

Granger, S. (2012). Towards a phraseological model of lexical proficiency: Evidence from learner corpus research. In *Proceedings of the Applied Linguistics Association of Korea Conference (ALAK 2012)*, October 2012, Seoul (Korea), 10-15.

Towards a phraseological model of lexical proficiency Evidence from learner corpus research

**Sylviane Granger
(Université catholique de Louvain)**

I. Introduction

Under the impetus of John Sinclair phraseology has acquired a much more central place in linguistics and language studies in general than was ever the case before. Sinclair's (1991) 'idiom principle', which underscores the prefabricated nature of language, has brought phraseology to the forefront and considerably extended its scope. The combined use of large corpora and powerful corpus analysis techniques has resulted in the uncovering of a much wider range of phraseological units than had been investigated in traditional studies of phraseology (the so-called "Russian tradition", cf. Cowie 1998). One major corpus finding is that the types of units that had traditionally been focused on are relative rare (cf. Moon 1998). As a result, interest progressively shifted to other types of units, which display a higher degree of semantic compositionality, in particular collocations and lexical bundles.

From the start the relevance of this new approach to phraseology for foreign language teaching has been highlighted, by Sinclair himself and by several scholars inspired by him (Nattinger & DeCarrico 1992, Lewis 1993 and many others). However, most studies tended to be based exclusively on native speaker data (cf. Barfield's review of Lewis 2000 in which he deplores that "the voices of typical language learners are largely omitted in this volume") and this is not a sufficient base to design efficient pedagogical resources. Native speaker data supply information on what is typical in English but have nothing to say on the degree of difficulty of phraseological patterns for language learners. It is therefore essential to combine the information about typicality provided by native corpora with information on degree of difficulty gained from the careful analysis of learner corpora, i.e. electronic collections of authentic language data produced by foreign or second (L2) language learners. The aim of my presentation is to give an overview of learner-corpus-based studies of the learner phrasicon, link them up to the current situation in materials design and vocabulary assessment, and provide directions for the future.

II. Corpus-based investigations of the learner phrasicon

Collocations and lexical bundles have been studied extensively in learner corpus research.

Collocations are usage-determined syntagmatic relations between two lexemes in a specific syntactic pattern (e.g. adjective + noun: *heavy rain*; adverb + adjective: *vitally important*; verb + adverb: *apologize profusely*; verb + noun: *to file a complaint*). Both lexemes make an isolable semantic contribution to the word combination but they do not have the same status: the 'base' of the collocation is selected first by a language user for its independent meaning, while the second element, i.e. the 'collocate' or 'collocator', is selected by and semantically dependent on the 'base' (Granger & Paquot 2008). Lexical bundles are the most frequent recurring sequences of words in a register (Biber et al 1999: ch. 13). A large number of studies, mostly focused on academic settings, have highlighted the major role that these prefabricated units play in discourse. Biber et al (2004) distinguishes between referential bundles which make direct reference to physical or abstract entities (*a lot of people, in the United States*), discourse organizers which reflect relationships between prior and coming discourse (*with this in mind, this is why*) and stance bundles which express attitude or assessment of certainty (*I don't want to, it is possible to*).

Learner-corpus-based studies of collocations have focused more particularly on verb + object patterns, with special focus on light verb constructions (e.g. Liu 2002, Nesselhauf 2005) and adverb + adjective patterns, with special focus on intensifying adverbs (e.g. Granger 1998, Lorenz 1999, Lee 2006). The difficulties that learners prove to have with collocations are two-fold:

(1) misuse: learners often choose incorrect collocates. Many of these errors are due to transfer from the learners' mother tongue (L1) (cf. Nesselhauf 2005: 48%, Warner et al in press: 67%).

(2) over- and underuse: learners tend to use default 'all-round' collocates (e.g. the intensifier 'very' instead of collocationally restricted ones like *vitally* or *closely*). This finding points to a lack of lexical diversity which should be catered for in targeted lexical expansion exercises.

Lexical bundles have also been the topic of many learner-corpus-based studies (De Cock 2000, Allen 2009, Juknevičienė 2009, Ping 2009, Chen & Baker 2010, Tribble 2011). In this case learner difficulties do not result in downright errors, but rather in clumsy formulations and lack of adherence to register conventions, in particular in academic writing. Here too L1 influence is clearly noticeable with many organizational and stance markers directly calqued from the learners' L1. Learners also prove to have little awareness of the register specificity of lexical bundles and tend to overuse speech-like bundles in formal writing.

III. Pedagogical integration of learner corpus findings

Both collocation and lexical bundle range and accuracy are strong indicators of L2 proficiency (cf. Crossley & Salisbury 2011, Granger & Bestgen in preparation). It is therefore essential to give them a prominent place in foreign language learning and teaching. This is especially important as these types of units will not simply be picked up by learners. Krishnamurthy (2002) rightly points out that a chunk is

“very difficult to notice if somebody hasn’t drawn your attention to it”. In the second part of my presentation I will investigate whether phraseological insights gained from the analysis of native and learner corpora have been implemented in vocabulary assessment and materials design.

Vocabulary assessment is still largely single-word-based. Vocabulary tests tend to rely on lists of single words or word families (i.e. base words + inflectional and/or derivational forms) and include very few multiword units, if any (cf. e.g. Nation & Laufer’s levels test¹ and Cobb’s VocabProfile²). The impact of corpora on *materials design* has been uneven. Dictionaries have undeniably benefited most from native and learner corpus insights. Monolingual learners’ dictionaries of English in particular have been significantly enhanced by the addition of collocation boxes and error notes designed to meet learners’ attested difficulties. However, the usefulness of these new features is limited by the fact that most of these dictionaries are generic, i.e. target all categories of learners whatever their mother tongue background, and inclusion of L1-specific information is therefore ruled out. In addition, lexical bundles are still largely absent from dictionaries or difficult to access. Compared to dictionaries, the impact on coursebooks is still quite limited. The coursebooks that are explicitly corpus-informed tend to incorporate findings from native rather than learner corpora and as a result, contain very few exercises that cater for learners’ authentic difficulties with phraseology. One type of resource which has allowed corpus research - and in particular learner corpus research - to show its full potential is the electronic writing aid. Some corpus-informed applications include *Check My Words*³ (Milton 2006) and *CARE (Concordancer of Academic wRitten English)*⁴ (Hsieh & Liou 2008). In the final part of my presentation I will describe the *Louvain English for Academic Purposes Dictionary (LEAD)* which aims to help non-native writers write academic texts (Granger & Paquot 2010). *LEAD* is fully corpus-based: it relies on a large corpus of academic texts written by professional writers and a large learner corpus, the *International Corpus of Learner English* (Granger et al 2009), which contains texts written by learners from 16 different mother tongue backgrounds. It provides information on the words that are typical of academic discourse and pose difficulty to learners mainly in terms of phraseology (collocations and lexical bundles) but also in terms of meaning, frequency, register, grammar and punctuation. It is a hybrid tool that combines the functions of a dictionary, a writing aid, a CALL program and a corpus tool. One of the most innovative features of the tool is that it can be customized in function of learners’ mother tongue and discipline.

IV. Directions for the future

¹ <http://www.er.uqam.ca/nobel/r21270/levels/>

² <http://www.lextutor.ca/vp/>

³ <http://www.compulang.com/cmw/>

⁴ <http://candle.fl.nthu.edu.tw/care/>

In spite of the active research that has been carried out in phraseology in recent years, the lexical model that still dominates foreign language learning and teaching is still largely a single-word-based model. Findings from both native and learner corpus research strongly suggest that we should move away from this model and give multiword lexis a place on a par with – if not greater than – that afforded to single words. Careful corpus-based investigations of the learner phrasicon hold great potential for L2 learning and teaching. One of the main lessons we can draw from studies of the learner phrasicon is that pedagogical resources need to be at least partly customized to the learners' mother tongue. This is essential as a large proportion of learner difficulties is transfer-related. Electronic tools are clearly the way forward as they afford various ways of integrating the learner's mother tongue. The data are there: it is now up to researchers and publishers to work hand in hand/join forces to translate this view into action.

References

- Allen, D. (2009). Lexical Bundles in Learner Writing: An Analysis of Formulaic Language in the ALESS Learner Corpus. *Komaba Journal of English Education*, Vol.1, 105-127. <http://park.itc.u-tokyo.ac.jp/eigo/KJEE/001/105-127.pdf>
- Barfield, A. (2001). Review of M. Lewis (ed.) "Teaching Collocation: Further Developments in the Lexical Approach. *English Language Teaching* 55(4), 413-415.
- Biber, D., Conrad, S. & Cortes, V. (2004). If you look at ... Lexical bundles in university lectures and textbooks. *Applied Linguistics* 25, 2004, 371-405.
- Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education Ltd.
- Chen, Y.-H. & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology*, 14(2), 30-49.
- Cowie, AP. (1998). Introduction. In Cowie, A.P. (ed.) *Phraseology: Theory, Analysis, and Applications*. Clarendon Press, 1-20
- Crossley, S. & Salsbury, T.L. (2011). The development of lexical bundle accuracy and production in English second language speakers. *IRAL - International Review of Applied Linguistics in Language Teaching* 49(1), 1-26.
- De Cock, S (2000). Repetitive phrase chunkiness and advanced EFL speech and writing. In C. Mair & M. Hundt (eds) *Corpus Linguistics and Linguistic Theory: Papers from the Twentieth International Conference on English Language Research on Computerized Corpora*. Amsterdam: Rodopi, 51-68.
- Granger, S. (1998). Prefabricated patterns in advanced EFL writing: collocations and formulae. In A. Cowie (ed.) *Phraseology: Theory, Analysis and Applications*. Oxford: Oxford University Press, 145-160.
- Granger, S. & Bestgen, Y. (in preparation). The use of collocations by intermediate vs. advanced non-native writers: A bigram-based study.

- Granger, S., Dagneaux, E., Meunier, F., Paquot, M. (2009). *The International Corpus of Learner English*. Handbook and CD-ROM. Version 2. Louvain-la-Neuve: Presses universitaires de Louvain.
- Granger S. and Paquot M. (2008) Disentangling the phraseological web. In Granger S. and F. Meunier (eds) *Phraseology: An Interdisciplinary Perspective*. Amsterdam & Philadelphia: Benjamins, 27-49.
- Granger, S. & Paquot, M. (2010). Customising a general EAP dictionary to meet learner needs. In Granger, S. & Paquot, M. (eds.) *eLexicography in the 21st century: New challenges, new applications*. Louvain-la-Neuve, Presses universitaires de Louvain, 87-96.
- Hsieh, W.-M. & Liou, H.-C. (2008). A Case Study of Corpus-Informed Online Academic Writing for EFL Graduate Students. *CALICO Journal* 26 (1), 28-47.
- Juknevičienė, R. (2009). Lexical bundles in learner language: Lithuanian learners vs. native speakers. *KaLBOTYRa* 61/3, 61-72.
- Krishnamurthy, R. (2002). Language as chunks, not words. *JALT2002 Proceedings*, 288-294. <http://jalt-publications.org/archive/proceedings/2002/288.pdf>
- Lee, S. (2006). A corpus-based analysis of Korean EFL learners' use of amplifier collocations. *English Teaching* 61(1), 3-17.
- Lewis, M. (1993). *The Lexical Approach: The State of ELT and a Way Forward*. Hove: Language Teaching Publications.
- Lewis, M. (ed.) (2000). *Teaching Collocation: Further Developments in the Lexical Approach*. Language Teaching Publications.
- Liu, L. E. (2002). A corpus-based lexical semantic investigation of verb-noun miscollocations in Taiwan learners' English. Unpublished master's thesis, Tamkang University, Taipei.
- Lorenz, G. (1999). *Adjective intensification - Learners versus native speakers. A corpus study of argumentative writing*. Amsterdam & Atlanta: Rodopi.
- Milton, J. (2006). Resource-Rich Web-Based Feedback: helping learners become independent writers. In Hyland, K & Hyland F. (eds), *Feedback in Second Language Writing: Contexts and Issues*. Cambridge University Press, 123-137.
- Moon, R. (1998). *Fixed Expressions and Idioms in English: a Corpus-based Approach*. Oxford: Oxford University Press.
- Nattinger, J.S. & DeCarrico, J.R. (1992). *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press.
- Nesselhauf, N. (2005). *Collocations in a Learner Corpus*. Amsterdam: John Benjamins.
- Ping, P. (2009) A study on the use of four-word lexical bundles in argumentative essays by Chinese English majors – A comparative study based on WECCCL and LOCNESS. *CELEA* 32 (3), 25-45. <http://www.celea.org.cn/teic/85/85-25.pdf>
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Tribble, C. (2011). Revisiting apprentice texts. Using lexical bundles to investigate expert and apprentice performances in academic writing. In: Meunier, F., De Cock, S., Gilquin, G. & Paquot, M. (eds.), *A Taste for Corpora: In honour of Sylviane Granger*, 85–108.

Wanner, L., Alonso Ramos, M., Vincze, O., Nazar, R., Ferraro, G., Mosqueira, E. & Prieto, G. (in press). Annotation of collocations in a learner corpus for building a learning environment. In Granger, S., Gilquin, G. & Meunier, F. (eds.) *Proceedings of the Learner Corpus Research 2011 Conference*. Presses universitaires de Louvain: Louvain-la-Neuve.

Author

Sylviane Granger is Professor of English Language and Linguistics at the University of Louvain (Belgium). She is the director of the Centre for English Corpus Linguistics, where research activity is focused on the compilation and exploitation of learner corpora and bilingual corpora. In 1990 she launched the *International Corpus of Learner English (ICLE)* project, which has grown to contain learner writing by learners of English from a wide range of mother tongue backgrounds and is the result of collaboration from a large number of universities internationally. Since then she has played a key role in defining the different facets of the field of learner corpus research. Her current research interests focus on issues of learner corpus design and annotation, the analysis of phraseology in learner language, the integration of learner corpus insights into electronic writing aids and the improvement of bilingual dictionaries afforded by the combined use of learner and bilingual data. She can be contacted at sylviane.granger@uclouvain.be