

## STRUCTURAL PATTERNS OF AMBIGUOUS LEXICAL BUNDLES IN FORMULAIC ACADEMIC LANGUAGE

Andressa Gomide\*, Carolina Bohórquez\*, Luciana Aguiar de Oliveira\* e Deise P. Dutra\* carolinaboho@gmail.com

\*Universidade Federal de Minas Gerais, Faculdade de Letras, Linguística Aplicada, Ensino/Aprendizagem de Língua Estrangeira 01/03/2012 à 01/03/2013

**ABSTRACT** Research of high frequent recurrent sequences of words, or lexical bundles (Biber et al, 1999) has demonstrated the formulaic trade of language and the importance of including this kind of knowledge as part of the EFL/ ESL curriculum (Bal, 2010). Biber (2004) created functional-pragmatic classification taxonomies for lexical bundles that were, later, expanded by Simpson-Vlach (2010). The main purpose of this classification is to serve for pedagogical use in academic speech and writing and as a consequence, a demand for ways of classifying these units has become evident. This research aims at clarifying ambiguous cases of bundles classification by exploring their larger context of use in order to satisfy the mentioned demand. To conduct this research, we chose to use as reference corpora the Brazilian Portuguese Sub-corpus of ICLE (Br-ICLE) - a corpus with almost 159,000 words that is still being compiled and is composed of argumentative essays written by Brazilian undergraduate students - and the Louvain Corpus of Native English Essay (LOCNESS) – built with academic texts from American and British students and composed of 324,006 words. We have chosen LOCNESS as a reference corpus and Br-ICLE as a representative corpus of learners of English. Our work is developed in the following steps. Step one: each researcher will, individually, classify 7587 bundles of 3 and 4 words that occur in Br-ICLE and 5357 bundles of 3 and 4 words that occur in LOCNESS according to the Academic Formulas List (Simpson-Vlach, R. & Ellis, N. 2010), all of which with a normalized frequency equal or higher than 12,345. Step two: the three researchers will compare / contrast their results. Classification which do not present total agreement between researchers and in which double classification is found possible will be selected for further analysis. Step three: we will examine the context of occurrence of the ambiguous cases (cases in which total agreement was not reached) checking for structural patterns that would help to better describe the occurrences of these bundles. We expect that the structural patterns of occurrences can contribute for an additional manner of considering bundles other than their functions when the same are borderline cases. We also expect that these patterns of occurrences may be used to assist the development of automatic taggers, improving their level of accuracy. In addition, the classification of the bundles of the two chosen corpora can also enable future studies aiming at comparing / contrasting native production to Brazilian learners production and inform curriculum development. In the present conference we will show the results of steps one and two (above mentioned) and the structural analysis of the context of about 20 ambiguous / borderline bundles.

KEYWORDS: lexical bundles classification, academic writing, Corpus Linguistics

## REFERENCES

BAL, B. Analysis of Four-word Lexical Bundles in Published Research Articles Written by Turkish Scholars. **Applied Linguistics and English as a Second Language Theses**, Georgia State University, paper 2, 2010.

BIBER, D.; JOHANSON, S.; LEECH, G.; CONRAD, S.; FINEGAN, E. Longman Grammar of Spoken and Written English. Essex: Longman, 1999.

BIBER, D.; CONRAD, S; CORTES, V. If you look at...: Lexical Bundles in University Teaching and Textobooks. **Applied Linguistics**, Oxford University Press, v.25, n.3, p.371-405, 2004.

SIMPSON-VLACH, R., ELLIS, N. An academic formulas list: New methods in phraseology research. **Applied Linguistics**, University of Michigan, v. 31, n. 4, p. 1-26, 2010.