



***Traduire le polar* : une étude textométrique
comparée de la phraséologie du
roman policier en français source et cible**

Olivier Kraif

De la phraséologie aux genres textuels : état de recherches et perspectives méthodologiques - Cracovie - 18-19 Oct. 2016

Etat des lieux

- Au départ de notre enquête, une interrogation : la traduction est-elle une activité **suspecte** ? Faut-il bannir les traductions de nos études sur le genre ?

*Il y a une objection essentielle aux corpus parallèles. Les traductions, quelles que soient leur qualité et leur quasi-perfection (ce qui est cependant rare), ne peuvent donner **qu'une image déformée de la langue** qu'elles représentent. Les linguistes ne devraient jamais se fier à des traductions lorsqu'ils décrivent une langue. C'est pourquoi **les traductions n'ont pas leur place dans les corpus de référence**. Plus qu'elles ne représentent la langue dans laquelle elles sont écrites, elles donnent **une image en miroir de leur langue source**. [nous soulignons]*

(Wolfgang Teubert, 1996 : 247)

Etat des lieux

*De même, les traducteurs ont tendance à normaliser la ponctuation, même lorsqu'elle est de nature expérimentale dans le texte original (May 1997; Malmkjær 1997), et les interprètes ont tendance à terminer les phrases laissées en suspens, à grammaticaliser les énoncés agrammaticaux et à omettre les hésitations et les faux départs même si l'orateur les produit volontairement (Shlesinger 1991 : 150; BerkSeligson 1990). **Cette tendance à «normaliser» la langue de traduction et à suivre de près les conventions de la langue cible** laisse croire, encore une fois, que les traducteurs répondent inconsciemment à la perception qu'ils ont du statut du texte ou de l'énoncé qu'ils produisent (Mona Baker, 1998 : 4)*

Enquête préliminaire

- La langue traduite plus pauvre ?
 - Deux exemples (Teubert, Ibid.) :
 - Au plan lexical : *Schadenfreude* → pas d'équivalent lexical en anglais, donc sans doute plus rare dans des traductions allemandes de textes anglais
 - Au plan syntaxique : passif impersonnel « Es wurde viel getrunken » sans équivalent en anglais, et donc sans doute absent ou plus rare de la langue traduite (en → de)

Enquête préliminaire

- La langue traduite plus pauvre ?

Observations tirées d'un corpus parallèle fr-de

Langue source	Occurrences de <i>Schadenfreude</i>	Taille du corpus parallèle
de	1	628 029
en	5	87 596
fr	21	17 079 606
sw	1	444 076

Enquête préliminaire

- La langue traduite plus pauvre ?

Recherche du passif impersonnel : *es + wurde + ADV + ParticipePassé* et *es + wurde + ParticipePassé*. On trouve seulement deux occurrences, dans des **textes traduits** :

Vorbeigehenden Besatzungsmitgliedern entging das nicht, und es wurde bereits getuschelt.

Des hommes d'équipage s'en rendirent compte au passage, et les murmures commencèrent.

Tom Clancy (1984) *The Hunt for Red October*

Wir stritten uns über den jeweiligen Kurs, es wurde gelacht und gesungen.

On disputa des courses, il y eut des rires, des chansons.

Simone de Beauvoir (1958) *Mémoires d'une jeune fille rangée*

La piste textométrique

- La situation est sans doute plus complexe et les biais traditionnels à nuancer
- Nous proposons une confrontation entre deux corpus de romans policiers :
 - romans policiers français
 - romans policiers écrits en anglais (GB/US) traduits en français
- Etude textométrique sur critères lexicaux, syntaxiques, et phraséologiques

Pièces à conviction

- Corpus FR :
 - 68 romans publiés entre 1956 et 2012 (63 ultérieurs à 1990)
 - 5 757 352 tokens
 - Auteurs : Belin, Aubert, Bannel, Brussolo, Chattam, Dard, de Villiers, Dijan, Giesbert, Grangé, Halter, Izzo, Jarrige, Lemaitre, Macouin, Magnan, Mauwls, Morillon, Oppel, Parot, Simenon, Thilliez, Vargas, Viau
- Corpus FR-trad :
 - 48 romans publiés entre 1955 et 2007 (40 ultérieurs à 1990)
 - 8 003 923 tokens
 - Auteurs : Clancy, Connelly, Cornwell, Crais, Crichton, Devon, Ellis, Evanovich, Grisham, Harris, Higgins Clack, Highsmith, Follett, Ludlum, Patterson, Reichs, Winstead Jones
- Sous-genres variés : roman noir, thriller, espionnage, judiciaire, historique

Indices lexicaux

Au plan lexical

- Pour comparer les distributions lexicales, calcul d'une mesure de similarité (cosinus, cf. Salton,)
 - Echantillonnage de POL-fr x POL-fr-trad en 17x17 blocs de tailles similaires. Les blocs contiennent des auteurs différents
 - Comparaison de tous les blocs 2 à 2. Pour chaque bloc, on retient les 3 blocs les plus similaires
 - On compte les appariements POL-fr/POL-fr, POL-fr/POL-fr-trad, POL-fr-trad/POL-fr, POL-fr-trad/POL-fr-trad
 - Un test de Khi2 permet de déterminer si les deux variables sont indépendantes

Au plan lexical

- Non rejet de l'hypothèse d'indépendance pour l'ensemble du lexique ($p=0,16$) hors entités nommées.
- Mais une situation contrastée en fonction des parties du discours :
 - pour les noms, association très marquée entre POL-fr x POL-fr, et POL-fr-trad x POL-fr-trad ($p(\text{khi}2)=8,87 \cdot 10^{-7}$)

	POL-fr	POL-fr-trad
POL-fr	31	20
POL-fr-trad	7	44

Au plan lexical

- Association significative pour les adjectifs ($p=0,0017$) et dans une moindre mesure les adverbes ($p=0,03$), tendancielle pour la ponctuation.
- Pas d'association pour le verbes et les autres parties du discours (pronoms, prépositions, déterminants, conjonctions, etc.). Pour les verbes, $p(\text{khi}^2)= 0,16$

	POL-fr	POL-fr-trad	tot
POL-fr	32	19	51
POL-fr-trad	25	26	51
tot	57	45	102

Au plan lexical

- Comment expliquer ces similarités lexicales partielles ?
 - Calcul des noms spécifiques (loglike)
 - On relève tous les noms apparaissant dans la liste des 100 lemmes les plus spécifiques, pour chaque corpus
 - Presque tous les noms peuvent se ranger dans moins d'une dizaine de grandes classes sémantiques
 - Ces classes sont liées à des classes thématiques en rapport avec les sous-genres et les spécificités culturelles

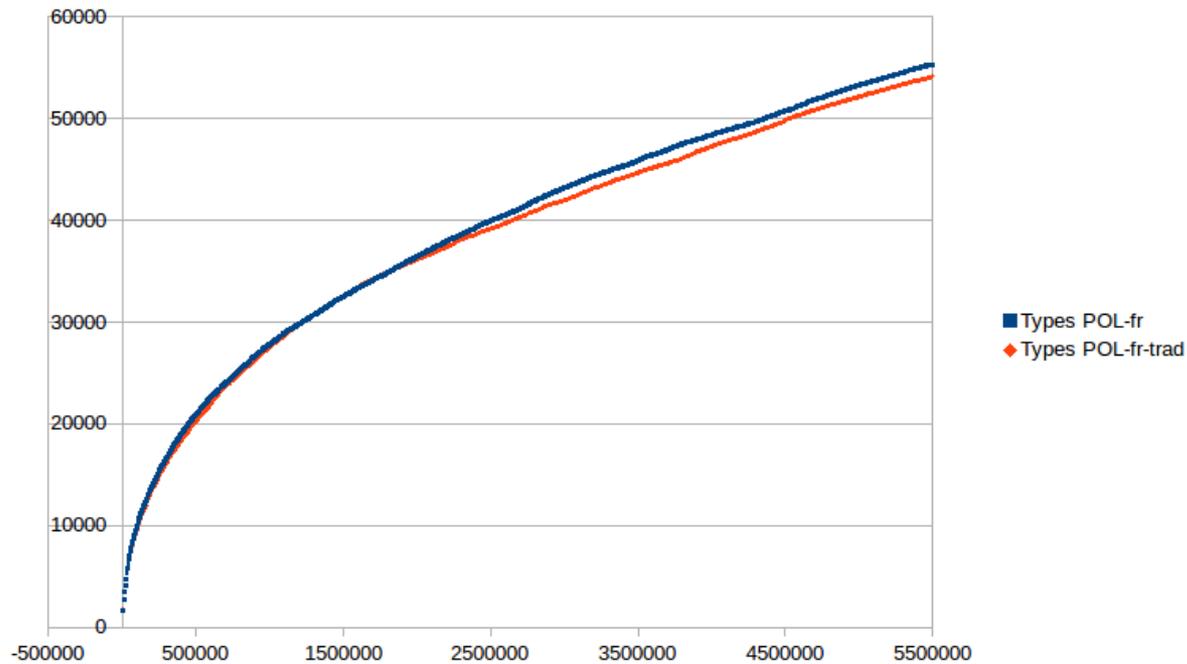
Au plan lexical

POL-fr	POL-fr-trad
<p>POLICE : commissaire, flic, gendarme</p> <p>NATURE : forêt, pierre, chien, loup, bête, nuit, araignée, animal, poussière, ombre, vent, gueule, bois</p> <p>CORPS : chair, corps, ventre, cœur, peau, sang, visage, trait, bouche, lèvres, odeur,</p> <p>HUMAIN : homme, professeur, sœur, femme, vieillard, même, vieux, bourreau, patron</p> <p>EMOTION : peur, cri, haine, délire, souffrance</p> <p>EGYPTE : roi, momie, sarcophage, prêtre</p> <p>MORT : mort, peste</p> <p>CONSTRUCTION : château, muraille, paroi, village</p> <p>ROMAN : manuscrit, roman, romancier</p>	<p>JURIDIQUE : avocat, affaire, tribunal, dossier, procès, jury, audience, société, associé, appel, accord</p> <p>MILITAIRE : commandant, officier, radar, sous-marin, marine, poste, division, armée, air</p> <p>ETAT : bureau, président, secrétaire, sénateur, réunion, cabinet, gouverneur, maison</p> <p>SERVICES : agent, renseignement, service, sécurité</p> <p>ARGENT : dollar, argent</p> <p>ARTEFACTS : téléphone, numéro, caméra, écran, radio, télévision, voiture, avion, appareil,</p> <p>AUTRE : virus, grand-mère, monsieur, journaliste, problème, merci, besoin, endroit, question, moment, gens</p>

Au plan lexical

– Et sur le plan de la richesse lexicale ?

De plus, les deux rapports types/occurrences et densité lexicale sont considérablement plus faibles en anglais de traduction (Mona Baker, 1998)



Au plan lexical

- l'hypothèse de l'homogénéité

*La traduction crée **un troisième code** parce qu'elle est une forme de communication unique, et non parce qu'elle est une forme de communication fautive, déviante ou non conforme à la norme [...] En fait, il existe des preuves de cette tendance dans le Translational English Corpus , mais ce dernier ne comporte pas encore assez de données pour vraiment corroborer des hypothèses fortes : le corpus ne compte qu'environ trois millions de mots, ce qui est très peu pour la linguistique du corpus. Mais il en ressort un trait spécifique : il y a une **homogénéité** frappante entre les textes traduits vers l'anglais lorsqu'on les compare aux textes rédigés d'abord en anglais et portant sur le même sujet. (Mona Baker, 1998 : 3) [nous soulignons]*

Au plan lexical

- Test : on calcule la distance moyenne des textes à l'intérieur du groupe POL-fr, et à l'intérieur du groupe POL-fr-trad
- Si l'hypothèse de l'homogénéité est vérifiée, la moyenne pour POL-fr-trad devrait être supérieure. On obtient :
 - $MoyCos(POL-fr)=0,947$
 - $MoyCos(POL-fr-trad)=0,953$
- D'après le test t de Student, cette différence n'est pas significative
- De même les densités lexicales sont très proches :
 - $d(POL-fr)=44,9 \%$ $d(POL-fr-trad)=45,1 \%$

Séquençage syntaxique

Au plan syntaxique

- Pour comparer les propriétés syntaxiques, on extrait des segments répétés au niveau des parties du discours (p.ex. DET NOUN PREP DET NOUN). Chaque bloc est ainsi caractérisé par une liste de segments répétés assortis de leur fréquence.
- En réitérant les comparaisons 2 à 2 via le cosinus, on obtient une matrice de contingence pour laquelle on peut rejeter l'hypothèse d'indépendance ($p(\chi^2)=0,0055$):

	POL-fr	POL-fr-trad	tot
POL-fr	33	18	51
POL-tr-trad	19	32	51
tot	52	50	102

Au plan syntaxique

- Néanmoins si on exclut tous les motifs incluant une ponctuation, l'hypothèse d'indépendance regagne du terrain, puisque $p(khi2)=0,047$. On note que certains motifs sont des artefacts liés à la constitution du corpus (PRON NEG PRON VERB ADV ; *je ne n ai pas, elle ne n pouvait plus*).
- La ponctuation joue donc un rôle qu'il convient d'éclaircir.
- En outre l'hypothèse homogénéité s'affaiblit également :
 - $MoyCos(POL-fr)=0,954$
 - $MoyCos(POL-fr-trad)=0,946$

Traces phraséologiques

Au plan phraséologique

- En comparant les segments répétés de longueur 5 (sans ponctuation), on peut plus clairement rejeter l'hypothèse d'indépendance ($p(\chi^2)=0,017$) :

	fr	fr-trad	tot
fr	34	17	51
fr-trad	22	29	51
tot	56	46	102

- Cette fois on peut noter un certain effet "langue de traduction"

Au plan phraséologique

- Peu de segments spécifiques de POL-fr (une vingtaine pour $LL > 10,83$). Souvent liés à un suremploi chez un ou deux auteurs, bien qu'avec une dispersion assez élevée :
 - *y avoir plus que* à (17 vs 1) : *y avait plus qu'à, y a plus qu'à...*
Fréquent chez de Villiers et Vargas
 - *il y en avoir un* (44 vs 16) : *Il y en a un, Il y en avait une, Il y en avait des, ...* Fréquent chez Aubert et Vargas. Registre oral familial. De même : *il y avoir aussi le* (13 vs 1) typique de Grangé, *ne y avoir plus que* (40 vs 15) fréquent chez de Villiers.
 - *ne avoir pas le temps* (74 vs 36) chez Brussolo, Chattam, de Villiers, *ne avoir pas le droit* (41 vs 15) chez Brussolo
 - *ne vous en faire pas* (22 vs 4) : *ne vous en faites pas*, fréquent chez Chattam et Halter

Au plan phraséologique

- *poser son main sur le* (16 vs 2) : *posa sa main sur* (l'épaule/bras/jambe/genou/poignet), typique de Chattam
- *ne savoir plus ce que* (13 vs 1) : *elle ne savait plus ce qu'elle désirait*, fréquent chez Brussolo
- *ce que ce être que* (20 vs 4) : *(savoir) ce que c'est que*, typique de Magnan.
- *je vous avoir dire que* (12 vs 1) : *je vous avais dit que*
- *il ne parvenir pas à* (37 vs 14) *Elle ne parvenait pas à* (démêler), très fréquent chez Brussolo et Chattam
- *il y avoir un silence* (21 vs 11) : *il y eut un silence*, fréquent chez Grangé, Halter, Vargas

Au plan phraséologique

- Pour POL-fr-trad, on compte beaucoup plus de segments spécifiques (une cinquantaine pour $LL > 10,83$) :
 - *acquiescer de un signe de* (52 vs 2) : (il) *acquiesça d'un signe de (tête)*. Très fréquent chez Connelly et Clancy.
 - *ce qui se être passer* (114 vs 30) : *ce qui s'est passé (ce soir)*, très fréquent chez Connelly, mais bien dispersé.
 - *il jeter un coup de* (33 vs 4) : *elle jeta un coup d'(oeil à son mari)*, très fréquent chez Connelly mais bien dispersé.
 - *le poche de son veste* (28 vs 3) : *(il glissa la carte) dans la poche de sa veste*, fréquent chez Connelly et Highsmith
 - *le autre côté de le* (99 vs 42) : *de l'autre côté de la (route/rue/table/frontière)*. Très fréquent chez Connelly et Clancy

Au plan phraséologique

- *dire il de un ton* (36 vs 6) : *dit-elle d'un ton* (*résolu/songeur/amer/railleur/admiratif/ferme/léger/glacial/maussade,...*) très fréquent chez Follet mais bien dispersé. On note une très grande variété de qualificatifs (49 différents !).
- *vouloir rentrer à le maison* (20 vs 1). *je veux rentrer à la maison*. Surtout chez Higgins Clark (dispersion faible).
- *avoir pas le intention de* (49 vs 15). (*Je n'ai pas l'intention d'intervenir*). Presque toujours à la première personne. Fréquent chez Harris, Crais, Ludlum, mais bien dispersé.
- *pour le amour de ciel* (16 vs 1). Fréquent chez Ludlum mais bien dispersé.

Au plan phraséologique

- Autres expressions marquées :
 - *il secoua/hocha la tête d'un air + ADJ*
 - *je n'arrive pas à croire que*
 - *dit-il d'une voix + ADJ*
 - *d'une façon ou d'une autre + PONCT*
 - *l'espace d'un instant*
 - *d'un hochement de tête*
 - *sans l'ombre d'un doute / hésitation / sourire*
 - *au bout d'un certain temps*
 - *il était plongé dans l'obscurité*
 - *il appuya sur la touche*

Au plan phraséologique

- Certaines expressions découlent directement de la traduction : expressions phraséologiques ou unités simples (*to nod, to glance at, anyway, ...*) → traductions récurrentes qui ne correspondent pas nécessairement à des expressions récurrentes en français. Absence d'équivalence lexicalisée → effet périphrastique.
- Certaines expressions sont plus fréquentes pour des raisons stylistiques (un auteur) ou culturelles - p.ex. *for god sake*.

Conclusions de l'enquête

Conclusion

- L'hypothèse de la langue de traduction n'est que partiellement vérifiée :
 - pas d'appauvrissement lexical avéré
 - pas de normalisation aboutissant à une homogénéité marquée en traduction
 - une empreinte thématique sensible au niveau des noms et adjectifs : sous-genres et contextes socio-culturels différents
 - pas de différence notable au plan syntaxique - mais un usage spécifique de la ponctuation à étudier plus avant
 - c'est au plan phraséologique que l'empreinte traductionnelle est plus visible

Conclusion

- Quelques expressions françaises sous-représentées en traduction :
 - registre oral familier : *y avait plus qu'à, Il y en avait des..., Il sait ce que c'est que...*
 - pragmatèmes : *ne vous en faites pas*
 - clichés liés au genre : *il y eut un silence*
 - motifs typiques d'un auteur : *il posa sa main sur son épaule*

Conclusion

- Quelques traces d'ADN de l'anglais source :
 - périphrases : *elle jeta un coup d'oeil à ... , il acquiesça d'un signe de tête, je n'avais pas l'intention de ... , d'une façon ou d'une autre..., je n'arrive pas à croire que ... , au bout d'un certain temps ...*
 - clichés liés au genre : *dans la poche de sa veste, il appuya sur la touche*
 - empreinte culturelle : *pour l'amour du ciel*
 - motifs typiques d'un auteur : *elle voulait rentrer à la maison*

Verdict (provisoire)

- Pas d'étrangeté, mais seulement des variations de fréquence
- Enquête préliminaire à poursuivre pour délimiter précisément le périmètre de ces effets traductionnels
- Bref... l'affaire n'est pas classée !

Merci pour votre attention !