Is phraseology the third articulation of language? Fresh insights into a theoretical conundrum

Jean-Pierre Colson
University of Louvain (Louvain-la-Neuve, Belgium)

Although the notion of *phraseology* is now used across a wide range of linguistic disciplines, its definition and the classification of *phraseological units* remain a subject of intense debate. It is generally agreed that phraseology implies *polylexicality*, but this term is problematic as well, because it brings us back to one of the most controversial topics in modern linguistics: the definition of a *word*.

On the other hand, another widely accepted principle of language is the double articulation or duality of patterning (Martinet 1960): the first articulation consists of morphemes and the second of phonemes. The very definition of *morphemes*, however, also poses several problems, and the situation becomes even more confused if we wish to take phraseology into account.

In this contribution, I will take the view that a corpus-based and computational approach to phraseology may shed some new light on this theoretical conundrum. A better understanding of the basic units of meaning is necessary for more efficient language learning and translation, especially in the case of machine translation. Previous research (Colson 2011, 2012, 2013, 2014), Corpas Pastor (2000, 2007, 2008, 2013, 2015), Corpas Pastor & Leiva Rojo (2011), Leiva Rojo (2013), has shown the paramount importance of phraseology for translation.

A tentative step towards a coherent explanation of the role of phraseology in language has been proposed by Mejri (2006): it is postulated that a third articulation of language intervenes at the level of words, including simple morphemes, sequences of free and bound morphemes, but also phraseological units.

I will present results from experiments with statistical associations of morphemes across several languages, and point out that (mainly) isolating languages such as Chinese are interesting for a better understanding of the interplay between morphemes and phraseological units. Named entities, in particular, are an extreme example of intertwining cultural, statistical and linguistic elements. Other examples show that the many borrowings and influences that characterize European languages tend to give a somewhat blurred vision of the interplay between morphology and phraseology.

From a statistical point of view, the *cpr-score* (Colson 2016) provides a methodology for adapting the automatic extraction of phraseological units to the morphological structure of each language. The results obtained can therefore be used for testing hypotheses about the interaction between morphology, phraseology and culture.

Experiments with the *cpr-score* on the extraction of Chinese phraseological units show that results depend on how the basic units of meaning are defined: a morpheme-based approach yields good results, which corroborates the claim by Beck and Mel'čuk (2011) that the association of morphemes into words may be similar to the association of words into

phraseological units. A cross-linguistic experiment carried out for English, French, Spanish and Chinese also reveals that the results are quite compatible with Mejri's hypothesis (2006) of a *third articulation of language*.

Such findings, if confirmed, also corroborate the notion of statistical semantics in language. To illustrate this point, I will present the *PhraseoRobot* (Colson 2016), a computational tool for extracting phraseological associations around key words from the media, such as *Brexit*. The results confirm a previous study on the term *globalization* (Colson 2016): a significant part of sociolinguistic associations prevailing in the media is related to phraseology in the broad sense, and can therefore be partly extracted by means of statistical scores.

References

Beck, D. & I. Mel'čuk (2011). Morphological phrasemes and Totonacan verbal morphology. *Linguistics* 49/1: 175-228.

Colson, J.-P. (2011). La traduction spécialisée basée sur les corpus : une expérience dans le domaine informatique. In : Sfar, I. & S. Mejri*, La traduction de textes spécialisés : retour sur des lieux communs.* Synergies Tunisie n° 2. Gerflint, Agence universitaire de la Francophonie, p. 115-123.

Colson, J.-P. (2012). Traduire le figement en langue de spécialité : une expérience de phraséologie informatique. In : Mogorrón Huerta, P. & S. Mejri (dirs.), *Lenguas de especialidad, traducción, fijación / Langues spécialisées, figement et traduction.* Encuentros Mediterráneos / Rencontres Méditerranéennes, N°4. Universidad de Alicante, p. 159-171.

Colson, J.-P. (2013). Pratique traduisante et idiomaticité : l'importance des structures semi-figées. In : Mogorrón Huerta, P., Gallego Hernández, D., Masseau, P. & Tolosa Igualada, M. (eds.), *Fraseología, Opacidad y Traduccíon. Studien zur romanischen Sprachwissenschaft und interkulturellen Kommunikation* (Herausgegeben von Gerd Wotjak). Frankfurt am Main, Peter Lang, p. 207-218.

Colson, J.-P. (2014). La phraséologie et les corpus dans les recherches traductologiques. Communication lors du colloque international Europhras 2014, Association Européenne de Phraséologie. Université de Paris Sorbonne, 10-12 septembre 2014.

Colson, J-P. (2016). Set phrases around globalization : an experiment in corpus-based computational phraseology. In: F. Alonso Almeida, I. Ortega Barrera, E. Quintana Toledo and M. Sánchez Cuervo (eds.), *Input a Word, Analyse the World: Selected Approaches to Corpus Linguistics.* Newcastle upon Tyne: Cambridge Scholars Publishing, p. 141-152.

Corpas Pastor, G. (2000). Acerca de la (in)traducibilidad de la fraseología. In: G. Corpas Pastor (ed*.), Las lenguas de Europa: Estudios de fraseología, fraseografía y traducción.* Granada: Comares, p. 483-522.

Corpas Pastor, G. (2007). Europäismen - von Natur aus phraseologische Äquivalente? Von blauem Blut und sangre azul. In: M. Emsel y J. Cuartero Otal (eds.), *Brücken: Übersetzen und interkulturelle Kommunikationen.* Festschrift für Gerd Wotjak zum 65. Geburtstag, Fráncfort: Peter Lang, p. 65-77.

Corpas Pastor, G. (2008). *Investigar con corpus en traducción: los retos de un nuevo paradigma* [Studien zur romanische Sprachwissenschaft und interkulturellen Kommunikation, 49], Fráncfort: Peter Lang.

Corpas Pastor, G. (2013). Detección, descripción y contraste de las unidades fraseológicas mediante tecnologías lingüísticas. In Olza, I. & R. Elvira Manero (eds.) *Fraseopragmática.* Berlin: Frank & Timme, p. 335-373.

Leiva Rojo, J. (2013). La traducción de unidades fraseológicas (alemán-español/español-alemán) como parámetro para la evaluación y revisión de traducciones. In: Mellado Blanco, C., Buján, P, Iglesias N.M., Losada M.C. & A. Mansilla (eds), *La fraseología del alemán y el español: lexicografía y traducción. ELS, Etudes Linguistiques / Linguistische Studien, Band 11*. München: Peniope, p. 31-42.

Leiva Rojo, J. & G. Corpas Pastor (2011). Placing Italian idioms in a foreign milieu: a case study. In: Pamies Bertrán, A., Luque Nadal, L., Bretana, J. &; M. Pazos (eds), (2011). *Multilingual phraseography. Second Language Learning and Translation Applications.* Baltmannsweiler: Schneider Verlag (Colección: Phraseologie und Parömiologie, 28), p. 289-298.

Martinet, A. (1966). Eléments de linguistique générale. Paris: Colin.

Mejri, S. (2006). Polylexicalité, monolexicalité et double articulation. *Cahiers de Lexicologie* 2: 209-221.