

Résumé français

Le projet ALICIA dont nous voulons rendre compte dans le cadre de cette communication trouve son origine dans une difficulté fréquemment rencontrée par les traducteurs pragmatiques : le manque de sources de référence fiables et à jour dans le domaine de la cyberdéfense. Pour traduire la cyberdéfense, et être opérationnel, traducteurs et experts du domaine se montrent pragmatiques. Ils gagneraient pourtant à travailler ensemble. Nous proposons de présenter comment la traductologie de corpus peut les y aider.

Abstract

When translating cyberdefense material, experts and translators alike need up to date and accurate references. We believe that comparable corpora and corpus-based translation studies are part of the answer. The ALICIA project can help, and this communication aims at explaining what is at stake and how linguists and experts can join forces in the cyberworld.

Sandrine Peraldi et Julie Remfort, *Intégration des outils de traitement de corpus en contexte professionnel : étude de cas dans la traduction juridique et financière*

Résumé français

Les formations en traduction intègrent de plus en plus systématiquement les outils et les méthodologies de la linguistique de corpus en vue d'une optimisation des compétences clés des futurs traducteurs (Bowker 1998, Bennett 2010, Frérot 2010, Kübler 2001, 2011). En effet, de nombreux chercheurs ont établi depuis plusieurs années les nombreux apports des corpus électroniques en traduction, et plus particulièrement ceux des corpus parallèles, lorsqu'il s'agit d'identifier des équivalents, des normes ou des stratégies de traduction ou de mettre en exergue des différences stylistiques ou d'usage entre la langue source et la langue cible (Aston 1999, Beeby & al. 2009, Looock 2014, 2016). Pourtant, l'utilisation des concordanciers demeure sporadique, notamment au regard des outils de TAO ou de traduction automatique. Ce constat nous a amenés à lancer il y a deux ans une étude terrain en partenariat avec le Centre de Traduction du Ministère de l'Economie et des Finances afin de déterminer si ces outils peuvent être aisément intégrés dans l'environnement de travail des traducteurs du centre tout en répondant à des objectifs d'amélioration de la qualité de traductions hautement techniques. En effet, les traducteurs du centre sont souvent contraints de traduire des textes touchant à des domaines à la fois extrêmement spécialisés et très peu récurrents, pour lesquels les mémoires de traduction n'ont pas d'utilité. Ainsi, il a été décidé d'évaluer de manière très pragmatique à la fois la pertinence et la facilité d'emploi des corpus comparables comme outils d'aide à la traduction : comment exploiter, dans un contexte de traduction professionnelle, des outils conçus initialement pour des linguistes ? Quelles fonctionnalités privilégier ? Pour quels usages et quels problèmes de traduction ? Sur un plan méthodologique, nous avons éprouvé en conditions réelles d'exercice un double corpus, spécialisé et explicatif, compilé selon des critères de constitution classiques (Pearson, Bowker 2002, Peraldi 2011) dans le domaine des licences de marque. Nous proposons donc d'exposer les premiers résultats de cette étude de cas, en vue de favoriser largement le recours aux outils de corpus en traductologie.

Abstract

Corpus-based tools and methodologies are increasingly becoming a part of training programmes for translators as an effective way to optimise several key competencies for future translators (Bowker 1998, Bennett 2010, Frérot 2010, Kübler 2001, 2011). Indeed, many researchers have demonstrated, for several years now, the relevance of corpora, and especially parallel corpora, in order to help students find appropriate equivalents, identify and understand translation standards and strategies and highlight differences in use and style between the source and the target language (Aston 1999, Beeby & al. 2009, Looock 2014, 2016). However, the use of corpus-based tools alongside CAT and MT tools remains an exception in a professional context. For this reason, a field experiment was launched two years ago at the translation centre of the French Ministry of Finance in order to determine if these new methodologies can be integrated into the work environment of the translators without disrupting their work habits, while enhancing the quality of highly technical translations. Indeed, the translators at the Centre are often required to translate non-recurring and non-documented topics where translation memory banks are of no use. This very pragmatic issue led us to assess the relevance and the ease of use of comparable corpora in order to help translators in their daily activity. How can these tools, initially conceived for linguists, be integrated in a translation environment? What are their most relevant features and what type of translation problems can they address? From a methodological point of view, a double corpus, that is both specialised and of an explanatory nature, was compiled according to specific linguistic criteria (Pearson, Bowker 2002, Peraldi 2011) in the field of trademark licences, and tested in real-life professional situations. This paper aims at describing the first results of this study with a view to encouraging the integration of corpus-based tools during the translation process in a professional setting.

Laurent Gautier, Cyril Nguyen Van, *Utiliser les corpus parallèles pour approcher la langue de la traduction en contexte institutionnel : le cas de la BCE*

Résumé français

Les corpus parallèles (au sens de W. Teubert) alignés sont devenus un outil quasi incontournable à la fois pour la formation et pour la pratique professionnelle de la traduction, a minima sous la forme (transparente pour l'utilisateur final) des mémoires de traduction. Ils sont toutefois nettement moins l'objet d'approches en termes de corpus comparables (au sens de M. Baker cette fois), à tout le moins pour les discours spécialisés. La communication proposée visera donc à s'inscrire dans cette lacune en posant la question, dans un contexte de traduction institutionnelle très contraint, de savoir ce que les traductions dans plusieurs langues-cibles données (ici français, allemand et néerlandais) d'un texte original en anglais-source permettent de mettre en évidence en matière de « langue de la traduction ». Présente-t-elle des traits saillants en fonction de la langue ? En particulier par rapport aux corpus étalons disponibles ? Quelles convergences/divergences apparaissent entre ces traductions lorsqu'elles sont considérées en bloc ?

Le corpus interrogé est composé des conférences de presse de la Banque Centrale Européenne des années 2015 et 2016. Il représente 19.883 mots dans le texte original EN et est aligné avec les trois traductions sous LF Aligner qui comptent chacune 23.931 mots en FR, 19.810 en DE et 21.324 en NL. Les textes individuels sont étiquetés et traités quantitativement sous TXM.

Suite à une première étude exploratoire (Gautier / Nguyen Van 2016), portant, de façon comparée, sur les marqueurs de l'ambiguïté, reconnue comme catégorie 'moteur' de ces textes, la contribution vise à dresser le portrait linguistique du genre discursif dans chacune des langues traduites, tant en termes de lexicque que de catégories grammaticales de base. Il s'agira en particulier de se demander dans quelle mesure ces « portraits robots » permettent d'approcher la qualité de la traduction.

Abstract

Aligned parallel corpora (as defined by W. Teubert) have become almost inescapable for both training and professional practice in translation, at least in a form of translation memory databases (transparent for the final consumer). They are, all the same, hardly a subject of study in terms of comparable corpora (as defined this time by M. Baker) or for specialised discourse. In the context of a very constrained institutional translation, this paper will therefore focus on this loophole by questioning to what extent the translations of an English source text into several given target languages (in this case French, German and Dutch) allow us to highlight from the point of view of the translated language. Does such a translation present the salient linguistic features, and particularly in relation to the available standard corpora? What convergences/divergences appear between these translations when they are considered as a whole?

The corpus under investigation is composed of press conferences from the European Central Bank in 2015 and 2016. It contains 19,883 words in the source text (EN) and is aligned with three translations using LF Aligner, each of which contains 23,931 words in French, 19,810 (GER) and 21,324 (NL). Each text is labelled and weighted quantitatively using TXM.

Following the first comparative and exploratory study (Gautier/Nguyen Van 2016) of the markers of ambiguity (that are recognised as a 'driving' category in these texts), this paper aims at compiling a linguistic portrait of the discursive register in each of the translated languages both in terms of lexis and basic grammatical categories. In particular, we will ask ourselves to what extent these 'profiles' facilitate the production of a quality translation.

Tiziri Bachir, *Constitution des corpus parallèles bilingues entre une langue de science et une langue littéraire : les défis en milieux professionnels*

Résumé français

La traduction spécialisée n'a pas bénéficié d'une théorisation propre à elle, mais tire ses enseignements des différentes théories développées dans le cadre de la traduction générale. Néanmoins, des chercheurs et praticiens de la traduction ont fait en sorte que cette catégorie soit munie d'un arsenal qui atténuerait sa complexité tant au niveau linguistique qu'opérationnel.

Dans la présente étude, nous aborderons la question des logiciels de Traduction Assistée par Ordinateur (CATtools), nous avons toutefois choisi de réduire leur champ qui en demeure plus vaste que d'être traité dans une seule étude. Il sera alors question des corpus parallèles tel que l'indique le titre sus-énoncé, des challenges auxquels les professionnels font face lors de la constitution mais aussi de l'utilisation de ces corpus, notamment lorsque la combinaison linguistique dont il est question, comporte l'arabe comme langue de départ ou langue d'arrivée. La langue arabe demeure encore aujourd'hui prisonnière d'un littéraire qui refuse de la remettre au goût du jour en sus d'un dialectal qui vient accentuer une problématique pas des moindres. En effet, au milieu des développements

qui s'opèrent tous les jours dans les domaines scientifiques et techniques, l'arabe se voit dans les derniers rangs pour des raisons que nous développerons ci-après – contrairement à la langue française connue pour être une langue de savoir et de recherche. Dans ce cas, quelles sont les difficultés auxquelles est heurté un traducteur professionnel lorsqu'il est en phase de constitution d'un corpus parallèle bilingue opposant une langue de science telle que le français à une langue purement littéraire et de moindre diffusion telle que l'arabe?

Afin de répondre à ce questionnement, nous étudierons à titre illustratif, un texte en guise de base d'un corpus parallèle bilingue Français-Arabe, il s'agira d'un compte rendu médical comportant une terminologie souvent représentée sous forme d'acronymes. Ce texte a préalablement fait l'objet d'une traduction vers l'arabe dans le cadre de notre travail en tant que traductrice au sein d'un cabinet de traduction officielle.

Tiziri Bachir, *Creating Parallel Bilingual Corpora for the Language of Science and Literature: Challenges in the Professional World*

Abstract

Specialised translation has not benefited from its own specific theorisation, but it is derived from various theories that have been developed within the framework of general translation. Nevertheless, translation researchers and practitioners have ensured that this category has its own tools that could reduce its complexity both linguistically and operationally.

In this paper, we will begin with the issue of software for computer-assisted Translation (CATtools), but by reducing the scope of such a vast study, before discussing parallel corpora (as mentioned in the title above), the challenges faced by professionals during the constitution and use of such corpora, especially when the linguistic combination in question includes Arabic as a source or target language.

Arabic today still remains a prisoner of its literary character, in addition to its dialectal nature that threatens its development as a scientific language. Indeed, in the midst of the developments taking place every day in scientific and technical fields, Arabic is lowly ranked for reasons we will discuss below, while French is known as a language of knowledge and research.

In this case, what are the difficulties encountered by a professional translator when creating a parallel bilingual corpus opposing a language of science such as French to a literary and less widely-spread language such as Arabic? In order to answer this question, we will study a text as the basis of a parallel bilingual (French-Arabic) corpus. We will present a medical report that includes terminology often in the form of acronyms. This text was previously translated into Arabic as part of our work as a translator in an officially certified translation office.

Résumé français

La traduction spécialisée n'a pas bénéficié d'une théorisation propre, mais elle tire ses enseignements des différentes théories développées dans le cadre de la traduction générale. Néanmoins, des chercheurs et praticiens de la traduction ont fait en sorte que cette catégorie soit munie d'un arsenal qui atténuerait sa complexité tant au niveau linguistique qu'opérationnel.

Dans la présente étude, nous aborderons la question des logiciels de Traduction Assistée par Ordinateur (CATtools). Nous avons toutefois choisi de réduire l'étude d'un champ aussi vaste afin d'aborder la question des corpus parallèles tel que l'indique le titre sus-énoncé, les défis auxquels les professionnels font face lors de la constitution mais aussi de l'utilisation de ces corpus, notamment lorsque la combinaison linguistique dont il est question, comporte l'arabe comme langue de départ ou langue d'arrivée.

La langue arabe demeure encore aujourd'hui prisonnière de son caractère littéraire en plus de sa situation de dialecte qui menace son développement en tant que langue scientifique. En effet, prise au milieu des développements qui s'opèrent tous les jours dans les domaines scientifiques et techniques, l'arabe se voit classée dans les derniers rangs pour des raisons que nous développerons ci-dessous, contrairement à la langue française connue pour être une langue de savoir et de recherche.

Quelles sont, dans ce cas, les difficultés auxquelles se heurte un traducteur professionnel lorsqu'il est en phase de constitution d'un corpus parallèle bilingue opposant une langue de science telle que le français à une langue purement littéraire et de moindre diffusion telle que l'arabe?

Afin de répondre à ce questionnement, nous étudierons à titre illustratif un texte servant de base à un corpus parallèle bilingue français-arabe. Il s'agira d'un compte rendu médical comportant une terminologie souvent représentée sous forme d'acronymes. Ce texte a préalablement fait l'objet d'une traduction vers l'arabe dans le cadre de notre travail en tant que traductrice au sein d'un cabinet de traduction officielle.

Adriano Ferraresi, Silvia Bernardini, Marie-Aude Lefer, Maja Miličević, *Investigating the Language of Written Translation and Simultaneous Interpretation: Simplification in a Tri-Directional, Monolingual Comparable, and Intermodal*

Abstract

The aim of this paper is to propose a framework for corpus-based comparative investigation of interpreting and translation, as exemplified by a study of simplification across different modes of language production and across different language pairs. This study is based on EPTIC, an intermodal corpus currently being constructed at the University of Bologna, UCLouvain and the University of Belgrade. EPTIC features plenary speeches by members of the European Parliament in their interpreted and translated versions that are aligned with each other as well as with their source texts in three languages (English, Italian, French). The overall size of the trilingual corpus is approximately 350,000 words. The corpus will soon be available through the NoSketch Engine platform offering multi-directional text alignment and text-to-video alignment, as well as advanced research features. Our aim is to shed light on lexico-syntactic simplification in different mediation modes, and compare interpretations and translations with each other and their corresponding non-translated and non-interpreted counterparts (original speeches and written versions). Specifically, we will compare lexico-syntactic features (lexical density and core vocabulary) in translation and interpreting into English from French and Italian, both from a comparable (monolingual) point of view and an intermodal perspective. Findings are discussed in the light of the debate on broadly conceived and typical features of translation with reflexions on the methodological potential of this multifaceted paradigm for research.

Résumé français

Dans cette présentation, située en traductologie de corpus, nous proposons un cadre pour l'analyse comparative de l'interprétation et de la traduction. Ce cadre d'analyse est illustré par le biais d'une étude de la simplification au travers différentes modalités (langue orale vs. écrite) et combinaisons de langues. Notre étude repose sur EPTIC, un corpus intermodal collecté conjointement par l'Université de Bologne, l'UCLouvain et l'Université de Belgrade. Le corpus EPTIC comprend, dans trois langues (anglais, italien et français), les interprétations et traductions des interventions des députés lors des séances plénières du Parlement européen, alignées entre elles et à leur texte source (discours oral et compte rendu écrit). Ce corpus trilingue, qui contient environ 350.000 mots, sera prochainement disponible sur la plateforme NoSketch Engine (alignement multidirectionnel et texte/vidéo ; requêtes avancées). Concrètement, l'objectif de notre étude est de mettre au jour les phénomènes liés à la simplification lexico-syntaxique dans deux types de médiation interlinguistique (traduction écrite vs. interprétation simultanée). Pour ce faire, nous comparons (1) les interprétations aux traductions des mêmes textes sources, (2) les interprétations/traductions à leurs textes sources respectifs. Plus précisément, nous comparons les caractéristiques lexico-syntaxiques (telles que la densité lexicale) en anglais traduit/interprété à partir du français et de l'italien, à la fois d'un point de vue comparable (monolingue) et d'un point de vue intermodal. Les résultats sont interprétés à la lumière du débat plus général sur les caractéristiques typiques de la langue traduite (au sens large). Pour conclure, nous offrons quelques réflexions méthodologiques sur le potentiel de ce nouveau paradigme de recherche.

Mali Saththachai, Dorothy Kenny, *Translation Features in an English-Thai Legal Translation Corpus*

Abstract

The broad hypothesis that translations differ from the original has encouraged many attempts to seek generalisations about translation. To date, two views on the distinct features of translations, i.e. features which set them apart from original texts, have been presented: (i) translated texts can reflect artifacts of the translation process and (ii) translated texts exhibit traces of the original language from which the texts were translated (Lembersky 2013: 1-2). The first view is derived from the idea of translation universals proposed by Baker (1993); the second from Toury's (1995) law of interference and Teich's (2003) idea of 'shining through'. Most existing research is based on European languages, and very little research has been conducted on legal texts. The current research thus sets out to investigate whether legal translations from English into Thai show evidence of 'translation features' as identified in a number of studies inspired by Baker, Toury and Teich, but drawing especially on Biel (2014). The study is data-driven, and findings regarding textual features such as text expansion, cohesive conjunctions, hypotaxis, parataxis, the passive voice as well as typical and atypical collocation, are used as a starting point for analysis. The corpora employed in this research are: a self-built parallel corpus (English legal texts aligned with their Thai translations), a self-built monolingual comparable corpus of Thai legal texts, and the Thai National Corpus, which is used as a general reference corpus. Sketch Engine is used to manage and interrogate the self-built corpora.

Résumé français

De nombreux chercheurs s'intéressent aux traits qui distinguent les traductions des autres textes écrits dans la langue cible en émettant l'hypothèse que : soit les textes traduits révèlent des artéfacts dus à la nature essentielle de la

traduction, soit ils révèlent des traces de la langue source (Lembersky 2013: 1-2). Pourtant, la plupart des travaux existants sont basés sur des langues européennes et il existe très peu de recherche sur les textes juridiques (sauf notables exceptions comme Biel 2014). Cette communication vise donc à déterminer si les traductions juridiques de l'anglais vers le thaï montrent des «traits de traduction» identifiés dans plusieurs travaux de recherche inspirés de Baker (1993), Toury (1995) et Teich (2003). En s'appuyant surtout sur Biel (2014) cette étude utilise (i) un corpus parallèle construit par les auteures et qui contient des textes juridiques en anglais avec leurs traductions en thaï, (ii) un corpus monolingue comparable de textes juridiques en thaï (également construit par les auteures) et (iii) le corpus national thaïlandais, qui sert de corpus de référence. Le logiciel Sketch Engine est utilisé pour gérer et interroger les deux premiers corpus. L'analyse de corpus menée dans ce travail s'intéresse surtout aux phénomènes textuels tels que l'expansion du texte, l'utilisation de la conjonction cohésive, l'hypotaxie, le parataxis, l'utilisation de la voix passive et des collocations typiques et atypiques.

Pauline De Baets, *Cognates in Translation: a Corpus-Based Behavioral Profile Approach*

Abstract

One of the main research topics within corpus-based translation studies is how translated texts differ from non-translated texts. Research has shown that language use in translated texts indeed differs linguistically from non-translated texts, but these differences are mostly ascribed to either translation universals or external language factors, such as register, source language, prestige, etc. (Baker, 1993 ; Capelle & Loock, 2013 ; Delaere, De Sutter & Plevoets, 2012 ; Kruger, 2012). However, considerably less attention is paid to cognitive explanations (see Halverson, 2003; Halverson 2007; Kroll & Stewart, 1994). By applying a usage-based method (i.e. behavioral profiling, see below) to semantic fields that contain both cognates and non-cognates, we hope to provide cognitive explanations for the attested translational features and thereby help to close the gap between cognitive linguistics and corpus-based translation studies. In this paper we will apply a behavioral profiling method to investigate the influence of cognates on the semantic structure of the translated Dutch semantic field of *beginnen* (compared to its use in Dutch). Behavioral profiling is a usage-based method that aims at capturing the complexity of the meaning of a word by looking at contextual features of the words under study (Gries & Divjak, 2009; Szymor, 2015). Firstly, the lexemes (in this case: Dutch verbs expressing inchoateness) are selected by means of the SMM++ (Vandevoorde, 2017), a method based on back translation. Secondly, the sentences containing one of those lexemes are extracted from the Dutch Parallel Corpus (Macken et al., 2011) and are annotated for both semantic and syntactic features, such as transitivity, verb mode and concreteness of the subject. In a final stage, the enriched corpus data is subjected to multiple correspondence analyses, and consequently the behavioral profile of the lexemes in translated and non-translated Dutch can be visualised. This visualisation can display even subtle changes in a semantic field, and thus enables us to thoroughly compare the semantic structure of non-translated and translated language. Consequently, in this paper we will argue that these differences can be indirectly interpreted as evidence of cognitive mechanisms in translation.

Résumé français

L'un des sujets les plus centraux en traductologie de corpus est l'influence de la traduction sur la langue : la traduction, change-t-elle la structure d'un texte? La recherche traductologique a déjà prouvé que le langage dans les textes traduits diffère de celui des textes non-traduits. Ces différences linguistiques ont été attribuées soit aux universaux de la traduction soit à des facteurs externes à la langue, comme le registre, la langue source, le prestige d'une langue, etc. (Baker, 1993 ; Capelle & Loock, 2013 ; Delaere, De Sutter & Plevoets, 2012 ; Kruger, 2012) Cependant, relativement peu d'attention a été donnée aux explications cognitives. En appliquant une méthode basée sur l'usage (i.e. le profilage comportemental [Gries & Divjak, 2009]), à des champs sémantiques contenant aussi bien des mots apparentés que des mots non-apparentés, nous espérons fournir des explications cognitives pour les traits linguistiques trouvés dans les textes traduits. Ainsi, nous voudrions combler l'écart entre la linguistique cognitive et la traductologie de corpus. Dans cette étude, nous utilisons la méthode du profilage comportemental pour examiner l'influence des mots apparentés à la structure et le champ sémantique de *beginnen* en néerlandais non-traduit et néerlandais traduit. Le profilage comportemental vise à capter la complexité de la signification des mots en regardant les traits contextuels des mots étudiés (Gries & Divjak, 2009; Szymor, 2015). D'abord, les lexèmes (dans notre cas, des verbes néerlandais exprimant l'inchoativité) sont sélectionnés au moyen de SMM++ (Vandevoorde 2017), une méthode fondée sur la retrotraduction. Ensuite, les phrases qui contiennent un de ces lexèmes sont extraites du Corpus parallèle néerlandais (Macken et al., 2011), et ensuite annotées pour des caractéristiques syntactiques ainsi que sémantiques (par exemple la transitivité et le mode du verbe, l'abstraction du sujet,...). Finalement, ces données de corpus enrichies sont soumises à des analyses de correspondances multiples et par conséquent, les profils comportementaux des lexèmes en néerlandais traduit et non-traduit peuvent être visualisés. Cette visualisation est apte à montrer des nuances même subtiles dans un champ sémantique et nous permet donc d'examiner de manière minutieuse la structure de la langue traduite avec celle de la langue non-traduite. En guise de conclusion, nous soutiendrons que les différences entre langue traduite et non-traduite peuvent être expliquées à travers des mécanismes cognitifs qui sont en jeu pendant le processus de traduction.

Lore Vandevoorde, *Register-Sensitivity on the Semantic Level: a Register-Specific Comparison of the Semantic Field of *Beginnen* in Translated and Non-Translated Dutch*

Abstract

One of the major research paradigms in corpus-based translation studies revolves around the idea that there exists a number of universally viable characteristics of translated texts. However, a particular characteristic of translated language can only be categorised as universal if it is invariable across all translation-relevant parameters such as source and target language, specific language pairs, genres etc. (Chesterman, 2004). Recent studies have indeed shown that specific language pairs (e.g. Xiao, 2010) or text genres (e.g. Delaere et al., 2012) have an impact on the presence, directionality and magnitude of differences accounted for in translated texts compared to non-translated texts, thus unsettling the universality of translation universals. In this paper, we aim to tackle two under-researched aspects of corpus-based translation studies. Firstly, and in line with research conducted by inter alia Kruger & van Rooy (2012) and Delaere et al. (2012), we will study the relationship between genre and the characteristics of the translated language. Secondly, we will explore these characteristics from a semantic point of view – another somewhat neglected area in translation studies. According to a recent study (Vandevoorde, 2016), differences between translated and non-translated language also exist at a semantic level. It was shown that semantic variation was levelled out in translation irrespective of the source language. In the current follow-up study, and in line with previous research by Kruger and Van Rooy on register-sensitivity, we want to investigate whether translated language is less genre sensitive (“levelled out”) from a semantic point of view. We will therefore further explore the semantic field of the word *beginnen* in Dutch. By using statistical visualisation methods such as hierarchical cluster analysis, we are able to visualise semantic fields of inchoateness by text type – the Dutch Parallel Corpus (Macken et al. 2011) allows us to analyse and compare both translated and non-translated language for the following text types: journalistic texts, external communication, non-fictional literature and administrative texts. In this way, we can see whether translated Dutch is less register-sensitive than non-translated language – a hypothesis put forward by Kruger and Van Rooy but unsupported for translated English.

Résumé français

L'un des principaux paradigmes de recherche en traductologie de corpus tourne autour de l'idée que les textes traduits possèdent un certain nombre de caractéristiques universellement viables. Cependant, une caractéristique particulière de la langue traduite ne peut être classée comme universelle que si elle est invariable à travers tous les paramètres liés à la traduction tels que la langue source et la langue cible, les paires de langues, le genre textuel, etc. (Chesterman, 2004). Des études récentes ont en effet montré que les paires de langues spécifiques (Xiao, 2010) ou les genres textuels (Delaere et al., 2012) ont un impact sur la présence, la directivité et l'ampleur des différences attestées dans les textes traduits par rapport aux textes non traduits, déstabilisant ainsi l'universalité des universaux de la traduction. Dans cet exposé, nous voulons aborder deux aspects peu recherchés en traductologie de corpus. Tout d'abord et conformément aux recherches menées par, entre autres, Kruger et van Rooy (2012) et Delaere et al. (2012), nous voulons étudier la relation entre le genre textuel et les caractéristiques de la langue traduite. Deuxièmement, nous explorerons ces caractéristiques sur le niveau sémantique, un aspect qui est également peu étudié en traductologie. Selon une étude récente (Vandevoorde, 2016), il existe également des différences entre la langue traduite et la langue non-traduite au niveau sémantique. Il a été montré que la variation sémantique est « nivelée » en langue traduite, un phénomène qui a été observé indépendamment de la langue source. Dans la présente étude de suivi, nous voulons, conformément aux recherches antérieures de Kruger et Van Rooy sur l'impact du genre textuel sur la langue traduite, examiner si les particularités des genres textuels sont moins apparentes en langue traduite qu'en langue non-traduite et ce au niveau sémantique. Nous explorerons pour cela davantage le champ sémantique du verbe néerlandais *beginnen* [commencer]. En utilisant des méthodes de visualisation statistique (classification hiérarchique), nous visualiserons les champs sémantiques d'inchoativité par genre textuel. Le Dutch Parallel Corpus (corpus parallèle du néerlandais) (Macken et al., 2011) nous permet de comparer la langue non traduite à la langue traduite pour des textes appartenant aux genres textuels suivants : textes journalistiques, textes de communication externe, textes de littérature non-fiction et textes administratifs. Nous pouvons ainsi voir si le néerlandais traduit est moins sensible aux différences liées au genre textuel que le néerlandais non traduit, une hypothèse mise en avant mais non corroborée pour l'anglais traduit par Kruger et Van Rooy (2012).