

Utiliser les syntagmes nominaux complexes anglais pour évaluer la robustesse des systèmes de traduction anglais-français en langue de spécialité

Maud Bénard

Laboratoire CLILLAC-ARP, université Paris Cité, 8 place Paul Ricœur, 75013 Paris, France
maud.benard@u-paris.fr

RÉSUMÉ

Nous défendons l'idée que l'analyse des erreurs faites lors de la traduction des syntagmes nominaux complexes présente un intérêt pour évaluer la robustesse des systèmes de traduction automatique anglais-français en langue de spécialité. Ces constructions syntaxiques impliquent des questions de syntaxe et de lexique qui constituent un obstacle important à leur compréhension et leur production pour les locuteurs d'anglais non natifs. Nous soutenons que ces analyses contribueraient à garantir que les systèmes de TA répondent aux exigences linguistiques des utilisateurs finaux auxquels ils sont destinés.

ABSTRACT

Using English complex noun phrases to evaluate the robustness of specialized English to French machine translation systems

We argue that the analysis of errors made in the translation of complex noun phrases might be pertinent to evaluate the robustness of specialized English to French machine translation systems. These syntactic constructions involve syntactic and lexical issues that hinder their use and their understanding by non-native English speakers. We argue that such analyses would help ensure that MT systems meet the requirements of their intended end-users.

MOTS-CLÉS : traduction automatique ; évaluation linguistique ; syntagmes nominaux complexes ; langue de spécialité ; discours scientifique.

KEYWORDS: machine translation ; linguistic evaluation ; complex noun phrases ; English for specific purposes ; academic discourse.

1 Introduction

Nous défendons l'idée que l'analyse des erreurs faites lors de la traduction des syntagmes nominaux complexes présente un intérêt pour évaluer la robustesse des systèmes de traduction automatique (TA) en langue de spécialité. Dans cet article, nous justifions tout d'abord l'intérêt d'une analyse linguistique des systèmes de TA (partie 2), puis nous présentons les syntagmes nominaux complexes et les difficultés qu'ils soulèvent (partie 3). Par l'analyse de l'état de l'art, nous montrons que cette construction syntaxique ne semble pourtant pas assez prise en compte aujourd'hui dans l'évaluation des systèmes de TA (partie 4). Puis, nous présentons une typologie d'erreurs adaptées à l'analyse de ce phénomène (partie 5). Nous illustrons ensuite l'intérêt de cette analyse par quelques exemples tirés

de nos travaux de thèse (partie 6).

2 Pourquoi une analyse linguistique des erreurs ?

Les progrès récents de la traduction automatique neuronale s'accompagnent d'attentes importantes de la part des usagers potentiels (de Sinety, 2019). Les progrès en matière de fluidité font que cet outil est de plus en plus perçu comme une aide potentielle à la rédaction en anglais (Goulet *et al.*, 2017), et comme un outil pour accéder à des connaissances scientifiques dans des langues non maîtrisées (Yvon, 2019; Cros & Kübler, 2019). En parallèle, l'usage de la postédition se développe en milieu professionnel. Dès 2017, Moorkens (Moorkens, 2017) supposait que son utilisation devrait certainement s'accroître en raison de la pression sur les coûts et la productivité. L'étude annuelle European Language Industry Survey menée par un groupement d'associations et d'organisations relevant des services linguistiques semble prouver cette tendance : la part des entreprises et des indépendants déclarant utiliser la traduction automatique est ainsi passée de plus de 50 % en 2018 à 58 % des entreprises et plus de 70 % des indépendants en 2022 (ELIS survey 2022¹).

Or, si la traduction neuronale s'est accompagnée d'un bond qualitatif indiscutable (meilleure fluidité, moins d'erreurs lexicales et morphologiques, etc.) et d'une réduction de certains efforts de postédition (Bentivogli *et al.*, 2016), elle n'est pas exempte d'erreurs (Castilho *et al.*, 2017; Esperança-Rodier & Becker, 2018). Les métriques automatiques (BLEU, METEOR...) classiquement utilisées pour évaluer les traductions automatiques ne suffisent plus à comprendre la capacité des systèmes neuronaux au regard de la fluidité des sorties (Burlot & Yvon, 2018).

Dans ce cadre, une analyse de phénomènes linguistiques précis, connus pour poser des difficultés aux utilisateurs finaux, prend tout son sens. En effet, si le système produit des erreurs similaires à celles des humains, il est probable que leur identification soit plus difficile dans le cas d'une postédition professionnelle (tâche par laquelle un être humain vient corriger un texte traduit automatiquement pour le rendre conforme aux attentes du client final) ou d'une utilisation de ces systèmes comme aide à la rédaction.

Depuis 2017, cette approche linguistique est de plus en plus utilisée pour évaluer les systèmes de TA. Macketanz *et al.* (Macketanz *et al.*, 2017) ont montré l'intérêt d'une telle approche pour identifier les faiblesses des systèmes neuronaux en utilisant un jeu de test contenant des phénomènes linguistiques connus pour poser des difficultés de traduction en anglais-allemand. Isabelle *et al.* (Isabelle *et al.*, 2017) ont effectué un travail similaire pour la paire anglais-français. Le traitement des noms composés du type < *Noun Noun* > (*juice filter*, *butter knife*...) a par exemple été étudié et les traductions se sont avérées décevantes. Loock (Loock, 2018) a mené une évaluation similaire sur une série de phénomènes linguistiques ayant la réputation de poser problème aux traducteurs anglais-français du fait d'une différence de fréquence d'utilisation entre les deux langues originales. Il a mis en évidence une surreprésentation de ces phénomènes lexicaux et grammaticaux en français traduit automatiquement par rapport au français original. Burlot et Yvon se sont intéressés à la compétence morphologique des systèmes : leurs travaux suggèrent que certains phénomènes grammaticaux sont moins bien modélisés par la traduction neuronale (Burlot & Yvon, 2018).

1. <https://elis-survey.org>

3 Les syntagmes nominaux complexes : une structure caractéristique du discours scientifique anglais

Si le syntagme nominal « simple » est constitué d'un nom tête et éventuellement d'un déterminant, le syntagme nominal complexe résulte alors de l'ajout d'unités grammaticales à ces structures simples (syntagme prépositionnel, prémodificateurs adjectivaux, nom adjectival...).

Le discours scientifique en anglais se caractérise par le recours important et croissant aux syntagmes nominaux prémodifiés. Ils permettent l'identification très précise des référents (Biber & Conrad, 2009; Biber & Gray, 2016). Ces groupes sont notamment employés pour compacter ou condenser l'information, et leur usage, comparativement, s'inscrit également dans des choix communicationnels spécifiques à la construction du discours scientifique, comme, par exemple, l'alternance de la pré-et postmodification dans le domaine médical (Gledhill & Pecman, 2018).

Au sein de ces syntagmes nominaux prémodifiés, les syntagmes comprenant une prémodification nominale sont particulièrement complexes. Ces groupes impliquent des questions de syntaxe et de lexique qui entraînent fréquemment des difficultés de découpage syntaxique. Leur complexité peut être accrue par la multiplication des éléments de complexification sur un même groupe considéré ((Rouleau, 2006; Berlage, 2014).

Considérons les exemples suivants tirés d'articles du domaine de la Traduction Automatique des Langues (TAL).

TABLE 1: Exemples de SNC (en gras)

(1)	An attention mechanism implemented by a feedforward neural network is then used to attend specific parts of the input and to generate an alignment between input and output sequence .
(2)	While both publications report results of an extensive analysis and comparison of NMT and PBMT approaches, neither of publications deals with language related issues based on the source and the target language properties and their differences.
(3)	This scenario is relevant for the so-called computer-assisted translation (CAT) framework , which now represents the standard operating environment in the translation industry.
(4)	A PBSMT and an NMT system were compared across four translation directions (i.e. from English (EN) into German (DE), Greek (EL), Portuguese (PT), and Russian (RU) in a series of extensive assessment tasks.

Les exemples ci-dessus illustrent la manière dont la complexité n'est pas seulement lexicale (présence d'un terme du domaine, c'est-à-dire d'un mot ou groupe de mots renvoyant à une notion définie dans un domaine de spécialité), mais peut également résulter de la longueur du syntagme en raison du phénomène de récursivité dans la construction des SNC (exemple (3)) ou de l'enchâssement des modificateurs avec par exemple la présence d'une coordination nominale (exemples (1) et (2)).

L'exemple (4) combine deux éléments de complexification avec une difficulté lexicale (présence de deux termes du domaine du TAL : *PBSMT* et *NMT*, et d'une coordination des prémodificateurs de la tête, *system*).

L'exemple (3) illustre la difficulté de découpage syntaxique qui résulte de la longueur potentielle des syntagmes : *so-called* doit-il s'interpréter comme un modifieur du seul *computer-assisted translation (CAT)* ou de *computer-assisted translation (CAT) framework* ?

Ces particularités d'usage et de construction constituent un obstacle important à leur compréhension, leur production et leur traduction pour les locuteurs d'anglais non natifs (Chuquet & Paillard, 1987; Biber & Conrad, 2009; Kübler *et al.*, 2022).

Dans ce contexte, une analyse de la capacité des systèmes de traduction automatique (TA) à traiter de ces syntagmes nominaux complexes comprenant une prémodification nominale (appelés "SNC" dans la suite de cet article) tirés de textes spécialisés authentiques prend tout son sens. Il s'agit de garantir que les textes scientifiques traduits automatiquement et postédités répondent aux exigences linguistiques de la communauté de discours à laquelle ils sont destinés.

4 Un phénomène peu abordé dans les évaluations des systèmes de traduction

Malgré l'importance des SNC pour une part des utilisateurs finaux, l'efficacité des systèmes de traduction automatique sur ce point n'a été que peu abordée (Nakov 2013). De nombreuses études n'abordent la question des SNC qu'à travers soit les mots ou termes composés (*noun compound*) (Lauer, 1995, Baldwin and Tanaka, 2004, Cao and Li, 2002, Moldovan et al., 2004, Nakov, 2013, Balyan and Chatterjee, 2015), et plus particulièrement les seuls syntagmes binominaux (structure du type < Nom Nom >), soit la question de l'identification des mots complexes (*complex word*) lexicalement dans le cadre des recherches sur la simplification des textes. Pour une étude de SNC plus complexes syntaxiquement, il faut se tourner vers les études des unités phraséologiques, collocations ou lexies composées (*multi-word units*) portant sur les n-grammes. La question des SNC y restent cependant très centrée sur les binominaux et les études se focalisent surtout sur la traduction des collocations et des idiomes.

Cette approche des SNC est réductrice, car en se concentrant sur la seule complexité lexicale et plus particulièrement celle portant sur les seuls syntagmes binominaux, une partie des difficultés de traduction est occultée.

5 Une typologie des erreurs permettant de confronter les difficultés des systèmes de traduction et les besoins des utilisateurs finaux

La définition d'une typologie d'erreurs est une tâche complexe : il faut déterminer les classes d'erreurs et leurs niveaux de détails adaptés aux phénomènes que l'on souhaite étudier (Popović, 2018; Chatzikoumi, 2020). Aucune typologie existante dans le domaine de la recherche en traduction

automatique ne semble s'être particulièrement concentrée sur l'identification des erreurs liées aux syntagmes nominaux complexes (SNC).

Les études répertoriées, se concentrant en majorité sur les syntagmes binominaux et l'ambiguïté lexicale qui les caractérisent, n'ont développé des typologies que pour décrire les relations sémantiques entre ces termes. Les autres typologies d'erreurs classiquement utilisées en TAL s'avèrent souvent trop générales pour prendre en compte les spécificités des erreurs des SNC, ou au contraire trop spécialisées sur l'étude d'un phénomène linguistique précis ou une paire de langues données (Vilar *et al.*, 2006; Stymne & Ahrenberg, 2012; Kirchhoff *et al.*, 2012; Costa *et al.*, 2015). Ainsi, une catégorie classique comme "Ordre des mots" pourrait correspondre à des phénomènes très différents pour l'analyse d'un SNC comme une mauvaise attribution d'un modifieur ou une mauvaise identification de la tête. Or, les anglophones langue seconde éprouvent généralement plus de difficultés dans l'attribution des modifieurs aux bons constituants du SNC plutôt qu'à l'identification de la tête. Une erreur du premier type sera donc plus difficilement identifiée et corrigée par un utilisateur final de TA.

Dans le domaine de la traduction professionnelle (agences de traduction, indépendants...), si les approches linguistiques prédominent pour analyser les erreurs produites, elles font appel, ne serait-ce qu'implicitement, à la finalité des traductions pour évaluer la gravité des erreurs commises. Les typologies d'analyses des erreurs, comme la typologie Dynamic Quality Framework (DQF), la Multidimensional Quality Metrics (MQM) ou la LISA QA, combinent ainsi la prise en compte des erreurs langagières, les attentes du client (guide de rédaction, terminologie propre...), les nécessités de distinction entre différentes variantes linguistiques (par exemple, entre anglais britannique ou américain, ou entre français de France ou français canadien) et les contraintes du processus de production (temps passé, délai de livraison...); chacun de ces éléments est pondéré en fonction de sa gravité. Ces typologies professionnelles s'avèrent également inadaptées à l'analyse des erreurs linguistiques sur les SNC en raison du fort l'accent mis sur les erreurs de localisation pour répondre aux besoins d'un client spécifique, ce qui rend difficile leur utilisation sur une grande diversité de genres textuels. Concernant l'identification des erreurs, elles présentent les mêmes limites que les typologies utilisées dans le domaine de la TA.

Pour trouver une typologie adaptée, il faut se tourner vers les typologies utilisées dans l'enseignement de la traduction anglais-français ou de l'anglais comme langue seconde. Mais, si la question des termes complexes en anglais a fait l'objet de très nombreuses études, peu d'entre elles se sont intéressées à la question des syntagmes nominaux complexes dans les langues de spécialité et la grande majorité s'est concentrée sur les syntagmes binominaux et l'ambiguïté lexicale qui les caractérisent (Tournier, 1985). Maniez (Maniez, 2008, 2017, 2020) s'est ainsi intéressé aux stratégies de traduction des traducteurs français et des systèmes de TA face aux SNC anglais dans le domaine médical sans toutefois développer une typologie détaillée des erreurs.

La typologie d'erreurs développée par Kübler, Mestivier et Pecman (Kübler *et al.*, 2022) s'avère particulièrement intéressante : elle résulte d'un inventaire sur plusieurs années portant sur les erreurs faites par les apprenants en traduction lors de la postédition d'articles de recherche du domaine des sciences de la Terre. Cette typologie nous paraît parfaitement adaptée aux besoins présentés dans cet article, car elle permet de comparer les difficultés rencontrées par des utilisateurs finaux (non-identification ou sous-identification d'un type d'erreur, difficultés d'identification des composants d'un syntagme...) avec les difficultés éventuelles des systèmes de TA.

Afin de tenir compte des spécificités d'erreurs produites par les systèmes de TA, Bénard (Bénard, 2019) a adapté cette typologie d'erreurs à l'évaluation de textes produits par des systèmes neuronaux. Des erreurs spécifiques à la TA ont ainsi été ajoutées comme l'instabilité des choix de traduction d'un

mot au sein d'un même texte ou l'ajout d'unités lexicales lors de la traduction tendant à modifier le sens du syntagme nominal source.

La typologie finale obtenue est présentée dans la Table 2 page 6. C'est cette typologie que nous proposons d'utiliser pour l'analyse des erreurs.

Niveau 1	Niveau 2
Erreurs terminologiques, en genre ou en nombre	<ol style="list-style-type: none"> 1. Incohérence de la traduction intra- et extratextuelle 2. Mauvaise traduction d'une unité lexicale qui n'est pas un terme du domaine ou du genre 3. Traduction d'un terme par un non-terme 4. Unité lexicale non traduite 5. Traduction d'un nom propre ou d'un nom commun qui ne doit pas être traduit 6. Problème de détermination, d'erreur en nombre ou en genre
Erreurs d'analyse syntaxique du texte source	<ol style="list-style-type: none"> 7. Identification erronée de la tête 8. Attribution du modifieur au mauvais constituant 9. Factorisation non détectée (ou partiellement) d'un constituant dans la coordination 10. Ajout d'une factorisation n'existant pas en langue source 11. Mauvaise interprétation des liens sémantiques entre les constituants, hors factorisation
Adaptation erronée à la syntaxe de la langue cible	<ol style="list-style-type: none"> 12. Absence d'une explicitation pourtant nécessaire 13. Calque erroné sur la langue source ou influence de la langue source 14. Inadaptation des poids des constituants
Erreurs de transferts	15. Ajout injustifié d'une unité lexicale ou d'une ponctuation
	16. Omission d'un constituant, voire de tout le syntagme lors de la traduction

TABLE 2: Typologie des erreurs de Bénard (Bénard, 2019) sur la base de Kübler et al. (Kübler *et al.*, 2022)

6 Résultats préliminaires

6.1 Méthode

6.1.1 Choix du corpus

Nous avons appliqué cette approche linguistique dans le cadre de notre thèse en analysant un corpus d'articles de recherche, traduits simultanément par plusieurs systèmes. Afin d'illustrer l'intérêt d'une telle approche, nous présentons, dans cette partie, deux exemples tirés d'une analyse diachronique à l'aide de deux systèmes commerciaux, non spécialisés et accessibles en ligne : DeepL et Systran. Les textes du corpus de test ont été traduits une première fois en décembre 2019 et une seconde fois en août 2022.

L'ensemble du corpus de test de cette analyse diachronique est constitué de 4 articles de recherche rédigés en anglais, tirés de ACL Anthology² : 2 articles de synthèse (*review papers*) et 2 articles de recherche expérimentale (*research papers*).

Les textes retenus ont été écrits ou co-écrits par des chercheurs spécialistes du domaine de la traduction automatique : la majorité des auteurs ont un indice de Hirsch (indice h) supérieur à 10 – six d'entre eux ont un indice supérieur à 20 – et leur indice i10 de Google Scholar (nombre d'articles qui ont été cités au moins 10 fois) est généralement élevé. Les articles eux-mêmes ont été publiés dans des revues à comité de lecture. Ils ont également été cités de très nombreuses fois depuis leur publication en 2017, ce qui pourrait témoigner de la capacité des chercheurs du domaine à s'approprier ces textes. Nous pouvons donc supposer que ces articles suivent les conventions d'écriture, la phraséologie et la terminologie attendues par le public du domaine du TAL.

Les textes ont été modifiés avant d'être soumis à la traduction. Les images, les références, les en-têtes, les pieds de page et les éléments spécifiques de mise en page (italiques, saut de page...) ont été supprimés. Seuls l'abstract et le corps de l'article ont été conservés. Les phrases éventuellement coupées par une image ou un tableau ont été reconstituées.

En revanche, les textes n'ont pas été corrigés : les erreurs d'orthographe, de grammaire ou de typographie ont été conservées. En effet, ce sont des éléments que les systèmes de TA devront être capables de traiter lors de leur utilisation par des utilisateurs finaux (postéditeur, rédacteur...). De plus, les textes n'ont pas été segmentés avant d'être soumis aux systèmes de traduction.

6.1.2 Identification des syntagmes

Les SNC ont été extraits de manière semi-automatique. Les textes traduits ont été étiquetés à l'aide du logiciel Le Trameur³ qui intègre TreeTagger, un système d'étiquetage automatique des catégories grammaticales des mots avec lemmatisation sur la base du jeu d'étiquettes The Penn Treebank. Notre choix s'est porté sur cette combinaison d'outils en raison de leur bon équilibre entre l'efficacité de l'étiquetage (ratio rapidité/fiabilité) en recourant au système intégré, la souplesse offerte par les formats en sortie et l'affichage de corpus parallèle multilingue, et les possibilités d'analyse statistique et d'annotation intégrées.

2. Voir les références complètes en fin d'article

3. Version 12.176. <http://www.tal.univ-paris3.fr/trameur/>

Sur la base de cet étiquetage, les syntagmes candidats sont extraits par repérage de patrons syntaxiques, c'est-à-dire sur la base de l'ordre d'enchaînement des catégories grammaticales formant un SNC. La table ci-dessous présente un extrait des patrons considérés (Table 3).

TABLE 3: Exemples de patrons syntaxiques identifiés lors de l'analyse

Patrons syntaxiques	Exemples de syntagmes identifiés
NN NN et par récursivité : DT NP NP NN NN DT NN NN DT JJ NN NN JJ NN NN NNS etc.	MT systems the "a posteriori" adaptation strategy a product listing high quality submissions a large output vocabulary
JJ NN CC NN et par récursivité : JJ NN NNS CC NNS etc.	additional nouns and adjectives chemical patent titles and abstracts

Toutefois, des erreurs sont présentes en raison des limitations techniques du système d'étiquetage utilisé et de l'ambiguïté inhérente à certains termes anglais. Afin de réaliser une étude qualitative, il est nécessaire de corriger ces erreurs manuellement comme l'illustre l'exemple ci-dessous.

ID_partitio n-phrase	GNC_ID_2	list_GNC_POS_ID	list_GNC_Lemme	list_GNC_Forme	list_GNC_POS	Texte source
146_T2	1907.0	['13829', '13830', '13831', '13832', '13833', '13834', '13835']	leverage past gradient information	leverage past gradient information	NN JJ JJ NN	For this reason, optimizers that can leverage past gradient information are usually more reliable.

FIGURE 1 – Exemple de SNC identifiés automatiquement à corriger

L'extrait de notre base (Figure 1) montre l'un des syntagmes identifiés automatiquement (*leverage past gradient information*) et les informations données par l'étiqueteur Tree-Tagger. Nous observons que l'étiqueteur a identifié de manière erronée *leverage* comme un nom, comme le prouve la mention *NN* dans la colonne *list_GNC_POS*. Il s'agit en réalité de la forme verbale et le mot n'aurait jamais dû être identifié comme un composant du syntagme nominal à étudier. Il apparaît donc nécessaire de supprimer l'intégralité des éléments marqués en rouge dans les cellules afin d'obtenir le syntagme correct.

Ces syntagmes candidats ont donc été corrigés manuellement par un linguiste au fur et à mesure de l'identification des erreurs. Ces bases de données, créées automatiquement, contiennent de nombreuses informations indispensables à l'analyse des syntagmes (liste des catégories grammaticales, liste des lemmes, liste des formes, etc. constituant les syntagmes). Afin de limiter les biais dans l'analyse, un

second annotateur a également annoté une partie du corpus et un accord inter-annotateur a été calculé. Au total, les traductions de 1072 SNC ont été étudiées, soit 4288 syntagmes traduits.

Les sous-section suivantes, 6.2 et 6.3, présentent deux exemples tirés de ce corpus qui illustrent l'intérêt d'analyser les SNC pour l'évaluation des systèmes de traduction spécialisés.

6.2 Exemple 1 : Des différences de performances significatives

En 2019, Systran se démarquait par une difficulté à gérer la présence de coordination dans les prémodificateurs nominaux avec 23% des groupes de ce type faisant l'objet d'une erreur contre une seule occurrence pour DeepL. Considérons les exemples ci-dessous (Table 4).

TABLE 4: Exemples de traduction erronées produites en 2019 par le système commercial non spécialisé Systran (le syntagme analysé est en gras)

(1.a)	An attention mechanism implemented by a feedforward neural network is then used to attend specific parts of the input and to generate an alignment between input and output sequence .
(1.b)	Un mécanisme d'attention mis en place par un réseau neuronal de flux est ensuite utilisé pour assister à des parties spécifiques de l'entrée et générer un alignement entre l'entrée et la séquence de sortie .
(2.a)	While both publications report results of an extensive analysis and comparison of NMT and PBMT approaches, neither of publications deals with language related issues based on the source and the target language properties and their differences.
(2.b)	Bien que les deux publications fassent état des résultats d'une analyse et d'une comparaison approfondies des approches NMT et PBMT, aucune de ces publications ne traite des questions liées à la langue en fonction de la source et des propriétés linguistiques cibles et de leurs différences.

Les exemples (1.a) et (2.a) correspondent aux phrases d'origine soumises aux systèmes et les exemples (1.b) et (2.b) correspondent aux traductions.

Les erreurs portent essentiellement sur l'identification de la première partie du constituant : le moteur n'identifie généralement pas que le premier nom du syntagme coordonné (respectivement *input* et *the source*) est également un modifieur de la tête (respectivement *sequence* et *language properties*). Il traite chacun des constituants de la coordination comme un syntagme nominal différent.

Cette différence marquée entre deux systèmes commerciaux peut s'avérer intéressante pour un utilisateur final. En effet, il peut être difficile pour un traducteur ou un utilisateur final de déterminer l'intention du locuteur et donc d'identifier correctement les limites des syntagmes nominaux. Cela pourrait amener un utilisateur final à privilégier un système de TA plus efficace sur ce point, lorsque ce type de construction syntaxique coordonnée s'avère particulièrement utilisée dans son domaine de spécialité, et ce, afin de limiter le temps à passer en postédition.

6.3 Exemple 2 : Une pérennité des performances qui n'est pas garantie

Les exemples présents dans la Table 5 ci-dessous montrent la continuité des difficultés de traduction dans le temps.

TABLE 5: Exemples de traduction erronées produites par les systèmes commerciaux non spécialisés DeepL et Systran en 2019 et 2022 (les syntagmes analysés sont en gras)

(3.a)	Texte original	The goal of this evaluation was to compare the performance between the mature Chinese to English patent MT engines used in production at Iconic with novel NMT engines developed at the ADAPT Centre on an 'apples to apples' basis, trained on the same available data.
(3.b)	DeepL 2019	Le but de cette évaluation était de comparer les performances des moteurs de TA brevetés chinois à anglais utilisés dans la production à Iconic avec les nouveaux moteurs de TA développés au Centre ADAPT sur une base de " pommes à pommes ", formés sur la base des mêmes données disponibles.
(3.c)	DeepL 2022	L'objectif de cette évaluation était de comparer les performances des moteurs de TA du chinois à l'anglais utilisés en production chez Iconic avec les nouveaux moteurs de NMT développés au Centre ADAPT, sur une base de comparaison, entraînés sur les mêmes données disponibles.
(3.d)	Systran 2019	L'objectif de cette évaluation était de comparer la performance entre les moteurs chinois matures et anglais brevetés MT utilisés en production à Iconic avec les nouveaux moteurs NMT mis au point au Centre ADAPT sur une base de 'pommes à pommes', formés sur les mêmes données disponibles.
(3.e)	Systran 2022	Le but de cette évaluation était de comparer les performances des moteurs MT de brevets chinois à anglais utilisés dans la production chez Iconic avec les nouveaux moteurs NMT développés au Centre ADAPT sur une base 'pomme à pomme', formés sur la base des mêmes données disponibles.

L'exemple (3.a) correspond à la phrases d'origine soumise aux systèmes de traduction, les exemples (3.b) et (3.d) correspondent aux traductions produites respectivement par DeepL et Systran en 2019 et les exemples (3.c) et (3.e) correspondent aux traductions produites respectivement par DeepL et Systran en 2022.

De manière générale, nous constatons une difficulté à interpréter les liens sémantiques dans les SNC de grande taille. Si les solutions ou les types d'erreurs observées peuvent évoluer ou les erreurs ne pas concerner exactement les mêmes composants, on constate un certain maintien des erreurs observées

ou leur apparition sur d'autres SNC. En sens inverse, des SNC ou des éléments de SNC qui pouvaient être bien traités en 2019 s'avèrent mal traduits en 2022.

Considérons le syntagme *the mature Chinese to English patent MT engines* dans l'exemple (3a) dont la traduction attendue est du type *les moteurs aboutis pour la traduction automatique de brevets du chinois vers l'anglais*. Il faut noter que ce syntagme combine plusieurs niveaux de complexité, notamment une complexité lexicale (présence de terme du domaine du Traitement Automatique des Langues et de non termes du domaine) et une complexité syntaxique (8 éléments).

Nous constatons que le système DeepL conserve la difficulté à traduire *patent* : même si la solution adoptée est différente entre 2019 (erreur d'analyse syntaxique avec une mauvaise interprétation du lien sémantique modifieur/modifié) et 2022 (non traduction du terme, c'est-à-dire une erreur de transfert), la traduction reste erronée. De manière générale, le système éprouve des difficultés à interpréter les liens sémantiques entre les composants du SNC, ce qui entraîne une mauvaise attribution des modifieurs et une traduction fautive. Le système Systran continue également à faire une erreur terminologique en ne traduisant pas *MT* par *TA*, alors même que cette traduction est bien faite dans d'autres segments traduits du même texte.

Si nous considérons le système Systran, nous constatons une amélioration de l'identification des liens sémantiques entre 2019 et 2022 : le mot *patent* est bien considéré en 2022 comme limitant le sens de *engines* en indiquant le domaine de spécialité du moteur considéré (traduction par *de brevets*), alors qu'il avait été analysé comme qualifiant le statut du moteur (traduction par *brevetés*). Cette amélioration s'accompagne cependant d'une dégradation importante avec l'omission de la traduction du mot *mature* ce qui entraîne une distorsion du sens en français par rapport à l'original.

Cet exemple illustre la pérennité des difficultés dans le temps pour traduire les SNC et la difficulté à maintenir les performances observées à un moment donné. Or, dans le cadre des systèmes disponibles en ligne, les utilisateurs finaux n'ont aucune information sur les évolutions de version qui leur permettrait de juger des évolutions sur la qualité des traductions. Ils peuvent également manquer des compétences linguistiques ou informatiques et des ressources financières ou matérielles pour mener une telle évaluation. Il apparaît donc important de prendre en compte ce point lors de l'évaluation de la qualité des systèmes de TA dès leur développement.

7 Conclusion et perspectives

Nous défendons l'idée que l'analyse des erreurs faites lors de la traduction des syntagmes nominaux complexes présente un intérêt pour évaluer la robustesse des systèmes de traduction automatique (TA) en langue de spécialité. Plus particulièrement, l'analyse des syntagmes nominaux présentant une prémodification nominale pour le premier modifieur (noté "SNC" dans cet article) est particulièrement adaptée à la prise en compte des besoins des utilisateurs finaux, qu'il soit un professionnel de la langue (traducteur-postéditeur, rédacteur...) ou un spécialiste du domaine recherchant une aide à la rédaction vers l'anglais ou une compréhension rapide du contenu d'un texte.

Cette construction syntaxique ne semble pourtant pas assez prise en compte aujourd'hui dans l'évaluation des systèmes de TA. Au mieux, ces questions sont partiellement intégrées à l'étude des composés nominaux, des unités phraséologiques ou aux questions sur la complexité lexicale des textes en vue de leur simplification. Les analyses se limitent souvent à la question des syntagmes binominaux ou à l'aspect terminologique sans prendre en compte les autres facettes possibles de la complexité comme

la longueur ou l'enchâssement.

Nous avons mené une première analyse diachronique basée sur une typologie permettant de mettre en parallèle les difficultés des systèmes avec les besoins des utilisateurs finaux. Dans cet article, nous nous sommes concentrés sur la présentation de notre méthodologie. Nous présentons deux exemples illustrant qu'en appliquant cette méthodologie à des textes du domaine du Traitement Automatique des Langues, il est possible de montrer que les systèmes de TA produisent des erreurs similaires en partie à celles produites par un traducteur humain, ce qui peut rendre leur détection difficile par un utilisateur final. Considérant que la construction syntaxique de ces syntagmes varient d'une langue de spécialité à l'autre, nous pensons donc que l'efficacité des systèmes de TA en matière de traduction des SNC pourrait constituer un facteur dans le choix d'un utilisateur final d'utiliser l'un ou l'autre des systèmes à sa disposition.

Nos travaux actuels et futurs visent à présenter de manière approfondie les résultats de cette analyse et à étendre celle-ci à d'autres domaines de spécialité (pour prendre en compte la diversité de construction des SNC) et à des systèmes de TA spécialisés (pour voir si la spécialisation dans un domaine permet d'améliorer la traduction des SNC). Au-delà du développement logiciel, une analyse fine des erreurs produites par les systèmes neuronaux est essentielle dans une visée pédagogique pour la formation des postéditeurs professionnels ou occasionnels (comme, par exemple, les spécialistes d'autres disciplines souhaitant une aide à la rédaction pour l'écriture d'un article scientifique dans une langue autre que sa langue maternelle).

Les corpus annotés seront mis librement à disposition de la communauté à l'automne 2023.

Remerciements

Ma directrice de thèse, Natalie Kübler, professeur des universités, laboratoire CLILLAC-ARP, université Paris Cité (France).

Références des textes inclus dans le corpus de test

CASTILHO S., MOORKENS J., GASPARI F., CALIXTO I., TINSLEY J. & WAY A. (2017). Is Neural Machine Translation the New State of the Art? *The Prague Bulletin of Mathematical Linguistics*, 108(1), 109–120. DOI : [10.1515/pralin-2017-0013](https://doi.org/10.1515/pralin-2017-0013).

COSTA-JUSSÀ M. R., ALLAUZEN A., BARRAULT L., CHO K. & SCHWENK H. (2017). Introduction to the special issue on deep learning approaches for machine translation. *Computer Speech I& Language*, 46(Journal Article), 367–373. DOI : [10.1016/j.csl.2017.03.001](https://doi.org/10.1016/j.csl.2017.03.001).

POPOVIC M. (2017). Comparing Language Related Issues for NMT and PBMT between German and English. *The Prague Bulletin of Mathematical Linguistics*, 108(1), 209–220. DOI : [10.1515/pralin-2017-0021](https://doi.org/10.1515/pralin-2017-0021).

TURCHI M., NEGRI M., FARAJIAN M. A. & FEDERICO M. (2017). Continuous Learning from Human Post-Edits for Neural Machine Translation. *The Prague Bulletin of Mathematical Linguistics*, 108(1), 233–244. DOI : [10.1515/pralin-2017-0023](https://doi.org/10.1515/pralin-2017-0023).

Références

- BENTIVOGLI L., BISAZZA A., CETTOLO M. & FEDERICO M. (2016). Neural versus Phrase-Based Machine Translation Quality : a Case Study. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, p. 257–267, Austin, Texas : Association for Computational Linguistics. DOI : [10.18653/v1/D16-1025](https://doi.org/10.18653/v1/D16-1025).
- BERLAGE E. (2014). *Noun Phrase Complexity in English*. Studies in English Language. United Kingdom : Cambridge University Press.
- BIBER D. & CONRAD S. (2009). *Register, Genre, and Style*. Cambridge Textbooks in Linguistics. Cambridge University Press.
- BIBER D. & GRAY B. (2016). *Grammatical Complexity in Academic English : Linguistic Change in Writing*. Studies in English Language. United Kingdom : Cambridge University Press.
- BURLLOT F. & YVON F. (2018). Évaluation morphologique pour la traduction automatique : adaptation au français (Morphological Evaluation for Machine Translation : Adaptation to French). In *Actes de la Conférence TALN. Volume 1 - Articles longs, articles courts de TALN*, p. 61–74, Rennes, France : ATALA.
- BÉNARD M. (2019). Mémoire de Master 2 Recherche LSCT : Évaluation de la qualité des systèmes neuronaux en matière de traduction des groupes nominaux complexes à prémodification nominale.
- CASTILHO S., MOORKENS J., GASPARI F., SENNRICH R., SOSONI V., GEORGAKOPOULOU P., LOHAR P., WAY A., VALERIO MICELI BARONE A. & GIALAMA M. (2017). A Comparative Quality Evaluation of PBSMT and NMT using Professional Translators. In *Proceedings of MT Summit XVI*, volume Vol. 1 : Research Track, p. 116–131, Nagoya, Aichi, Japan. https://www.researchgate.net/profile/Vilemini-Sosoni/publication/320016264_A_Comparative_Quality_Evaluation_of_PBSMT_and_NMT_using_Professional_Translators/pdf
- CHATZIKOUMI E. (2020). How to evaluate machine translation : A review of automated and human metrics. *Natural Language Engineering*, **26**(2), 137–161. Publisher : Cambridge University Press, DOI : [10.1017/S1351324919000469](https://doi.org/10.1017/S1351324919000469).
- CHUQUET H. & PAILLARD M. (1987). *Approche linguistique des problèmes de traduction*. Paris, France : Éditions Ophrys. Réimpression 2015 d'un ouvrage de 1987.
- COSTA A., LING W., LUÍS T., CORREIA R. & COHEUR L. (2015). A linguistically motivated taxonomy for Machine Translation error analysis. *Machine Translation*, **29**, 127–161. DOI : [10.1007/s10590-015-9169-0](https://doi.org/10.1007/s10590-015-9169-0).
- CROS I. & KÜBLER N. (2019). Du FOU au FLA à l'Université française ou de la pédagogie universitaire à l'écrit scientifique spécialisé. *FIU Francophonie et innovation à l'université*.
- DE SINETY P. (2019). "Pour un plurilinguisme de la pensée scientifique". Library Catalog : www.culture.gouv.fr.
- ESPERANÇA-RODIER E. & BECKER N. (2018). Comparaison de systèmes de traduction automatique, probabiliste et neuronal, par analyse d'erreurs. Nancy, France.
- GLEDHILL C. & PECMAN M. (2018). On alternating pre-modified and post-modified nominals such as aspirin synthesis vs. synthesis of aspirin : Rhetorical and cognitive packing in English science writing. *Fachsprache : Internationale Zeitschrift für Fachsprachenforschung- didaktik und Terminologie*, **40**(1), 24–46.

- GOULET M.-J., SIMARD M., PARRA ESCARTÍN C. & O'BRIEN S. (2017). La traduction automatique comme outil d'aide à la rédaction scientifique en anglais langue seconde : résultats d'une étude exploratoire sur la qualité linguistique. *ASp. la revue du GERAS*, (72), 5–28. Number : 72 Publisher : Groupe d'étude et de recherche en anglais de spécialité, DOI : [10.4000/asp.5045](https://doi.org/10.4000/asp.5045).
- ISABELLE P., CHERRY C. & FOSTER G. (2017). A Challenge Set Approach to Evaluating Machine Translation. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP 2017)*, p. 2476–2486, Copenhagen, Denmark. arXiv : 1704.07431.
- KIRCHHOFF K., CAPURRO D. & TURNER A. (2012). Evaluating User Preferences in Machine Translation Using Conjoint Analysis. In *Proceedings of the 16th EAMT Conference*, p. 119–126, Trento, Italy : European Association for Machine Translation.
- KÜBLER N., MESTIVIER A. & PECMAN M. (2022). Using Comparable Corpora for Translating and Post-Editing Complex Noun Phrases in Specialised Texts : Insights from English-to-French Specialised Translation. In S. GRANGER & M.-A. LEFER, Édts., *Extending the Scope of Corpus-Based Translation Studies*, Bloomsbury Advances in Translation. Bloomsbury Academic. <https://www.bloomsbury.com/us/extending-the-scope-of-corpusbased-translation-studies-9781350143258> 24/02/2022.
- LOOCK R. (2018). Traduction automatique et usage linguistique : une analyse de traductions anglais-français réunies en corpus. *Meta : journal des traducteurs / Meta : Translators' Journal*, **63**(3), 786–806. DOI : <https://doi.org/10.7202/1060173ar>.
- MACKETANZ V., AVRAMIDIS E., BURCHARDT A., HELCL J. & SRIVASTAVA A. (2017). Machine Translation : Phrase-Based, Rule-Based and Neural Approaches with Linguistic Evaluation. *Cybernetics and Information Technologies*, **17**(2), 28–43. Publisher : Sciendo Section : Cybernetics and Information Technologies, DOI : [10.1515/cait-2017-0014](https://doi.org/10.1515/cait-2017-0014).
- MANIEZ F. (2008). Traduction automatique et ambiguïté syntaxique : le cas de la coordination dans les groupes nominaux complexes en anglais médical. p. 765–776, Lyon. DOI : [10.4000/asp.500](https://doi.org/10.4000/asp.500).
- MANIEZ F. (2017). An appraisal of recent breakthroughs in machine translation : the case of past participle-based compound adjectives in ESP. *ASp*, **72**, 29–48. DOI : [10.4000/asp.5059](https://doi.org/10.4000/asp.5059).
- MANIEZ F. (2020). Chapter 18 : The identification of potentially ambiguous noun phrases in scientific english : a crucial aspect of translator. In *Strategies and Analyses of Language and Communication in Multilingual and International Contexts*, p. 187–195. Cambridge Scholars Publishing.
- MOORKENS J. (2017). Under pressure : translation in times of austerity. *Perspectives*, **25**, 464–477. DOI : [10.1080/0907676X.2017.1285331](https://doi.org/10.1080/0907676X.2017.1285331).
- POPOVIĆ M. (2018). Error Classification and Analysis for Machine Translation Quality Assessment. In *Translation Quality Assessment : From Principles to Practice.*, volume 1 de Machine Translation : Technologies and Applications, p. 129–158. Springer International Publishing, 1st ed édition. https://doi.org/10.1007/978-3-319-91241-7_7.
- ROULEAU M. (2006). Complexité de la phrase en langue de spécialité : mythe ou réalité ? Le cas de la langue médicale. *Panace@*, **VII**(24), 298–306.
- STYMNE S. & AHRENBERG L. (2012). On the practice of error analysis for machine translation evaluation. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, p. 1785–1790, Istanbul, Turkey : European Language Resources Association (ELRA).

TOURNIER J. (1985). *Introduction à la lexicogénétique de l'anglais contemporain*. Genève : Édition Slatkine. Réimpression de l'édition de 1985.

VILAR D., XU J., D'HARO L. & NEY H. (2006). Error Analysis of Machine Translation Output. p. 697–702, Genoa, Italy.

YVON F. (2019). Les deux voies de la traduction automatique. *Hermes, La Revue*, n° 85(3), 62–68. Publisher : C.N.R.S. Editions.

Timothy Baldwin and Takaaki Tanaka. 2004. Translation by machine of complex nominals. In *Proceedings of the Workshop on Multiword Expressions Integrating Processing - MWE '04*, pages 24–31. Association for Computational Linguistics.

Renu Balyan and Niladri Chatterjee. 2015. Translating noun compounds using semantic relations. *Comput Speech Lang.*, 32(1) :91–108.