

2010

# eLexicography in the 21<sup>st</sup> Century: New Challenges, New Applications

Proceedings of eLex 2009, Louvain-la-Neuve, 22-24 October 2009

*Sylviane Granger and Magali Paquot (eds)*

UCL PRESSES  
UNIVERSITAIRES  
DE LOUVAIN

## **Cahiers du CENTAL**

### **Comité scientifique**

|                       |  |
|-----------------------|--|
| Sergio Bolasco        | Università di Roma « La Sapienza », IT               |
| Laurence Danlos       | Université Paris 7, FR                               |
| Guy Deville           | Facultés universitaires Notre-Dame de la Paix, BE    |
| Ray Dougherty         | University of New York, US                           |
| Thierry Dutoit        | Faculté Polytechnique de Mons, BE                    |
| Louissette Emirkanian | Université du Québec à Montréal, CA                  |
| Cédric Fairon         | Université catholique de Louvain, BE                 |
| Thierry Fontenelle    | Centre de traduction de la Communauté européenne, LU |
| Sylviane Granger      | Université catholique de Louvain, BE                 |
| Guy Lapalme           | Université de Montréal, CA                           |
| Eric Laporte          | Université de Marne-la-Vallée, FR                    |
| Denis Maurel          | Université François-Rabelais de Tours, FR            |
| Damon Mayaffre        | Université de Nice – Sophia Antipolis, FR            |
| Sébastien Paumier     | Université de Marne-la-Vallée, FR                    |
| Alain Polguère        | Université de Montréal, CA                           |
| Antoinette Renouf     | University of Central England in Birmingham, UK      |
| Jean Senellart        | Systran, FR  |
| Anne-Catherine Simon  | Université catholique de Louvain, BE                 |
| Agnès Tutin           | Université Stendhal – Grenoble 3, FR                 |
| Pierre Zweigenbaum    | CHU Pitié-Salpêtrière, Université Paris 6, FR        |

### **Comité de rédaction**

|                           |  |
|---------------------------|--|
| Anne Dister               | CENTAL, Université catholique de Louvain, BE |
| Cédric Fairon (directeur) | CENTAL, Université catholique de Louvain, BE |

### **Secrétariat de rédaction**

|                     |   |
|---------------------|---|
| Bernadette Dehottay | CENTAL, Université catholique de Louvain, BE<br>Place Blaise Pascal, 1 – B-1348 Louvain-la-Neuve<br>tél. : +32 10 47 37 86 – fax : +32 10 47 26 06<br>email : cental@tedm.ucl.ac.be – <a href="http://www.uclouvain.be/cental">http://www.uclouvain.be/cental</a> |
|---------------------|---|

Membre de l'Association des Revues Scientifiques et Culturelles – A.R.S.C. (<http://www.arse.be>)

Graphisme de la couverture Olivier Vereecken <http://aphine.com>

© Presses universitaires de Louvain, 2010

Dépôt légal : D/2010/9964/10  
ISBN 978-2-87463-211-2  
ISSN : 1783-2845

Imprimé en Belgique

Tous droits de reproduction, d'adaptation ou de traduction, par quelque procédé que ce soit, réservés pour tous pays, sauf autorisation de l'éditeur ou de ses ayants droit.

*Diffusion :*

[www.i6doc.com](http://www.i6doc.com) l'édition universitaire en ligne

*Sur commande en librairie ou à :*

Diffusion universitaire CIACO  
Grand-Place, 7  
1348 Louvain-la-Neuve, Belgique  
Tél. +32 10 47 33 78 - Fax +32 10 45 73 50  
[duc@ciaco.com](mailto:duc@ciaco.com)

*Diffusion pour la France :*

Librairie Wallonie-Bruxelles  
46 rue Quincampoix  
75004 Paris  
Tél. +33 1 42 71 58 03 - Fax +33 1 42 71 58 09  
[libwubr@club-internet.fr](mailto:libwubr@club-internet.fr)

# Table of Contents

|  |          |
|--|----------|
| Table of Contents .....  | v        |
| Acknowledgements .....   | ix       |
| <b>Papers</b> .....  | <b>1</b> |
| Andrea ABEL, <i>Towards a systematic classification framework for dictionaries and CALL</i> .....  | 3        |
| James BREEN, <i>Identification of neologisms in Japanese by corpus analysis</i> .....  | 13       |
| Annelen BRUNNER, Kathrin STEYER, <i>Wortverbindungsfelder – Fields of multi-word expressions</i> .....   | 23       |
| Louise-Amélie COUGNON, Richard BEAUFORT, <i>SSLD: a French SMS to Standard Language Dictionary</i> .....   | 33       |
| Patrick DROUIN, <i>Extracting a bilingual transdisciplinary scientific lexicon</i> .....   | 43       |
| Isabel DURÁN-MUÑOZ, <i>Specialised lexicographical resources: a survey of translators' needs</i> .....   | 55       |
| Carolina FLINZ, <i>DIL: a German-Italian online specialized dictionary of linguistics</i> .....  | 67       |
| Nuria GALA, Véronique REY, <i>Acquiring semantics from structured corpora to enrich an existing lexicon</i> .....  | 77       |
| Sylviane GRANGER, Magali PAQUOT, <i>Customising a general EAP dictionary to meet learner needs</i> .....   | 87       |
| Antton GURRUTXAGA, Igor LETURIA, Eli POCIELLO, Xabier SARALEGI, Iñaki SAN VICENTE, <i>Evaluation of an automatic process for specialized web corpora collection and term extraction for Basque</i> ..... | 97       |
| Patrick HANKS, <i>Elliptical arguments: a problem in relating meaning to use</i> .....   | 109      |
| Thomas HERBST, Peter UHRIG, <i>Valency information online – research and pedagogic reference tools</i> .....   | 125      |
| Kristina HMELJAK SANGAWA, Tomaž ERJAVEC, Yoshiko KAWAMURA, <i>Automated collection of Japanese word usage examples from a parallel and a monolingual corpus</i> .....                                    | 137      |
| Olga KARPOVA, Mikhail GORBUNOV, <i>Cultural values in a learner's dictionary: in search of a model</i> .....   | 149      |
| Annette KLOSA, <i>On the combination of automated information and lexicographically interpreted information in two German online dictionaries...</i>   | 157      |

|   |     |
|---|-----|
| Jette Hedegaard KRISTOFFERSEN, Thomas TROELSGÅRD, <i>Making a dictionary without words: lemmatization problems in a sign language dictionary</i> .....  | 165 |
| Vincent LANNOY, <i>Free online dictionaries: why and how?</i> .....   | 173 |
| Godelieve LAUREYS, <i>The Hub&amp;Spoke Model put into practice. Report on the semi-automatic extraction of a pre-version of a Finnish-Danish dictionary from a multilingual interlinkable database</i> ..... | 183 |
| Robert LEW, Patryk TOKAREK, <i>Entry menus in bilingual electronic dictionaries...</i>  | 193 |
| Marie-Claude L'HOMME, <i>Designing specialized dictionaries with natural language processing. Examples of applications in the fields of computing and climate change</i> .....                                | 203 |
| Hanhong LI, <i>Word frequency distribution for electronic learner's dictionaries</i> ...  | 217 |
| Marc LUDER, <i>Building an OLIF-based lexical database for representing constructions</i> .....   | 229 |
| Cédric MESSIANT, Thierry POIBEAU, <i>Automatic lexical acquisition from corpora: some limitations and tentative solutions</i> .....   | 241 |
| Mojca PECMAN, Claudie JULLIARD, Natalie KÜBLER, Alexandra VOLANSCHI, <i>Processing collocations in a terminological database based on a cross-disciplinary study of scientific texts</i> .....                | 249 |
| Bálint SASS, Júlia PAJZS, <i>FDVC – Creating a corpus-driven frequency dictionary of verb phrase constructions for Hungarian</i> .....  | 263 |
| Stefania SPINA, <i>The Dici project: towards a dictionary of Italian collocations integrated with an online language learning platform</i> .....  | 273 |
| Daniela TISCORNIA, <i>An ontology-based approach to the multilingual complexity of law</i> .....  | 283 |
| Sabine TITTEL, <i>Dynamic access to a static dictionary: a lexicographical “cathedral” lives to see the twenty-first century – the Dictionnaire étymologique de l'ancien français</i> .....                   | 295 |
| Lars TRAP-JENSEN, <i>Access to multiple lexical resources at a stroke: integrating dictionary, corpus and Wordnet data</i> .....  | 303 |
| Agnès TUTIN, <i>Showing phraseology in context: onomasiological access to lexico-grammatical patterns in corpora of French scientific writings</i> .....  | 313 |
| Isabel VERDAGUER, Elisabet COMELLES, Natàlia J. LASO, Eva GIMÉNEZ, Danica SALAZAR, <i>SciE-Lex : an electronic lexical database for the Spanish medical community</i> .....                                   | 325 |
| Serge VERLINDE, <i>The Base lexicale du français: a multi-purpose lexicographic tool</i> .....  | 335 |
| Alexandra VOLANSCHI, Natalie KÜBLER, <i>Building an electronic combinatory dictionary as a writing aid tool for researchers in biology</i> .....  | 343 |
| Michael ZOCK, Tonio WANDMACHER, Ekaterina OVCHINNIKOVA, <i>Are vector-based approaches a feasible solution to the “tip-of-the-tongue” problem?</i> .....  | 355 |

|   |     |
|---|-----|
| <b>Posters and software demonstrations</b> .....  | 367 |
| Margarita ALONSO RAMOS, Alfonso NISHIKAWA, Orsolya VINCZE, <i>DiCE in the web. An online Spanish collocation dictionary</i> .....   | 369 |
| Margarita ALONSO RAMOS, Leo WANNER, Nancy VAZQUEZ VEIGA, Orsolya VINCZE, Estela MOSQUEIRA SUAREZ, Sabela PRIETO GONZALEZ, <i>Tagging collocations for learners</i> .....                | 375 |
| James BREEN, <i>WWWJDIC – A feature-rich WWW-based Japanese dictionary</i> .....  | 381 |
| Elisa CORINO, Cristina ONESTI, <i>Have I got the wrong definition of ...? How to write simple technical definitions on the basis of examples taken from Newsgroup discussions</i> ..... | 387 |
| Nathalie GASIGLIA, <i>Some editorial orientations for a multi-tier electronic monolingual school dictionary</i> .....   | 393 |
| Fadila HADOUCHE, Marie-Claude L'HOMME, Guy LAPALME, <i>Automatic annotation of actants in specialized corpora</i> .....   | 399 |
| Jakob HALSKOV, Pia JARVAD, <i>Automated extraction of neologisms for lexicography</i> .....   | 405 |
| Adam KILGARRIFF, Vojtěch KOVAR, Pavel RYCHLÝ, <i>Tickbox Lexicography</i> .....   | 411 |
| Vera KUZMINA, Anna RYLOVA, <i>ABBY Lingvo electronic dictionary platform and Lingvo Content dictionary writing system</i> .....   | 419 |
| Margit LANGEMETS, Andres LOOPMANN, Ülle VIKS, <i>Dictionary management system for bilingual dictionaries</i> .....  | 425 |
| Héctor MARTÍNEZ, Marta VILLEGAS, Núria BEL, Santiago BEL, Francesca ALEMANY, <i>Lexicography in the grid environment</i> .....  | 431 |
| Carolin MÜLLER-SPITZER, Christine MÖHRS, <i>The “Online Bibliography of Electronic Lexicography” (OBELEX)</i> .....   | 439 |
| Cornelia TSCHICHOLD, <i>From lexical database to intelligent vocabulary trainers</i> .....  | 445 |
| Eveline WANDL-VOGT, <i>Multiple access routes. The dictionary of Bavarian dialects in Austria / Wörterbuch der bairischen Mundarten in Österreich (WBÖ)</i> .....                       | 451 |

# Tagging Collocations for Learners

Margarita Alonso Ramos<sup>1</sup>, Leo Wanner<sup>2</sup>, Nancy Vázquez Veiga<sup>1</sup>,  
Orsolya Vincze<sup>1</sup>, Estela Mosqueira Suárez<sup>1</sup>, Sabela Prieto González<sup>1</sup>  
<sup>1</sup>University of A Coruña, <sup>2</sup>ICREA and Pompeu Fabra University

## Abstract

Collocations play a significant role in second language acquisition. In order to be able to offer efficient support to learners, an NLP-based CALL environment for learning collocations should be based on a representative collocation error annotated learner corpus. We are currently working on such a corpus for Spanish, starting from a fine-grained typology of collocation errors and drawing upon an existing learner corpus, namely CEDEL2 from the Autonomous University of Madrid. In this paper, we present this typology and discuss the first findings obtained from our annotation work.

**Keywords:** collocation, learner corpus, error typology, Spanish as second language

## 1. Introduction

The importance of *collocations* in second language acquisition is increasingly recognized in the community (Lewis 2000; Granger 1998b; Howarth 1998; Nesselhauf 2003, 2005; Alonso Ramos 2006; Higuera 2006; Martelli 2006). To adequately support students in learning collocations, it is crucial to identify and classify the collocation errors made by them and then offer targeted exercises and adequate illustrative material. This presupposes the availability of collocation tagged learner and general corpora: a learner corpus allows us to identify the most common collocation errors; a general corpus is needed as a source of illustration and training material.

We aim at the development of an advanced NLP-based computer assisted language learning (CALL) environment for learning collocations in Spanish. In this paper, we focus on the problem of processing Spanish learner corpora, which consists of three stages: (i) analysis of the corpus and derivation of a collocation error typology; (ii) definition of a tag set to annotate the corpus; and (iii) tagging the corpus.

## 2. Towards a collocation error typology

---

<sup>1</sup> University of A Coruña, lxalonso@udc.es

<sup>2</sup> ICREA and Pompeu Fabra University, leo.wanner@upf.edu. This paper was written within the framework of a research project: FFI2008-06479-C02-01 (Ministerio de Ciencia and partially FEDER).

A detailed analysis of learner corpora has proved to be essential (Dagneaux *et al.* 1998; Granger 1998a, 2007; Tono 2003). Such an analysis requires a predefined error tag set or error typology (Granger 2007). This is also true for the analysis of a collocation learner corpus. Currently available general learner error typologies tend to group collocation errors into a single subclass of lexical errors (Aldabe *et al.* 2005; Milićević and Hamel; 2007; Granger 2007; Díaz-Negrillo and García-Cumbreras 2007). Occasionally, collocation errors are also discussed referring to the POS of the collocation elements (Philip 2007). A closer look at a learner corpus reveals, however that a more detailed typology is needed. For the purpose of the present study, we used the *Corpus Escrito del Español L2* (CEDEL2) from the Autonomous University of Madrid<sup>3</sup>, which consists of short compositions written by native speakers of English (L1). Consider some examples from CEDEL2:

- (1) *deseo* lograr el gol *de ser bilingual*, lit. ‘I desire achieve the goal of being bilingual’
- (2) [...] llenar un puesto [*de trabajo*], lit. ‘fill a position [of work]’
- (3) recibí un llamo *de Brad*, lit. ‘I received a call from Brad’.
- (4) *Algunos* tienen prejuicio *por edad*, lit. ‘Some have prejudice for age’

Apart from errors not related to collocations (e.g. *bilingual* instead of *bilingüe*), which we ignore, the following collocation construction errors stand out<sup>4</sup>:

- (1') error in the base resulting from the projection of a word in L1 (English) to L2 (Spanish), as e.g. *goal* → *gol*: *lograr* [*el*] *gol* – instead of *lograr* [*el*] *objetivo*;
- (2') error in the collocate resulting from a literal translation of a word from L1 to L2, as e.g. *fill* → *llenar*: *llenar* [*un*] *puesto* – instead of *ocupar* [*un*] *puesto*;
- (3') error in the base resulting from a wrong morphological derivation and an inappropriate use of the collocation as a whole in the given context, as e.g. *llamar* → *llamo*: *recibí un* *llamo* *de Brad* – instead of *recibí una* *llamada* *de Brad*; or, better: *me llamó Brad*;
- (4') error in the number of the base and in the governed preposition, as e.g. *prejuicio*: *tienen prejuicio* [*por algo*], instead of *tienen prejuicios* [*hacia algo*].

<sup>3</sup> CEDEL2, which has been compiled by the group directed by Amaya Mendikoetxea, contains about 400,000 words of essays written in Spanish by native speakers of English. The essays are classified with respect to the proficiency level of the authors. The essays underlying our study were written by learners with intermediate or advanced level of Spanish. For more information, see <http://www.uam.es/proyectosinv/woslac/cedel2.htm>.

<sup>4</sup> We interpret collocations in the sense of Hausmann (1979) as idiosyncratic word co-occurrences consisting of a base and a collocate.

The errors are very different. Therefore, a fine-grained collocation error typology is needed to capture these differences and be able to offer adequate didactic means to address them.

In the present stage of our work, we distinguish three main types of collocation errors: lexical errors, grammatical errors and register errors. Lexical errors concern either the whole collocation or one of its elements. In the first case, we find inexistent collocations in Spanish whose meaning would be correctly expressed by a single lexical unit (LU) (e.g. *\*hacer de cotilleos*, lit. '[to] make of gossip' instead of *cotillear* '[to] gossip'), and inexistent single LUs used instead of collocations (e.g. *\*escaparatar* instead of *ir de escaparates*, lit. '[to] go of shop window'). In the second case, we distinguish between errors concerning paradigmatic lexical selection (e.g. *\*lograr un gol* lit. '[to] achieve a goal (in football)' instead of *lograr un objetivo*, lit. '[to] achieve a goal') and errors concerning syntagmatic lexical selection (e.g. *\*escribir el examen*, lit. '[to] write the exam' instead of *hacer el examen*, lit. '[to] do the exam'); the former concern the base, the second the collocate.

Most lexical errors are literal translations from L1. Although a finer distinction is necessary later on to determine the source of errors, as a first approximation, the distinction between “transfer by importation”, i.e., adoption of an inexistent form in L2 – *recibir un llamo*, lit. '[to] receive a call', instead of *recibir una llamada* – and “transfer by extension”, i.e., extension of the meaning of an L2 lexical unit – *salvar dinero*, lit. '[to] save money', instead of *ahorrar dinero* – is valid.

Grammatical errors in our typology are directly linked to collocations. They concern information that a learner cannot derive from the grammar of L2 and that must be described in the entry for the base of the collocation (e.g. *\*hablar al teléfono* lit. '[to] speak to the phone' instead of *hablar por teléfono* '[to] speak through the phone').

In the class of register error, we group collocations that are pragmatically inappropriate. Thus, *tengo el deseo de ser bilingüe*, lit. 'I have the desire of being bilingual' sounds odd in an informal context – better: *me gustaría ser bilingüe* 'I would like to be bilingual'.

### 3. The process of tagging collocations in CEDEL2

Apart from a collocation error typology, a detailed semantic typology of collocations is crucial in order to be able to offer the learner examples of similar collocations. The most detailed and systematic semantically-oriented typology of collocations we know of are the Lexical Functions (Mel'čuk, 1996), from now on referred to as LFs.

With the collocation error and the LF typologies at hand, we tag all collocations in CEDEL2. In the case of collocation errors, we also annotate the correct version of the erroneous collocation and the corresponding LF. Consider the following examples.



- (1'') *lograr [el] gol*: lexical error in the base; extension of the meaning of Sp. *gol* 'goal (in football)' due to phonetic similarity with Eng. *goal*; LF: Real1; correct: *lograr [el] objetivo*
- (2'') *llenar [un] puesto*: lexical error in the collocate; extension of the meaning of Sp. *llenar* 'fill' based on the English collocation [*to*] *fill a position*; LF: Oper1; correct: *ocupar [un] puesto*
- (3'') *recibí un llamo [de Brad]*: lexical error in the base; erroneous derivation based on the first person singular form of the verb Sp. *llamar*, possibly analogous with forms like *paseo*<*pasear*, *canto*<*cantar*, etc.; LF: Oper2; correct: *recibí una llamada [de Brad]*

Example (4'') shows that, on the one hand, a single collocation may show more than one error, and, on the other hand, that the determination of the source of an error is not always straightforward.

- (4'') *tienen prejuicio [por algo]*: 1. grammatical error in the government of the base; intralingual; 2. grammatical error in the number of the base; intralingual or possibly interlingual since Eng. *prejudice* can be used both as a countable or an uncountable noun; LF: Oper1; correct: *tienen prejuicios [hacia algo]*.

The tagging of the learner corpus is currently being performed manually, supported by an interactive annotation tool, *Knowtator*, which is realized as a plug-in of the knowledge acquisition framework Protégé. The application allows us to define an annotation schema used in the process of annotation to give information on the semantics of the combinations— through LFs —, and, in the case of erroneous collocations, to describe the errors and propose a correction. We are also about to develop a collocation tagger that will tag both LFs and collocation errors. The work on the LF-tagger draws upon the work described in Wanner *et al.* (2006).

## 4. Conclusion

The preliminary evaluation of the corpus we annotated so far in accordance with the schema presented above, reveals that 39% of the collocations used by learners contain some error. 62% of the erroneous collocations contain lexical errors, 33% show grammatical errors, whereas 5% have both lexical and grammatical errors. In a more fine-grained analysis of the more prominent lexical errors, we find that 54% of these represent an incorrect choice of the collocate, 20% the use of an incorrect base, 16% the use of an existing collocation with a different sense, while 10% are cases of using collocation-type constructions instead of single LUs. As for the possible source of errors, we can establish that the great majority of lexical errors – 70% – represent clear cases of lexical transfer from L1 to L2. However, further investigation based on a larger annotated corpus is needed to draw more fine-grained conclusions. We are thus currently working on the extension of our collocation error annotated learner corpus.

## References

- ALDABE, I., ARRIETA, B., DÍAZ DE ILARRAZA, A., MARITXALAR, M., ORONÓZ, M. and URÍA, L. (2005). Propuesta de una clasificación general y dinámica para la definición de errores. *Revista de Psicodidáctica*, 10/2: 47-60.
- ALONSO RAMOS, M. (2006). Towards a dynamic way of learning collocations in a second language. In E. Corino, C. Marengo and C. Onesti (eds) *Proceedings XII EURALEX International Congress*, Torino, Italy, September 6th-9th 2006. Alessandria: Edizioni Dell'Orso: 909-921.
- DAGNEAUX, E., DENNESS, S. and GRANGER, S. (1998). Computer-aided error analysis. *System*, 26: 163-174.
- DÍAZ-NEGRILLO, A. and GARCÍA-CUMBRERAS, M.A. (2007). A tagging tool for error analysis on learner corpora. *ICAME Journal*, 31/1: 197-203.
- GRANGER, S. (ed.) (1998a). *Learner English on Computer*. Oxford: Oxford University Press.
- GRANGER, S. (1998b). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. P. Cowie (ed.) *Phraseology. Theory, Analysis, and Applications*. Oxford: Clarendon Press : 145-160.
- GRANGER, S. (2007). Corpus d'apprenants, annotation d'erreurs et ALAO: une synergie prometteuse, *Cahiers de lexicologie*, 91/2 : 465-480.
- HAUSMANN, F. J. (1979). Un dictionnaire des collocations est-il possible? *Travaux de littérature et de linguistique de l'Université de Strasbourg*, 17/1: 187-195.
- HIGUERAS, M. (2006). *Las colocaciones y su enseñanza en la clase de ELE*. Madrid: Arco Libros.
- HOWARTH, P. (1998). The phraseology of learners' academic writing'. In A. P. Cowie (ed.) *Phraseology. Theory, Analysis, and Applications*. Oxford: Clarendon Press: 161-186.
- LEWIS, M. (2000). *Teaching collocation. Further developments in the lexical approach*. London: Language Teaching Publications.
- MARTELLI, A. (2006). A corpus-based description of English lexical collocations used by Italian advanced learners. In E. Corino, C. Marengo and C. Onesti (eds) *Proceedings XII EURALEX International Congress*, Torino, Italy, September 6th-9th 2006. Alessandria: Edizioni Dell'Orso: 1005-1012.
- MEL'ČUK, I. (1996). Lexical Functions: A tool for the Description of Lexical Relations in the Lexicon. In L. Wanner (ed.). *Lexical functions in lexicography and natural language processing*. Amsterdam and Philadelphia: John Benjamins: 37-102.
- MILIĆEVIĆ, J. and M-J., HAMEL. (2007). Un dictionnaire de reformulation pour les apprenants du français langue seconde. In G. Chevalier, K. Gauvin and D. Merkle (eds) *Actes du 29<sup>e</sup> Colloque annuel de l'ALPA tenu a l'Université de Moncton*, Moncton, Canada, November 4th-5th 2005. *Revue de l'Université de Moncton*, numéro hors série: 145-167.
- PHILIP, G. (2007). Decomposition and delexicalisation in learners' collocational (mis)behaviour. In M. Davies, P. Rayson, S. Hunston and P. Danielsson (eds) *Online Proceedings of Corpus Linguistics 2007*, Birmingham, United Kingdom, July 27th-30th 2007. Birmingham: University of Birmingham: 1-11.
- NESSELHAUF, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics*, 24/2: 223-242.
- NESSELHAUF, N. (2005). *Collocations in a learner corpus*. Amsterdam and Philadelphia: John Benjamins.
- NESSELHAUF, N. and TSCHICHOLD, C. (2002). Collocations in CALL: An investigation of vocabulary-building software for EFL, *Computer Assisted Language Learning*, 15/3: 251-279.

- TONO, Y. (2003). Learner corpora: Design, development and applications. In D. Archer *et al.* (eds.) *Proceedings of the Corpus Linguistics 2003*, Lancaster, United Kingdom, March 28th-31th 2003. Lancaster: Lancaster University, University Centre for Computer Corpus Research on Language: 323-343.
- WANNER, L., BOHNET, B. and GIERETH, M. (2006). Making sense of collocations. *Computer Speech & Language*, 20/4: 609-624.