

Lessia M. Kotsyuk,
Ostroh Academy National University, Ostroh

LEARNER CORPORA AS A FEEDBACK ON CURRICULUM

The paper presents a sample research into how the analysis of learner corpora can influence syllabus design. The investigation is conducted on the basis of the Ostroh Academy Corpus of Learner English (OACLE) – a corpus of essays of different types written in English by the students of the College of Romance and Germanic Languages of the Ostroh Academy National University. Due to the on-site character of the corpus the administration of the college can use corpus data to control problematic areas in the learning process, evaluate the syllabus and initiate changes into the planning of certain courses, influence the choice of textbooks.

Key words: learner corpus, OACLE, essay, syllabus, punctuation skills.

Стаття наводить приклад того, як аналіз учнівського корпусу текстів може бути використаний при створенні чи покращенні навчальних програм. Дане дослідження проведено на основі Корпусу Учнівської Української Англійської Мови (КУУАМ) – корпусу англомовних есе різного типу, написаних студентами факультету романо-германських мов Національного університету «Острозька академія». Зважаючи на локальний характер корпусу, адміністрація факультету може використовувати його дані для контролю за проблемними зонами у навчальному процесі, оцінювати навчальні програми та ініціювати зміни до планування певних курсів, впливати на вибір підручників та навчальних матеріалів.

Ключові слова: учнівський корпус, КУУАМ, есе, навчальні програми, пунктуаційні навички.

Статья приводит пример того, как анализ ученического корпуса текстов может быть использован при создании или улучшении учебных программ. Данное исследование было проведено на основе Корпуса Ученического Украинского Английского Языка (КУУАЯ) – корпуса англоязычных эссе разного типа, написанных студентами факультета романо-германских языков Национального университета «Острожская академия». Принимая во внимание локальный характер корпуса, администрация факультета может использовать его данные для контроля за проблемными зонами учебного процесса, оценки учебных программ и инициации изменений касающихся планирования определенных курсов, влиять на выбор учебников и учебных материалов.

Ключевые слова: ученический корпус, КУУАЯ, эссе, учебные программы, пунктуационные навыки.

Introduction

Being a young independent state, Ukraine looks for the effective ways of integration into the European community. English has become an integral part of educational process in every university. Ostroh Academy National University chose English as a second working language along with the official Ukrainian language. The students use English during the classes, communicate with foreign professors and guests, write scientific papers, apply for programs abroad. The research group of lexicographic laboratory LEXILAB initiated creating of a Learner English Corpus to become a part of ICLE project and to represent samples of the Ukrainian English. While the data is being collected and organized, the research group of professors and students already make use of the Corpus to inform curriculum design at the College of Romance and Germanic Languages.

1. Learner Corpora in Language Teaching

1.1 State of Arts

Modern corpus linguistics appears to be a powerful source of language data accumulation. The developments in this sphere effect both theoretical and applied linguistic studies. Language pedagogy is not an exception. Teachers and linguists consider a wide range of issues related to direct or indirect using of learner corpora in language teaching (Leech, 1997; Keck, 2004). Unfortunately, this methodology of foreign language teaching is among the last in the list of great variety of modern teaching techniques to be used in the process of training of future linguists and philologists in Ukraine. The reason for that is probably the existence of much more elaborated methods, already developed teaching packages, provided with authentic audio and video materials and recommendations on how to use them. The introduction of corpus linguistics resources into the process of learning and teaching turns out to be outside of traditional understanding of foreign language didactics and presupposes reaching a new qualitative level of training of future linguist, teacher or translator. In Ukraine the publications that address the aspects of teaching and language corpora are not numerous, they are mainly theoretical and do not deal with the analysis of a certain learner corpora (Demaska-Kulchytska, O. (2006); Buk, S. (2007); Kunych, L. (2008); Diduk-Stupiak, G. (2010). Some attempts to create learner corpus and to describe its applications in language teaching though were made in Kyiv National Linguistic University (Ukrainian Corpus of Learner English, UCLE) and Ostroh Academy National University (Ostroh Academy Corpus of Learner English (OACLE) (Kolomiyets, V, Kotyk, S. (2012), Pampurak, A. (2012)).

This direction of corpus study most actively develops in the countries that already have some compiled learner corpus: Belgium (ICLE, FRIDA, LINDSEI, LONGDALE, and VESPA), UK (BAWE, BALC, and CLC), Spain (BELC), China (BICCEL, CLEC, and ESCCL), Germany (CALE, TCEEE), Japan (CEEAUS), and France (ENGLISH), Czech Republic (AKCES/CZESL). Among the most significant works that have recently been represented on the topic of teaching and language corpora are the following: Barker, F. (2010), Collentine, J., & Asencion-Delaney, Y. (2010), Gilquin Gaëtanelle, Granger Sylviane (2010, 2011), Paquot Magali (2010), Gilquin, G. (2011), Granger, S., & Paquot, M. (2011), Hawkins, J., & Buttery, P. (2010), Huang, C.-R., Cheung, W., Harada, Y., Hong, H., Skoufaki, S., & Chen, H. (2010), Lozano, C., & Mendikoetxea, A. (2010), Mendikoetxea, A., Murcia, S., & Rollison, P. (2010), Meunier

Fanny (2010), Racine, I., Detey, S., Bühler, N., Schwab, S., Zay, F., & Kawaguchi, Y. (2010), Thewissen, J., Bestgen, Y., & Lories, G. (2010).

The above mentioned scientific studies deal with the questions of how the learner corpora are used in both second language acquisition and in foreign language teaching. The joint interest of both directions of didactics is the difference in language acquisition of a native and a foreign languages, universal processes in mastering a foreign language, the role of native language in foreign language learning, language variability, individual peculiarities and strategies in teaching. Learner's speech samples, reports, introspections, experiments make up the data to help solve these problems.

The most extensively learner corpora have been used to provide accurate descriptions of learners language use. The tradition of language error analysis of the texts produced by learners goes back to 1960-70s. Some not very big collections of texts, non-electronic, variegated were looked through. Only certain types of mistakes were fixed and extracted, the rest of the material was not taken into account. Modern learner corpora in contrast have a whole set of essential advantages: the amount of data is much bigger, electronic form, accurate criterion of data collection, etc. They provide the possibility not only to fix language errors but to make conclusions as to learners' speech in general and look critically at existing teaching methods, syllabuses and teaching materials. Some works prove that corpora may provide data, which may further alter what is taught (Mindt, D. (1996), Hunston, S. (2002), McEnery, T. and Xiao, R (2011).

1.2 Learner corpus

As stated by Granger, S (2002: 124), learner corpora are defined as 'electronic collections of spoken or written texts produced by foreign or second language learners in a variety of language settings.' A learner corpus provides data for a deep quantitative and qualitative analysis of language skills of a learner, which leads to working out of new teaching means and methods and language acquisition facilitation. Such type of a corpus can also be used for the error analysis to display the main lexical and syntactical mistakes of a foreign language learner. The facts about the quantity of different language use errors, the types of contexts, in which they were encountered, can be provided only by a learner corpus.

The website of Learner Corpus Association (<http://www.learnercorpusassociation.org/>) represents a list of 125 learner corpora around the world. The most famous one is the *International Corpus of Learner English (ICLE)* – an English learner corpus, which contains argumentative essays written by higher intermediate to advanced learners of English from several mother tongue backgrounds. The creation of the ICLE was initiated by The Louvain Centre for English Corpus Linguistics, which pioneered learner corpus investigations. The computerized database of the Centre is the result of more than ten years work of a great number of universities from around the world. Now the Corpus consists of more than 4 million words, pronounced or written by the representatives of more than 20 countries.

2. OACLE for the College Needs

The research group of lexicographic laboratory LEXILAB initiated creating of *Ostroh Academy Corpus of Learner English* to become a part of ICLE project and to represent samples of the Ukrainian English. The final amount of words that will be included in OACLE is 200,000. The process of corpus creation is divided into several phases. The first step is to collect the essays of the students of the College of Romance and Germanic Languages. Every student can donate only one 500-1000 word essay. The essays collected were of the same type – argumentative (are written as a home task and allow using additional materials). At the moment 165 essays are taken from the students, whose native language is Ukrainian, they are mainly at their third and fourth year of studying at the College. All the samples were coded according to the system proposed by ICLE project.

The development of such a corpus appeared to be useful not only to participate in the international project but to inform the administration of the College as to typical mistakes, peculiarities of vocabulary choice, grammatical organization of the writings of the students.

One of the surveys was conducted to identify problematic punctuation areas in the writings of the College students and to think over possible treatment of the problem. The main questions to answer were: which punctuation mark appeared to be the most frequent; what were the most typical punctuation errors and which of them could be considered as transfer (as the result of the mother tongue influence) or intralingual (influence coming from the target language).

At the first stage of the investigation the corpus was manually tagged. Each punctuation mark was coded according to the system of codes developed earlier. Three types of elements could be contextualized for each of the punctuation mark:

a) Correct use of the punctuation mark – full stop (P.P), question mark (P?P), exclamation mark (P!P), comma (P,P), semicolon (P;P), colon (P:P), dash (P-P), parentheses (P()P), suspension points (P...P), and apostrophe (P'P);

b) Incorrect use of punctuation mark: full stop (PI.P), question mark (PI?P), exclamation mark (PI!P), comma (PI,P), semicolon (PI;P), colon (PI:P), dash (PI-P), parentheses (PI()P), suspension points (PI...P), and apostrophe (PI'P);

c) The punctuation mark is missing: full stop (P0.P), question mark (P0?P), exclamation mark (P0!P), comma (P0,P), semicolon (P0;P), colon (P0:P), dash (P0-P), parentheses (P0()P), suspension points (P0...P), and apostrophe (P0'P).

According to the concordances run for each of the element it was found out that the exclamation mark, the question mark and the suspension points were not used at all, the colons were rare, which can be explained by the type of essays included in the corpus. The most frequent were the full stop and the comma. Taking into account the nature of the essays (argumentative), it can be concluded that such poor use of the variety of punctuation marks testifies students' fear of complex sentence constructions.

The typical punctuation mistakes found in the essays were in the use of coma and semicolons. Students chose to use coma instead of semicolon in composite extended sentences. The most frequent coma mistakes are: parenthetic

words in sentences; in sentences with *and*, *because*, *but*; defining/non-defining clauses; conditional sentences. Some of the mistakes are typical for all the essays; some appear in the same essay for several times. Most cases of punctuation mark misuse can be considered as the result of students' mother tongue influence.

Having compared college curricula, the programs and the textbooks of the courses of Academic Writing, English Language and the results of corpora analysis the research group came to conclusion, that punctuation is paid not enough attention in the process of studying. The students' knowledge of punctuation rules is superficial; they lack practice and transfer their punctuation experience from the Ukrainian language.

Changes into curricula as well as refinements to teaching techniques should be incorporated. The research group proposed to the faculty of the College to check with the curricula and developed a set of exercises and additional materials that can be used by teachers and students in complex with the textbooks.

3. Conclusions

Corpus linguistics is in direct contribution to language teaching. Textbooks do not usually consider learners peculiarities coming from their mother tongue, they are designated to heterogeneous audiences and global markets [13]. Learner corpus can help teachers visualize the results of their work; develop learning materials and exercises to the direct needs of the students. On-site learner corpora can influence the development of adequate courses and learning materials, evaluation of the curricular within the frames of a university, college, or school.

References:

1. Barker, F. (2010). How can corpora be used in language testing? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics* (Routledge Handbooks in Applied Linguistics) (pp. 633-46). New York: Routledge.
2. Collentine, J., & Asencion-Delaney, Y. (2010). A corpus-based analysis of the discourse functions of 'ser/estar' plus adjective in three levels of Spanish as FL learners. *Language Learning*, 60(2), 409-445.
3. Gilquin Gaëtanelle, Granger Sylviane, How can data-driven learning be used in language teaching?, In: *The Routledge Handbook of Corpus Linguistics*, London, Routledge, 2010, p. 359-370.
4. Gilquin, G. (2011). Lexical infelicity in causative constructions. Comparing native and learner collocations. In J. Leino & R. von Waldenfels (Eds.), *Analytical Causatives* (). München: Lincom Europa.
5. Granger, S. (2002), «A bird's-eye view of learner corpora research,» in *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*, Granger, S., Hung, J. & Petch-Tyson, S. eds., John Benjamins, Amsterdam, pp. 3-33.
6. Granger, S., & Gilquin, G. (2011). From EFL to ESL: evidence from the International Corpus of Learner English. In J. Mukherjee & M. Hundt (Eds.), *Exploring second-language varieties of English and learner Englishes: bridging a paradigm gap* (). Amsterdam; Philadelphia: John Benjamins.
7. Granger, S., & Paquot, M. (2011). Language for specific purposes learner corpora. In T. Upton & C. Chapelle (Eds.), *Language for specific purposes. The encyclopedia of applied linguistics* (). Oxford: Blackwell-Wiley.
8. Hawkins, J., & Buttery, P. (2010). Criterial features in learner corpora: theory and illustrations. *English Profile Journal*, 1(1), 1-23.
9. Huang, C.-R., Cheung, W., Harada, Y., Hong, H., Skoufaki, S., & Chen, H. (2010). English learner corpus: global perspectives with an Asian focus. Paper presented at the A new look at language teaching and testing: English as subject and vehicle. Selected papers from the 2009 LTTC international conference on English Language Teaching and Testing, March 6-7, Taipei (Taiwan).
10. Hunston, S. (2002) *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
11. Keck, C. (2004) 'Corpus linguistics and language teaching research: Bridging the gap'. *Language Teaching Research* 8/1: 83-109.
12. Leech, G. (1997) 'Teaching and language corpora: a convergence' in A. Wichmann, S. Fligelstone, T. McEnery and G. Knowles (eds.) *Teaching and Language Corpora*, pp. 1-23. London: Longman.
13. López, Willelmira Castillejos. Error analysis in a learner corpus. What are the learners' strategies? – <http://www.scribd.com/doc/73162791/Error-Analysis-in-a-Learner-Corpus>
14. Lozano, C., & Mendikoetxea, A. (2010). Interface conditions on postverbal subjects: a corpus study of L2 English. *Bilingualism: Language and Cognition*, 13(4), 475-497.
15. McEnery, T. and Xiao, R. (2011) What Corpora Can Offer in Language Teaching and Learning. In: Hinkel, E. (ed). *Handbook of Research in Second Language Teaching and Learning: Volume 2*. Routledge.
16. Mendikoetxea, A., Murcia, S., & Rollison, P. (2010). Focus on errors: learner corpora as pedagogical tools. In M. C. Campoy-Cubillo, B. Bellés-Fortuño & M. L. Gea-Valor (Eds.), *Corpus-based approaches to English language teaching*. London; New-York: Continuum.
17. Meunier Fanny, *Learner Corpora and English Language Teaching: Checkup Time*, *Anglistik: International Journal of English Studies*, 21, 1, 2010, p. 209-220.
18. Mindt, D. (1996) 'English corpus linguistics and the foreign language teaching syllabus' in J. Thomas and M. Short (eds.) *Using Corpora for Language Research*, pp. 232-247. Harlow: Longman.
19. Paquot Magali, *Academic vocabulary in learner writing: from extraction to analysis*, Continuum, 2010, *Corpus and Discourse*
20. Racine, I., Detey, S., Bühler, N., Schwab, S., Zay, F., & Kawaguchi, Y. (2010). The production of French nasal vowels by advanced Japanese and Spanish learners of French : a corpus-based evaluation study. Paper presented at the New Sounds 2010 – Sixth International Symposium on the Acquisition of Second Language Speech, Poznan: Adam Mickiewicz University.
21. Thewissen, J., Bestgen, Y., & Lories, G. (2010). Using latent semantic analysis to measure coherence in essays by foreign language learners?. In S. Bolasco, I. Chiari & L. Giuliano (Eds.), *Proceedings of 10th International Conference on statistical analysis of textual data (JADT2010)* (pp. 385-95). Rome (Italy): LED.

22. Бук С. Учніські корпуси в методиці викладання іноземної мови / С. Бук // Теорія і практика викладання української мови як іноземної : збірник наукових праць. – Львів : Видавничий центр ЛНУ імені І. Франка, 2007. – Випуск 2. – С. 19–23.
23. Демська-Кульчицька О. Корпусні підходи в методиці мови // Теорія і практика викладання української мови як іноземної. – 2006. – Вип. 1. – С. 22–27.
24. Дідук-Ступ'як, Г. І. Лінгводидактичні можливості корпусної лінгвістики [Текст] / Г. І. Дідук-Ступ'як // Наукові записки ТНПУ ім. В. Гнатюка. Сер. Педагогіка. – Тернопіль, 2010. – № 1. – С. 105-109.
25. Коломієць В, Котик С. Спеціальний навчальний корпус текстів UCLE: сучасний стан і перспективи використання. Комп'ютерна лінгвістика: сучасне та майбутнє. Матеріали Міжнародної науково-практичної конференції – К. : КНЛУ, 2012.– С. 29-32.
26. Кунич Л. Інноваційні технології англійської мови у світлі ідей корпусної лінгвістики / Л. Кунич // Імідж сучасного педагога. – 2008. – № 9–10. – С. 93–95.
27. Пампурак, А. М. (A. Pamprak) (2012) Застосування учнівських корпусів текстів при навчанні лексики іноземної мови. Студентські наукові записки. Серія «Філологічна» (Вип. 5). с. 144-146.