

DÉTECTION AUTOMATIQUE DES COUPES EN TRADUCTION OUTILLER LA GÉNÉTIQUE ÉDITORIALE DES FICTIONS CRIMINELLES

VALENTIN CHABAUX

UNIVERSITÉ PARIS NANTERRE (MSH MONDES)

ADRIEN FRENAY

UNIVERSITÉ PARIS NANTERRE (CRPM/CSLF)

LUCIA QUAQUARELLI

UNIVERSITÉ PARIS NANTERRE (CRPM)

v.chabaux@parisnanterre.fr ; af@parisnanterre.fr ; lquaquarelli@parisnanterre.fr

Citation: Chabaux, Valentin, Adrien Frenay, Lucia Quaquarelli (2025) “Détection automatique des coupes en traduction. Outiller la génétique éditoriale des fictions criminelles”, in Adrien Frenay, Lucia Quaquarelli, Benoît Tadié (eds.) *Pratiques et politiques de traduction dans les fictions criminelles*, *mediAzioni* 48: A58-A77, <https://doi.org/10.60923/issn.1974-4382/23660>, ISSN 1974-4382.

Abstract: In this article, we present the usefulness and effectiveness of the automatic tool for detecting cuts in translation developed by the Pôle DHUNE of the MSH Mondes in collaboration with the CRPM (Centre de Recherches Pluridisciplinaires et Multilingues, Université Paris Nanterre). This tool enables quantitative analyses of texts through the comparison of similarity between textual sequences written in different languages, in order to identify the modifications introduced during the translation process. We first describe how the tool operates (segmentation, similarity computation, alignment, filtering, and verification). We then apply it to a quantitative study designed to test the hypotheses on editorial translation practices (cuts) formulated in Frenay and Quaquarelli (2020a; 2020b; 2021; 2022), using a corpus of 310 novels from the “Série Noire” collection and their English translations. Finally, we show that the results produced by the tool confirm and reinforce these hypotheses, before discussing the limitations of this analytical method as well as the translational perspectives it opens.

Keywords: crime fiction; translation; editorial work; textual genetics.

1. Un nouvel outil pour l'étude des coupes dans les projets de traduction

La pratique des coupes dans les productions narratives de grande consommation (fictions criminelles, littératures de l'imaginaire, *etc.*) est connue des lecteurs et des lectrices passionnés-es comme des chercheurs et chercheuses. Souvent expliquée par l'urgence des contraintes matérielles (formats de la collection, capacité des présentoirs...) ou par la nécessité de réorienter le contexte temporel ou culturel du récit, cette pratique n'a pas fait l'objet jusqu'à présent d'une analyse approfondie à partir d'une étude de grand corpus.

À la suite des travaux de Clem Robyns (1990) et Jean-Marc Gouanvic (2018) et des recherches menées depuis 2018 portant sur le rôle de la traduction dans la création et la stabilisation de l'identité narrative de la « Série Noire » (Frenay et Quaquarelli 2020a, 2020b, 2021 et 2022) à partir des fonds d'archives présents à la BiLiPo, Bibliothèque des Littératures Policières, nous avons mis au point un outil de détection automatique des coupes, en collaboration avec le pôle DHUNE de la MSH Mondes, et réalisé ainsi une analyse quantitative des coupes sur un corpus de 310 romans de la collection, publiés entre 1945, année de sa création, et 1972 (n°1 à n°1496).

L'objectif était multiple. Nous voulions tout d'abord mettre à l'épreuve d'un grand nombre de récits les résultats de nos recherches précédentes, menées sur un corpus plus réduit, à partir d'une analyse qualitative des coupes et empruntant ses outils à la génétique, à la théorie de la traduction d'orientation descriptive et socio-culturelle et aux études littéraires.

Feuilleter les ouvrages conservés à la Bibliothèque des Littératures Policières de Paris qui ont servi de support pour les traductions de la « Série Noire », en inventorier et classer les marques (de coupes, d'injonctions traductives, stylistiques ou linguistiques, d'ajouts), lire les notes de services ou les communications externes, parcourir un certain nombre d'ouvrages en analyse contrastive, nous avait permis de relever l'existence d'une « machine traductive » au sein de laquelle les coupes jouaient un rôle essentiel dans la réorientation narrative de récits et répondaient à une nécessité de mise en conformité générico-sérielle qui dépassait largement les exigences purement matérielles ou culturo-spécifiques. Cette découverte nous avait alors permis d'esquisser une première liste de typologies narratives de coupes (Frenay et Quaquarelli 2020a, 2020b), qui demandait à être vérifiée et complétée à partir d'un corpus plus large.

Nous voulions aussi, par un travail quantitatif, affiner l'analyse (et la compréhension) du processus de traduction en « Série Noire » dans son intégralité, produisant, par exemple, des données par période, par auteur-e et par traducteur-ice. Enfin, il nous semblait utile, en retour, de tester l'outil de détection des coupes mis au point par Valentin Chabaux du pôle DHUNE (MSH Mondes) sur un corpus en partie connu, afin d'en mesurer les limites et les avantages et de pouvoir ainsi préfigurer d'autres usages du dispositif en traductologie, en génétique textuelle ou en histoire de l'édition.

Nous présentons dans la suite de cet article le corpus et la méthode d'analyse de notre étude quantitative. Nous avons fait le choix d'illustrer les différentes étapes de notre protocole en suivant pas à pas l'analyse d'un cas d'étude : le roman *Seven Lies South* de William P. McGivern (1960), traduit par Jean

Rosenthal en 1962 (*Vol en Vol*, « Série Noire », n° 742). Nous présentons ensuite les résultats plus généraux de notre étude, ainsi que ses limites et les perspectives qu'elle ouvre.

2. Le corpus

Le corpus de l'étude quantitative est constitué en fonction de la double disponibilité des versions numériques des textes à la fois en français et en anglais, soit un corpus de 310 romans anglophones répartis de manière relativement homogène entre le premier numéro de la collection et le n° 1496. Entre ces deux extrêmes, 1269 romans sont traduits de l'anglais (tous pays d'origine confondus, avec une forte prédominance des États-Unis et du Royaume Uni), 227 sont écrits en français, 1 est traduit du suédois. Ce corpus nous semble représentatif puisqu'il représente 24,7% des romans anglophones parus pendant la période considérée et 21% du total de la production de la collection.

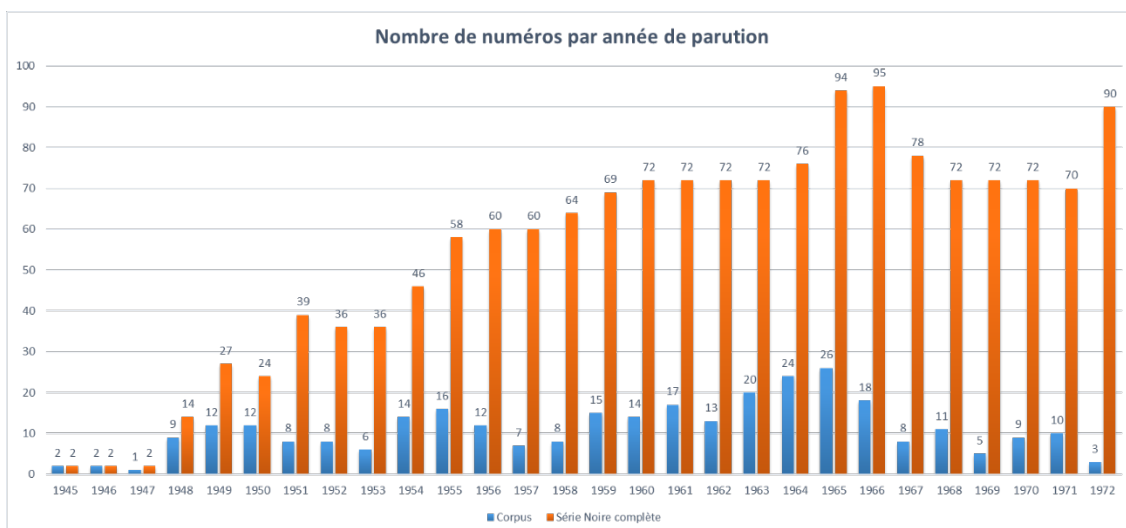


Figure 1. Nombre de romans pris en compte par année de publication

Ce corpus résulte également d'un effet de notoriété : les numérisations réalisées soit par des passionnés, soit par des universitaires se concentrent sur les œuvres et auteurs les plus connus. Il présente donc l'intérêt de différer du corpus d'étude de notre recherche initiale, réalisée à partir des archives du fonds Duhamel/Gallimard de la BiLiPo. Ce fonds d'archive est en fait caractérisé par le manque fréquent de documents servant à la traduction des œuvres les plus connues. La présente étude offre donc l'opportunité de mettre à l'épreuve nos conclusions sur un segment encore inexploité des productions éditoriales de la « Série Noire ».

Les recherches que nous avons menées jusqu'ici établissent l'existence d'une étape de coupe qui s'insère au sein du processus éditorial de traduction dans la collection « Série Noire » avant que le texte ne soit remis à la personne chargée de la traduction. Elle est matérialisée par la présence d'instructions de coupes dans les ouvrages anglophones servant à la traduction, sous la forme

d'encadrement, de biffures, de rayures ou de croix de Saint-André. C'est donc l'équipe éditoriale, pour des raisons à la fois matérielles (taille des romans) et narratives (cohérence requise par la contrainte sérielle), qui procède au caviardage. Nous avons donc mis à l'épreuve d'une analyse quantitative cette reconstitution du travail éditorial de traduction en recourant à un outil de détection automatique des coupes, conçu de la manière suivante.

3. La méthode

La détection automatique de coupes dans la version traduite en français d'un roman écrit en anglais présente plusieurs difficultés. Il faut tout d'abord se mettre en capacité de comparer automatiquement le texte des deux versions. Cette comparaison doit permettre d'identifier si un extrait issu de la version française a une correspondance dans la version anglaise du texte. Si un segment du roman en version anglaise n'a pas de correspondance dans la traduction française, alors il s'agit possiblement d'une coupe.

Pour réaliser cette tâche, nous avons recours à deux techniques successives. Premièrement, nous découpons les deux textes en plusieurs segments puis, grâce à un modèle entraîné à cet effet, nous obtenons une valeur de similarité entre chacun d'entre eux, deux à deux. Mais cette méthode seule ne suffit pas à identifier les segments qui ont été coupés. Pour ce faire, nous réalisons ensuite, sur la base des valeurs de similarités de chacun des segments, les deux textes afin de mettre en évidence des « ruptures de similarités », et donc de possibles coupes.

3.1. Le cas d'étude

Nous avons choisi *Seven Lies South* de William P. McGivern ainsi que sa traduction en français pour illustrer pas à pas la méthode de traitement des textes dans le cadre de la détection automatique des coupes parce qu'il est exemplaire du traitement des romans américains par l'équipe de la « Série Noire » et ses traductrices et traducteurs. Les coupes sont importantes et particulièrement bien identifiées par l'outil. De plus, elles transforment le roman de manière significative : le récit de McGivern met en scène un vétéran de la Seconde guerre mondiale et de la guerre de Corée en proie à la difficulté de trouver sa place et son identité dans un monde qui ignore ou préfère ne pas ouvrir les yeux sur la Shoah et qui commence seulement à saisir les conséquences de la seconde guerre mondiale sur les empires coloniaux. La traduction en « Série Noire » en fait l'aventure épicée d'un expatrié américain dur à cuir et séduisant.

Pour présenter en détail le fonctionnement de notre outil, nous allons prendre l'exemple d'un court extrait de ce roman dans ses versions anglaises et françaises :

<i>Seven Lies South</i>	<i>Vol en vol</i>
<p>He found himself smiling. “Well, welcome to the casa. Welcome to Spain. And how is Bunny? Did she send any messages?”</p> <p>“No. It was the oddest thing. I hadn’t seen her in years, until I bumped into her shopping in New York just a month or so ago. I told her I was coming to Spain, and she told me you were living here. She was going to write you, but I begged her not to. I know what a bore things like that are.” She smiled at him. “To set your mind at ease, I don’t need bullfight tickets, plane or train reservations, or someone to help move a big trunk. Isn’t that a relief?”</p> <p>Beecher felt pleasantly stimulated. “Well, I hope you need a drink at least.”</p> <p>“Thanks, I’d love one.”</p> <p>Beecher called to Encarna who must have been waiting just outside the door, for she appeared instantly with a tray of bottles, glasses, and ice. She moved in a hush of dignity, eyes and face grave, and her slippers barely whispering on the tiled floor. The tray was set with the villa’s best linen and silver, and everything looked calm and pretty, with the firelight sparkling on the shining glasses and deepening the colors of the fresh flowers. Beecher was grateful to Encarna for sensing that this was a special occasion.</p> <p>Sit by the fire,” he said to Laura Meadows. “The nights cool off pretty fast.”</p> <p>“I love your place. I walked down to the swimming pool before you came in. The sun was setting and the fishing boats were bobbing around in the water. It looked like... I don’t know... like something out of a Fifth Avenue toy shop.” She sat down and crossed her lovely legs and put her hands out to the fire. “No wonder Bunny can’t entice you home.”</p> <p>“Ye gods!” Beecher said. “Did she send you over with the heavy artillery?”</p> <p>“No, please,” she said, looking up at him quickly. She seemed to sense that she had blundered into a personal area. “She just mentioned that she misses you. I’m sorry.”</p> <p>Beecher smiled. “Bunny lives in pictures, as you may know. Firelight flickering on family and friends, Christmas carols sounding faintly, even if it’s the middle of July, and everybody rosy and happy and gay. If someone is missing, she’s mildly upset. The picture isn’t complete.”</p> <p>“I understand what you mean.”</p> <p>Beecher made two drinks with care, and as he measured gin and lemon juice into ice-filled glasses, he tried to analyze the current of emotion that was running through him. It must be that he was lonely for someone who looked and talked like Laura Meadows, he thought.</p>	<p>Il se surprit à sourire.</p> <p>— Comment va cette chère Bunny ? Vous a-t-elle chargée d’une commission pour moi ?</p> <p>— Non. Nous nous sommes rencontrées de façon très bizarre. Je ne l’avais pas vue depuis des années, et nous nous sommes heurtées dans un magasin à New York, il y a environ un mois. Je lui ai dit que je partais pour l’Espagne et elle m’a appris que vous habitiez ici. Elle allait vous écrire, mais je l’ai suppliée de n’en rien faire. Je sais combien ces choses-là sont ennuyeuses.</p> <p>Beecher se sentit tout ragaillardi.</p> <p>— Voulez-vous prendre quelque chose ?</p> <p>— Volontiers.</p> <p>Beecher prépara deux cocktails avec soin, et tout en versant le gin et le jus de citron sur les cubes de glace, il essaya d’analyser l’émotion qui l’envahissait. Était-ce parce qu’il rêvait depuis longtemps d’une femme qui aurait la voix et la silhouette de Laura Meadows ?</p>

Figure 2. Comparaison entre un passage de la version anglaise et sa traduction dans la version française

Nous pouvons remarquer au premier coup d'œil que des passages ont été coupés dans la version française. Nous allons voir comment, de manière automatisée, nous pouvons identifier précisément ces coupes. La méthodologie que nous présentons ici concerne un court extrait du roman, dans un souci de simplicité, mais elle est normalement appliquée à la totalité d'un roman et sa traduction.

3.2. Segmenter

La première étape consiste à segmenter les deux textes en unités plus petites, comparables sémantiquement entre elles. Nous choisissons de découper les textes par phrase. Cette approche a l'avantage d'une plus grande précision des résultats par rapport à une découpe par paragraphe. Pour réaliser cette segmentation, pour chaque paragraphe, nous utilisons SpaCy, une bibliothèque *python* de traitement automatique du langage, pour repérer les phrases. SpaCy permet ce découpage grâce à son composant de segmentation. Le texte est analysé pour identifier les frontières de phrases, en se basant sur la ponctuation et la structure syntaxique. Les règles de segmentation doivent être identiques pour les deux versions. Cela est assuré par l'utilisation de la même méthode pour les deux textes.

<i>Seven Lies South</i>		<i>Vol en vol</i>	
EN_0	He found himself smiling.	FR_0	Il se surprit à sourire.
EN_1	“Well, welcome to the casa.	FR_1	— Comment va cette chère Bunny ?
EN_2	Welcome to Spain.	FR_2	Vous a-t-elle chargée d'une commission pour moi ?
EN_3	And how is Bunny?	FR_3	— Non.
EN_4	Did she send any messages?”	FR_4	Nous nous sommes rencontrées de façon très bizarre.
EN_5	“No.	FR_5	Je ne l'avais pas vue depuis des années, et nous nous sommes heurtées dans un magasin à New York, il y a environ un mois.
EN_6	It was the oddest thing.	FR_6	Je lui ai dit que je partais pour l'Espagne et elle m'a appris que vous habitiez ici.
EN_7	I hadn't seen her in years, until I bumped into her shopping in New York just a month or so ago.	FR_7	Elle allait vous écrire, mais je l'ai suppliée de n'en rien faire.
EN_8	I told her I was coming to Spain, and she told me you were living here.	FR_8	Je sais combien ces choses-là sont ennuyeuses.
EN_9	She was going to write you, but I begged her not to.	FR_9	Beecher se sentit tout ragaillardi.
EN_10	I know what a bore things like that are.”	FR_10	— Voulez-vous prendre quelque chose ?
EN_11	She smiled at him.	FR_11	— Volontiers.
EN_12	“To set your mind at ease, I don't need bullfight tickets, plane or train reservations, or someone to help move a big trunk.	FR_12	Beecher prépara deux cocktails avec soin, et tout en versant le gin et le jus de citron sur les cubes de glace, il essaya d'analyser l'émotion qui l'envahissait.
EN_13	Isn't that a relief?”	FR_13	Était-ce parce qu'il rêvait depuis longtemps d'une femme qui aurait la voix et la silhouette de Laura Meadows ?
EN_14	Beecher felt pleasantly stimulated.		
EN_15	“Well, I hope you need a drink at least.”		
EN_16	“Thanks, I'd love one.”		

EN_17	Beecher called to Encarna who must have been waiting just outside the door, for she appeared instantly with a tray of bottles, glasses, and ice.
EN_18	She moved in a hush of dignity, eyes and face grave, and her slippers barely whispering on the tiled floor.
EN_19	The tray was set with the villa's best linen and silver, and everything looked calm and pretty, with the firelight sparkling on the shining glasses and deepening the colors of the fresh flowers.
EN_20	Beecher was grateful to Encarna for sensing that this was a special occasion.
EN_21	"Sit by the fire," he said to Laura Meadows.
EN_22	"The nights cool off pretty fast."
EN_23	"I love your place.
EN_24	I walked down to the swimming pool before you came in.
EN_25	The sun was setting and the fishing boats were bobbing around in the water.
EN_26	It looked like...
EN_27	I don't know... like something out of a Fifth Avenue toy shop."
EN_28	She sat down and crossed her lovely legs and put her hands out to the fire.
EN_29	"No wonder Bunny can't entice you home."
EN_30	"Ye gods!"
EN_31	Beecher said.
EN_32	"Did she send you over with the heavy artillery?"
EN_33	"No, please," she said, looking up at him quickly.
EN_34	She seemed to sense that she had blundered into a personal area.
EN_35	"She just mentioned that she misses you.
EN_36	I'm sorry."
EN_37	Beecher smiled.
EN_38	"Bunny lives in pictures, as you may know.
EN_39	Firelight flickering on family and friends, Christmas carols sounding faintly, even if it's the middle of July, and everybody rosy and happy and gay.
EN_40	If someone is missing, she's mildly upset.
EN_41	The picture isn't complete."
EN_42	"I understand what you mean."
EN_43	Beecher made two drinks with care, and as he measured gin and lemon juice into ice-filled glasses, he tried to analyze the current of emotion that was running through him.
EN_44	It must be that he was lonely for someone who looked and talked like Laura Meadows, he thought.

Figure 3. Segmentation automatique de l'extrait dans sa version anglaise et française

Dans le cas de notre extrait, nous obtenons 45 segments en anglais pour 14 segments dans la version française.

Notre objectif est ensuite de déterminer automatiquement, parmi ces segments, lesquels sont des traductions, lesquels sont des coupes (pour la version anglaise) ou des ajouts (pour la version française). Cela repose sur deux étapes successives : le calcul de similarités entre chacune des phrases puis la recherche du meilleur alignement possible entre les deux versions.

3.3. Repérer la traduction : mesurer la similarité

Pour repérer automatiquement les segments qui se correspondent entre la version française et la version anglaise d'un texte, nous calculons d'abord leur similarité sémantique. Cette étape repose sur un grand modèle de langage multilingue de type XLM-RoBERTa (Conneau et al 2020). Ce modèle est capable d'encoder des phrases de différentes langues dans un même espace vectoriel. La proximité du segment anglais et français dans cet espace à 768 dimensions est fonction de leur proximité sémantique et non lexicale. Pour chaque paire de phrases, le modèle nous permet de produire une similarité cosinus (comprise entre -1 et 1) à partir des deux vecteurs obtenus : une valeur élevée indique une forte proximité sémantique, tandis qu'une valeur faible ou proche de zéro signale l'absence de correspondance.

	FR_0	FR_1	FR_2	...	FR_4	FR_12	FR_13
EN_0	0,86	0,02	-0,07	...	0,17	-0,07	-0,03
EN_1	0,21	0,04	0,04	...	0,19	-0,09	-0,14
EN_2	0,16	0,01	-0,03	...	0,03	-0,15	-0,11
EN_3	-0,05	0,93	0,07	...	-0,08	-0,26	0,03
...
EN_43	-0,02	-0,21	-0,11	...	-0,12	0,97	-0,08
EN_44	-0,02	0,04	-0,07	...	-0,04	0,04	0,71

Figure 4. Extrait de la matrice de similarité sémantique entre les segments de la version anglaise et française de notre exemple

En appliquant ce calcul à l'ensemble des segments du texte, nous construisons une matrice de similarité (Figure 4), où chaque ligne correspond à un segment anglais et chaque colonne à un segment français. Cette matrice permet de visualiser les correspondances potentielles. Par exemple, dans le cas de notre extrait, la première phrase anglaise présente une similarité élevée avec la première phrase française (0,86), mais quasiment aucune avec la deuxième (0,02). Cela nous indique donc que, selon notre modèle et le corpus de textes sur lequel il a été entraîné, le

sens de la première phrase des deux versions est très proche. Nous pouvons déduire que le premier segment français est bien la traduction du premier segment anglais. De la même manière, la similarité des derniers segments de chaque texte (EN_44 et FR_13) est de 0,71. Il semble donc que le dernier segment du texte anglais soit traduit dans la version française et cela est détecté par notre modèle.

Pour généraliser et automatiser cette méthode afin de détecter les coupes dans la version française, nous devons identifier les phrases qui doivent être considérées comme des traductions. Cela revient alors à aligner les deux textes. Pour construire un alignement cohérent, il est nécessaire d'introduire des trous, c'est-à-dire des positions sans correspondance, dans l'une ou l'autre des séquences : un trou dans la version française correspond alors à une coupe tandis qu'un trou dans la version anglaise correspond à un ajout.

Ce problème est équivalent au problème bien connu de l'alignement de séquences, largement étudié en bioinformatique. Nous nous appuyons donc sur un algorithme conçu précisément pour optimiser ce type d'alignement.

3.4. Aligner les séquences : l'algorithme de Needleman-Wunsch

L'algorithme de Needleman-Wunsch (1970), initialement conçu pour l'alignement global de séquences d'acides aminés dans le cadre de la comparaison de protéines, s'avère particulièrement adapté à notre objectif : déterminer automatiquement l'alignement optimal entre les segments des deux textes.

À partir de la matrice de similarité construite précédemment, l'algorithme calcule progressivement, de proche en proche, les valeurs d'une matrice d'alignement. Chaque cellule de la matrice représente le score optimal obtenu jusqu'à cette position. Chaque terme est déduit de l'un de ses trois voisins (diagonale, haut, gauche), en tenant compte soit d'un alignement direct, soit de l'introduction d'un trou (Figure 5). Parallèlement, une matrice de directions enregistre la provenance de chaque valeur, ce qui permet par la suite de reconstruire le chemin, et donc l'alignement, optimal (Figure 6).

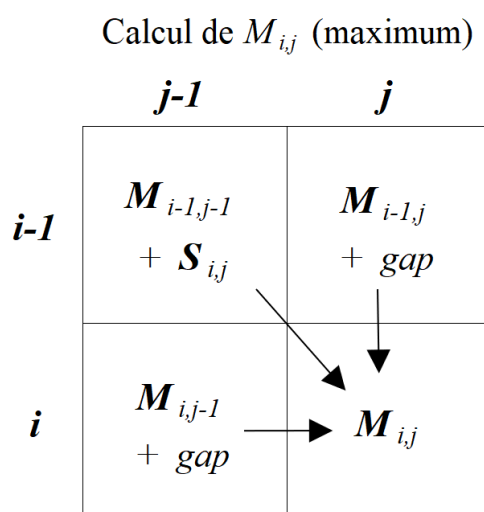


Figure 5. Visualisation du calcul de proche en proche pour chaque terme $M_{i,j}$ de la matrice d'alignement dans l'algorithme de Needleman-Wunsch. Avec *gap* la pénalité de trou, et *S* la matrice de similarité. $M_{i,j}$ est le maximum des trois cases adjacentes

	FR_0	FR_1	FR_2	FR_3	FR_4	FR_5	FR_6	FR_7	FR_8	FR_9	FR_10	FR_11	FR_12	FR_13
EN_0	↖	←	←	←	←	←	←	←	←	←	←	←	←	←
EN_1	↑	←	←	←	←	←	↖	←	←	←	←	↖	←	←
EN_2	↑	←	←	←	←	←	↖	←	←	←	←	←	←	←
EN_3	↑	↖	←	←	←	←	←	←	←	←	←	←	←	←
EN_4	↑	↑	↖	←	←	←	←	←	←	←	←	←	←	←
EN_5	↑	↑	↑	↖	←	←	←	←	←	←	←	←	←	←
EN_6	↑	↑	↑	↑	↖	←	←	←	←	←	←	←	←	←
EN_7	↑	↑	↑	↑	↑	↖	←	←	←	←	←	←	←	←
EN_8	↑	↑	↑	↑	↑	↑	↖	←	←	←	←	←	←	←
EN_9	↑	↑	↑	↑	↑	↑	↑	↖	←	←	←	←	←	←
EN_10	↑	↑	↑	↑	↑	↑	↑	↑	↖	←	←	←	←	←
EN_11	↑	↑	↑	↑	↑	↑	↑	↑	↑	↖	←	←	←	←
EN_12	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↖	←	←	←
EN_13	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↖	←	←
EN_14	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↖	←
EN_15	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↖
EN_16	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
EN_17	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
EN_18	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
EN_19	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
EN_20	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
EN_21	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
EN_22	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
EN_23	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
EN_24	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
EN_25	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
EN_26	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
EN_27	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
EN_28	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
EN_29	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
EN_30	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
EN_31	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
EN_32	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
EN_33	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
EN_34	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
EN_35	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
EN_36	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
EN_37	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
EN_38	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
EN_39	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
EN_40	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
EN_41	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
EN_42	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
EN_43	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
EN_44	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑

Figure 6. Visualisation de la matrice de direction permettant d'obtenir le meilleur alignement global des deux séquences à partir de la matrice d'alignement M.

L'algorithme repose également sur un paramètre clé : la pénalité de trou, qui ajuste le coût associé à l'ajout d'une coupe ou d'un ajout dans l'une des deux séquences. Une pénalité trop élevée réduirait excessivement le nombre de trous possibles, tandis qu'une pénalité trop faible en introduirait trop. Après plusieurs tests empiriques, nous retenons une valeur de 0,1, qui offre un bon compromis pour nos données.

Une fois la matrice complétée, l'alignement optimal est obtenu en remontant le parcours fléché depuis la dernière cellule jusqu'à l'origine. Dans ce parcours :

- une diagonale indique un alignement direct entre deux phrases, correspondant à une traduction,
- une flèche vers le haut correspond à une coupe,
- une flèche vers la droite signale un ajout.

Cet alignement peut ensuite être transposé dans un format plus lisible, permettant de visualiser immédiatement les phrases alignées, celles qui ont été coupées, et celles qui ont été ajoutées. Ce rendu final constitue une base solide pour l'analyse comparative des deux versions du texte (Figure 7).

EN_0	He found himself smiling.”	FR_0	Il se surprit à sourire.
EN_1	Well, welcome to the casa.	COUPE	
EN_2	Welcome to Spain.		
EN_3	And how is Bunny?		
EN_4	Did she send any messages?”	FR_1	— Comment va cette chère Bunny ?
EN_5	“No.	FR_2	Vous a-t-elle chargée d’une commission pour moi ?
EN_6	It was the oddest thing.	FR_3	— Non.
EN_7	I hadn’t seen her in years, until I bumped into her shopping in New York just a month or so ago.	FR_4	Nous nous sommes rencontrées de façon très bizarre.
EN_8	I told her I was coming to Spain, and she told me you were living here.	FR_5	Je ne l’avais pas vue depuis des années, et nous nous sommes heurtées dans un magasin à New York, il y a environ un mois.
EN_9	She was going to write you, but I begged her not to.	FR_6	Je lui ai dit que je partais pour l’Espagne et elle m’a appris que vous habitiez ici.
EN_10	I know what a bore things like that are.”	FR_7	Elle allait vous écrire, mais je l’ai suppliée de n’en rien faire.
EN_11	She smiled at him. “	FR_8	Je sais combien ces choses-là sont ennuyeuses.
EN_12	To set your mind at ease, I don’t need bullfight tickets, plane or train reservations, or someone to help move a big trunk.	COUPE	
EN_13	Isn’t that a relief?”		
EN_14	Beecher felt pleasantly stimulated. “		
EN_15	Well, I hope you need a drink at least.”	FR_9	Beecher se sentit tout ragaillardi.
EN_16	“Thanks, I’d love one.”	FR_10	— Voulez-vous prendre quelque chose ?
EN_17	Beecher called to Encarna who must have been waiting just outside the door, for she appeared instantly with a tray of bottles, glasses, and ice.	FR_11	— Volontiers.
EN_18	She moved in a hush of dignity, eyes and face grave, and her slippers barely whispering on the tiled floor.	COUPE	
EN_19	The tray was set with the villa’s best linen and silver, and everything looked calm and pretty, with the firelight sparkling on the shining glasses and deepening the colors of the fresh flowers.		
EN_20	Beecher was grateful to Encarna for sensing that this was a special occasion.		
EN_21	“Sit by the fire,” he said to Laura Meadows. “		
EN_22	The nights cool off pretty fast.”		
EN_23	“I love your place.		
EN_24	I walked down to the swimming pool before you came in.		
EN_25	The sun was setting and the fishing boats were bobbing around in the water.		
EN_26	It looked like...		
EN_27	I don’t know... like something out of a Fifth Avenue toy shop.”		
EN_28	She sat down and crossed her lovely legs and put her hands out to the fire.”		
EN_29	No wonder Bunny can’t entice you home.”		
EN_30	“Ye gods!”		
EN_31	Beecher said. “		
EN_32	Did she send you over with the heavy artillery?”		
EN_33	“No, please,” she said, looking up at him quickly.		
EN_34	She seemed to sense that she had blundered into a personal area. “		
EN_35	She just mentioned that she misses you.		
EN_36	I’m sorry.”		
EN_37	Beecher smiled. “		
EN_38	Bunny lives in pictures, as you may know.		
EN_39	Firelight flickering on family and friends, Christmas carols sounding faintly, even if it’s the middle of July, and everybody rosy and happy and gay.		
EN_40	If someone is missing, she’s mildly upset.		
EN_41	The picture isn’t complete.”		
EN_42	“I understand what you mean.”		

EN_43	Beecher made two drinks with care, and as he measured gin and lemon juice into ice-filled glasses, he tried to analyze the current of emotion that was running through him.	FR_12	Beecher prépara deux cocktails avec soin, et tout en versant le gin et le jus de citron sur les cubes de glace, il essaya d'analyser l'émotion qui l'envahissait.
EN_44	It must be that he was lonely for someone who looked and talked like Laura Meadows, he thought.	FR_13	Était-ce parce qu'il rêvait depuis longtemps d'une femme qui aurait la voix et là silhouette de Laura Meadows ?

Figure 7. Transposition de la matrice de direction (Figure 6) qui résulte de l'algorithme d'alignement dans une visualisation et permet l'analyse qualitative des coupes.

3.5 Affiner les résultats : les filtres

Après une étude qualitative de nos premiers résultats, nous remarquons que, dans une très grande majorité des cas, la détection automatique d'une coupe d'un faible nombre de segments consécutifs résulte d'un défaut de segmentation (ou plus rarement d'alignement) et est à considérer comme un faux positif. Dans le cadre d'une approche quantitative, afin d'ignorer ces faux positifs et de ne pas surestimer le phénomène de coupe, nous avons adopté une stratégie reposant sur un système de filtres.

Nous posons deux conditions filtrantes pour valider une coupe. Premièrement, nous considérons comme une coupe seulement les sous séquences où la différence entre le nombre de segments anglais et français est supérieure à un seuil $S = 4$ (dont la valeur est définie empiriquement). Cette approche est plus fine qu'un seuil de trous consécutifs dans la version française car elle permet de détecter les passages mêlant coupes et ajouts. Ainsi, une coupe n'est pas détectée uniquement par une suite de trous consécutifs dans la version française.

Une trop grande réécriture de certains segments peut poser un problème dans le cadre de notre approche de la détection des coupes par similarité sémantique. Si notre modèle ne parvient pas à rapprocher sémantiquement deux segments que nous identifions pourtant comme une traduction, cela s'exprimera par une coupe suivie d'un ajout. Si ce phénomène se répète sur plusieurs segments cela peut générer un faux positif. Pour filtrer ce cas nous établissons un second seuil : une coupe est détectée seulement si le passage compte au moins deux fois plus de segments anglais que français.

Par exemple, un passage composé de trois trous dans la version française, une traduction, un trou dans la version anglaise, trois nouveaux trous dans la version française, puis une suite de traductions, serait bien détectée comme un passage contenant une coupe. Nous avons, dans cette sous-séquence de 8 segments présentant des trous, 2 segments français pour 7 segments anglais. Comme $7 - 2 = 5 > 4$ et $7 > 2 \times 2$, nos deux seuils sont franchis et ce passage est donc identifié comme comptant une ou plusieurs coupes. Étant donné que nous comptabilisons par la suite, dans notre étude quantitative, seulement les caractères de la version anglaise réellement coupés, les segments traduits de notre sous-séquence pourront être ignorés.

Si nous éliminons ainsi un grand nombre de faux positifs, nous prenons le risque d'éliminer de vraies coupes de petite taille, inférieures à nos seuils. Nous considérons que le rapport entre efficacité de l'approche et qualité des résultats est pertinent vis-à-vis de notre problématique. Puisque nous cherchons à identifier des coupes qui résultent de pratiques éditoriales, nous priorisons la détection des coupes significatives, dépourvues de faux positif.

3.6. Contrôle des résultats de détection automatique par vérification manuelle

Enfin, nous effectuons une vérification manuelle des résultats afin de mesurer l'efficacité de notre méthode. À ce jour, cette étape est toujours en cours. Nous avons effectué dans un premier temps huit vérifications exhaustives (Figure 8). Elles confirment que le traitement automatique est efficace et précis.

Numéro	Nombre de caractères coupés (détection automatique)	Nombre de caractères coupés (vérification manuelle)	Nombre de caractères total de la version anglaise	Faux positifs (caractères)	Faux positifs (% sur total détection automatique)	Caractères coupés - Vérification manuelle (% sur nombre de caractères total)	Caractères coupés - Détection automatique (% sur nombre de caractères total)
55	6719	10695	318191	83	1%	3%	2%
183	35544	43474	337130	149	0%	13%	11%
212	10100	10171	264792	41	0%	4%	4%
334	19197	27767	340851	227	1%	8%	6%
383	590	590	274767	0	0%	0%	0%
402	57913	57703	380609	0	0%	15%	15%
488	807	81	285007	227	28%	0%	0%
535	22237	32169	320377	1323	6%	10%	7%

Figure 8. Tableau récapitulatif du contrôle des résultats de détection automatique par vérification manuelle

Nous trouvons quelques cas de légère sur- et/ou sous-détection, qui ne sont pas de nature à modifier les conclusions d'une analyse des coupes à l'échelle du roman en question. Nous remarquons que la segmentation de la version anglaise influe sur le résultat de la détection automatique : les cas qui présentent une différence entre les résultats de l'analyse manuelle et les résultats de la détection automatique concernent des romans constitués de phrases courtes, réunies dans la traduction en français dans des phrases plus longues. Cette étape de vérification nous permet par ailleurs d'améliorer les filtres.

Nous validons également notre stratégie consistant à minimiser les fausses détections en priorité. Sept sur huit des romans vérifiés présentent 6% ou moins de faux positifs. Un hapax à 28% peut quant à lui s'expliquer notamment par le très faible nombre de caractères coupés dans le roman, seulement 807.

Nous prévoyons à l'avenir de mener ce travail de vérification de la détection automatique sur un échantillon plus important afin de quantifier plus formellement, à l'aide d'outils statistiques, notre marge d'erreur. Cette étape sera conjointe à une amélioration de la méthode, notamment en termes de segmentation, afin d'augmenter les performances de notre détection.

3.7. Exemple d'exploitation qualitative de l'étude quantitative

Dans quelle mesure ce protocole de détection automatique de coupes est-il utile aux chercheurs et aux chercheuses en traduction ? Revenons d'abord au roman qui nous a servi d'exemple afin de montrer comment les résultats de la détection

automatique peuvent être prolongés par une analyse qualitative, avant d'aborder les résultats de l'analyse quantitative en eux-mêmes.

Il est possible, à partir des résultats obtenus par notre méthode, d'obtenir un pourcentage de coupes. Celui-ci permet d'identifier des cas d'étude significatifs ou représentatifs. C'est le cas de *Seven Lies South / Vol en vol*. La version anglaise comporte 481 723 caractères ; le total du nombre de caractère des segments coupés est de 163 407. Notre méthode de détection estime donc le pourcentage de coupe à 33,9%. La visualisation des coupes permet de passer alors aisément à l'étude qualitative.

Par l'intermédiaire de la mise en scène d'un protagoniste américain désabusé et pétri de doute sur son identité et sa place dans la société, *Seven Lies South* propose une réflexion sur la place et le rôle des États-Unis dans l'ordre mondial qui se recompose après la seconde guerre mondiale. Dans le roman, l'Espagne et le Maroc sont les terrains de rencontre de personnages américains, espagnols, français, allemands, autrichiens, britanniques, canadiens et berbères. Leurs identités nationales stéréotypiques tantôt expliquent leurs actions tantôt masquent leurs motivations, et font de l'idée même d'une concorde internationale une forme de mascarade.

Le fichier de sortie de l'outil permet de montrer que le processus de traduction de la collection opère des coupes nombreuses, longues, et qui correspondent à la typologie proposée dans nos travaux : transformer le rythme, amoindrir les considérations psychologiques et favoriser l'action sur les pensées des personnages, effacer ou amoindrir les questions politiques, amoindrir ou supprimer les références culturelles trop spécifiques (Figure 9).

Once a terrible thing had happened to Don Willie in the village.
While sitting with friends on the terrace of the Bar Central, a small, poorly dressed man had attacked him violently and suddenly, raining impotent blows against his broad beefy chest and shoulders.
Don Willie had stared at him for an instant in hideous disbelief, and then had leaped to his feet and fled across the plaza.
From there he had run down the narrow street that led to his villa, eyes rolling like a stallion in panic, leather coattails flapping madly.
That same night he had flown his airplane to Madrid, and had spent the rest of the summer attending to his various interests in Barcelona and Valencia.
Meanwhile the little man had been taken into custody by the police.
He claimed that Don Willie had murdered his wife and children in a concentration camp during the war.

Figure 9. Exemple de coupe. La coupe de l'analepse permet de prioriser l'action principale tout en supprimant les allusions à la Shoah

La traduction en français est certes plus courte, mais elle correspond surtout mieux à l'idéal que se fait Marcel Duhamel du roman policier : un roman d'action au sein duquel les crimes s'enchainent et où ne subsiste que ce qui est filmable. Il s'approche du synopsis. Nous observons de surcroît des coupes qui portent spécifiquement sur l'amuïssement des références à l'homosexualité et

des modifications liées à ces coupes qui traitent le sujet de manière caricaturale (Figure 10).

Beecher bought a newspaper and sipped his drink.
When he first came to Spain this was the hour of the day he had liked best of all.
The siesta was over, the breezes were cool, and the promise of the long evening ahead filled everyone with a sense of significance and excitement.
And there had been a lot of good evenings, Beecher thought, putting his paper aside.
All sorts of people were available in Mirimar, all national, political, and sexual complexions, and it was an easy matter to find friends.
You drifted with the crowd.
Everything was informal.
You went to parties where everyone drank red wine or Spanish brandy, and someone had a guitar and sang folk songs.
In the daytime you soaked up sun on the beach and drank a few bottles of cold beer to minimize the hangovers, which were the most staple topic of conversation in the village.
People were always going off on trips.
To Gibraltar to change money; to Tangier to eat couscous and hear jazz and watch the slim and epicene boys who danced in the tourist restaurants; to somewhere over the mountains to see the bullfighter everyone was comparing to Manolete and Belmonte.

Figure 10. Exemple de coupe. La coupe supprime les pensées du personnage, les allusions à la sexualité, à l'orientation sexuelle et plus précisément à l'homosexualité.

Prenons un exemple : la version française est notamment le résultat de la suppression des deux premiers chapitres du roman. Ces chapitres tissent le portrait d'un héros désabusé, qui observe et analyse les comportements et les paroles des personnages avec lesquels il interagit. Après une partie de golf dans le chapitre 1, il analyse, dans le chapitre 2, sa position d'expatrié en Espagne ainsi que celle des autres expatriés américains dont il se distingue par l'âge et l'occupation. Le héros à la dérive, chez lui nulle part, réfléchit à la question de la présence américaine en dehors des États-Unis.

Ce sont ces pensées et les faits qu'il observe (la partie de golf du chapitre 1) qui sont coupés. Ces coupes longues permettent à la fois de transformer le rythme du roman, d'éviter les considérations politiques et psychologiques, la question de la guerre et de la Shoah et réduire les spécificités culturelles liées à la situation géographique de la ville où débute l'action du roman. Dans la version française, le protagoniste est mis dans une situation bien plus stéréotypique : dans un tête-à-tête avec une jeune américaine séduisante qu'il emmène par la suite en balade. Ces deux chapitres sont par la suite résumés pour conserver la cohérence de la diégèse. L'équipe éditoriale ou le traducteur a procédé à un raccord (en bleu).

Beecher n'était pas d'humeur à recevoir des visites ; il avait joué au golf cet après-midi avec une grande perche d'Anglais, qui avait traîtreusement réussi à lui faire perdre la partie en marchant sur sa balle avant le dix-huitième trou.
Cet incident l'avait déprimé : c'était une tricherie gratuite et absolument inattendue.
L'Anglais, qui s'appelait Jimmy Lynch, lui avait semblé pourtant un modèle des vertus sportives britanniques : bonne famille, bonne éducation, brave type.
Tout cela, Beecher l'avait déduit des hasards de la conversation, truffée d'allusions à la R.A.F. où il avait servi pendant la guerre, et de noms de clubs et de collègues anglais très sélects.
Et pour gagner une partie amicale, il avait expédié du talon la balle de Beecher dans le gazon... Beecher était navré.
Parce qu'il avait envie de gagner.
Il commençait à se dire que ça devenait une habitude chez lui de perdre.
Perdre sur tous les tableaux...
Beecher entra dans le salon qui donnait sur le jardin et sur la mer.

Figure 11. Exemple d'ajout sous la forme d'un raccord dans la version française (en bleu). Le chapitre 1 est ici résumé

L'outil de détection des coupes permet donc d'obtenir un fichier qui fait apparaître clairement les coupes et les ajouts, et en facilite l'analyse. Ce résultat est d'autant plus intéressant que la version anglophone du roman se trouve dans le fonds Duhamel à la BiLiPo mais ne présente aucune marque visible de coupes. Cela vient donc renforcer les analyses que nous avons développées dans Frenay et Quaquarelli (2020a, 2020b, 2021 et 2022). Mêmes lorsque les exemplaires anglophones des romans traduits par la « Série Noire » du fonds Duhamel sont vierges de toute annotation, ils subissent le même traitement traductif de coupe, de domestication, de réorientation narrative et de « mise en série » éditoriale, ce dont nous avons fait l'hypothèse.

4. Résultats de l'étude quantitative concernant les coupes en général

4.1. L'ampleur des coupes

Sur 310 romans analysés, 50 (soit 16,1% du corpus) présentent moins de 0,1% de coupes. 127 (41%) présentent des coupes comprises entre 0,1% et 4,9% du total du nombre de caractères ; 75 (24,2%) sont coupés à hauteur de 5% à 14,9% ; 58 (18,7%) sont coupés à plus de 15%. De plus, 32 romans (soit 10,3% du corpus) sont coupés à hauteur de 20 à 60%.

L'outil de détection automatique des coupes permet donc de montrer que la majorité des romans traduits en « Série Noire » sont coupés et que ces coupes atteignent dans plus de 10% des cas des proportions significatives.

4.2. Corrélation significative entre année de publication et pourcentage de coupes

		Pourcentage de caractères coupés			
		1% ou moins	Plus de 1% à 8%	Plus de 8%	Total
Année de publication dans la « Série Noire »	1945 à 1955	48	38	14	100
	1956 à 1963	33	41	26	100
	1964 à 1972	27	23	50	100

Chi2 : 33.77 / P-value: 8.33e-07

Figure 12. Pourcentage de caractères coupés en fonction de l'année de publication dans la Série Noire (en %).

Nous observons une corrélation fortement significative entre l'année de publication dans la « Série Noire » et le pourcentage de coupes effectuées (Figure 12). Nous constatons que la pratique des coupes s'intensifie avec le temps, conformément à l'hypothèse formulée dans Frenay et Quaquarelli (2022), où la période 1945-1955 est conçue comme un moment de mise en place progressive du processus éditorial de traduction. Entre 1945 et 1955, nous observons que 48% des romans sont peu ou pas coupés (entre 0% et 1%) ; 38% sont coupés à hauteur de 1 à 8% et 14% à plus de 8%. La répartition s'équilibre pendant la période 1956-1963 puis s'inverse entre 1964 et 1972 (respectivement 27%, 23%, 50%), alors même qu'entre 1945 et 1972, le nombre moyen de caractères des romans anglophones baisse (dans notre corpus) : il passe de plus de 300 000 à moins de 300 000. Entre 1964 et 1972, un roman sur deux présente plus de 8% de coupes, ce n'est le cas que pour 14% entre 1945 et 1955.

4.3. Réfutation de l'hypothèse de la question matérielle comme unique motivation des coupes

Il existe bien une relation causale évidente entre la taille du roman anglophone et la présence des coupes (Figure 13). Plus un roman est volumineux, plus il a de chance d'être coupé et plus le nombre de caractère retranché est important, mais la contrainte matérielle n'explique pas tout.

Pourcentage de caractères coupés					
		1% ou moins	Plus de 1% à 8%	Plus de 8%	Total
Nombre de caractères dans la version anglaise	Tiers bas	51	35	14	100
	Tiers médian	28	43	29	100
	Tiers haut	26	22	51	100

Chi2 : 40.45 / P-value: 3.49e-08

Figure 13. Pourcentage de caractères coupés en fonction Nombre de caractères dans la version anglaise (en %)

En effet, de nombreux romans de taille modeste sont eux aussi coupés. Si la corrélation entre taille du roman et part de caractères coupés est très forte, nous constatons également que sur la période 1945-1972 la taille des romans en version anglaise tend à diminuer au cours du temps de près de 2000 caractères par année en moyenne (Figure 14).

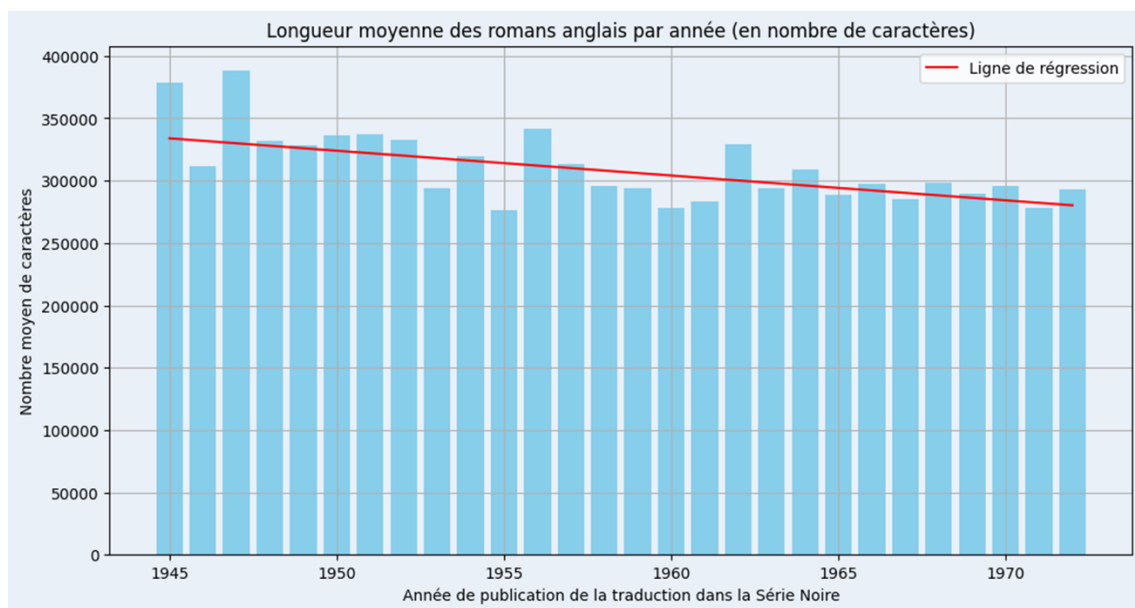


Figure 14. Évolution du nombre moyen de caractères par année pour les romans anglais de notre corpus.

4.4. Vérification de l'hypothèse d'une mise en place progressive d'un processus éditorial de traduction qui s'appuie sur les coupes

Si la corrélation entre année de publication et nombre de caractères des romans à traduire n'est pas significative, cela conforte notre résultat selon lequel la période de publication, toutes choses égales par ailleurs, est une variable fortement explicative des pratiques de coupes.

L'augmentation du recours aux coupes et de la quantité de ces coupes résultent donc bien de l'affirmation d'une pratique éditoriale qui s'exerce aussi

bien sur les romans les plus courts, qui ne nécessitent pas de coupe pour correspondre au format de la collection, qu'aux romans plus longs.

5. Résultats de l'étude quantitative concernant les auteur-es et les traducteur-ices

Pour identifier et isoler l'effet des auteur-es et des traducteur-ices sur les pratiques de coupes, nous effectuons une régression des moindres carrés ordinaires (OLS) pour expliquer le pourcentage de coupes en fonction de l'année de publication, du nombre de caractères, de l'auteur-es et du/de la traducteur-ice. Nous ne conservons que les romans dont le/la traducteur-ices et l'auteur-e apparaît au moins trois fois dans le corpus.

Cette régression confirme d'emblée le rôle très significatif de l'année de publication et de la taille de la version anglophone sur le pourcentage de coupe de la traduction. Toutes choses égales par ailleurs, chaque année qui passe donne une augmentation moyenne de 0.46 points de pourcentage de coupes ($p \leq 0.021$). De même, une augmentation de 1000 caractères donne une augmentation moyenne de 0.08 points ($p < 0.000$).

Pour ce qui est de l'effet des auteur-es et traducteur-ices, nous obtenons peu de résultats significatifs. Dans la majorité des cas, et sur la base des données de notre corpus, les auteur-es et traducteur-ices d'un roman n'influent pas de manière significative sur le taux de coupes.

Cela montre que l'équipe éditoriale pense cette collection comme une série, en dehors de toute politique d'auctorialité exclusive. Le matériel textuel, quelle que soit sa provenance, est traité dans la plupart des cas de la même manière. L'opération de traduction est uniformisée au point que, du point de vue des coupes, les plus de 200 traducteurs et traductrices, semblent traduire de la même façon. – ce qui est tout à fait improbable.

6. Limites et perspectives

Les résultats obtenus doivent être considérés en fonction des limites inhérentes à la méthode employées et au corpus choisi. Nous avons testé l'outil de détection automatique des coupes sur un corpus de romans de la collection « Série Noire » déjà numérisés. Il est donc possible que, malgré nos vérifications, des erreurs de numérisation demeurent.

D'autre part, il est possible que les versions numérisées de la traduction française soient établies à partir de retraductions ou rééditions ou, encore, que la première traduction ait été complétée, sans qu'aucune mention n'en prévienne le lecteur. Par conséquent, une partie des romans qui présentent moins de 1% de coupes pourraient en réalité être coupés davantage.

Enfin, notre choix de segmentation et d'analyse s'est porté sur la phrase. Les résultats sont probants à ce niveau, mais ne permettent pas de déceler des coupes qui seraient effectuées à l'intérieur de celles-ci. Nous pouvons donc considérer que la détection des coupes est de manière générale légèrement sous-évaluée, sans qu'il soit possible de quantifier précisément ce phénomène.

Cependant, nous sommes convaincu-es de l'importance et l'utilité du dispositif tant pour notre recherche spécifique, qui pourra à l'avenir s'appuyer sur de nouvelles numérisations des ouvrages de la « Série Noire » afin d'obtenir des résultats encore plus précis et concluants, que pour d'autres recherches encore, dans le champ de la littérature de grande consommation et jeunesse par exemple, au sein desquels la pratique de coupes est généralement très répandue. Nous pourrions y inclure également un travail d'enquête quantitative sur les retraductions, les rééditions et les republications, afin de vérifier, et sans doute déjouer, un certain nombre d'appellations (« traduction intégrale », « nouvelle traduction »...) ou d'imaginaires justifiant ou motivant une nouvelle traduction, comme celui du « vieillissement » des traductions par exemple.

BIBLIOGRAPHIE

- Conneau, A., K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, V. Stoyanov (2020) *Unsupervised Cross-lingual Representation Learning at Scale*, arXiv:1911.02116.
- Frenay, A. et L. Quaquarelli (2020a) « Faire américain. Fabbricare testi e autori: tradurre », *mediAzioni* 27 : A166-A190, https://mediazioni.sitlec.unibo.it/images/stories/PDF_folder/document-pdf/27-2020/quaquarelli-frenay.pdf.
- Frenay, A. et L. Quaquarelli (2020b) « Ce n'est pas tout à fait Série Noire ! Quand la traduction produit du roman noir », in D. Cailleux, C. Denti, L. Quaquarelli (éds.) *Expériences de traduction*, Bruxelles : Peter Lang, 71-89.
- Frenay, A. et L. Quaquarelli (2021) « “Nettoyage par le bide”. La Série Noire à l'épreuve des archives », *Belphégor* 19(2) : 1-18, en ligne, <https://doi.org/10.4000/belphegor.4342>.
- Frenay, A. et L. Quaquarelli (2022) « Sérialiser la Série Noire. Roman noir, traduction et génétique sérielle », *Génésis. Revue internationale de critique génétique* 54: 57-70, <https://doi.org/10.4000/genesis.7025>.
- Gouanvic, J.-M. (2018) *Hard-boiled fiction et Série Noire*, Paris : Classiques Garnier.
- Needleman, S. B., C. D. Wunsch (1970) « A general method applicable to the search for similarities in the amino acid sequence of two proteins », *Journal of Molecular Biology* 48(3) : 443-453.
- Robyns, Clem (1990) « The Normative Model of Twentieth Century Belles Infidèles: Detective, Novels in French Translation », *Target. International Journal of Translations Studies* 1 : 23-42.