

« Vous pouvez entendre les mâchoires de l’arme de l’ange ».  
Le traducteur humain face à la traduction automatique  
dans le contexte littéraire :  
*Les Livres de Jakób* d’Olga Tokarczuk

“You can hear the jaws of the angel’s weapon”.  
The human translator versus machine translation  
in the literary context:  
the case of Olga Tokarczuk’s novel *The Books of Jacob*

Magdalena Mitura

Université Marie Curie-Skłodowska de Lublin

[magdalena.mitura@mail.umcs.pl](mailto:magdalena.mitura@mail.umcs.pl)

<https://orcid.org/0000-0002-2871-6719>

**Abstract**

The dynamic development of automatic translators raises questions about the future of human translators. The aim of this article is to evaluate the French versions of an excerpt from *The Books of Jacob* by Olga Tokarczuk, made by Systran, Google Translate and DeepL, and to compare them with the reference translation made by a human translator. The translations were carried out in 2019 as part of the ATLAS association project ‘L’Observatoire de la traduction automatique’. The analyses show that the inability to interpret the original in a complex way leads to weaknesses in machine translators. This proves that the position of the human translator remains unthreatened in the field of translating high-quality literary texts.

**Keywords:** neural machine translation, literary translation assessment, Olga Tokarczuk, Maryla Laurent

« Dans une traduction,  
choisir un mot – un seul –,  
c'est déjà proposer  
une vision du monde.  
Et c'est un mal salubre »  
(Defert, 2022, résumé).

## 1. INTRODUCTION

Les profanes de la traductologie peuvent avoir l'impression que l'histoire de la traduction automatique (TA) a commencé il y a à peine quelques années et qu'elle coïncide avec l'apparition de versions gratuites de moteurs de traduction disponibles sur l'internet. Certes, les progrès les plus spectaculaires sont notés à partir de l'introduction des réseaux de neurones artificiels, mais les recherches ont commencé aux États-Unis dans les années 50 du siècle précédent. À l'époque, l'intérêt pour l'automatisation de la traduction était lié aux systèmes de décryptage mis au point pendant la Deuxième Guerre mondiale. L'impulsion donnée à son développement était due à une atmosphère politique tendue de part et d'autre du rideau de fer (Melby & Warner, 1995, p. 15).

Le format court de cet article exclut la présentation de l'état des recherches dans ce domaine. Tel n'est pas d'ailleurs notre objectif, d'autant plus que les synthèses diachroniques sur la TA des textes généraux et sur la traduction automatique littéraire (TAL) ont déjà été soigneusement réalisées (Loffler-Laurian, 1996 ; Wilks, 2009 ; Hansen, 2021 ; Marzi, 2021 ; Baillot, 2022)<sup>1</sup>. La première étude pilote sur la TAL pour la paire anglais – français (Besacier, 2014) a été suivie des travaux consacrés à d'autres paires linguistiques, souvent périphériques où intervenaient par exemple le gaélique, l'écossais, l'irlandais, le catalan (Hansen, 2021, p. 32 ; Toral & Way, 2018).

L'arrivée en 2014 du logiciel basé sur des algorithmes d'intelligence artificielle (IA) de type *apprentissage neuronal profond* a marqué une étape cruciale quant aux performances de la TA. Les capacités des systèmes de traduction automatique neuronale (TAN) ont de loin dépassé les outils offerts par ses prédécesseurs : les systèmes de la TA à base de règles ou les systèmes statistiques basés sur de grands corpus (Koehn, 2010). Quant aux textes littéraires, le potentiel de la TAN est constamment mis à l'épreuve et donne lieu à des débats entre opposants farouches et partisans ardents de cet outil (Froeliger, 2022). En effet, l'engouement croissant pour l'IA s'exprime, entre autres, par une liste de professions qui vont bientôt disparaître en raison des performances de ChatGPT. À cet égard, le développement de la TAN laisse présager que, dans ce domaine également, l'IA remplacera l'intelligence humaine.

---

<sup>1</sup> La TA fait l'objet d'un nombre toujours croissant de travaux et de colloques suite à un essor fulgurant de l'IA. La bibliographie citée dans l'article est donc nécessairement sélective. Sans prétendre à l'exhaustivité, elle ancre nos analyses dans le panorama des travaux théoriques menés jusqu'à présent.

D'autant plus que l'expérience des traducteurs de textes informatifs confirme que les outils d'aide à la traduction réduisent considérablement le temps du processus traductif. Depuis que l'accès à des logiciels de la TA s'est démocratisé grâce à leur mise en ligne gratuite, chacun a sa petite idée de leur utilité qui peut osciller entre la critique des lacunes linguistiques ou la crainte<sup>2</sup> de voir la machine rendre l'homme inutile d'une part, et l'euphorie que de nouvelles technologies épargneront l'effort du traducteur humain, de l'autre part (Loock, 2019).

## 2. CORPUS DE TRAVAIL ET MÉTHODOLOGIE

Bien que l'évaluation de la TA soit un domaine des recherches très dynamique, les travaux consacrés à la TAL sont beaucoup plus rares que ceux concernant des textes informatifs. La réflexion reste en friche, au moins autant que nous le sachions, quant à la TA de la littérature polonaise. Pour cette raison, il nous a paru intéressant d'observer les problèmes traductifs soulevés dans la paire polonais-français. À cette fin, nous évaluons trois versions françaises d'un extrait des *Livres de Jakób* d'Olga Tokarczuk, toutes effectuées le 4 novembre 2019 par trois moteurs de la TA : Systran, Google Traduction et DeepL<sup>3</sup>. Réalisées dans le cadre du projet *L'Observatoire de la traduction automatique* de l'association ATLAS<sup>4</sup>, elles présentent les sorties brutes, autrement dit qui n'ont été soumises ni à la post-édition automatique ni à la révision humaine.

Précisons que l'extrait de l'original fait partie du *Livre de la comète* (Tokarczuk, 2017, pp. 472-469), compte 5500 signes, espaces compris et il a été sélectionné par la traductrice à la demande de Jörn Cambreleng, directeur de l'ATLAS et coordinateur de l'*Observatoire*<sup>5</sup>.

L'analyse du corpus s'est déroulée en deux étapes. Premièrement, la confrontation des trois versions avec l'original nous a permis de juger l'adéquation de la TA. Deuxièmement, leur comparaison avec la « traduction de référence » (Koehn, 2010, p. 222), faite par Maryla Laurent (la traductrice professionnelle et à la fois professeur de traductologie et de littérature) a donné lieu à des observations au niveau de la fluidité du texte cible. Cette double démarche nous a mené à repérer les erreurs générées par les logiciels qui témoignent de leurs défaillances face au texte littéraire.

La méthodologie adoptée pour systématiser ces points faibles s'inscrit dans le cadre de « l'évaluation humaine » (Denkowski & Lavie, 2010). Notre corpus étant très limité par rapport au nombre d'unités habituellement traitées dans l'évaluation de

<sup>2</sup> Cronin (2016, p. 9) parle même de « la désorientation effrayante et extrême » de l'homme moderne face à un répertoire d'outils de la traduction numérique en constante expansion.

<sup>3</sup> En 2019, ces trois logiciels utilisent déjà la TAN.

<sup>4</sup> <https://www.atlas-citl.org/observatoire-de-la-traduction-automatique/>.

<sup>5</sup> Je tiens à exprimer ma gratitude à Maryla Laurent pour m'avoir fourni sa correspondance privée avec Jörn Cambreleng, les versions de la TA et sa propre traduction du fragment choisi.

la TA, nous avons opté pour une méthode de classification et d'analyse linguistique des erreurs engendrées par les moteurs de la TA. Plusieurs typologies d'erreurs ont été proposées jusqu'à présent en fonction du corpus étudié et des objectifs de l'étude (Popović, 2018, p. 14 ; Grass, 2022, pp. 17-21). En ce qui concerne l'établissement de la taxonomie appropriée, M. Popović (2018) remarque :

Elle peut être effectuée manuellement, automatiquement ou en utilisant une méthode combinée (semi-automatique). Différentes sources d'information (en plus de la traduction analysée) peuvent être utilisées telles que les textes en langue source, les traductions de référence ou, plus récemment, les traductions post-éditées (p. 131, trad. M.M.).

En nous inspirant des typologies existantes, mais surtout en accord avec la spécificité des résultats obtenus pendant l'examen détaillé, nous avons dressé la classification d'erreurs pertinentes pour notre corpus. Celles-ci peuvent être généralisées en quatre catégories principales qui sont : le lexique (le choix de mots isolés), la morphosyntaxe (le choix au niveau intraphrastique), la textualité (la cohésion et la cohérence) et l'orthographe. Il est indispensable de noter que les frontières ne sont pas étanches entre ces domaines. Souvent, les erreurs classées dans chacun d'eux se conditionnent mutuellement. Comme nous allons le voir, par exemple les distorsions syntaxiques entraînent une rupture de cohésion et provoquent des nonsens.

Notre hypothèse de travail part du principe que la complexité du texte artistique, due à l'inséparabilité de la vision subjective du monde construite et des moyens linguistiques utilisés pour y parvenir, constitue une entrave majeure au développement rapide de la TA des œuvres littéraires.

L'objectif de l'article n'est pas d'évaluer la qualité des trois moteurs de la TA dans leur ensemble. Le corpus limité rendrait une telle entreprise trop ambitieuse, mais il offre sans aucun doute des opportunités pour ouvrir la discussion sur la TAL du polonais vers d'autres langues. L'analyse des erreurs permettra de saisir les points spécifiques à la langue polonaise qui résistent au traitement automatique du texte littéraire. Le profil des erreurs face aux solutions proposées par Maryla Laurent incitera à réfléchir dans quelle mesure le traducteur humain reste (ir)remplaçable dans la traduction littéraire de haute qualité. La réponse à cette question est au cœur de la présente étude.

### 3. ANALYSE DES ERREURS

#### 3.1. LES ERREURS LEXICALES

Les erreurs lexicales constituent une catégorie la plus représentée dans le corpus analysé. Elles affectent les formes verbales et nominales. La première sous-catégorie peut être illustrée par l'exemple du groupe verbal *zadzierać głowę* qui fait partie de la

phrase « Kometa przypomina kosę wymierzoną w ludzi, nagie, lśniące ostrze gotowe w każdej chwili obciąć miliony **głów**, nie tylko tych **zadartych** w górę w Iwaniu [...] » (p. 472<sup>6</sup>). Le verbe *zadzierać* a été rendu par M. Laurent comme « têtes [...] **tournées** vers ciel »<sup>7</sup>. DeepL propose « têtes [...] **déchirées** », ce qui prouve la non prise en considération du contexte et un mauvais choix parmi les significations potentiellement ouvertes. Le verbe *zadzierać* est notoirement polysémique en polonais. La première signification (choisie par le moteur) exprime le fait de *déchirer légèrement quelque chose*, la deuxième celui de *s'en prendre à quelqu'un*. L'original contient en fait la troisième signification du verbe, à savoir *soulever quelque chose*, en l'occurrence *soulever la tête*<sup>8</sup>. Par contre, il nous serait difficile de voir une justification quelconque pour la proposition de Systran « têtes [...] **condamnés** », déformant complètement le sens de l'original. La solution de Google « têtes [...] **relevées** » peut être considérée correcte hors contexte. Il n'en reste pas moins que le verbe n'est pas un choix lexical naturel en français pour décrire la posture des personnes regardant le ciel. En effet, il entre dans des collocations avec des noms autres que *tête*, tels que par exemple « relever son col, ses lunettes, sa moustache, sa robe, son voile, le front, les sourcils, les yeux, l'ancre » (Dictionnaire CNRTL, 2.09.2023).

Les mauvaises traductions des substantifs donnent souvent lieu à des solutions complètement absurdes, comme le montre la TA de deux phrases qui suivent : « Za horyzontem szykują się już wojska archaniołów. Jeżeli wyteży się słuch, słyhać **szczęk** anielskiego oręża » (p. 471). Le lexème *szczęk* est polysémique<sup>9</sup> en dehors du contexte. Il peut signifier, entre autres, le son produit lorsque des objets en métal se heurtent l'un contre l'autre. Deuxièmement, c'est une forme du génitif du nom *szczęki/mâchoires*. La traductrice interprète correctement le sens et choisit l'équivalent approprié : « À l'horizon, des cohortes d'archanges se préparent déjà. Qui tend l'oreille entend le **tintement** de leurs armes ». En revanche, tous les moteurs automatiques choisissent la deuxième lecture, ce qui aboutit à une image apocalyptique : « vous entendez **la mâchoire d'une arme** d'ange » (S) ; « vous pouvez entendre **les mâchoires de l'arme** de l'ange » (G) ; « on peut entendre **les mâchoires de l'arme** de l'ange » (D).

La qualité des traductions baisse dramatiquement lorsque la densité des erreurs ne se borne plus à altérer les sens ponctuels, mais déforme l'image décrite. L'incongruité des équivalents erronés peut entraîner une perturbation quasi totale de la lecture en raison de l'absence de liens logiques entre les faits présentés. La TA du passage suivant en est un exemple flagrant :

<sup>6</sup> Toutes les citations de l'original proviennent de l'édition figurant dans la bibliographie.

<sup>7</sup> Pour la commodité du lecteur francophone, nous mettons la traduction de référence avant la TA.

<sup>8</sup> Le verbe *zadzierać* entre aussi dans la locution verbale *zadzierać nosa* (littéralement *tourner le nez vers le haut*) et veut dire *être arrogant*.

<sup>9</sup> La polysémie constitue un problème fondamental dans la description de la langue à des fins de la TA (cf. Śmigielska, 2011).

Jenta widzi na przykład, że Iwanie ma szczególny status w **hierarchii bytów**; **nie jest do końca usadowione na ziemi**, nie jest do końca realne. **Chaty** stoją **zgarbione jak żywe istoty, starodawne tury**, z pyskami do ziemi, grzeją zmarzniętą ziemię oddechem (p. 471).

La traductrice apporte ici :

Jenta voit par exemple qu'Iwanie bénéficie d'un statut particulier dans **la scala naturæ**<sup>10</sup>, **la grande chaîne de la vie**. Le village **n'est pas tout-à-fait sur terre**, pas tout à fait réel. **Les chaumières** sont **voûtées comme des êtres vivants**, elles sont pareilles aux **anciens aurochs** qui, le museau au sol, réchauffent de leurs souffles la terre gelée.

Examinons les trois versions de la TA :

Par exemple, **il** voit qu'Ivan a un statut spécial dans **la hiérarchie des animaux** ; ce **n'est pas complètement assis sur le sol**, ce n'est pas tout à fait réel. **Les cottages** sont **ø des créatures mortes**, de **vieilles excursions**, avec **leur bouche au sol**, et ils réchauffent la terre gelée avec de l'haleine (S).

Par exemple, Jenta voit qu'Ivan a un statut spécial dans **la hiérarchie des êtres**; il **n'est pas complètement assis**, ce n'est pas réel. **Les huttes** se tiennent **comme des êtres vivants, d'anciennes tournées**, la **bouche bouche bée**, réchauffant le sol gelé (G).

Jenta voit, par exemple, qu'Ivan a un statut spécial dans **la hiérarchie des êtres** ; il **n'est pas pleinement établi sur terre**, il n'est pas entièrement réel. **Les huttes** sont **affamées comme des êtres vivants, d'anciens aurochs, la bouche sur le sol**, réchauffent le sol gelé par le souffle (D).

La version de Systran présente clairement la qualité la plus faible. Tout d'abord, le substantif *Jenta*, désignant l'héroïne principale, de la phrase précédente est anaphorisé par le pronom masculin *il*, ce qui rompt le lien cohésif et dérouté le lecteur. Ensuite, la décision de traduire le mot *êtres* par *animaux* restreint considérablement l'étendue de *l'échelle des êtres* dans laquelle les animaux ne sont qu'un élément à côté des hommes ou du monde végétal. L'omission (signalée par  $\emptyset$ ) de la comparaison dans la phrase suivante aboutit à un aplatissement stylistique. Nous obtenons ainsi une personification incompréhensible des chaumières devenues *mortes*, et donc un contresens par rapport à l'original. Le substantif choisi comme équivalent du mot *chaty* est également erroné. Tokarczuk dépeint des habitations modestes, recouvertes de paille. Leur équivalent le plus proche en français est précisément le mot *chaumières*<sup>11</sup>, proposé par la traductrice. Les choix des mots *cottage* et *hutte* sont en revanche inappropriés, car

<sup>10</sup> Ce terme a été annoté par la traductrice comme « Conception médiévale à dimension théologique relayée par Copernic puis Linné dans son *Systema naturae* en 1753, réactualisée au 18e par Leibniz. (N.d.T.) ».

<sup>11</sup> « Maison rurale couverte de chaume [...]. Maison simple et pauvre » (CNRTL, 2.09.2023).

ils définissent des structures différentes, respectivement trop ou pas assez raffinées<sup>12</sup>. Une autre erreur lexicale concerne la traduction du mot *tury*. Manifestement, il est absent des ressources lexicales à disposition de Systran et Google. Le logiciel l'identifie donc en se basant sur la forme la plus proche graphiquement, à savoir *tours* et donne comme équivalent *excursions* et *tournées*. DeepL s'en sort beaucoup mieux avec la première phrase et ne tombe pas dans le piège de *tours*. La traduction absurde « Les huttes sont affamées » et l'ambiguïté du groupe verbal « réchauffent le sol gelé par le souffle » dans la deuxième phrase n'en sont que plus surprenantes.

### 3.2. LES ERREURS MORPHOSYNTAXIQUES

Les erreurs morphosyntaxiques constituent une catégorie très hétérogène. Elle englobe par exemple les fautes relatives au nombre, au genre, aux parties de la phrase ou à la réction du verbe. Dans l'exemple cité auparavant et concernant le tintement d'armes des archanges, nous avons le syntagme adjectival *anielski oręż*. En polonais, le substantif *oręż* représente singularia tantum. Morphologiquement au singulier, il désigne pourtant un ensemble d'armes. La forme *anielski* (littéralement *angélique*) recouvre aussi bien le singulier (*les armes de l'ange*) que le pluriel (*les armes des anges*). Seul le contexte de la phrase précédente (*wojska archaniołów*) désambiguïse le sens en puissance et impose une interprétation unique au pluriel. Conformément à ce raisonnement, M. Laurent donne « leurs armes », vu que l'adjectif possessif se réfère aux *cohortes d'archanges*. Toutefois, une telle analyse dépasse les capacités de l'IA. Systran opte alors pour le singulier « une arme d'ange », malgré le pluriel « les archanges » dans la phrase précédente. Google et DeepL proposent une traduction erronée dès la première phrase, en remplaçant le pluriel par le singulier : « Les troupes de l'Archange » et « Les armée d'archange ». Les singuliers « les mâchoires de l'arme de l'ange » dans la deuxième phrase n'en est que la conséquence.

L'exemple suivant montre les difficultés rencontrées par la TA en raison des divergences au niveau de l'ordre des constituants phrastiques dans les deux langues : « Przyglądają jej się **kobiety**, które z niepokojem lada chwila spodziewają się porodu, i ściśnięte pod pokładem **rodziny**, które filigranowymi stateczkami przemierzają ocean w poszukiwaniu nowego życia po tamtej stronie (p. 472) ». La flexibilité syntaxique du polonais, aidée par une riche flexion, est un moyen précieux pour construire la progression thématique dans le texte d'une manière qui marque la perspective subjective de son auteur. Ainsi, observons-nous le bouleversement de l'ordre canonique SVE (*Les femmes et les familles regardent la comète*). Les deux noms additionnés (*kobiety* et *rodziny*), qui forment un groupe sujet composé, ont été séparés et placés

<sup>12</sup> Cottage : « Petite maison élégante de la campagne anglaise » (CNRTL, 2.09.2023). Hutte : « Abri ou habitation rudimentaire, construite généralement à l'aide de matériaux divers grossièrement assemblés » (CNRTL, 2.09.2023).

après le prédicat et à la fin de la phrase. La traductrice restaure la structure classique en français où le sujet se situe en tête des propositions : « **Les femmes** qui vont accoucher à tout instant l'observent avec anxiété, tout comme **les familles** qui, entassées à bord de fragiles esquifs, traversent l'océan en quête d'une nouvelle vie ». En termes de la perspective fonctionnelle, il est possible de voir dans ces deux versions la réalisation de la progression thématique à thème constant (Maingueneau, 2000, p. 161). En polonais, l'élément thématique (*la comète*) a une tendance nette à se positionner au début de la phrase, les éléments rhématiques (*kobiety* et *rodziny*) viennent ensuite. En français, la perspective fonctionnelle de la phrase est inverse, les éléments rhématiques (*les femmes* et *les familles*) occupent la position initiale, mais tous les deux réalisent la fonction grammaticale sujet. Chaque version de la TA introduit la diathèse passive qui permet de simplifier les nuances syntactico-cohésives mentionnées à travers la scission du sujet grammatical (*Elle*) avec le thème (*la comète*) dans la première proposition. Cependant, la postposition du deuxième élément du sujet (*les familles*) dans l'original s'avère être un obstacle à l'identification correcte par les moteurs de la TA :

Elle est surveillée par **des femmes** qui s'attendent à une naissance anxieuse, et coincées sous le pont de **la famille**, qui fuient l'océan à la recherche d'une nouvelle vie de l'autre côté (S).

Elle est surveillée par **des femmes** qui attendent avec impatience l'accouchement et est coincée sous le pont de **la famille** qui, dans des navires en filigrane, sillonne l'océan à la recherche d'une nouvelle vie de ce côté (G).

Elle est surveillée par **des femmes** qui attendent anxieusement l'accouchement d'un moment à l'autre et qui se pressent sous le pont de **leur famille**, alors qu'elles traversent l'océan sur des bateaux filigranes à la recherche d'une nouvelle vie de ce côté-là de l'océan (D).

En raison de sa proximité sur la ligne de la phrase, le mot *familles* est interprété comme complément du nom du substantif *pokład/le pont du navire*. Cette solution, dépourvue de sens, s'accompagne dans la version de Google d'une anaphorisation erronée de l'adjectif verbal *coincé*. Ses nombre et genre indiquent notamment qu'il instaure le lien anaphorique avec le sujet de la phrase (*la comète*), alors qu'il se réfère en réalité aux familles entassées sous le pont du navire.

### 3.3. LA TEXTUALITÉ

La première phrase de l'original introduit comme sujet le substantif *kometa/une comète* : « Roku 1759, 13 marca, na niebie pojawiła się **kometa** i jakby na **jej** komendę śniegi nagle stopniały [...] » (p. 472). La traductrice la rend ainsi : « En 1759, le 13 mars, **une comète** apparut dans le ciel, et alors, comme si elles obéissaient à **son** ordre, les neiges fondent soudain [...] ». La comète constitue l'hypertexte autour

duquel s'articule la progression thématique et, par conséquent, la cohérence de la première moitié de l'extrait analysé. Elle joue le rôle du pivot anaphorique, repris dans la suite du texte notamment à travers des anaphores pronominales, mais aussi des anaphores nominales et une position vide :

Wiele dni **o** wisi nad całym wilgotnym i rozległym światem [...]. Widać **ją** nad całym światem. Widać **ją** nawet w Chinach. Patrzą na **nią** żołnierze [...]. Przyglądają **jej** się kobiety [...]. **Kometa** przypomina kosę wymierzoną w ludzi. [...] **to** jest **znak końca świata** [...] (pp. 472-471).

M. Laurent reproduit correctement cette grille anaphorique, en apportant parfois des modifications dans la répartition des éléments thématiques et rhématiques, comme nous l'avons signalé dans l'un des exemples précédents :

[...] **elle** reste en suspension au-dessus du vaste monde humide pendant de nombreux jours. **Elle** est visible de partout. On **la** voit même de Chine. [...] les soldats [...] **la** regardent [...]. Les femmes [...] l'observent [...]. **La comète** rappelle une faux prête à s'abattre sur les hommes. [...] il s'agit là d'**un signe de la fin des temps** [...].

Dans les sorties successives de la TA, on constate une légère amélioration de la qualité de la traduction. Néanmoins, même la version de DeepL ne peut être considérée comme satisfaisante. Commençons par les résultats de Systran et Google :

**De nombreux jours** sont suspendus sur tout le monde humide et vaste [...]. Vous pouvez **le** voir partout dans le monde. Vous pouvez **le** voir même en Chine. **Elle** est vue par les soldats [...]. **Elle** est surveillée par des femmes [...]. **La comète** ressemble à une escroquerie dirigée contre les gens [...] c'est **un signe de la fin du monde** [...] (S).

**De nombreux jours** couvrent tout le monde humide et vaste [...]. Vous pouvez **le** voir partout dans le monde. Vous pouvez **le** voir même en Chine. Des soldats [...] **la** regardent [...]. **Elle** est surveillée par des femmes [...]. **La comète** ressemble à une faux dirigée contre les hommes [...] c'est **un signe de la fin du monde** [...] (G).

Vu la position initiale du syntagme *wiele dni/de nombreux jours*, ce complément circonstanciel est reconnu mécaniquement comme sujet. Ainsi, cette deuxième phrase, outre l'absence de lien grammatical, ne présente-t-elle non plus de connexité au niveau de la cohérence sémantique avec la phrase d'ouverture. La reprise pronominale du sujet dans les deux phrases successives (qui n'en est d'ailleurs pas une à cause des différences de nombre grammatical entre le référant *le* et le référé *jours*) fait s'égarer de plus en plus le lecteur potentiel. En raison de contraintes grammaticales, il établira un lien entre le pronom personnel *le*, interprété alors comme neutre, et toute la phrase précédente, rompant ainsi la progression thématique que développe la description de la comète en polonias. Le mot *comète* ne réapparaît qu'en septième phrase et c'est

seulement à partir de ce moment-là qu'il devient possible d'instaurer le lien de corréférence avec sa première occurrence.

DeepL identifie correctement le complément circonstanciel et le sujet dans l'original. Malheureusement, il perturbe la corrélation dans les deux phrases consécutives, comme cela s'est produit pour les TA précédentes :

Pendant de nombreux jours, **elle** plane sur l'ensemble du monde humide et vaste [...]. On peut **le** voir dans le monde entier. On peut **le** voir même en Chine. **Elle** est regardée par des soldats [...]. **Elle** est surveillée par des femmes [...]. **La comète** ressemble à une faux pointée sur les gens [...] c'est **un signe de la fin du monde** [...] (D).

### 3.4. L'ORTHOGRAPHE

La présence de fautes d'orthographe a été constatée principalement dans les noms propres. Examinons-les en prenant comme exemple quatre occurrences sélectionnées de l'original, à savoir deux noms de personnes *Jakub*, *Jenta*, un nom de l'immanence divine dans le monde *Szechina* et *Iwanie*, le nom du village de Podolie. Dans la traduction humaine, ils sont systématiquement rendus par : « Jakób, Yenta, *Shekhina*, Iwanie ». Systran apporte deux versions assez divergentes pour le premier nom : « Yakub, Jacob », une seule pour le deuxième : « Jent », trois versions pour le troisième : « Shakhina, Shekhin, Shakhin », et deux versions pour le nom du village : « Ivanie, Ivan ». Dans les versions de Google et DeepL, nous avons respectivement : « Jacob, Jakub, Jenta, Jent, Szechina, Széchin, Ivan » (G) et « Jacob, Jent, Jencie (le datif de *Jenta*), Shekhinah, Shechina, Ivan » (D).

Si ces déformations ne sont pas surprenantes dans le cas du premier système, elles le sont déjà fortement dans les deux moteurs automatiques suivants, plus avancés dans l'interprétation de sens complexes. Les fautes d'orthographe peuvent sembler anodines car elles ne concernent que des lettres isolées. Cependant, leur impact ne se limite pas au niveau graphique car les noms propres constituent les nœuds les plus forts du réseau de cohésion dans le texte. Leurs déformations perturbent donc la perception de l'unicité du monde présenté dans l'œuvre. Par conséquent, elles introduisent le risque d'une lecture erronée au cours de laquelle les formes fautives ouvrent l'identification potentielle des nouvelles entités extralinguistiques.

## 4. CONCLUSION

La méthode de classification et d'analyse des erreurs choisie a permis de mettre en évidence et d'explorer en profondeur les spécificités du corpus et d'identifier au mieux les obstacles auxquels se heurte la TAL du polonais vers le français. La première observation qui s'impose au vu des résultats obtenus corrobore les attentes découlant du

savoir commun non professionnel. Notamment, Systran a donné la traduction la plus faible en termes d'adéquation. La version effectuée par DeepL est celle qui comporte le moins d'erreurs par rapport à l'original. Les résultats sont répartis de la même manière pour le deuxième critère, celui de fluidité. Dans le version de Systran, le nombre d'erreurs commises rend le texte traduit pratiquement illisible dans son ensemble. Considérée comme un texte autonome, la version de DeepL fournit la lecture relativement la plus cohérente et alors la plus compréhensible. Ceci dit, même dans le cas de cette traduction, nous sommes toujours loin d'un texte satisfaisant d'un point de vue linguistique et encore moins stylistique.

Les faiblesses identifiées à différents niveaux de trois versions automatiques permettent de conclure que toute post-édition serait ici une solution insatisfaisante, voire improductive, en raison de l'énormité des modifications requises.

Comme il sied pour chaque conclusion, nous avons esquissé les observations de synthèse découlant des analyses comparatives effectuées. Cependant, nous avons une conviction tenace qu'elles n'épuisent pas l'essentiel de la question se résumant en fait au rôle du traducteur du texte littéraire. Un traducteur humain n'est pas irremplaçable parce qu'une machine choisit mal les équivalents, ne gère pas du tout la polysémie ou ne tient pas compte du contexte intra- et interphrastique. Il en est ainsi parce que la traduction constitue à chaque fois un acte unique de recréation du monde en langue cible, le monde construit auparavant par l'auteur de l'original, ce qui confirme l'hypothèse de travail que nous avons formulée au début de l'article. Seuls l'imagination et l'intellect du traducteur humain en tant que sujet cognitif est capable non seulement de saisir les sens internes du texte, mais surtout de prévoir comment les façonner linguistiquement et stylistiquement pour s'assurer que le lecteur cible ait droit au même monde que le lecteur de l'œuvre source.

Notre article est une étude pilote pour cette paire de langues. En tant que tel, il aborde des aspects sélectionnés, concernant un spéciment du texte littéraire de haute qualité. Contrairement à la plupart des travaux existants sur la TAL (Toral & Way, 2018), les analyses que nous avons effectuées sont basées sur un ensemble de données très limité. Par conséquent, leurs résultats ont pu être soumis à une évaluation humaine. Bien que très chronophage, cette méthode garantit, pour des corpus peu volumineux, la précision et l'intégralité des observations.

Nous espérons avoir montré les perspectives prometteuses qui s'ouvrent dans le domaine de l'évaluation des œuvres. Par exemple, il faut soumettre l'original aux performances actuelles des moteurs de la TAN, qui ont considérablement augmenté. Cela permettrait de vérifier si les mêmes types d'erreurs persistent ou bien si l'alimentation des bases de données neuronales avec de nouveaux textes littéraires permet d'éliminer certains types de difficultés identifiées<sup>13</sup>. Prolonger les investigations sur d'autres corpus, plus étendus de la TAL donnerait sûrement lieu à d'intéressantes observations.

---

<sup>13</sup> Signalons à cet égard l'existence d'un excellent corpus littéraire français-polonais (Bralewski, 2012) qui attend toujours à être traité avec les outils perfectionnés de la TAN.

## BIBLIOGRAPHIE

- Baillot, A., Carter, E., Grass, T. & Fabo, P.R. (2022). Vers une robotique du traduire – Introduction. *Journal of Data Mining & Digital Humanities*. <https://doi.org/10.46298/jdmdh.10445>.
- Besacier, L. (2014). Traduction automatisée d'une œuvre littéraire: une étude pilote. *Traitement Automatique du Langage Naturel (TALN)*. Marseille. <https://inria.hal.science/hal-011003944/document>.
- Bralewski, D. (2012). *Od przekładu do słownika. Korpus równoległy w redakcji słowników tłumaczeniowych*. Łask : Oficyna Wydawnicza LEKSEM.
- Cronin, M. (2016). *Przekład w epoce cyfrowej*. Kraków : Wydawnictwo Uniwersytetu Jagiellońskiego.
- Defert, D. (2022). Dan Brown, Patricia Cornwell et John Grisham à l'épreuve de DeepL. *Journal of Data Mining & Digital Humanities*. <https://doi.org/10.46298/jdmdh.9084>.
- Denkowski, M. & Lavie, A. (2010). Choosing the Right Evaluation for Machine Translation : an Examination of Annotator and Automatic Metric Performance on Human Judgment Tasks. In *Proceedings of the 9th Conference of the Association for Machine Translation in the Americas : Research Papers*, Denver. <https://aclanthology.org/2010.amta-papers.20/>.
- Dictionnaire CNRTL en ligne. <https://www.cnrtl.fr>.
- Froeliger, N. (2022). *Le mépris (envers la traduction automatique) est-il une attitude traductologique ?*. <https://hal.science/hal-03507329/document>.
- Grass, T. (2022). L'erreur n'est pas humaine. *Traduire*, 246, 10-23. <https://doi.org/10.4000/traduire.2763>.
- Hansen, D. (2021). Les lettres et la machine : un état de l'art en traduction littéraire automatique. In P. Denis et al. (éds.), *Actes de la 28e Conférence sur le Traitement Automatique des Langues Naturelles*, vol. 2 : 23e REcontres jeunes Chercheurs en Informatique pour le TAL (RECITAL). Lille (pp. 28-45). <https://aclanthology.org/2021.jeptalnrecital-recital.3/>.
- Koehn, P. (2010). *Statistical Machine Translation*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511815829>.
- Loffler-Laurian, A.-M. (1996). *La traduction automatique*. Villeneuve d'Ascq : Presses Universitaires du Septentrion. <https://doi.org/10.4000/books.septentrion.74824>.
- Loock, R. (2019). La plus-value de la biotraduction face à la machine. *Traduire*, 241, 54-65. <https://doi.org/10.4000/traduire.1848>.
- Maingueneau, D. (2000). *Éléments de linguistique pour le texte littéraire*. Paris : Nathan.
- Marzi, E. (2021). La traduction automatique neuronale et les biais de genre : le cas des noms de métiers entre l'italien et le français. *Synergies Italy*, 17, 19-36. <https://gerflint.fr/Base/Italie17/marzi.pdf>.
- Melby, A.K., Warner, T. (1995). *The Possibility of Language. A discussion of the nature of language, with implications for human and machine translation*. Amsterdam-Philadelphia : John Benjamins Publishing. <https://doi.org/10.1075/btl.14>.
- Popović, M. (2018). Error Classification and Analysis for Machine Translation Quality Assessment. In J. Moorkens, S. Castilho, F. Gaspari & S. Doherty (éds.), *Translation Quality Assessment. Machine Translation : Technologies and Applications*, vol. 1 (pp. 129-158). Cham : Springer. [https://doi.org/10.1007/978-3-319-91241-7\\_7](https://doi.org/10.1007/978-3-319-91241-7_7).
- Śmigielńska, B. (2011). Rôle et description du contexte dans la traduction automatique des textes – approche orientée objets. *Romanica Cracoviensia*, 11 (1), 422-432. <https://doi.org/10.4467/20843917RC.11.047.0115>.
- Tokarczuk, O. (2017). *Księgi Jakubowe*. Kraków : Wydawnictwo Literackie.
- Toral, A. & Way, A. (2018). What Level of Quality can Neural Machine Translation Attain on Literary Text ? In J. Moorkens, S. Castilho, F. Gaspari & S. Doherty (éds.), *Translation Quality Assessment : From Principles to Practice* (pp. 263-287). Springer. <https://doi.org/10.48550/arXiv.1801.04962>.
- Wilks, Y. (2009). *Machine Translation. Its Scope and Limits*. New York : Springer.