

### **Résumé français**

La « diplomatie » du discours est polymorphe dans la mesure où elle renvoie aussi bien aux discours émis par les diplomates qu'aux comportements discursifs entre les États ou les acteurs transnationaux, convoquant un certain « art » empreint de tact et d'habileté (Vilar, 2006 : 12). Après avoir caractérisé sommairement le discours diplomatique, nous verrons en quoi la traduction institutionnelle de ces discours constitue un acte de communication diplomatique à part entière. Pour ce faire, nous appuierons notre réflexion sur les conférences de presse du président de la république et du premier ministre espagnol, lors des trois derniers sommets franco-espagnols (2012, 2013, 2014), et sur leurs traductions par les institutions qu'ils incarnent (Élysée, Présidence du gouvernement espagnol). Ce matériau discursif est constitué en corpus parallèle (Bowker, Pearson, 2002 : 92) comptabilisant pour le volet langue-source 6 839 occurrences, et 6 839 occurrences pour le volet langue-cible. Il s'agit d'un corpus bidirectionnel dans la mesure où le français et l'espagnol peuvent être langue-source et langue-cible. Le corpus ainsi élaboré suit les préceptes et objectifs de la linguistique de corpus (Sinclair, 1996 ; Tognini-Bonelli, 2001) pour des fins traductologiques (Baker, 1995 ; Kübler, 2014) moyennant le déploiement d'outils informatiques (Zimina, 2013), *Le Trameur* (Fleury, 2016) et *Lexico 3* (Lamalle et al., 2003).

### **Abstract**

The « diplomatic » aspect of discourse is polymorphous insofar as it refers back to both the discourse of diplomats and the discursive conduct between transnational States or players, bringing together a certain « art » that calls for tact and skills (Vilar, 2006 : 12). After having briefly described the diplomatic discourse, we will see how the institutional translation of these discourses constitutes a diplomatic communicative act in its own right. For this purpose, we will go back to our reflections on press conferences by the President of the Republic and the Spanish Prime Minister, during the last three Franco-Spanish Summit Meetings (2012, 2013, 2014), and their translations by the institutions they represent (Élysée, Presidency of the Spanish Government...). This discursive material is transformed into a parallel corpus (Bowker, Pearson, 2002: 92) accounting for 6,839 occurrences for the source language component, and 6,839 occurrences on the target language side. This is a bidirectional corpus to the extent that French and Spanish can be either the source language or the target language. It was developed following the precepts and objectives for corpus linguistics (Sinclair, 1996 ; Tognini-Bonelli, 2001) in translation studies (Baker, 1995 ; Kübler, 2014) through the use of computer tools (Zimina, 2013), *Le Trameur* (Fleury, 2016) and *Lexico 3* (Lamalle et al., 2003).

### **Christophe Réthoré, *Approche textométrique et comparative du discours de segmentation du marché automobile : implications pour la traduction publicitaire anglais > français***

#### **Résumé français**

Notre point de départ est le travail pionnier en linguistique descriptive de Leech (*The Language of Advertising*, 1966). D'autres recherches ont ensuite été publiées sur le discours publicitaire (Bruthiaux 1995) ou sur la traduction publicitaire (Guidère 2000). Il ressort de ces recherches que :

1. les mots de la publicité constituent bien un discours spécialisé
2. la publicité ressemble à la fois à la poésie et à la langue courante
3. l'intertextualité et la participation active d'un récepteur ciblé sont deux données clés de la communication publicitaire
4. la langue publicitaire est émotionnelle, ambiguë, indirecte et implicite
5. de façon simpliste et ludique, la publicité nous dépeint un monde de bonheur parfait
6. simplicité et conventionnalisation du discours publicitaire, et invariants lexicaux

Pour explorer la publicité, nous avons choisi une approche quantitative, à savoir l'approche lexicométrique (ou statistique textuelle, devenue récemment textométrie), développée notamment par le « groupe de Saint-Cloud » et ses collaborateurs, dont André Salem (concepteur de *Lexico*, le logiciel utilisé pour nos analyses). Grâce à un corpus parallèle (bitexte) d'environ 800 000 occurrences en anglais et un million en français (170 brochures publicitaires compilées à partir de sites Web canadiens de 20 marques, 10 constructeurs, 229 véhicules au total), nous avons réussi à déterminer un profil lexical distinct pour chacun des sept segments du marché automobile : AB, CD, EF, MPV (véhicules multifonctions), PK (pickups/camions), SP (voitures de sport) et SUV (VUS : véhicules utilitaires sports) et montré que dans l'ensemble, les publicités sont traduites assez littéralement.

#### **Abstract**

Our research starts with the pioneering work by Leech (*The Language of Advertising*, 1966). Other authors have subsequently investigated the advertising discourse (Bruthiaux, 1995) or translation for advertising (Guidère, 2000). Our literature review highlighted the following issues:

1. advertising is indeed a specialised discourse, or a language for special purpose (LSP)
2. advertising shares some common features with both poetry and the language we speak everyday
3. intertextuality and active participation of the targeted reader are two key factors in a publicity message
4. the language of advertising is emotional, ambiguous, indirect and implicit
5. in a simple and playful way, advertising depicts a perfect, happy world
6. simplicity and conventionalisation of the advertising discourse, with lexical constants.

To further explore advertising, we chose a quantitative approach, i.e., the lexicometric methodology (statistical linguistics, more recently called textometry) developed by the Saint-Cloud group in France, including André Salem (founder and developer of *Lexico*, which we used to analyse our corpus). With our parallel corpus of roughly 800,000 items in English and one million in French (170 advertising brochures downloaded from 20 brands/10 carmakers' Canadian websites, for a total of 229 vehicles), we found distinctive lexical profiles for each of the seven car market segments: AB, CD, EF, MPV, PK (multifunctional vehicles), SP (sports cars) and SUV. We also showed that, overall, advertisements are translated rather literally.

### Olivier Kraif, *Sous-genre littéraire et traduction : une approche textométrique*

#### Résumé français

Notre étude s'inscrit dans le cadre du projet *PhraséoRom* (ANR-FDG), qui a pour objectif d'étudier la relation entre sous-genres romanesques et phraséologie, à travers un corpus multilingue français, anglais et allemand. Dans une approche de linguistique de corpus outillée, nous nous demandons s'il est pertinent de s'appuyer sur des textes traduits pour observer des phénomènes langagiers idiomatiques ainsi que des motifs textuels (au sens de Mellet et Longrée, 2013) – les textes traduits étant traditionnellement bannis des approches sur corpus.

Pour ce faire, nous avons constitué un corpus comparable de romans policiers en français langue source et de *detective novels* anglophones traduits en français (environ une 115 œuvre au total). À travers une étude textométrique, nous proposons de comparer les œuvres sur le plan lexical, syntaxique et phraséologique pour déterminer les spécificités linguistiques de chacun de nos 2 sous-corpus, et évaluer les proximités génériques entre textes sources et textes traduits.

#### Abstract

Our study is part of the *PhraséoRom* Project (ANR-FDG), which aims at studying the phraseological specificities of novel subgenres, through a multilingual French, English and German corpus. Using a corpus-driven methodology, we plan to assess the reliability of translated texts to observe idiomatic phenomena as well as textual motifs (in the sense of Mellet and Longrée, 2013) – the translated texts being traditionally banned from corpus-driven approaches.

For this purpose, we have created a comparable corpus of French "romans policiers" and English detective novels translated into French (a total of about 115 novels). Through a textometric study, we propose to compare the texts on the lexical, syntactic and phraseological grounds, in order to determine the linguistic specificity of each sub-corpora, and to evaluate the generic similarities between source texts and translated texts.

### Jun Miao, *Exploitation textométrique des stratégies dans la traduction des noms de personnes du français vers le chinois*

#### Résumé français

Le problème de la traduction des noms propres est récurrent, il constitue toujours un défi, quelle que soit la langue de départ et la traduction effectuée (Newmark, 1988 ; Ballard, 2001 ; Grass, 2002). Cependant, lorsqu'il s'agit de noms de personnes, les anthroponymes, la difficulté dépend largement des langues concernées. Dans la plupart des langues occidentales, le fait que les noms de personnes ne se traduisent pas est incontestable (Vaxelaire, 2006 : 729), mais il peut parfois exister certaines adaptations. Par contre, dès qu'il s'agit de langues dont les systèmes d'écriture sont éloignés, telles que le français et le chinois, ou l'anglais et l'arabe, une modification graphique est inévitable. Notre étude se limite à la traduction des noms de personnes du français vers le chinois. À l'aide de la textométrie et de son outil (*Lexico3*), nous avons pu mener une recherche méthodique en comparant les trois traductions chinoises d'une même œuvre littéraire française : *Jean-Christophe* de Romain Rolland. Les premiers constats font apparaître qu'en dehors des aspects phonétique et significatif, les traducteurs s'intéressent également au nombre de syllabes/caractères et à la graphie avant de retranscrire les noms en chinois. Ex. le trisyllabique et la combinaison « yin-partant-yin » occupent une place relativement importante dans chaque traduction. Par ailleurs, les analyses statistiques sur le corpus (ex. spécificité et spécificité chronologique) nous montrent que le procédé d'un traducteur reste relativement stable au fil de son œuvre (qui comprend 10 tomes), ce qui nous permet de nous interroger sur la notion de style du traducteur. Enfin, l'affichage simultané grâce à la cartographie dans *Lexico3* nous aide à examiner les noms dans leur contexte, notamment à la lumière du caractère des personnages.

## Abstract

The translation of proper nouns is a recurring challenge, regardless of either the source language or the translation (Newmark, 1988; Ballard, 2001; Grass, 2002). The specific issue of translating people's names, however, is largely a function of the languages concerned. In most Western languages, the generally accepted practice is not to translate personal names (Vaxelaire, 2006: 729), but sometimes adaptations are made. However, with regard to translations between languages based on different distant systems, such as French and Chinese, or English and Arabic, a graphic modification is inevitable. Our research is limited to the translation of personal names from French to Chinese. Using textometric analysis and its tool (*Lexico3*) to support our methodology, we compared three Chinese translations of the same French literary work: *Jean-Christophe* by Romain Rolland. Our initial findings showed that apart from aspects of phonetics and meaning, the translators also take into account the number of syllabics/characters and the graph when creating the Chinese name. For example, the tri-syllabic and the "yin-falling tone-yin" combination occupy a relatively important place in each translation. Further statistical analyses (specificity and chronological specificity measures) on the corpus suggested that the method translators used in their work (containing 10 volumes) remained stable. This point allowed us to query the styles of the translators. Finally, the simultaneous display using cartography in *Lexico3* allowed us to examine the names in their context, by associating them with the character of the persona.

**Bendinelli Marion, *Exploiter un corpus comparable multilingue de taille déséquilibrée : réponses textométriques et statistiques à un enjeu méthodologique. Illustration par la recherche d'équivalents de traduction aux termes delicacy et configuration en LSF***

## Résumé français

Notre communication porte sur les enjeux méthodologiques que posent les corpus comparables multilingues de taille réduite et/ou déséquilibrée. Notre réflexion se fonde sur un corpus de textes scientifiques présentant la linguistique systémique fonctionnelle en anglais et en français (volets LSF-EN et LSF-FR). Seuls deux articles de Michael A.K. Halliday, son auteur, ayant été traduits dans l'espace francophone, tout chercheur intéressé par la terminologie de ce modèle et par sa traduction, doit construire un corpus comparable, réunissant publications du linguiste britannique et de chercheurs francophones (D. Banks notamment). Outre le fait que l'on puisse s'interroger sur la comparabilité de ces écrits, deux questions se dégagent : que faire en cas de fort déséquilibre entre les deux volets (LSF-EN > 350 000 mots vs LSF-FR < 30 000 mots) ? Et comment travailler avec des fréquences d'emploi basses (inhérentes à un corpus ou volet de petite taille) ? Ces enjeux sont illustrés par la recherche d'équivalents de traduction aux termes *delicacy* et *configuration*. Notre démarche s'inscrit dans le cadre d'une linguistique de corpus outillée (cf. Zimina, 2014 ; Kraif, 2016). Nous revenons sur les fondements statistiques de méthodes quantitatives promues dans le cadre de la démarche textométrique (e.g. calcul des segments répétés et des cooccurrents) et sur les possibilités de visualisation offertes (graphes des cooccurrents, courbes de progression). Nous comparons également le recours à divers tests statistiques (écart réduit, log-likelihood ratio, Fisher-Yates exact test) et soulignons l'importance de tenir compte de la dispersion des items dans le calcul de leur fréquence d'emploi.

## Abstract

Our paper focuses on a methodological issue raised by the use of unbalanced or quite small multilingual comparable corpora. It stems from an ongoing research project dealing with the terminology and phraseology of the Systemic Functional Linguistics, a theoretical framework which has been developed since the 1960s by Michael A.K. Halliday. As only two of his papers have been translated so far into French, the research for translation equivalents must rely on comparable papers in French (e.g., D. Banks in particular). Indeed, in contrast with the numerous publications by Halliday, papers in French are quite rare. Hence our comparable corpus appears highly unbalanced with one part rather small (Eng. > 350,000 words; Fr. < 30,000 words). While we may discuss the actual comparability of these texts, we would like to address two important issues: how are we to deal with an unbalanced and/or small corpus, and consequently how are we to deal with low word frequencies? We illustrate these through the research of translation equivalents for the terms *delicacy* and *configuration*. Following previous works (Zimina, 2014; Kraif, 2016), we believe that textometrics (i.e. a set of quantitative and statistical methods to explore corpora) provides answers to such methodological issues. We first consider the statistics that stand at the core of some methods used to identify (dis)continuous multi-word expressions and cooccurrence networks. We also compare the results of different statistical tests (z-score, log-likelihood ratio, Fisher-Yates exact test) and stress the necessity to compute a statistical spread of measurements.

**Ilaria Cennamo, *Unité textométrique et unité de traduction : une réflexion pédagogique***

## Résumé français

La textométrie a pour objectif principal de compter les éléments textuels à partir de différents types d'unités élémentaires analysables dans un corpus ou dans des portions de ce dernier. À son tour, la traduction est une

opération de saisie et restitution d'un contenu s'avérant exécutable grâce au repérage et au traitement d'unités de sens. Étant donné que les unités textométriques et les unités de traduction sont strictement dépendantes du contexte discursif qui les abrite, pourraient-elles premières faciliter l'identification des deuxièmes ? C'est dans le cadre du projet de recherche post-doctorale qui fait l'objet de cette communication que l'on essaiera d'apporter les premières réponses à cette question. Notre étude est centrée sur l'adoption des méthodes d'exploration textométrique de corpus comparables en français et italien, pensés en tant que ressources pédagogiques pour l'apprentissage de la traduction. Nos corpus sont notamment représentatifs du discours institutionnel public actuel, diffusé par la rubrique des actualités du site web de la mairie de Paris : [www.paris.fr](http://www.paris.fr) (côté français) et du site de la mairie de Rome : <http://www.comune.roma.it/pcr/> (côté italien). Par le biais d'exemples tirés de nos corpus, nous montrerons les différentes unités observables : celles qui renvoient aux thématiques abordées, celles qui relèvent des spécificités discursives du genre (« Exprimez-vous ci-dessous, des experts de la Mairie de Paris vous répondront personnellement »), et les unités porteuses d'informations identitaires (« les Franciliens de la petite/grande couronne »). Notre réflexion concernera l'apport de la textométrie au processus d'apprentissage de la traduction.

### **Abstract**

The main goal of textometrics is to count textual elements starting from different types of elementary units, which can be analysed in a corpus, or portions of it. Translation, in its turn, is an operation of content capture and restitution, based on semantic unit identification and processing. Since textometric units and translation units are strictly context-dependent, can the first ones indicated here help identify the second? The research project presented in this communication will try to give some answers to this question. Our study is focused on the use of textometric exploration methods applied to comparable corpora. Those corpora have been thought of as pedagogical resources for translation teaching and learning and are representative of present public institutional discourse communicated through the Web. More precisely, corpora have been constituted on the basis of the news heading of the website for the municipality of Paris: [www.paris.fr](http://www.paris.fr) (French side) and the website for the municipality of Rome, <http://www.comune.roma.it/pcr/> (Italian side). With the use of some examples extracted from our corpora, we will show different observable units: units belonging to the subject discussed, units containing discourse genre-specific features ("Exprimez-vous ci-dessous, des experts de la Mairie de Paris vous répondront personnellement"), and identity informative units ("les Franciliens de la petite/grande couronne"). Our presentation will address the issue of textometric contribution to translation learning process.

### **Maria Zimina, *Vers une Mémoire de Traduction (MT) dynamique et multidimensionnelle***

#### **Résumé français**

Proposé à la fin des années 80s par Brian Harris (1988), le terme *bi-text* désigne l'ensemble d'équivalents de traduction unissant les textes source et cible dans l'esprit du traducteur. Dès la fin des années 90s, le succès rapide des méthodes d'alignement des phrases a amené un rapprochement important entre le bi-texte et les corpus alignés. Cette tendance s'est confirmée en Traitement Automatique des Langues (TAL) avec le développement des formats d'échange, comme le TMX (Translation Memory eXchange) pour le recyclage des données de traductions existantes dans la production de nouvelles traductions. Les Mémoires de Traduction (MT) actuelles peuvent être assimilées à des tables de correspondances. Cette représentation a ses limites. La démarche textométrique permet une modélisation complémentaire de l'alignement textuel. La textométrie repose principalement sur l'observation des variations de fréquences d'unités textuelles (formes, lemmes, segments, co-occurrences, patrons, motifs, etc.) appelées *contenus textuels*, dans les différentes parties d'un ensemble de textes, considérées comme des *contenants textuels* (parties, sections, zones, chapitres, paragraphes, phrases, séquence, etc.). Cette spécificité de la démarche textométrique permet d'utiliser des modèles de données qui préservent des liens unissant les textes source et cible à plusieurs niveaux (mots, syntagmes, phrases, cadres discursifs, etc.). On peut ainsi envisager une transition vers des MT qui deviennent dynamiques et multidimensionnelles et reflètent davantage les besoins en traduction.

#### **Abstract**

At the end of the 80s, Brian Harris coined the term *bi-text* (1988) referring to source and target text translation correspondences present in the translator's mind. In the late 90s, rapid success of sentence alignment methods brought together aligned corpora and the concept of bi-text. This tendency was further reinforced in Natural Language Processing (NLP) with the development of TMX (Translation Memory eXchange). This format allows sharing and reusing existing translation resources in new translations. For the moment, Translation Memories (TM) are more or less static tables matching translation correspondences. A complementary model can be developed on the basis of textometric methods for studying frequency variations of different text units, or *contents* (tokens, lemmas, part-of-speech patterns, etc.), in text *containers* (corpus parts, sections, zones, chapters, sentences, sequences, etc.). Textometric analysis relies upon data models preserving links between source and target texts at different levels of analysis (words, expressions, sentences, discourse frames, etc.). Following this approach, it becomes possible to conceive a dynamic multidimensional TM that better suits translation needs.