent étudiés, cinq (*angoisse, bonheur, colère, joie, peur*) apparaissent le plus fréquemment dans les CVN. En revanche, en russe les verbes correspondant aux N\_sent *vosxiščenie (admiration), ljubov (amour), trevoga (angoisse), gnev (colère), radost' (joie) et strax (peur)*,<sup>29</sup> présentent des taux de fréquences plus élevées que les CVN. Ce fait confirme notre hypothèse sur l'emploi plus fréquent de verbes que de CVN en russe. Pour *sčast'e* et *bonheur* il n'existe pas de verbes correspondants dans les deux langues.

### 3.2 L'extraction des équivalents fonctionnels (corpus parallèle)

Nous allons maintenant tester notre hypothèse sur le corpus parallèle. Celui-ci nous a permis d'établir la liste des équivalents fonctionnels des CVN françaises en russe : CVN, verbes distributionnels et constructions impersonnelles (Tableau 4) :

<sup>28</sup> Pour le calcul des verbes russes on a pris en compte les verbe imperfectifs et perfectifs les plus fréquents : *vosxišat'sja / vosxitit'sja (admirer), ljubit' / poljubit' (aimer), trevožit'sja / vstrevožit'sja (s'angoisser), serdit'sja / rasserdit'sja* (« être en colère »), *radovat'sja / obradovat'sja (se réjouir), bojat'sja / ispugat'sja* (« avoir peur » qui n'a pas d'équivalents verbaux fréquents en français).

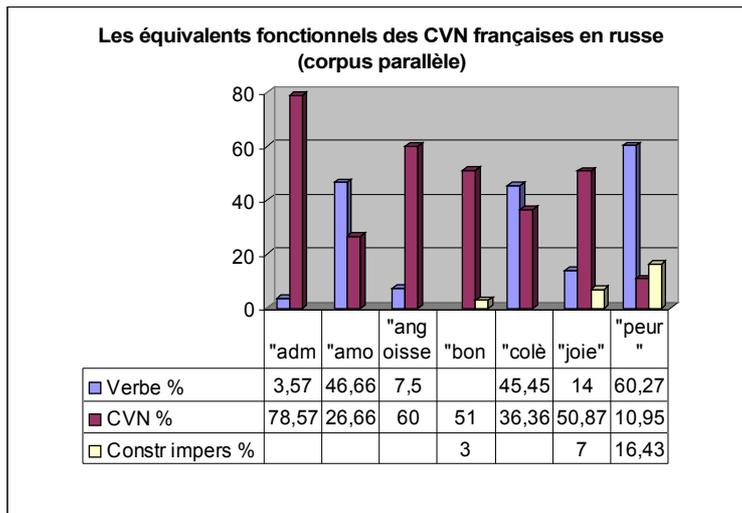
<sup>29</sup> On observe que *amour* et *ljubov'* sont le plus souvent utilisés dans les corpus sous forme de verbes (*aimer* et *ljubit'*) et ce, dans les deux langues.

En français		En russe		
CVN	Verbes distributionnels	CVN	Constructions impersonnelles	Autres
Avoir de l'admiration	Vosxiščat'ja (admirer) <b>3,57%</b>	Ispytyvat' vosxiščenie (Eprouver admiration (acc)) <b>78,57%</b>	∅	17,86%
Avoir de l'amour	Ljubit' (aimer) <b>46%</b>	Čuvstvovat' ljubov' (Sentir amour (acc)) <b>27%</b>	∅	27%
Avoir de l'angoisse	Trevožit'sja, bojat'sja (s'angoisser) <b>7,5%</b>	Počuvstvovat' trevogu (sentir angoisse (acc)) <b>60%</b>	∅	32,5%
Avoir le bonheur	∅	Imet' sčast'e (avoir bonheur (acc)) <b>51%</b>	Emu posčastlivilos' (à lui avoir du bonheur) <b>3%</b>	46%
Être en colère	Serdit'sja (se colérer) <b>45%</b>	Byt' v gneve (être en colère (loc)) <b>36%</b>	∅	19%
Avoir de la joie	Radovat'sja (se réjouir) <b>14%</b>	Ispytyvat' radost' <b>51%</b>	Emu radostno (à lui (dat) joyeusement) <b>7%</b>	28%
Avoir de la peur	Bojat'sja (s'apeurer) <b>60%</b>	Počuvstvovat' strax (Sentir peur (acc)) <b>11%</b>	Emu strašno (à lui peureusement) <b>16%</b>	13%

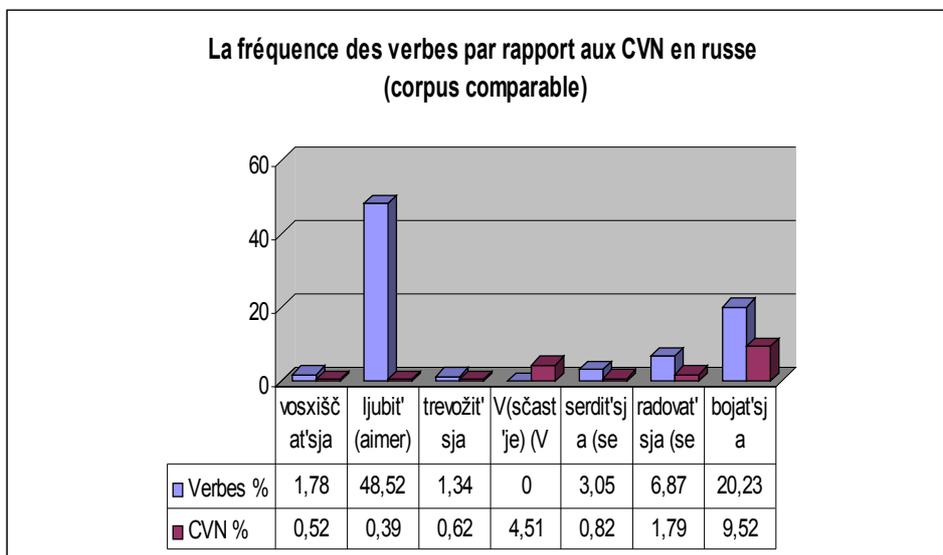
Tableau 4 : Les équivalents fonctionnels des CVN françaises en russe

Les décomptes effectués sur le corpus de traductions (Tableau 4) montrent que *admiration* (78,57%), *angoisse* (60%), *bonheur* (51%) et *joie* (51%) apparaissent souvent dans des CVN (plus de 50%) en russe. En revanche, pour *amour*, *colère* et *peur*, le taux d'équivalents verbaux augmente sensiblement : 46% de verbes contre 27% de CVN (*amour*), 45% de verbes contre 36% de CVN (*colère*), 60% d'équivalents verbaux contre 11% de CVN (*peur*). Le N\_sent *sčast'e* (*bonheur*), quant à lui, fait souvent partie des CVN, car il n'existe pas d'équivalents verbaux pour ce N\_sent ni en français, ni en russe. Les constructions impersonnelles ont un taux important de fréquence pour *peur* (16%) et *joie* (7%).

Si l'on compare les données des deux graphiques (figure 7 et 8), nous pouvons voir les similitudes et les divergences dans les tendances calculées à partir des différents corpus.



*Figure 7<sup>30</sup>*



<sup>30</sup> Les pourcentages des autres cas d'équivalents non-réguliers (données du Tableau 4) n'ont pas été introduits dans le graphique. Ce choix a été fait pour mieux mettre en évidence les principaux équivalents.

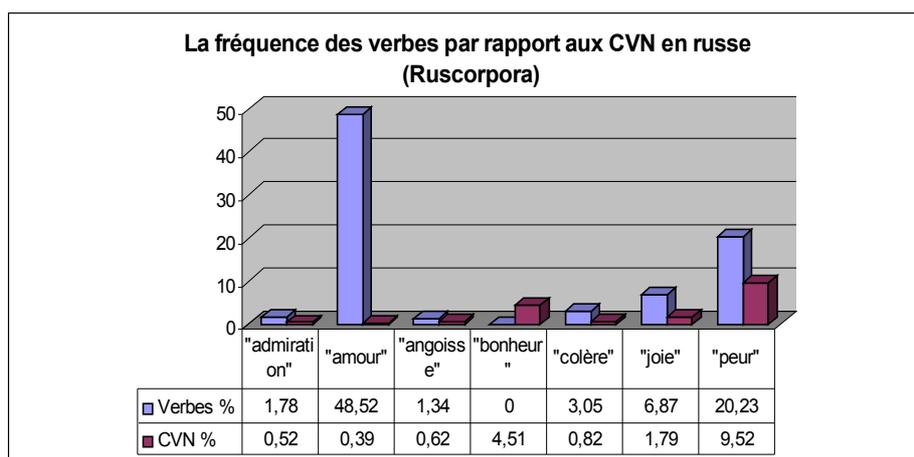


Figure 8<sup>31</sup>

Ainsi, le taux des fréquences élevées des équivalents verbaux pour *ljubov'* (*amour*), *gnev* (*colère*) et *strax* (*peur*) sont confirmées dans les deux corpus. En revanche, les données observées dans Ruscorpora (Figure 8) nous permettent de constater que *vosxiščenie* (*admiration*), *trevoga* (*angoisse*) et *radost'* (*joie*) ont également, en majorité, des équivalents verbaux. Ces résultats sont contrastés par rapports au corpus parallèle. Cette différence entre les données de corpus parallèle et celles du corpus russe (comparable) peut être due, à notre avis, aux aléas du corpus parallèle (les artefacts de la traduction des textes du XIX<sup>e</sup> siècle, une forte influence de la langue source, les époques différentes du texte original et de sa traduction, la petite taille du corpus parallèle, ce qui augmente le risque de fluctuations liées à des phénomènes aléatoires).

Voici quelques exemples pour illustrer ce que nous venons d'exposer précédemment. La CVN *avoir peur* en français est traduite fréquemment en russe par un verbe *bojat'sja* (imperf.) ou *ispugat'sja* (perf.), (1) :

- (1) Texte orig. :- Elle eut peur d'être battue ;... (E. Zola. *Thérèse Raquin*)  
 Trad. :- Она испугалась, что он побьет ее ; ...  
 Translitt. :- *Ona ispugalas', čto on pob'ët eë, ...*  
 Litt. :- Elle avoir peur (passé, perf.), qu' il battre (futur, perf.) la (gén.), ...

L'équivalent de CVN\_amour est souvent un verbe *ljubit'* (imperf.) – *poljubit'* (perf.) ou *razljubit'* (perf. négatif) comme en (2):

- (2) Texte orig. :- C'en est fait, se dit-il avec désespoir, elle n'a plus d'amour pour moi, ... (Stendhal. *La chartreuse de Parme*)  
 Trad. :- "Конец! - подумал он с ужасом. - Она разлюбила меня...  
 Translitt. :- *Konec ! – podumal on s užasom – Ona razljubila menja...*  
 Litt. :- « Fin ! – penser (passé, perf.) lui (nom.) avec terreur (instr.). – Elle ne plus aimer (passé, perf.) moi (acc.)

Le corpus parallèle nous a permis d'identifier un autre type d'équivalent russe des CVN françaises, à savoir les constructions impersonnelles. Cette construction est présentée souvent

<sup>31</sup> Les données de Figure 8 reflètent les données de la Figure 6. Les pourcentages de ce graphique ont été calculés automatiquement par le système Excel qui a pris pour 100% la totalité des verbes et des CVN de tous les N\_sent.

par le pronom personnel au datif et un adverbe de sentiment (*mne radostno – à moi joyeusement*) comme en (3):

(3) Texte orig. : - ... *j' éprouve de la joie à entendre le bruit franc et net que font nos semelles sur le fond de terre dure ...* (H. Barbusse. *Le feu*)

Trad. :- ... *мне радостно слышать отчётливый звук наших шагов по твердому грунту...*

Translitt. :-... *mne radostno slyšat' otčëtlivyj zvuk našix šagov po tverdomu gruntu*

Litt. :- ... *moi (dat.) joyeux entendre distinct son (acc.) nos pas (gén.) sur dur sol (dat.)...*

Pour résumer, nous avons réuni les résultats des deux corpus dans le tableau ci-dessous :

N_sent	Corpus russe (Ruscorpora)		Corpus parallèle	
	Verbe	CVN	Verbe	CVN
Vosxiščenie (admiration)	1,78% +	0,52% -	3,57% -	78,57% +
Ljubov' (amour)	48,52% +	0,39% -	46,66% +	26,66% -
Trevoga (angoisse)	1,34% +	0,62% -	7,5% -	60% +
Sčast'e (bonheur)	0	4,51%	0	51%
Gnev (colère)	3,05% +	0,82% -	45,45% +	36,36% -
Radost' (joie)	6,87% +	1,79% -	14% -	50,87% +
Strax (peur)	20,23% +	9,52% -	60,27% +	10,95% -

Tableau 5 : Taux des verbes et des CVN dans les deux corpus en russe

Dans le corpus parallèle trois N\_sent (*ljubov (amour), gnev (colère) et strax (peur)*) apparaissent plus fréquemment sous la forme du verbe correspondant, plutôt qu'au sein d'une CVN. Les trois autres *vosxiscenie (admiration), trevoga (angoisse) et radost' (joie)* présentent des fréquences élevées dans des CVN par rapport aux verbes correspondants<sup>32</sup>. Nous expliquons ce résultat contrasté surtout par l'influence sur la traduction russe du texte source français, où les N\_sent apparaissent souvent sous forme de CVN. Le corpus comparable russe révèle, comme nous l'avons vu, une prédominance de verbes et ce, pour pratiquement tous les N\_sent étudiés.

## 4 CONCLUSIONS

Les corpus comparables et parallèles sont complémentaires et présentent plusieurs avantages, mais aussi des inconvénients (Tableau 6) :

<sup>32</sup> Nous avons enlevé ici *bonheur* de nos analyses car il n'existe pas de verbes correspondants dans les deux langues pour ce nom.

Corpus comparables		Corpus parallèles	
+	-	+	-
L'accès de manière directe aux occurrences de N_sent, la constitution des corpus de taille importante et le <b>calcul des fréquences</b> d'un N_sent, des verbes et des CVN.	ne permettent pas d'établir des équivalents fonctionnels	permettent <b>d'établir les équivalents fonctionnels</b> des CVN_sent français en russe et inversement	La langue des textes traduits contient souvent des traces de la langue source

Tableau 6 : La complémentarité des deux corpus

L'hypothèse émise est vérifiée en partie. Selon le corpus comparable, six N\_sent (à l'exception *bonheur* et *sčast'e*) ont des équivalents verbaux en russe. En revanche, selon le corpus parallèle, c'est uniquement trois N\_sent (*ljubov'* (*amour*), *gnev* (*colère*) et *strax* (*peur*)) ont des équivalents verbaux en russe par rapport aux CVN françaises. Le russe, malgré l'aspect verbal morphologisé, utilise aussi bien des CVN\_sent pour *vosxiščenie* (*admiration*), *trevoga* (*angoisse*), *bonheur* (*sčast'e*), *radost'* (*joie*). Ceci est surtout révélé par le corpus parallèle, ce qui peut être dû à l'influence de la langue source. Par ailleurs, il s'ensuit que le russe, malgré l'aspect verbal morphologisé, emploie aussi bien les verbes de sentiment que les CVN\_sent. Les spécificités des deux langues se révèlent à travers les calculs sur les deux corpus: 1) le corpus comparable et le corpus parallèle ont laissé découvrir les équivalents « absolus » des trois N\_sent (*ljubov'* (*amour*), *gnev* (*colère*), *strax* (*peur*)); 2) le corpus parallèle a permis d'identifier un autre type d'équivalents, les constructions impersonnelles en russe.

Les résultats provenant des deux types de corpus ne sont pas identiques et peuvent être liés à de nombreuses causes : 1) l'artefact de la traduction ; 2) une forte influence de la langue source ; 3) les époques différentes des textes de deux corpus ; 4) les époques différentes des textes originaux et de leurs traductions dans le corpus parallèle ; 5) la petite taille du corpus parallèle, plus difficile à obtenir du fait des droits d'auteur liés à la traduction ; 6) les textes issus d'un seul genre littéraire, et un large emploi de temps de narration (l'imparfait et le passé simple) ; 7) les systèmes différents du français et du russe.

La comparaison des résultats des deux types de corpus nous a amenés à réfléchir sur le statut des équivalents fonctionnels établis: sont-ils des épiphénomènes liés à l'idiosyncrasie de l'auteur/traducteur (« posture sourcière ») ou bien des correspondances relevant du système des deux langues sur le plan sémantique, syntaxique et morphologique. Nous estimons, à l'état actuelle de notre recherche que certes, l'influence de la « posture sourcière » est assez forte, mais que certainement l'hypothèse avancée suite aux recherches de Gak (1983, [2006]) et à nos propres intuitions mérite d'être nuancée.

Cette étude laisse encore des pistes à explorer. Notamment, vérifier la fiabilité des données issues du corpus parallèle en croisant les données dans les deux directions (du français vers le russe et du russe vers le français), et observer les fréquences d'emploi de différents équivalents fonctionnels dans les corpus comparables, par un va-et-vient constant entre les deux types de corpus. En outre, il serait également intéressant d'observer l'apport des deux corpus pour l'étude de l'aspectualité des N\_sent au sein des CVN\_sent. Le calcul des paramètres aspectuels (l'aspect lexical et grammatical des verbes, et aussi les classifieurs, les modifieurs et les déterminants des N\_sent) sera effectué à partir de la combinatoire syntaxique et lexicale des CVN\_sent dans les deux langues.

## 5 RÉFÉRENCES

- Celle A. (2006). *Temps et modalité. L'anglais, l'allemand, le français en contraste*. Allemagne : Presses scientifiques européennes.
- Dalbera J-Ph. « Le corpus entre données, analyse et théorie », *Corpus* [En ligne], n°1 | novembre 2002, mis en ligne le 15 décembre 2003, Consulté le 24 juin 2009. URL : <http://corpus.revues.org/index10.html>
- Degand L. (2005). « De l'analyse contrastive à la traduction : Le cas de paire puisque – aangezien ». *La linguistique de corpus*. Rennes : Presses universitaires de Rennes. p. 155-168.
- Gak V. ([1983] (2006). *Sravnitelnja tipologija francuzskogo i russkogo jazykov (La typologie contrastive du français et du russe)*. Moskva : URSS.
- Kraif O. (2006). « Qu'attendre de l'alignement de corpus multilingues ? ». *Revue Traduire, 4e Journée de la traduction professionnelle*, Société Française des Traducteur, N° 210, p. 17-37.
- Lewis D. M. (2005). « Corpus comparable et analyse contrastive : l'apport d'un corpus de français / anglais de discours politique à l'analyse des connecteurs adversatifs ». Dans G. Williams *La linguistique de corpus*. Rennes : Presses universitaires de Rennes.
- Williams G. (2005). *La linguistique de corpus*. Rennes : Presses universitaires de Rennes.
- Frantext*. CNRS, ATILF (Analyse et traitement informatique de la langue française), UMR CNRS-Université Nancy2, <http://atilf.atilf.fr/categ.htm>
- Ruscorpora*. ([www.ruscorpora.ru](http://www.ruscorpora.ru))