

**Kyo Kageura, *Applying a Corpus Evaluation Method to Assess the Status of Materials for Translation Practice in Relation to Technical Terms***

**Abstract**

From the point of view of practice or application, several links between textual corpora and terms have been explored so far, i.e. a) using the corpora as sources from which terms are extracted, b) using the words as key features for classification in the corpora, c) using terms to assess the characteristics or the status of the corpora. A range of methods developed and features identified in these tasks constitute a rich pool of knowledge for us to explore the links between textual data and terms. Unlike the corpora used in most applications, textual material in translation classes are characterised by their small size. Partly due to this, the potential of the methods has not been fully examined when assessing the nature and status of the textual material. This paper examines how, in technical terms, methods for evaluating corpora can be used to characterise textual material in teaching translation. For this purpose, we have prepared a total of 12 short English texts from six domains: law, medicine, politics, physics, technology and philosophy. Two texts were chosen from each domain, and their length ranged from 300 to 1,150 words. We extracted terms manually from each text, and they were used to assess the nature and status of the textual material. The analysis shows that even in the case of short texts, certain methods that were put forward to characterise the corpus, such as the self-referential measurement of the occurrence of terminology and the semantic range of the discourse, provide useful insights into the textual material.

**Résumé français**

Plusieurs relations entre les corpus textuels et les termes ont été explorées à ce jour, e.g. en utilisant les corpus comme sources à partir desquelles les termes sont extraits, en utilisant les termes comme éléments-clés pour classer les corpus, en utilisant les termes pour évaluer les caractéristiques des corpus. Les méthodes développées et des caractéristiques identifiées constituent un riche ensemble de connaissances pour explorer les relations entre les données textuelles et les termes. Contrairement à la plupart des corpus, les documents utilisés dans les classes de traduction se caractérisent par leur petite taille, et le potentiel des méthodes existantes n'a pas été entièrement examiné dans l'évaluation de la nature de ces documents. Nous examinons comment les méthodes d'évaluation de corpus en termes techniques peuvent être utilisées pour caractériser les documents utilisés dans l'apprentissage de la traduction. Nous avons préparé 12 textes courts en anglais dans les six domaines du droit, de la médecine, de la politique, de la physique, de la technologie et de la philosophie (deux textes choisis pour chaque domaine), de 300 à 1150 mots. Nous avons extrait manuellement les termes de chaque texte, et, en utilisant les termes, nous avons évalué la nature des documents. L'analyse montre que, même pour les textes courts, certaines méthodes avancées de caractérisation de corpus, telles que la mesure auto-référente de l'apparition des termes et de l'échelle sémantique du discours, donnent un aperçu utile des matériaux textuels.

**Chedi Bechikh Ali, *Crosslingual Approaches to Terminology Extraction and Application to Crosslingual Retrieval***

**Abstract**

We will propose and assess a model for the retrieval of inter-language information based on a method for bilingual lexicon extraction using comparable corpora. The model uses Formal Analysis to extract closed bilingual concepts that will serve as a resource to create bilingual documents and their indexes. The correspondence between closed bilingual concepts makes it possible to construct a bilingual lexicon. This bilingual lexicon contains complex in addition to simple terms, which will be used either for translation or in a process for the retrieval of inter-language information. The retrieval of the bilingual lexicon will be carried out using a comparable corpus, CLEF 2003. We will apply the results of this evaluation to an Italian-French task and an English-French task.

**Résumé français**

Nous proposons et évaluons un modèle de recherche d'information inter-langues qui se base sur une méthode d'extraction de lexique bilingue à partir de corpus comparables. Le modèle utilise l'Analyse formelle de concepts pour extraire des concepts fermés bilingues qui seront utilisés comme ressource pour créer des documents bilingues et leurs index. La correspondance entre les Concepts fermés bilingues permet de construire un lexique bilingue. Ce lexique bilingue contient des termes complexes en plus des termes simples, qui seront utilisé soit pour la traduction, soit dans un processus de recherche d'information inter-langue. L'extraction du lexique bilingue sera effectuée sur le corpus comparable CLEF 2003. Nous rapportons les résultats de l'évaluation pour une tâche italien-français et une tâche anglais-français.

## Mojca Pecman, *Corpus as a Reservoir for Term Management in Emerging Fields of Knowledge: Looking into the Discourse and the Terminology of Fractal Geometry*

### Abstract

Nowadays, given the rapid growth of information technologies and the constant renewal of specialised knowledge, the corpus is often the only source that provides the necessary data for terminology management. With corpus linguistics, terminologists have gained access to up-to-date language resources and thereby a deeper understanding of terminology-in-the-making (Pavel 1993). Terminology databases strive to keep up with the pace of change, and provide information on terms, variants, definitions, and conceptual relations by reflecting the different tendencies. This presentation illustrates the underlying principles of corpus-based and corpus-driven terminology management for studying emerging fields of knowledge, as used in the ARTES terminology database which has now been developed for six years at the Paris Diderot University (Pecman and Kübler 2011). This DB supports the teaching of terminology to future specialised translators (Kübler et al. 2016). We will show how corpus analysis is at the heart of terminology management but also how essential it is to combine the corpora with discourse analysis. We will discuss these points using a number of examples taken from a particular scientific discourse, fractal geometry, which is a rich source for the analysis and the observation of how it spreads to other domains (Pavel et Boileau 2003[1994]). By focusing on English-French interaction, we will discuss some tendencies such as those for bicapitalisation (*FractInt*, *WinFract*, *FractalPro*), truncation (*p-adic*, *p-Laplace equation*, *PL-systems*), complex nominal groups (*self-similar p-adic fractal strings*, *construction of the 3-adic Cantor string via an IFS*), and the metaphoric framing of terms (*Appenzell cheese*, *Minkowski sausage*, *devil's staircase*, *fairy ring*, *twindragon*).

### Résumé français

L'évolution rapide des technologies d'information et le renouvellement constant de nos connaissances ont promu le corpus au rang du réservoir inépuisable pour la gestion terminologique. En effet, la linguistique de corpus a offert aux terminologues non seulement un réservoir de données authentiques, mais également de données actuelles permettant d'observer de près les processus de création terminologique à l'œuvre (Pavel 1993). Les bases de données terminologiques s'efforcent de suivre le rythme rapide d'innovation en fournissant [, et fournissent] les informations sur les termes, les variants, les définitions et les relations conceptuelles en reflétant les évolutions. Notre communication illustre les principes de la gestion terminologique des domaines émergents selon les deux approches, corpus-based et corpus-driven, utilisées dans la base ARTES depuis maintenant six ans (Pecman and Kübler 2011) et soutenant l'enseignement de la terminologie aux futurs traducteurs à l'Université Paris Diderot (Kübler et al. 2016). Le corpus est central pour la gestion terminologique, mais l'analyse du discours l'est autant. Nous discutons ces points à l'aide d'exemples pris dans le discours scientifique, notamment celui de la géométrie fractale qui offre un réservoir riche pour l'analyse terminologique tout en permettant d'observer sa propagation à d'autres domaines (Pavel et Boileau 2003[1994]). En explorant l'interaction entre l'anglais et le français, nous illustrerons quelques tendances, comme la bicapitalisation (*FractInt*, *WinFract*, *FractalPro*), l'apocope (*p-adic*, *p-Laplace equation*, *PL-systems*), la complexification des groupes nominaux (*self-similar p-adic fractal strings*, *construction of the 3-adic Cantor string via an IFS*) et le recours à la métaphore (*Appenzell cheese*, *Minkowski sausage*, *devil's staircase*, *fairy ring*, *twindragon*, etc.).

## Mavina Pantazara et Eleni Tziafa, *Les termes de la crise économique grecque dans les corpus*

### Résumé français

Les changements du vocabulaire suivent de près la dynamique de la société, dans le sens que le lexique, la composante la plus socialisée de la langue, essaie de refléter de son mieux et le plus vite possible son évolution. Depuis quelques années déjà, on ne parle que de la crise, qui de financière, bancaire et boursière au début est devenue économique et sociale ensuite (Mladin, 2009). Le point de départ de notre recherche est le glossaire multilingue « Greek Crisis Multilingual Term Project – GCMTMP » (disponible en ligne: <http://gcmtmp.herobo.com/>) qui a été réalisé dans le cadre du Programme de Master 2 « Traduction-Traductologie » de l'Université nationale et capodistrienne d'Athènes (2011-2013) (Seel & Pantazara, 2013). Il couvre un large éventail de termes liés à la crise économique grecque, dont le nombre total s'élève actuellement à 822 entrées. Notre objectif est de vérifier les occurrences de ces termes dans des corpus en trois langues (français, anglais, grec), d'étudier leur fréquence et de décrire leur comportement en contexte à l'aide de Sketch Engine (Kilgarriff, 2014). Les corpus nous permettent d'une part de valider des termes qui ne sont pas normalisés et encodés dans des ouvrages de référence et d'autre part de répertorier la variation des équivalents et des synonymes. Notre démarche s'inscrit dans la perspective de proposer un outil d'aide pour les traducteurs et une ressource de référence pour les professionnels travaillant dans des secteurs d'activité variés, tels que la rédaction technique ou la presse.

### Abstract

Vocabulary changes closely follow societal dynamics, in the sense that the lexicon, the most social linguistic component, tries to catch up with its evolution as efficiently and as quickly as possible. For a few years now, the crisis has been one of the most discussed subjects, starting out as a financial, banking and stock market crisis to finally

becoming economic and social (Mladin, 2009). The starting point of our research is the multilingual glossary “Greek Crisis Multilingual Term Project – GCMTP” (available online at <http://gcmtp.herobo.com/>). This glossary was created as part of the Master 2 Programme “Translation Studies” at the National and Kapodistrian University of Athens (2011-2013) (Seel & Pantazara, 2013). The glossary covers a large variety of terms related to the Greek financial crisis and comprises a total of 822 entries. Our objective is to verify the occurrences of these terms in three language corpora (French, English, Greek), to study their frequencies and to describe, in context, how they function using Sketch Engine (Kilgarriff, 2014). The corpora allow us to validate terms that are not yet standardised or included in reference resources. Moreover, we can extract and list the variations of equivalent terms and synonyms. The aim of our approach is to offer a useful tool for translators, along with a reference resource for professionals working in various sectors, such as technical writing or the press.

**Teresa Ortego-Antón, *L'utilisation d'anglicismes dans la terminologie des réseaux sociaux en espagnol et en portugais du Brésil : une étude basée sur corpus***

### **Résumé français**

Ces dernières années le nombre de notions et de termes a augmenté très rapidement dans le domaine des réseaux sociaux. Ce domaine a été créé dans un contexte anglophone, donc les personnes parlant d'autres langues ont besoin d'exprimer ces notions dans leur propre langue. Dans ce contexte, les traducteurs et les journalistes ont besoin de comprendre comment le transfert de ces concepts fonctionne dans des textes. Étant donné que l'anglais est devenu la lingua franca dans la terminologie des réseaux sociaux, dans cet étude on vérifie si l'usage d'anglicismes est parmi les procédés de traduction les plus utilisés pour faire le transfert des termes de ce domaine de l'anglais vers l'espagnol et le portugais du Brésil. Par conséquent, un corpus comparable en anglais, espagnol et portugais du Brésil est compilé. Avec l'aide de l'outil Termostat Web 3.0., les termes en espagnol et en portugais du Brésil sont extraits et classifiés en emprunts ou calques selon la forme linguistique. En plus, l'utilisation d'autres dénominations pour faire référence à la même notion est vérifiée ainsi que l'usage du terme dans le discours anglais. À partir des résultats obtenus les [des ?] conclusions sont établies, qui pourront nous aider à déterminer des modèles de comportement dans l'usage d'anglicismes pendant le transfert de la terminologie des réseaux sociaux de l'anglais vers l'espagnol et le portugais.

### **Abstract**

In the social media, the number of new concepts and terms has increased rapidly in recent years. As social media was first created in an English-speaking context, speakers of other languages need to express concepts from this field in their own languages. It is therefore relevant for translators and journalists to understand the ways in which the transfer of such concepts takes place linguistically. Given that English has become the lingua franca in social media terminology, we would like to confirm in this paper that Anglicisms constitute the main translation procedure used when transferring social media terminology from English into Spanish and Brazilian Portuguese. In order to do this, a comparable corpus of newspaper articles on social media (Facebook and WhatsApp) written in English, Spanish and Brazilian Portuguese was compiled. With the aid of an automatic term extractor, TermoStat Web 3.0., the sample for analysis in Spanish and Brazilian Portuguese was selected. The terms were then classified into borrowings or calques according to their linguistic form. In addition, the use of alternative terms to name the same concept is checked and set in the context of English discourse. Conclusions are drawn from the results of this analysis, which could determine patterns in the use of Anglicisms when transferring social media terminology from English into Spanish and Brazilian Portuguese.

**Fu'ad Al-Qaisi, *Corpus et collocations : l'arme de complémentarité et d'interaction massives au profit de la traduction***

### **Résumé français**

Le traducteur, ce passeur d'idées, se trouve au carrefour des pensées, des cultures et des systèmes linguistiques. Si le sens est de toute évidence le but ultime de la traduction, l'intérêt stylistique qui lui est portée n'en demeure pas moins essentiel. S'appuyant sur la collocation, cette étude vise à démontrer l'importance des corpus dans la compréhension et dans la traduction. Pour ce faire, trois corpus (monolingue français, parallèle français-arabe, monolingue comparable arabe) sont exploités. Le respect de la convention langagière constitue une préoccupation majeure en langue cible. Une convention langagière est, pour schématiser, l'empreinte linguistique d'une spécialité et son timbre stylistique. La collocation fait partie des phénomènes qui influent sur la convention langagière. La traduction de la collocation rime parfois avec calque ou avec omission des contraintes combinatoires en langue cible lorsque la traduction par une collocation est possible. C'est dans ce cadre que le corpus semble être une aide intéressante pour le traducteur afin de conserver, au-delà du sens, la convention langagière d'une communauté donnée. Cette étude est d'autant plus utile que les dictionnaires français-arabe sont loin de combler le besoin de leurs utilisateurs dans ce domaine. Ainsi, l'on est en droit de s'interroger sur la capacité du corpus à aider à fournir une traduction conventionnelle en langue cible. La complémentarité inter-corpus (comparable, parallèle, etc.) saura-t-elle suffire aux

besoins du traducteur et de la traductologie en matière de convention ?

### **Abstract**

This paper is in reference to the long felt concern by translators (as purveyors of ideas) to be faithful to source and target languages in both meaning and style. We will focus particularly on collocations to show how corpora can play an important part in understanding phraseological units, and in helping translators as they move between them. We will use three corpora, which include French and Arabic (a French monolingual corpus, an Arabic comparable corpus, and a French-Arabic parallel corpus). Collocations in the source language may or may not have a natural collocational counterpart in the target language. When they do, small adjustments may be necessary, e.g. French singular vs. Arabic plural. In this context, we will argue that corpus studies can be a very useful tool to help the translator remain faithful to both the meaning and linguistic "naturalness", and avoid the danger of unnatural-sounding calques. Indeed, for some languages, such as Arabic, analysis at corpus level can help clarify the very existence of collocations, which are often not reported in dictionaries, and so may have gone unnoticed. We will also show how, through judicious use, our three corpora can be of service to translators even though the Arabic corpus is not morpho-syntactically "tagged". The paper is therefore a case study on how corpus studies can usefully meet the needs of translators and translation.

### **Vincenzo Lambertini, *Corpus, parémiologie et interprétation : un trinôme gagnant***

#### **Résumé français**

Quel est le rôle des corpus de langue générale dans l'interprétation ? Peut-on utiliser ces corpus dans les études sur l'interprétation ? Si on considère un domaine comme celui de la parémiologie, on se rend compte que les corpus de langue générale peuvent aider le chercheur à fournir des réponses pratiques aux professionnels de l'interprétation. Tout d'abord, avant de procéder à l'étude des proverbes liés à l'interprétation, il faut se pencher sur leurs caractéristiques linguistiques. Pour ce faire, il est incontournable de se baser sur des données authentiques. En particulier, c'est grâce à l'utilisation des données linguistiques authentiques qu'il est possible de mieux comprendre les façons dont les proverbes sont utilisés, voire d'analyser les fins qui sont poursuivies grâce à leur énonciation. Une fois établies les bases théoriques de la parémiologie, il est possible de passer à la deuxième phase de l'analyse, qui vise notamment à repérer des régularités au sein des proverbes. C'est grâce à ces observations qu'on peut mieux saisir les points principaux concernant la pratique des proverbes, afin de produire des considérations intéressantes pour l'interprète. De cette façon, nous sommes en mesure de démontrer non seulement que nombre de considérations couramment attribuées aux proverbes ne sont pas forcément exactes, mais surtout que ce sont les données textuelles qui indiquent ce qu'il faut tenir en considération lors de la traduction ou de l'énonciation des proverbes. Voilà, donc, qu'on démontrera l'importance des corpus écrits pour l'interprétation, notamment pour ce qui est de la phase de formation de l'interprète.

#### **Abstract**

What is the role of reference corpora in interpreting? Is it possible to use these corpora in interpreting studies? If we consider a field such as paremiology, we can understand that reference corpora may be very useful for researchers to produce studies with a practical function for interpreters. First of all, before studying proverbs when interpreting, it is necessary to analyse their linguistic characteristics by using authentic data. With this information in mind, we can understand how and why they are used. Once we have established the theoretical basis for paremiology, we can detect regularities in proverbs that highlight the most important points concerning their use, and especially some useful applications for an interpreter. In this way, we can show that many theories commonly ascribed to proverbs are not necessarily true and that textual data indicate what must be taken into account when translating or using proverbs. We will also demonstrate that the analysis of written corpora is crucial for interpreting, especially when training interpreters.