

## **Le corpus *Per-Fide* et ses applications: une étude de cas sur les expressions idiomatiques**

Sílvia Araújo, Sílvia Sacadura, Ana Correia

Université du Minho, Portugal

Section 16: Projets en cours; ressources et outils nouveaux

La valeur ajoutée des corpus alignés pour l'étude de langues particulières dans les différents domaines de la linguistique appliquée a été soulignée dans de nombreuses publications (Williams, 2005). Cependant, les corpus parallèles ayant le portugais comme langue source et/ou comme langue cible sont peu nombreux et ceux qui existent sont bien souvent limités au niveau du nombre de langues mises en regard et/ou au niveau des types de textes considérés.

Le projet *Per-Fide* (Araújo *et al.*, 2010) dont nous voulons rendre compte dans le cadre de cette communication trouve précisément son origine dans la nécessité de construire, avec des textes appartenant à différents domaines (littéraire, technico-scientifique, journalistique, juridique et religieux), des corpus alignés à la phrase, qui mettent parallèle des textes originaux en portugais dans ses différentes variantes (portugais européen, portugais brésilien, portugais africain) et leurs traductions dans six autres langues (espagnol, russe, français, italien, allemand et anglais), ainsi que des textes originaux dans ces six langues étrangères et leurs traductions vers le portugais. En rendant ces corpus disponibles et gratuitement accessibles ainsi que d'autres ressources multilingues, telles que des dictionnaires probabilistes de traduction (Simões, 2008) et de la terminologie bilingue, ce projet financé par la Fondation pour la Science et la Technologie (FCT) et par le Fonds Européen de Développement Régional (FEDER) tient, bien évidemment, à servir de matériau de base indispensable à un grand nombre d'applications en aide à la traduction, la lexicographie, la terminologie ou la didactique des langues.

Après une brève présentation du corpus multilingue *Per-Fide* qui met en valeur ce qui distingue ce corpus d'autres incluant également le portugais, nous tâcherons de montrer l'utilité de cet outil dans les domaines touchant à l'étude du langage et de la traduction sur la base d'un exemple concret, celui des expressions idiomatiques. Ces expressions caractérisées par le figement constituent un phénomène extrêmement complexe du point de vue formel et sémantique (Lamiroy, 2008; Mejri, 2008) qui présente un intérêt certain sur le plan contrastif, puisque, comme nous essayerons de le montrer, la plupart d'entre elles échappent à une traduction directe lors de la recherche d'équivalents (figés ou non) dans une autre langue. Il s'avère, par ailleurs, que ces expressions dont regorgent aussi bien les textes littéraires que les textes technico-scientifiques résistent généralement à une traduction automatique, comme en témoignent, par exemple, les traductions portugaise et anglaise (beaucoup trop littérales)

obtenues par le biais de *Google Traduction* pour l'expression idiomatique française *couper les cheveux en quatre*:

FR Monsieur le President, il ne faut pas *couper les cheveux en quatre*, comme disent les Français.

PT \*Sr. Presidente, *não cortar os cabelos*, como dizem os franceses.

EN \*Mr. President, *do not cut the hairs*, as the French say.

Il est donc important de motiver l'intégration des corpus et des outils d'analyse de corpus dans l'enseignement de la traduction (outre celui de la linguistique et des langues) car le recours à ce type de ressources améliore la qualité des traductions par rapport à la seule utilisation de dictionnaires, notamment au niveau du choix des termes et de l'utilisation d'expressions idiomatiques (Frerot, 2010). Comme le note, à juste titre, Kraif (2006), les outils d'alignement et de concordance bilingue sont encore largement sous-exploités. Familiariser les utilisateurs — traducteurs, étudiants ou linguistes — avec les principaux concepts-clés du domaine et avec les techniques utilisées pour exploiter des corpus (Araújo *et al.*, 2011) se révèle donc essentiel.

## Références bibliographiques

- Araújo, S., Almeida, J. J., Dias, I., Simões, A. (2010) «Apresentação do projecto Per-Fide: Paralelizando o Português com seis outras línguas». *Linguamática*, v2 n2, pp. 71-74.
- Araújo, S., Oliveira, A. Dias, I., (2011) «Como pesquisar em corpora». Workshop présenté aux *I Journées Internationales Corpora & Traduction*, 22 et 23 novembre 2011, Université du Minho, <http://perfide.ilch.uminho.pt/site.pl/workshop.pt>.
- Frerot, C. (2010) «Outils d'aide à la traduction: pour une intégration des corpus et des outils d'analyse de corpus dans l'enseignement de la traduction et la formation des traducteurs», *Les Cahiers du GEPE*, N°2. Outils de traduction - outils du traducteur?
- Kraif, O. (2006) «Qu'attendre de l'alignement de corpus multilingues?». *Revue Traduire. 4<sup>e</sup> Journée de la traduction professionnelle*, n° 210, pp. 17-37.
- Lamiroy, B. (2008) «Les expressions figées: à la recherche d'une définition». In P. Blumenthal & S. Mejri (eds) *Les séquences figées: entre langue et discours*. Zeitschrift für Französische Sprache und Literatur, Beihefte 36, pp. 85-98.
- Mejri, S. (2008) «Figement et traduction: problématique générale». *Meta: journal des traducteurs*, Volume 53, numéro 2, pp. 244-252.
- Simões, A. (2008) *Extracção de recursos de tradução com base em dicionários probabilísticos de tradução*. Thèse de Doctorat en Informatique, Université du Minho.
- Williams, G. (ed.) (2005) *La Linguistique de corpus*. Rennes: Presses Universitaires de Rennes.