À la croisée des corpus et de la phraséologie : une proposition d'outil informatique

At the crossroads of corpora and phraseology: a proposal for a computer tool

Jean-Pierre Colson

Abstract: Corpora are currently enjoying ever-increasing success, and are no longer solely the domain of corpus linguists. They provide us with a picture of language as described by John Sinclair: on average about half of any text consists of phraseological units in the broadest sense. It would therefore be unrealistic to teach foreign languages without taking this aspect into account.

However, the exploitation of corpora for the didactics of phraseology is fraught with obstacles: the frequency of phraseological units is low, and huge corpora are necessary in order to guarantee sufficient numbers of relevant examples; simple concordances of phraseological units are sometimes contradictory or hard to interpret; finally, the available tools for manipulating such vast collections of texts are not always user-friendly, and require that learners should have a flair for computing and/or statistics.

In this paper we propose new techniques to build on a notion introduced by Michael Lewis in 1993: phraseological awareness raising, based on recurrent comparison with corpus examples. We also present an experimental tool, the *IdiomSearch* application (J. P. Colson 2016), that provides learners with immediate feedback on the presence of phraseological units in an input text, as well as a ranking on the basis of frequency and fixedness.

Key words: phraseology, foreign language teaching, corpora, statistical semantics, collocations.