



ISSN 1774-7988

ISSN en ligne : 2261-3455

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

**Olivier Kraif**

Université Grenoble Alpes, France  
 olivier.kraif@univ-grenoble-alpes.fr

### Résumé

Cet article propose une étude textométrique comparée entre un corpus de romans policiers écrits en français, et un corpus de romans s'inscrivant dans le même sous-genre, mais traduits depuis l'anglais. Nous faisons l'hypothèse que le corpus de traductions peut être exploité dans une perspective de linguistique de corpus, tout comme les textes originaux, afin d'étudier le sous-genre littéraire en question - contrairement à la tradition de la linguistique de corpus, qui préfère recourir à des corpus comparables plutôt qu'à des corpus de traductions, jugés peu fiables. Nous pensons en effet que certaines propriétés génériques sont interculturelles, et proviennent à la fois de la littérature traduite et de la littérature en langue originale. Pour étayer cette hypothèse, nous reprenons la liste des distorsions habituellement prêtées aux textes traduits, et nous mettons en œuvre une série d'observations textométriques afin d'en évaluer la portée. Nous montrons que dans le cas de nos deux corpus, la langue traduite ne présente pas de spécificités suffisamment saillantes pour considérer qu'il s'agirait d'un « troisième code » qu'il faudrait bannir de notre étude des sous-genres littéraires.

**Mots-clés :** traduction, romans policiers, textométrie

### Translating crime fiction: a comparison of textual statistics between original and translated-to-French novels

#### Abstract

This article presents a comparative study between a corpus of crime fiction novels written in French, and a corpus of novels belonging to the same sub-genre, but translated from English. We formulate the hypothesis that the translation corpus can be exploited just like the original texts in order to study a given literary sub-genre - contrary to the habits in corpus linguistics, who prefers to use comparable corpora rather than translated corpora, considered as unreliable. We believe that a lot of features of the literary sub-genres are inter-cultural, and come from both translated literature and literature in the original language. To support this hypothesis, we start from the list of distortions usually associated with translated texts, and we conduct a series of statistical observations in order to assess their relevance. We show that in the case of our two sub-corpora, the translated language does not present such marked peculiarities that it should be considered as a "third code" and should be banished from our study of literary sub-genres.

**Keywords :** translation, crime fiction novels, textual statistics

## Introduction<sup>1</sup>

Dans l'étude des sous-genres romanesques, il paraît difficile d'éluder le rôle de la traduction dans la circulation des œuvres d'un pays à l'autre. Certes, à l'intérieur d'une culture donnée, on peut sans doute identifier des œuvres fondatrices et des filiations entre les auteurs qui ont marqué le genre : pour le roman policier français, on citera naturellement Emile Gaboriau, Gaston Leroux, Maurice Leblanc, Léo Malet ou Georges Simenon. Dans le domaine anglophone, on pense à Edgard Poe ou Conan Doyle pour le roman à énigme, et plus tard Dashiell Hammett, Raymond Chandler ou James Hadley Chase pour le roman noir américain. Il serait cependant illusoire de croire que ces courants littéraires ont évolué parallèlement sans s'influencer ni se co-déterminer. Les œuvres fondatrices ont très tôt été traduites, et les collections spécialisées font la part belle aux traductions tout autant qu'aux œuvres originales. Par exemple, dans la plus ancienne collection policière française, *le Masque*, qui a notamment publié en France les œuvres d'Agatha Christie, sur 2546 ouvrages publiés entre 1927 et 2012 nous avons dénombré 1745 traductions (presqu'exclusivement de l'anglais) et 802 œuvres originales, avec des auteurs comme Charles Exbrayat ou Boileau Narcejac.

Comme le note Boyer (1992 : 97), « chaque genre comprend un certain nombre de sous-ensembles, des séries fondées sur la réutilisation de composantes identiques : réapparition d'un même protagoniste ou, du moins, recours à un typologie à peu près fixe de personnages ; fonctions déterminées attribuées à chacun d'entre eux ; constantes stylistiques ; utilisation d'un décor propre ; techniques spécifiques d'agencement des épisodes ; etc. » Ces caractéristiques aboutissent à des prototypes narratifs qui dépassent, selon nous, les frontières linguistiques, par le fonctionnement même de l'intertextualité, inhérente au texte littéraire. Comme l'écrivait joliment Barthes (1973), dans son article de l'*Encyclopedia Universalis*, « Tout texte est un intertexte ; d'autres textes sont présents en lui, à des niveaux variables, sous des formes plus ou moins reconnaissables : les textes de la culture antérieure et ceux de la culture environnante ; tout texte est un tissu nouveau de citations révolues. » Dans ces parcours intertextuels, rien ne permettrait de supposer qu'une frontière imaginaire condamnerait les textes traduits au bannissement - tout auteur étant avant tout lecteur, et tout lecteur étant également lecteur de traduction.

Dans cette étude, nous partons d'une hypothèse méthodologique assez forte : contrairement à une certaine tradition de la linguistique de corpus qui considère la traduction avec méfiance, comme le produit d'un acte de communication secondaire et artificiel, susceptible de transporter des propriétés de la langue source qui viendraient déformer les observations effectuées, nous pensons qu'il est tout

à fait possible de s'appuyer sur des textes traduits dans une approche *corpus driven*, guidée par le corpus, en restituant au texte traduit son statut de production langagière à part entière. Qui plus est, nous pensons que dans l'étude de la caractérisation des sous-genres littéraires sous un angle phraséologique, le recours à des traductions est indispensable, l'édification des sous-genres littéraires n'étant pas séparable de l'exercice de la traduction, comme le montre, sur le plan éditorial, l'important taux de traduction dans les collections spécialisées.

Dans une première partie de cet article nous examinerons quels sont les soupçons qui pèsent sur les textes traduits dans le domaine de la linguistique de corpus. Nous effectuerons ensuite une étude statistique au plan textométrique, basée sur un corpus de romans policiers originaux et traduits en français et en anglais, pour examiner la présence d'éventuels biais traductionnels. Nous chercherons ces biais à trois niveaux : lexical, syntaxique et phraséologique. Enfin, à la lumière des observations effectuées, nous concluons sur quelques principes méthodologiques concernant l'observation des corpus de traductions.

## 2. Pourquoi faut-il se méfier des traductions ?

*Traduttore, traditore* : on utilise en général cette expression italienne pour signifier que l'œuvre traduite ne respecte jamais tout à fait l'œuvre originale. Mais dans le domaine de la linguistique de corpus, on pourrait la réinterpréter autrement, en affirmant que la langue traduite ne respecte jamais tout à fait la langue cible, car elle y introduit subrepticement des déterminations qui ne lui appartiennent pas. Teubert (1996 : 247), dans une comparaison méthodologique entre corpus comparables et corpus de traductions, exprime ce point de vue sans ambiguïté :

*There is one essential objection to parallel corpora. Translations, however good and near-perfect they may be (but rarely are), cannot but give a distorted picture of the language they represent. Linguists should never rely on translations when they are describing a language. That is why translations have no place in reference corpora. Rather than representing the language they are written in, they give a mirror image of their source language.*

On peut citer deux types de déformations susceptibles d'altérer la langue cible : d'une part ce qui ressortit aux interférences avec la langue source, ou ce qu'on nomme parfois *translationese* (Johansson & Hofland, 1994 : 26), d'autre part ce que Baker (1993) nomme des universaux de la traduction, « des traits qui apparaissent spécifiquement dans les textes traduits plutôt que dans les textes sources, et qui ne sont pas le produit d'interférences liées à un système linguistique particulier »

(1993 :243, nous traduisons<sup>2</sup>). Concernant les interférences caractérisant le *translationese*, on pourra citer de nombreux travaux (Hopkinson, 2007 ; Teich, 2003). Par exemple Xiao (2015 : 159) met en évidence des différences significatives, sur un plan statistique, entre un corpus de textes chinois traduits de l'anglais et un corpus de textes originaux en chinois, sur différents plans : longueur moyenne des mots, utilisation des affixes, des pronoms et des unités polylexicales, longueur moyenne des phrases, voix passive, structures prédicatives... Pour un couple de langues moins typologiquement éloignées, Gellerstam (1986) énumère un certain nombre de caractéristiques observées sur des textes suédois traduits de l'anglais : présence de nuances sémantiques empruntées à l'anglais, sous-représentation du registre familier, fréquence supérieure d'emprunts.

Quant à ce que Mona Baker (1993 : 243) appelle « universal features of translation », elle identifie les traits suivants : 1/ un niveau supérieur d'explicitation ; 2/ une tendance à l'élimination des ambiguïtés et à la simplification ; 3/ un plus grand respect de la langue standard et de ses normes de grammaticalité ; 4/ une tendance à éviter les répétitions du texte source ; 5/ et enfin une tendance générale à exagérer certains traits de la langue cible. Ce dernier trait, constituerait un genre d'hypercorrection de la part des traducteurs qui chercheraient ainsi à maquiller le fait que les traductions n'étaient pas originellement destinées à leur environnement cible (Baker, 1993 : 244). Cet effort de « normalisation » (Baker, 1996 ; Ippolito, 2014) conduirait par ailleurs le traducteur à éviter les usages créatifs de la langue, allant jusqu'à normaliser la ponctuation même lorsque celle-ci est délibérément expérimentale (Baker, 1998 : 4). Baker (1996 : 184) note également une tendance à la convergence (*levelling-out*), la traduction opérant une forme de voie moyenne entre les deux extrêmes de la langue source et de la langue d'arrivée. Pour Laviosa (2002 :72), cela aboutit à un plus grand degré d'homogénéité des textes traduits, par rapport aux textes originaux. Xiao (2015 :140) cite enfin deux autres caractéristiques universelles : la sous-représentation, aussi appelée « unique items hypothesis », hypothèse selon laquelle des traits n'existant que dans la langue cible seront sans doute sous-représentés dans des traductions ; et, reprenant le terme de Kenny (1998), l'« assainissement » (en anglais *sanitisation*), aboutissant à l'élimination des contenus connotationnels et des double sens.

Toutes ces distorsions aboutissent à l'élimination *a priori* des textes traduits hors du champ de la linguistique de corpus, qui préfère recourir, pour l'étude contrastive et les comparaisons interlinguistiques, à des corpus comparables. Comme le note Sinclair (1996), « un corpus comparable doit permettre de comparer différentes langues ou variétés dans des circonstances de communication similaire, mais en évitant les inévitables distorsions dues à la traduction dans les corpus parallèles<sup>3</sup> ».

S'appuyant sur des corpus parallèles, nous avons nous-même constaté ce type d'altération. À partir d'un vaste corpus de textes parallèles de l'Union européenne, Torrellas Castillo (2009) a consacré sa thèse à une analyse minutieuse des interférences linguistiques dans les textes espagnols de l'UE. Le recours à des corpus massifs (le *JRC-Acquis*) et l'utilisation d'outils de traitement de corpus parallèles lui ont permis d'identifier des phénomènes d'interférence assez ténus, difficilement repérables par un dépouillement manuel. Dans une publication commune (Duchet, Kraif & Torrellas, 2008), nous mentionnons de nombreux types d'interférences :

- emprunt lexical » (ibid. : 138) : p.ex. *délocalisation* = \* *deslocalización* (avec glissement sémantique de *localizar*)
- « emprunt de collocations » (ibid. : 139-140) : p. ex. *prestataires de service* = \* *prestatario de servicios* (avec glissement sémantique de *prestatario*), *indemnité journalière* = ?*indemnización diaria* (au lieu de *dieta*, l'équivalent le plus conforme)
- « calques constructionnels » (ibid. : 141) : p. ex. *susceptible de + V*, comme dans *susceptible de provoquer* → *susceptibles de provocar*, alors qu'en espagnol *susceptible de* doit en principe être suivi par un verbe de sens passif ou une nominalisation avec déterminant zéro (*susceptible de recurso*).
- « calques syntaxiques » (ibid. : 142) : syntagme prépositionnel vs gérondif en apposition : *sur la base de* = *en base a*, au lieu du gérondif en apposition *basándose en* plus idiomatique.

Outre ces interférences, ces corpus sont par ailleurs marqués par une certaine forme d'appauvrissement, lié au caractère répétitif et mécanique des traductions :

*L'état de langue que nous avons observé, très fortement marqué par l'activité des traducteurs, manifeste aussi une restriction des choix lexicaux (...). Cette tendance est confortée par l'effet des mémoires de traduction (...) qui peuvent imposer pendant une longue période une traduction exclusive aux dépens de traductions équipossibles, pouvant aller jusqu'à la fossilisation d'erreurs reprises par tous les utilisateurs d'une même mémoire* (Torrellas Castillo, 2009 : 302).

Si toutefois on resitue ces phénomènes dans le champ plus large de la variation, le caractère inauthentique de ces productions peut être relativisé. Goffin (1994 : 642) remarque que « par sa nature, ses origines, ses modes de formation et son fonctionnement, ce langage - auquel on peut conférer la dignité d'*eurolécte* - ne se démarque aucunement des règles qui gouvernent toute langue de spécialité. » Certes la traduction impose des contraintes qui peuvent se manifester par des variations, telles que celles que l'on vient de mentionner, mais celles-ci suffisent-elles à

rejeter hors du champ de la langue tous ces textes traduits qui seraient en quelque sorte marqués du sceau de l'impureté ?

### 3. Observations textométriques

Ces effets de traduction se combinent et se mêlent à un autre niveau de contrainte, qui nous semble plus essentiel et plus profondément déterminant : celui du genre, voire du sous-genre. Il s'agit là d'un niveau de structuration tout à fait central, car « instance stratégique de normativité, le genre détermine l'essentiel de la sémiotique textuelle » (Rastier, 2006). Après une étude approfondie concernant des caractéristiques morphosyntaxiques d'un corpus de romans policier, Malrieu & Rastier (2001 : 564) résumant ainsi quelques traits saillants au plan linguistique :

*On peut ainsi caractériser, minimalement, le roman policier : caractère bref et ramassé de la phrase, compatible avec des propositions subordonnées complétives ; importance du dialogue ; importance de l'action (par opposition aux analyses et commentaires) ; caractère objectiviste du policier, par contraste avec le roman psychologique. La technique narrative diffère du roman « sérieux », où dominant l'imparfait, le présent et le passé simple, l'imparfait jouant un rôle clé dans la création du monde subjectif de la remémoration, alors que le récit policier privilégie le passé composé qui domine à l'oral et dans les dialogues, et présente le compte rendu d'événements « réels », non par un narrateur omniscient mais à travers le récit des personnages.*

Pour départager les déterminations liées au sous-genre de celles liées à l'exercice de la traduction, et donner une meilleure appréhension des biais traductionnels dans une méthodologie de linguistique de corpus, nous proposons donc de confronter deux corpus : 1/ un corpus de romans policier français (POL-FR) et 2/ un corpus de romans policiers écrits en anglais (britannique et américain) et traduit en français (POL-FR-trad). Ces corpus sont constitués de la manière suivante :

- Corpus POL-fr : 68 romans publiés entre 1956 et 2012 (63 ultérieurs à 1990) comportant 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 POL-fr-trad : 48 romans publiés entre 1955 et 2007 (40 ultérieurs à 1990) comportant 8 003 923 tokens

Auteurs : Clancy, Connelly, Cornwell, Crais, Crichton, Devon, Ellis, Evanovich, Grisham, Harris, Higgins Clack, Highsmith, Follett, Ludlum, Patterson, Reichs, Winstead Jones.

Ces corpus incorporent des sous-genres variés du roman policier : roman noir, thriller, roman d'espionnage, judiciaire, historique. Pour guider notre comparaison, nous proposons différents axes d'observations : lexical, syntaxique et phraséologique.

### 3.1 Observations lexicales et syntaxiques

Pour une comparaison globale du vocabulaire, nous avons segmenté chaque corpus en 17 blocs de tailles comparables (autour de 140 000 mots par blocs). Pour les textes de POL-FR-trad, chaque bloc correspond à un des 17 auteurs mentionnés ci-dessus. Pour les textes de POL-FR, nous avons également 1 bloc par auteur, sauf pour 4 blocs qui regroupent 2 à 3 auteurs pour lesquels nous avons moins de volume. De la sorte, nous avons obtenu 34 blocs comparables en taille, correspondant chacun à des auteurs différents (aucun auteur n'appartient à plus d'un bloc).

Ce premier test vise à répondre à la question suivante : les blocs de POL-FR sont-ils plus proches au niveau de leurs distributions lexicales des blocs de POL-FR ? De même, les blocs de POL-FR-trad se ressemblent-ils plus entre eux ? Ou bien, au contraire, observe-t-on le même niveau de ressemblance entre blocs français source et français traduits ?

On peut calculer la similarité lexicale entre 2 blocs en s'appuyant sur une mesure classique telle que le cosinus<sup>4</sup>, en assimilant chaque bloc à un vecteur dont les coordonnées sont les fréquences des vocables, dans un espace où chaque vocable constitue une dimension.

Pour réaliser la comparaison, nous calculons la similarité entre tous les blocs pris deux à deux. Pour chaque bloc, on retient les trois blocs les plus similaires. De la sorte on peut dénombrer les appariements entre blocs français source (POL-fr x POL-fr), blocs français traduits (POL-fr-trad x POL-fr-trad) ou bien entre le croisement des deux (POL-fr-trad x POL-fr) et (POL-fr x POL-fr-trad). A partir du tableau de contingence ainsi obtenu, on peut effectuer un test de Khi2 pour rejeter ou non l'hypothèse d'indépendance.

Globalement, en ignorant les mots en majuscules, qui correspondent à des entités nommées (anthroponymes, toponymes, etc.), le test de Khi2 aboutit à une p-valeur de 0,16, ce qui ne permet pas de rejeter l'hypothèse d'indépendance - c'est-à-dire que la distribution des ressemblances entre POL-fr et POL-fr-trad ne montre pas de tendance nette. Mais si l'on considère la nature des mots, on trouve des résultats très contrastés :

- pour les noms, on obtient 44 appariements de type POL-fr-trad x POL-fr-trad, 31 appariements de type POL-fr x POL-fr, 20 POL-fr x POL-fr-trad et 7 POL-fr-trad x POL-fr. La p-value de  $8,87 \cdot 10^{-7}$  montre une association très significative. L'association entre les blocs pris à l'intérieur de chaque corpus est également significative pour les adjectifs ( $p=0,0017$ ) et tendancielle pour les adverbes ( $p=0,03$ ).

- en revanche, pour les verbes et les autres parties du discours (pronoms, prépositions, déterminants, conjonctions, etc.), on ne relève pas d'association marquée à l'intérieur de chaque corpus.

Afin de mieux comprendre quelle type d'affinité lexicale on trouve à l'intérieur de POL-fr et POL-fr-trad, nous avons calculé une mesure de spécificité pour les noms, le *loglike* ou rapport de vraisemblance, utilisée pour identifier les mots clés dans de nombreux outils (Bertels, Speelman, 2013). Pour chaque corpus, nous avons extrait les 100 mots les plus spécifiques, puis nous avons retenu les noms. De la sorte, on obtient deux ensembles de noms spécifiques par rapport à l'autre corpus. Nous avons tenté de regrouper ces noms en fonction de classes sémantiques assez larges, et avons obtenu les deux listes ci-dessous :

*POL-fr :*

*POLICE : commissaire, flic, gendarme*

*NATURE : forêt, pierre, chien, loup, bête, nuit, araignée, animal, poussière, ombre, vent, gueule, bois*

*CORPS : chair, corps, ventre, cœur, peau, sang, visage, trait, bouche, lèvres, odeur*

*HUMAIN : homme, professeur, sœur, femme, vieillard, même, vieux, bourreau, patron*

*EMOTION : peur, cri, haine, délire, souffrance*

*EGYPTE : roi, momie, sarcophage, prêtre*

*MORT : mort, peste*

*CONSTRUCTION : château, muraille, paroi, village*

*ROMAN : manuscrit, roman, romancier*

*POL-fr-trad :*

*JURIDIQUE : avocat, affaire, tribunal, dossier, procès, jury, audience, société, associé, appel, accord*

*MILITAIRE : commandant, officier, radar, sous-marin, marine, poste, division, armée, air*

*ETAT : bureau, président, secrétaire, sénateur, réunion, cabinet, gouverneur, maison*

*SERVICES : agent, renseignement, service, sécurité*

*ARGENT : dollar, argent*

*ARTEFACTS* : téléphone, numéro, caméra, écran, radio, télévision, voiture, avion, appareil,

*AUTRE* : virus, grand-mère, monsieur, journaliste, problème, merci, besoin, endroit, question, moment, gens

Ces classes révèlent à la fois des aspects thématiques (p.ex. les services secrets pour les romans d'espionnage), des éléments du contexte (p.ex. un roman se situe dans l'Égypte ancienne, un autre au Moyen-âge) et des traits culturels (p.ex. *gendarme* en France, ou *gouverneur* aux Etats-Unis). On trouve par ailleurs des classes qui pourraient indiquer des isotopies au plan sémantique, concernant les éléments corporels ou naturels dans le corpus français source. Mais pour interpréter correctement ces spécificités, il faudrait comparer des œuvres proches au plan thématique et contextuel, afin de ne faire ressortir que des traits purement linguistiques, ce que nos deux corpus ne permettent pas en l'état.

Les affinités lexicales observées pour les noms, les adjectifs et les adverbes ne permettent donc pas de conclure à d'éventuelles distorsions traductionnelles.

Pour approfondir notre étude, nous avons examiné l'hypothèse de la simplification, qui se traduit notamment par une forme d'appauvrissement lexical. L'utilisation d'un vocabulaire moins varié se retrouve notamment dans un rapport type/token plus faible. Baker (1998 : 4), citant Laviosa-Braithwaite, l'exprime très clairement : « De plus, les deux rapports types/occurrences et densité lexicale sont considérablement plus faibles en anglais de traduction ». Nous avons donc comparé les courbes d'accroissement lexical, qui représentent le nombre de types (nombre de vocables différents) en fonction du nombre d'occurrences, pour nos deux corpus.

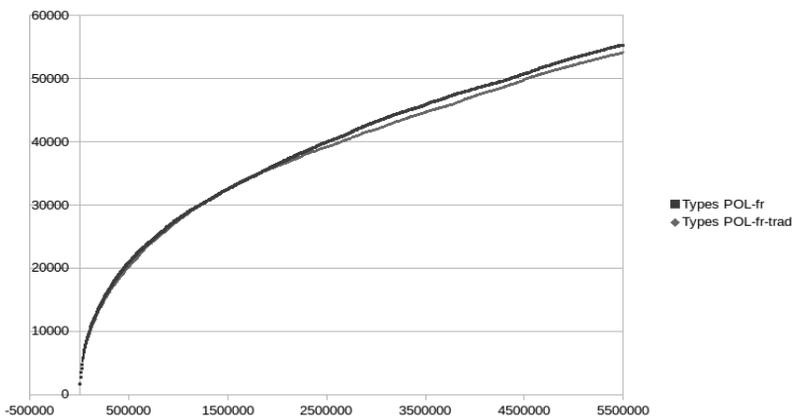


Figure 1 : Accroissements comparés du vocabulaire pour les corpus POL-fr et POL-fr-trad après lissage.

Afin de lisser les courbes, nous avons regroupé les textes par auteurs, puis pioché le premier mot de chaque groupe, puis le mot suivant pour chaque groupe, puis réitéré l'opération ainsi de suite jusqu'à arriver au dernier mot du groupe le plus volumineux. En effectuant ainsi un feuilletage des textes de notre corpus, on aboutit à un accroissement progressif global qui correspond en quelque sorte à la moyenne de tous les textes. Comme le montre la figure 1, on constate que les courbes sont pratiquement superposées, à part une légère augmentation de 900 vocables entre 2 165 000 et 3 795 000. Il faut noter que pour le corpus français source, pour un même nombre d'occurrences, on a à la fois plus d'auteurs différents (24 pour POL-fr contre 17 pour POL-fr-trad) et des œuvres plus courtes et plus nombreuses (pour arriver à 6 millions d'occurrences, on passe 59 œuvres différentes pour POL-fr, contre 41 pour POL-fr-trad). Cette diversité plus grande devrait conférer à POL-fr, a priori, une plus grande diversité lexicale (avec plus de sujets, de situations, de personnages différents), ce qui n'est presque pas le cas. On ne peut donc pas conclure, ici, sur une plus grande richesse lexicale du français source.

Une autre hypothèse évoquée précédemment est celle de convergence ou *levelling out* : autre effet de la normalisation, les traductions seraient plus homogènes entre elles que les textes sources, sur un plan linguistique :

*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 [...] 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. (Baker, 1998 : 3).*

Pour tester cette hypothèse, nous avons calculé la similarité moyenne des blocs à l'intérieur du groupe POL-fr d'une part, et à l'intérieur du groupe POL-fr-trad d'autre part. Si l'hypothèse de l'homogénéité est vérifiée, la moyenne pour POL-fr-trad devrait être significativement supérieure (d'autant que pour POL-fr on a plus d'auteurs et d'ouvrages différents, donc une plus grande hétérogénéité à priori).

On obtient :

Moyenne-Cosinus(POL-fr)=0,947

Moyenne-Cosinus(POL-fr-trad)=0,953

En utilisant un test t de Student<sup>5</sup>, on trouve que cette différence n'est pas significative (c'est-à-dire qu'il est probable qu'elle soit due au hasard). De même les densités lexicales moyennes (c'est-à-dire la proportion de mots pleins - noms, adjectifs, verbes et adverbes) sont très proches, contrairement à ce qu'affirme Baker (1998 :4) :

Densité(POL-fr)=44,9 %  
 Densité(POL-fr-trad)=45,1 %

Pour mesurer cette éventuelle homogénéité au plan syntaxique et non plus lexical, nous avons relevé pour chaque texte les suites de 5 parties du discours consécutives (du type DET NOUN PREP DET NOUN) assorties de leurs fréquences. On obtient ainsi, pour chaque bloc, des vecteurs dont chaque dimension représente la fréquence d'une séquence syntaxique particulière. A nouveau, nous avons comparé ces vecteurs à l'intérieur de chaque bloc. On obtient :

Moyenne-Cosinus-5grams(POL-fr)=0,936  
 Moyenne-Cosinus-5grams(POL-fr-trad)=0,928

Cette fois ce sont les blocs français sources qui paraissent légèrement plus homogènes, mais là encore le test t de Student nous indique que cette différence n'est pas significative. En tout état de cause, l'hypothèse de convergence n'est pas confirmée par ces observations.

### 3.2 Observations phraséologiques

Pour capturer d'éventuelles différences au plan phraséologique, nous avons extrait les séquences récurrentes de 3 à 7 lemmes consécutifs, apparaissant avec une fréquence supérieure ou égale à trois. Ces segments répétés (Salem, 1987), parfois nommés *ngrams*, *lexical bundles* ou *clusters*, présentent l'avantage d'identifier plusieurs types de phénomènes :

- des expressions figées : *un coup de œil*
- des collocations : *jeter un coup de œil, le téléphone sonner*
- des locutions : *ne avoir rien à voir avec*
- des routines : *ce être le premier fois que, je être prêt à parler que*
- des motifs (au sens de Longrée et Mellet, 2013) : *le joueur de flûte, le numéro de immatriculation de le camionnette, le veille de le disparition de, avoir être voir pour le dernier fois* (les exemples ci-dessus sont extraits du bloc Higgins-Clark).

Sur la base de ces extractions, nous avons à nouveau calculé un tableau de contingence des similarités entre POL-fr et POL-fr-trad. On obtient les valeurs suivantes : 27 associations pour POL-fr-trad avec POL-fr-trad, 40 associations pour POL-fr avec POL-fr, 11 pour POL-fr avec POL-fr-trad et 24 pour POL-fr-trad avec POL-fr. Le  $\chi^2$  donne une p-valeur de 0,001, ce qui indique une association statistiquement significative, particulièrement marquée à l'intérieur de POL-fr.

Pour interpréter ce résultat nous avons examiné plus en détail les segments répétés de longueur 5, intéressants car moins fragmentaires que les segments plus petits, et riches en motifs pertinents pour le roman policier. Nous avons retenu des segments apparaissant dans au moins 3 blocs, afin d'avoir des segments assez généraux pour être utilisés par plusieurs auteurs (et pour éviter les entités nommées trop spécifiques). Nous avons ensuite calculé une mesure de spécificité (le *loglike*, noté LL) pour comparer POL-fr par rapport à POL-fr-trad et réciproquement.

Pour POL-fr, on trouve assez peu de segments obtenant un score élevé (une vingtaine seulement avec LL>10,83). Ces segments sont souvent liés à un suremploi chez un ou deux auteurs :

- *y avoir plus que à* (17 occ. dans POL-fr vs 1 occ. dans POL-fr-trad) : p.ex. *y avait plus qu'à, y a plus qu'à...* Fréquent chez de Villiers et Vargas
- *il y en avoir un* (44 vs 16) : p.ex. *Il y en a un, Il y en avait une, Il y en avait des, ...* Fréquent chez Aubert et Vargas. On note que ces expressions appartiennent à un registre oral familier. De même, on peut noter : *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) fréquent chez Brussolo, Chattam, de Villiers, *ne avoir pas le droit* (41 vs 15) chez Brussolo
- *ne vous en faire pas* (22 vs 4) : p.ex. *ne vous en faites pas*, fréquent chez Chattam et Halter
- *poser son main sur le* (16 vs 2) : p.ex. *posa sa main sur (l'épaule/bras/jambe/genou/poignet)*, typique de Chattam
- *ne savoir plus ce que* (13 vs 1) : p.ex. *elle ne savait plus ce qu'elle désirait*, fréquent chez Brussolo
- *ce que ce être que* (20 vs 4) : p.ex. *(savoir) ce que c'est que*, typique de Magnan
- *je vous avoir dire que* (12 vs 1) : p.ex. *je vous avais dit que*
- *il ne parvenir pas à* (37 vs 14) p.ex. *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

Pour POL-fr-trad, on compte beaucoup plus de segments spécifiques (une cinquantaine avec LL>10,83) :

- *acquiescer de un signe de* (52 vs 2) : p.ex. *(il) acquiesça d'un signe de (tête)*. Très fréquent chez Connelly et Clancy.
- *ce qui se être passer* (114 vs 30) : p.ex. *ce qui s'est passé (ce soir)*, très fréquent chez Connelly.
- *il jeter un coup de* (33 vs 4) : p.ex. *elle jeta un coup d'oeil à son mari*, très fréquent chez Connelly.

- *le poche de son veste* (28 vs 3) : p.ex. (*il glissa la carte*) dans la poche de sa veste, fréquent chez Connelly et Highsmith.
- *le autre côté de le* (99 vs 42) : p.ex. *de l'autre côté de la (route/rue/table/frontière)*. Très fréquent chez Connelly et Clancy.
- *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. On note une très grande variété de qualificatifs (49 différents).
- *avoir pas le intention de* (49 vs 15). p.ex. (*Je n'ai pas l'intention d'intervenir*). Presque toujours à la première personne. Fréquent chez Harris, Crais, Ludlum.
- *pour le amour de ciel* (16 vs 1). Fréquent chez Ludlum.

On trouve par ailleurs d'autres expressions marquées comme spécifiques, utilisées par de nombreux auteurs :

*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...*

La comparaison de ces deux ensembles met en évidence des expressions aux statuts différents : pour le français traduit, on trouve beaucoup d'expressions qui semblent caractéristiques du genre romanesque, voire du sous-genre policier : descriptions d'attitudes de personnages (*acquiescer, jeter un coup d'œil, dire d'un ton ..., secouer la tête, d'un hochement de tête*), routines narratives (*au bout d'un certain temps, l'espace d'un instant, d'une façon ou d'une autre*), des clichés (*la poche de sa veste*), etc. Pour le corpus français source, ces expressions apparaissent comme sous-employées en comparaison avec POL-fr-trad, bien qu'elles apparaissent aussi avec une certaine fréquence. Quant à POL-fr, ce sont plutôt des expressions liées aux dialogues (*y a plus qu'à, il y en avait des, ne vous en faites pas, je vous avais dit que*) ainsi qu'à des réflexions des personnages (*elles ne parvenaient pas à démêler, elle ne savait plus ce qu'elle désirait, ...*) qui le caractérisent.

Parmi ces spécificités, quelques-unes semblent être des effets de traduction : certaines expressions familières, argotiques ou dialectales (p.ex. *il y en avait des ...*) sont plus rares voire absentes dans le corpus POL-fr-trad, ce qui correspond bien à la tendance normalisatrice de la traduction. Par ailleurs, certaines périphrases traduisent des expressions plus courtes et plus fréquentes en anglais. Afin de le vérifier, nous avons aligné les textes de POL-fr-trad avec les textes originaux, afin de retrouver les traductions dans leurs contextes. On trouve les équivalences suivantes :

- *acquiescer d'un signe de tête* correspond généralement à *to nod*, et parfois à *to nod agreement*. De même *hochement de tête* sert souvent à traduire *nod*, sous sa forme verbale ou nominale.
- *jeter un coup d'oeil* à traduit une grande variétés de constructions de la forme verbe + préposition : *look at, look in, look into, look over, look down, look around, look out, glance up, glance over, glance into, glance at, check, peer over, peek around, peek out, etc.*

Comme l'avaient déjà noté Vinay et Darbelnet (1958) l'anglais, pour exprimer le mouvement du regard, est à la fois plus nuancé et plus synthétique que le français, du fait de sa grande richesse lexicale et des possibilités combinatoires avec un grand nombre de prépositions. Là où le verbe et sa préposition suffisent, le français doit parfois recourir à des locutions adverbiales complexes, comme dans l'exemple suivant : (...) *then peered over the top riser* → *jeta un coup d'œil au ras de la dernière marche* (Robert Crais, *The last detective*)

- *dire d'un ton + ADJ* traduit la construction plus simple en anglais *to say +ADV*. On trouve par exemple : *dire d'un ton + de défi/languide/embarrassé/agacé/méditatif/mordant* → *to say + defiantly/languidly/uneasily/irritably/thoughtfully/waspishly*.
- *d'une façon ou d'une autre* correspond également à des expressions plus simples en anglais : *somehow, either way* - en français l'expression est longue et paraît plus lourde, ce qui explique peut-être sa moindre fréquence dans les textes en français source.
- enfin, la récurrence de la *poche de sa veste* correspond peut-être à un cliché du polar américain (on en tire des flacons de whisky, des cigares, et on y cache des lettres et des billets d'avion), mais cette récurrence vient peut-être aussi du fait que *jacket pocket* est lexicalisé en anglais, ce qui n'est pas le cas de son équivalent français.

Ces exemples montrent clairement ce qu'on pourrait appeler des effets de traductions - et à vrai dire, ils sont assez bien connus et ont été souvent observés : l'absence d'équivalents simples et directs aboutit presque systématiquement à des circonlocutions dont la fréquence devient mécaniquement plus élevée en français traduit qu'en langue originale. Par ailleurs, un certain registre oral, familier ou dialectal, parfois enraciné dans une culture régionale (comme l'argot parisien chez Dard ou les expressions provençales chez Izzo), est également sous-employé dans le français traduit de l'anglais ou de l'américain. Comme le note Bruneaud-Wheel (2010), la traduction des vernaculaires touche « aux limites de l'intraduisible », et peut donner lieu à des stratégies de traduction variables allant de la « standardisation » à l'« exotisation », en passant par la substitution d'un sociolecte par

un autre. D'une certaine manière, on est en droit de mettre sur le même plan - celui de la variation - la disparition ou la transposition des marques dialectales et l'apparition de ces marques d'étrangeté qui proviennent de la langue traduite. Dans un cas comme dans l'autre, elles n'ont rien d'accidentel et signent une forme d'altérité qui décrit tout un *continuum* dans la langue, sans solution de continuité.

## Conclusion

Ces quelques observations textométriques nous montrent que les romans policiers traduits de l'anglais présentent bien des spécificités linguistiques. Certaines, comme les fréquences lexicales pour les noms, adjectifs et adverbes sont liées à des aspects thématiques, contextuels et culturels, les romans étant enracinés dans leur culture d'origine. De même, certaines caractéristiques de la langue cible, comme certains usages familiers ou dialectaux n'apparaissent pratiquement pas dans les textes traduits. Ces constats étaient somme toute assez prévisibles. En revanche, certaines données textométriques de masse, telle que le rapport type/token, indice de la richesse lexicale, ou la densité lexicale, ne semblent pas être significativement différents dans les corpus traduits. De même, nos observations ne nous ont pas montré une plus grande convergence ou homogénéité dans les textes traduits, sur les plans lexicaux ou syntaxiques. Les hypothèses de simplification et de normalisation n'apparaissent pas de manière évidente dans ce type de traduction littéraire. Précisons toutefois que ces notions nécessiteraient d'être étudiées plus en profondeur : en manquant certaines caractéristiques idiolectales ou stylistiques de l'original, le traducteur peut effectuer une forme de normalisation - mais par ailleurs il introduit également des écarts à la norme, du fait même des effets de traductions signalés précédemment.

Concernant l'existence d'une « langue de traduction », Baker (1998 : 2) cite Frawley (1984) et sa notion de « troisième code » :

*Ce dernier suggère que la confrontation du texte source et de la langue cible pendant le processus de traduction crée ce qu'il appelle un « troisième code » (third code). En d'autres mots, le code (ou la langue) qui évolue pendant la traduction, et dans lequel le texte cible est rédigé, serait unique. Il s'agit d'un compromis entre les normes ou structures de la langue source et ceux de la langue cible.*

Cette notion nous semble très contestable, car elle fait de la traduction une activité à part, qui serait par essence différente de toutes les autres activités communicatives, au point que son résultat ne serait pas un produit langagier comme les autres, mais une forme artificielle à mettre de côté, à isoler dans sa singularité hors des frontières de la langue naturelle.

Or, tout ce qu'on l'on peut dire, c'est que la traduction, comme toute forme de communication, subit des contraintes spécifiques dont les messages traduits portent la marque - et que ces marques peuvent simplement s'étudier en termes de variations. Comme nous l'avons montré, certaines expressions sont surreprésentées dans les textes traduits. Mais qu'il s'agisse de calques, d'emprunts ou même d'interférences ayant échappé à la vigilance du traducteur, ces phénomènes de contact de langue n'ont rien d'exclusifs à la traduction, et apparaissent dans bien d'autres types de communications, tels que les textes scientifiques ou techniques.

Conscient de ces spécificités, il nous paraît plus intéressant de resituer le texte traduit dans son genre textuel, ou son sous-genre, dont les conventions et les normes dépassent selon nous les frontières linguistiques. Comme l'écrit Rastier (2006) « la question de la traduction spécifie une question générale qui concerne non les rapports de langue à langue, mais les rapports de texte à texte, puisque tout texte en transforme d'autres : quels sont les rapports sémiotiques entre deux textes qui dérivent l'un de l'autre, qu'il s'agisse de réécriture créatrice, de commentaire ou de traduction ? » Traduire, c'est écrire un texte qui s'inscrit dans un corpus intertextuel, et qui participe au devenir de la langue : « En outre, en élargissant le corpus, [la traduction] fait évoluer la langue : le corpus des textes traduits s'intègre au corpus de la langue. » (Ibid.)

La prochaine étape de notre recherche consistera donc à comparer les textes de différents sous-genres retenus pour notre projet (à savoir romans historiques, science-fiction, *fantasy*, sentimental, policier), et à vérifier si les textes en langue originale et les traductions partagent les mêmes propriétés lorsqu'ils s'inscrivent dans des sous-genres similaires.

### Remerciements

Cette étude préliminaire s'inscrit dans le cadre du projet Phraseorom, financé conjointement par l'ANR et la DFG, et dirigé par Iva Novakova et Dirk Siepmann.

### Bibliographie

- Baker, M. 1993. Corpus Linguistics and Translation Studies : Implications and Applications. In : M. Baker, G. Francis, E. Tognini-Bonelli (eds), *Text And Technology, In Honour of John Sinclair*. Philadelphia/Amsterdam : John Benjamins, p. 233-249.
- Baker, M. 1996. Corpus-based Translation Studies: The Challenges that Lie Ahead. In : H. Somers (Ed.), *Terminology, LSP and translation studies in language engineering: In honor of Juan C. Sager*. Amsterdam & Philadelphia: John Benjamins, p. 175-186.
- Baker, M. 1998. « Réexplorer la langue de la traduction : une approche par corpus ». *Meta : journal des traducteurs / Meta: Translators' Journal*, vol. 43, n° 4, p. 480-485.
- Barthes, R. 1973. Article «Texte (théorie du)». *Encyclopaedia universalis*.

- Bertels, A., Speelman, Dirk. 2013. « 'Keywords Method' versus 'Calcul des Spécificités' ». *International Journal of Corpus Linguistics* 18:4, p. 536-560.
- Boyer, A.-M. 1992. *La paralittérature* (1. éd.). Paris: Presses Universitaire de France.
- Bruneau-Wheal, K. 2010. « (M)oralité et traduction : les voix de Huck », In Yves Gambier & Olli Philippe Lautenbacher (Eds.), *Glottopol*, n°15, (URL: [http://glottopol.univ-rouen.fr/numero\\_15.html](http://glottopol.univ-rouen.fr/numero_15.html))
- Duchet, J.-L., Kraif, O., Torrellas Castillo, M. .2008. Corpus massifs et corpus bilingues alignés : leur impact sur la recherche linguistique. *Bulletin de la Société de Linguistique de Paris*, t. CIII, fasc. 1, p. 129-150.
- Frawley, W. 1984. Prolegomenon to a Theory of Translation. In : William Frawley (Ed.), *Translation: Literary, Linguistic, and Philosophical Perspectives*. London & Toronto :Associated University Presses, p. 159-175.
- Goffin, R. 1994. « L'eurolecte : oui, jargon communautaire : non ». *Meta : journal des traducteurs / Meta: Translators' Journal*, vol. 39, n° 4, p. 636-642.
- Gellerstam, M. 1986. Translationese in Swedish novels translated from English. In: Wollin, L. & Lindquist, H. (eds.) *Translation studies in Scandinavia. Lund Studies in English* 75. Malmö: CWK Gleerup, p. 88-95.
- Hopkinson, C. 2007. « Factors in linguistic interference: A case of study in translation ». *SKASE Journal of Translation and Interpretation*, 1, p. 13-23.
- Ippolito, M. 2014. *Simplification, Explicitation and Normalization: Corpus-Based Research into English to Italian Translations of Children's Classics*. Cambridge Scholars Publishing.
- Johansson, S., Hofland, K. 1994. Towards an English-Norwegian parallel corpus. In: Fries, U., Tottie, G. & Schneider, P. (eds.) *Creating and using English language corpora*. Amsterdam/ Atlanta: Rodopi, p. 25-37.
- Kenny, D. 1998. « Creatures of habit? What translators usually do with words ». *Meta*, 43(4), p. 515-523.
- Laviosa, S. 2002. *Corpus-based translation studies: Theory, findings, applications*. Amsterdam:Rodopi.
- Longrée, D., Mellet, S. 2013. « Le motif : une unité phraséologique englobante ? Étendre le champ de la phraséologie de la langue au discours ». *Langages*, 189, (1), p. 65-79.
- Malrieu, D., Rastier, F. 2001. « Genres et variations morphosyntaxiques ». *Traitement automatique des langues*, vol. 42, n°2, p. 548-577.
- Rastier, F. 2001. *Les styles en tant qu'usages singuliers du genre*, RASTIER, François. *Arts et Sciences du texte, Collection Formes sémiotiques*. Paris : Presses Universitaires de France.
- Rastier, F. 2006. La traduction : interprétation et genèse du sens. In : Marianne Lederer et Fortunato Israël, eds. *Le sens en traduction*. Paris : Minard.
- Salem, A. 1987. *Pratique des segments répétés*. Paris : Klincksieck.
- Scott, M., Tribble, C. 2006. *Textual Patterns: Keyword and Corpus Analysis in Language Education*. Amsterdam : John Benjamins.
- Torellas Castillo, M. 2009. *Les interférences linguistiques dans les textes en espagnol des institutions de l'Union Européenne : étude fondée sur le corpus bilingue massif aligné de l'acquis communautaire*. Thèse de doctorat, sous la dir. de J.L. Duchet, Université de Poitiers.
- Teich, E. 2003. *Cross-linguistic variation in system and text: A methodology for the investigation of translations and comparable texts*. Berlin: Mouton de Gruyter.
- Teubert, W. 2005. My version of corpus linguistics. *International Journal of Corpus Linguistics*, vol. 10-1, p. 1-13.
- Vinay, J.-P., Darbelnet, J. 1958. *Stylistique comparée du français et de l'anglais*. Paris: Didier.

Xiao, R. 2015<sup>6</sup>. Source Language Interference in English-to-Chinese Translation. In: Jesús Romero-Trillo (Ed.), *Yearbook of Corpus Linguistics and Pragmatics 2015, Current Approaches to Discourse and Translation Studies*, Springer International Publishing Switzerland.

## Notes

1. Une première version de cette recherche a fait l'objet d'une présentation lors de la journée d'étude « De la phraséologie aux genres textuels : état des recherches et perspectives méthodologiques » le 18 octobre 2016 à l'Université pédagogique de Cracovie.
2. « that is features which typically occur in translated text rather than original utterances and which are not the result of interference from specific linguistic systems. » (1993:243)
3. « The possibilities of a comparable corpus are to compare different languages or varieties in similar circumstances of communication, but avoiding the inevitable distortion introduced by the translations of a parallel corpus », URL: <http://www.ilc.cnr.it/EAGLES96/corpus/typ/node21.html>, [consulté en juin 2014].
4. La mesure du cosinus est une mesure classique de similarité intertextuelle. En assimilant chaque texte à la liste des fréquences des vocables qui le constituent (liste nommée *vecteur*), elle permet de mesurer quantitativement l'angle formé entre ces vecteurs : deux textes qui utilisent les mêmes vocables dans les mêmes proportions auront une mesure voisine de 1 ; si au contraire les deux textes utilisent un vocabulaire très différent, la mesure se rapproche de 0.
5. Le test de Student est utilisé pour comparer les moyennes d'une variable aléatoire (ici la similarité des textes pris deux à deux) sur deux échantillons différents (ici les textes de POL-fr d'une part, et ceux de POL-fr-trad d'autre part). Le test s'applique dans le cas d'une distribution supposée normale (comme par exemple la taille des individus d'une population). Il permet de mesurer la probabilité que l'écart obtenu entre les deux moyennes ne soit pas simplement due au hasard.