

Czy korpus prawdę ci powie?

O wykorzystaniu korpusów w nauczaniu języka do celów zawodowych

JAROSŁAW KRAJKA

Korpusy językowe, w przeszłości wykorzystywane jedynie do badań naukowych, stały się powszechnie dostępne w Internecie. Dzięki darmowym programom do obsługi tekstów nauczyciele języków specjalistycznych mogą konstruować własne minikorpusy, na ich podstawie wybierać częstsze lub rzadsze słowa, najbardziej interesujące konteksty, wreszcie konstruować materiały dydaktyczne w oparciu o autentyczne dane językowe.

Językoznawstwo korpusowe, a w szczególności procesy kompilowania korpusów i ich przeszukiwania przy pomocy programów konkordancyjnych, stanowi dziedzinę językoznawstwa stosunkowo mało wykorzystywaną w procesie kształcenia nauczycieli, co było prawdopodobnie spowodowane ograniczonym dostępem do korpusów oraz brakiem właściwego kształcenia informatycznego na studiach filologicznych w przeszłości. Wydaje się jednak, że wraz z upowszechnieniem Internetu oraz oprogramowania typu *freeware/open source*, główne przeszkody w używaniu gotowych korpusów online i tworzeniu własnych (ang. *do-it-yourself corpora*) przestały być aż tak istotne.

Celem niniejszego artykułu jest rozważenie możliwości upowszechnienia procedur językoznawstwa korpusowego w nauczaniu języka do celów zawodowych. Najpierw przywołamy wyniki badań wskazujących na zalety i wady uczenia się w oparciu o dane korpusowe, następnie zaś – wskazówki metodyczne dla nauczycieli planujących wzbogacenie procesu nauczania o ćwiczenia oparte na danych korpusowych.

Narzędzia językoznawstwa korpusowego – wprowadzenie

Wiele badań wskazuje na zalety stosowania procedur językoznawstwa korpusowego w nauczaniu języków obcych, a więc, jak definiuje je Barbara Lewandowska-Tomaszczyk (2005), jednej z części językoznawstwa komputerowego, zajmującej się analizą języka zgromadzonego w komputerowych zbiorach autentycznych tekstów językowych, mówionych i pisanych, reprezentujących różne odmiany, style

i typy tekstu (czyli korpusach językowych). Dla przykładu, badacze podkreślają efektywność użycia zarówno niewielkich korpusów, dostosowanych do potrzeb uczniów (Aston 1997), jak i przeszukiwania dużych lub bardzo dużych kolekcji tekstowych (Bernardini 2000; de Schryver 2002) dla rozwijania sprawności pisania u początkujących uczniów (Gaskell i Cobb 2004; Yoon i Hirvela 2004) i na wyższych poziomach zaawansowania (Chambers i O’Sullivan 2004). Wspomaganie nauczania gramatyki, przede wszystkim w fazie prezentacji struktur (Hadley 2002) oraz formułowania przez uczniów reguł gramatycznych (St. John 2001) może być realizowane za pomocą odpowiednio dobranych konkordancji (przy tzw. „miękkim” podejściu, kiedy uczniowie nie pracują sami z korpusem, ale analizują wybrane przez nauczyciela konkordancje – Gabrielatos 2005). Najbardziej obiecujące efekty pokazują badania w zakresie wspomaganie rozumienia tekstu czytanego i nabywania leksyki z kontekstu pisanego (Chodkiewicz 2000; Cobb 1997; Cobb 1998, Cobb i in. 2001; Horst i in. 2005) oraz wspomaganie autonomii ucznia poprzez realizację treningu strategicznego (Chambers 2005; Kennedy i Miceli 2001; St. John 2001).

Alternatywą dla gotowych korpusów mogą być zbiory tworzone przez nauczyciela lub grupy nauczycieli z wykorzystaniem gotowych tekstów w odpowiedzi na konkretne potrzeby klasy (Lee i Swales 2006). Takie *korpusy zrób-to-sam* (ang. *do-it-yourself corpora*) będą często jedyną możliwością w sytuacjach, gdy uczniowie mają specyficzne potrzeby wynikające z wąsko sprecyzowanego celu uczenia się, a istniejące korpusy nie odzwierciedlają w należyтым stopniu danych typów tekstów lub dziedzin życia

(Tribble 1997), lub też gdy celem nauczyciela jest wyraziste pokazanie języka potocznego, mówionego czy konkretnej odmiany geograficznej. Tego typu procedury są również zalecane nauczycielom języka do celów zawodowych, głównie z uwagi na fakt, że korpusy języka ogólnego w zbyt małym stopniu zawierają w sobie teksty z określonych domen specjalistycznych, co skutkuje niewielką ilością przykładów określonych użyć słów czy zwrotów, a nawet ich zupełną nieobecnością.

Punktem wyjścia dla nauczyciela przygotowującego własny minikorpus z danej dziedziny powinno być zebranie możliwie najbardziej reprezentatywnych autentycznych wzorców tekstowych dla danej domeny języka specjalistycznego. Zgromadzenie takich dokumentów (np. aktów prawnych z systemu Eur-Lex, instrukcji obsługi urządzeń, artykułów z medycznych czasopism naukowych itd.), według ściśle określonych kryteriów: powstania tekstu, źródła jego pozyskania, czasu utworzenia, celu utworzenia, długości, pozwoli na zgromadzenie własnego zasobu wzorców tekstowych jako minikorpusu języka specjalistycznego.

Wprowadzenie zgromadzonych tekstów do programu konkordancyjnego (np. TextSTAT, <http://neon.niederlandistik.fu-berlin.de/textstat/>) pozwoli na stworzenie listy frekwencyjnej, wyselekcjonowanie z listy frekwencyjnej słów kluczowych, określenie ich konkordancji i wybranie odpowiednich ilustracyjnych akapitów tekstu.

Po określeniu listy najczęstszych/najbardziej produktywnych słów/zwrotów specjalistycznych z danej dziedziny na podstawie wzorców tekstowych, skorzystać można z gotowych korpusów językowych, takich jak na przykład uznane korpusy referencyjne dla określenia najbardziej prototypowych przykładów użycia, najczęstszych kolokacji czy zbadania gramatyki słów, poprzez sprawdzenie współwystępowania wybranych sformułowań:

- język angielski – British National Corpus, Collins COBUILD, Corpus of Contemporary American English – <http://corpus.byu.edu>;
- język francuski – FRANTEXT (<https://artfl-project.uchicago.edu/content/artfl-frantext>), ELICOP/ELILAP/LANCOM (<http://bach.arts.kuleuven.be/pmertens/corpus/search/s.html>);
- język niemiecki – COSMAS (<http://www.ids-mannheim.de/cosmas2/>), DWDS (<http://www.dwds.de/>);
- język hiszpański – Corpus del Español (<http://www.corpusdelespanol.org/>);
- język portugalski – Corpus do Português (<http://www.corpusdoportugues.org/>);
- język rosyjski – Russian National Corpus (<http://www.ruscorpora.ru/index.html>).

Korzyści i problemy wynikające z wykorzystania elementów językoznawstwa korpusowego w praktyce nauczyciela języka obcego

Procedury językoznawstwa korpusowego, a w szczególności właściwie wybrany lub skompilowany korpus oraz odpowiedni program konkordancyjny, mogą stanowić istotną pomoc naukową zarówno w przygotowaniu zajęć przez nauczyciela, w jego procesie samokształcenia, jak i w kształtowaniu świadomości językowej uczniów („bepośrednie” użycie korpusów w nauczaniu – Leech 1997). Najważniejsze korzyści wynikające z nauczania wspomagane korpusami to:

- zwiększenie efektywności nauczania języka, zwłaszcza w zakresie nabywania słownictwa, poprzez osadzenie poznawanej leksyki w bogatym kontekście (Leńko-Szymańska 2001, 2005; Lewandowska-Tomaszczyk 1998, 2000, 2001, 2005);
- wprowadzenie elementu autentyczności (tekstu, celu, zadania) do procesu uczenia, przy przejściu przez uczniów części odpowiedzialności za kształtowanie własnej kompetencji językowej (Johns 1988);
- poszerzenie znajomości słownictwa poprzez obserwowanie słów w rozlicznych kontekstach, zwiększające skuteczność użycia leksyki (Cobb 1998);
- ukazanie słownictwa w szerszej gamie sytuacji i użyciu niż czasem stereotypowe przykłady z podręcznika (Cobb 1998; Hasselgård 2001);
- wspomaganie procesu zdobywania informacji ze słownika poprzez dostarczenie przykładów ilustrujących użycie słowa w zdaniu (Thomas 2003);
- weryfikowanie hipotez językowych przez nauczycieli, uczniów czy tłumaczy (St. John 2001; Thomas 2003);
- samodzielne kształtowanie wiedzy językowej w procesie indukcji, wpływające na jej większą użyteczność (Cobb 1999; Tribble i Jones 1990);
- wprowadzanie do świadomości uczniów/nauczycieli pojęć i rozróżnień gramatycznych, przez co buduje się formalną wiedzę o języku (Godwin-Jones 2001);
- uwrażliwienie na świadomość rejestru językowego, poprzez integrację konkordancji z dokumentami źródłowymi (Krieger 2003).

Z drugiej strony, warto przestrzec nauczycieli przed nadmiernym entuzjazmem przy użyciu procedur językoznawstwa korpusowego, w szczególności w związku z następującymi problemami:

- duża ilość trudnej do zrozumienia informacji językowej, z wieloma kontekstami nieznanymi uczniom (Cobb 1998), może doprowadzić do „utonięcia” w nadmiernej liczbie danych (Tribble 1997);

- mimo różnorodności i bogactwa kontekstów, konkordancje często są krótkie, niekompletne, nie tworzą zwartej całości (Cobb 1998);
- interfejs, składnia przeszukiwania czy terminologia użyta w narzędziach konkordancyjnych czyni je trudnymi do swobodnego opanowania przez nauczycieli i uczniów (Stevens 1995);
- bardziej wyrafinowane użycia, wykraczające poza proste sprawdzanie słowa kluczowego, wymagają odpowiedniego szkolenia i wiedzy specjalistycznej (Stevens 1995);
- nie wszyscy uczniowie pozytywnie przyjmują indukcyjne podejście do nauczania języka (Krieger 2003), a samodzielne formułowanie zapytań pozwalających na testowanie hipotez językowych może być poza zasięgiem przeciętnych możliwości uczniów (Stevens 1995);
- specyfika korpusu nie zawsze jest zrozumiała dla uczniów, traktujących go jako kolejny słownik (Stevens 1995).

Wskazania metodyczne dotyczące wykorzystania danych korpusowych na lekcjach języka do celów zawodowych

Osiągnięcia językoznawstwa korpusowego, zwłaszcza możliwości samodzielnej kompilacji tekstów z wybranej dziedziny i ich analizowania przy pomocy prostych narzędzi typu TextSTAT (<http://neon.niederlandistik.fu-berlin.de/textstat/>), umożliwiają nauczycielom języka do celów zawodowych tworzenie własnych minikorpusów poprzez zgromadzenie oczekiwanej liczby równoważnych tekstów (np. podobnych ustaw, rozporządzeń, transkryptów przemówień, instrukcji obsługi, dokumentacji technicznych itd.), „załadowanie” ich do programu TextSTAT oraz wyszukiwanie interesujących przykładów, zastosowań, związków frazeologicznych. Warto jednak przy tej okazji mieć na uwadze pewne istotne aspekty procesu badania języka z użyciem korpusów, wraz ze świadomością ograniczeń i problemów.

KORPUS JEST DLA NAUCZYCIELA, NIE DLA UCZNI

Interfejsy korpusów (często określane mianem programów konkordancyjnych czy konkordancerów) nie zawsze są intuicyjne i proste w obsłudze, czasem wykorzystywana jest też w nich bardziej zaawansowana terminologia językoznawcza w języku angielskim. Stąd bardziej efektywne jest korzystanie z korpusu (wyszukiwanie przykładów, wybieranie tych najbardziej odpowiednich, kopiowanie do edytora tekstu i przygotowywanie ćwiczeń) przez nauczyciela, niż oczekiwanie, że to sami uczniowie będą

samodzielnie „sprawdzać” słowa w korpusie, tak jak to robią ze słownikiem.

KORPUS NIE ZAWSZE PRAWDĘ CI POWIE...

Korpusy językowe rejestrują użycie języka w określonych sytuacjach komunikacyjnych w określonym miejscu i czasie, bez ingerencji badacza w poprawianie tekstów, dlatego znaleźć w nich można mniej stereotypowe, rzadsze czy po prostu niepoprawne użycia wyszukiwanych słów. Stąd istotną rolą nauczyciela jest odpowiednia interpretacja danych, pomijanie jednostkowych przykładów odmiennych od ustalonej normy, wybieranie do ćwiczeń tych konkordancji, które najlepiej ilustrują znaczenie danych słów.

SAM KORPUS TO ZA MAŁO

I nauczyciele, i uczniowie powinni mieć świadomość konieczności używania korpusu razem ze słownikiem (lub słownikami), dla lepszego zrozumienia znaczeń danych słów, wyselekcjonowania tych znaczeń, które są bardziej powszechnie używane, czy też tych, które występują w określonej dziedzinie specjalistycznej. Zwłaszcza uczniom mogłoby być trudno zrozumieć znaczenie nowych słów wyłącznie na podstawie konkordancji, które mogą być bardzo różne – mniej lub bardziej zrozumiałe.

SZUKAJ ODPOWIEDZI NA PROBLEMY JĘZYKOWE

W przeciwieństwie do słownika, który podaje często „jedynie poprawne” użycia, interfejsów wyszukiwawczych korpusów można użyć do formułowania zapytań mających na celu weryfikację problemów/hipotez/reguł. Dla przykładu, sprawdzić można częstotliwość i znaczenie wyrażen *arrive at*, *arrive in*, *arrive to*, dla określenia ich odcieni znaczeniowych. Inną możliwością jest sprawdzenie częstotliwości i zakresu użycia pewnych słów różniących się znaczeniem w brytyjskiej i amerykańskiej odmianie języka (np. *autumn* i *fall*). W podobny sposób formułowanie kilku zapytań o podobne słowa, „wyłuskiwanie” interesujących przykładów i tworzenie na ich podstawie zadań językowych (typu dopasowywanie części zdań, test luk, rozsypanka wyrazowa itd.) pozwoli na zaangażowanie uczniów w proces „odkrywania” języka bez niebezpieczeństwa „utonięcia” w danych językowych czy zniechęcenia mało przyjaznym środowiskiem nauki.

Zakończenie

Narzędzia leksykalne doby Internetu 2.0 to już nie tylko gotowe słowniki, w których można jedynie poszukiwać odpowiedzi na nurtujące nauczyciela czy uczniów pytania dotyczące znaczeń nowych słów. Internet drugiej generacji

wprowadził otwarty charakter aplikacji, dając użytkownikom kontrolę nad kształtem programów. W dziedzinie narzędzi leksykalnych widać to w przypadku korpusów językowych, które przy użyciu prostych programów konkordacyjnych mogą być kompilowane dla dowolnego obszaru tematycznego czy domeny tekstowej. W ten sposób nauczyciele języka do celów zawodowych zyskują znakomite narzędzie pozwalające na uzupełnienie materiału z podręcznika czy samodzielne dobieranie tekstów.

BIBLIOGRAFIA

- Aston G. (1997) Small and Large Corpora in Language Learning. W: J. Melia, B. Lewandowska-Tomaszczyk (red.) *PALC'97 Proceedings. The First International Conference: Practical Applications in Language Corpora (1997)*, University of Łódź, Poland. Łódź: Wydawnictwo Uniwersytetu Łódzkiego, 51-62. [online] [dostęp 18.07.2016] <<http://www.sslmit.unibo.it/~guy/wudj1.htm>>.
- Bernardini S. (2000) Systematising Serendipity: Proposals for Concordancing Large Corpora with Language Learners. W: L. Burnard, T. McEnery (red.) *Rethinking Language Pedagogy from a Corpus Perspective*. Frankfurt: Peter Lang, 225-234.
- Chambers, A. (2005) Integrating Corpus Consultation Procedures in Language Studies. W: *Language Learning & Technology*, nr 9(2), 111-125.
- Chambers, A., O'Sullivan, Í. (2004) Corpus Consultation and Advanced Learners' Writing Skills in French. W: *ReCALL*, nr 16(1), 158-172.
- Chodkiewicz H. (2000) *Vocabulary Acquisition from the Written Context*. Lublin: Wydawnictwo UMCS.
- Cobb, T. (1997) Is There Any Measurable Learning from Hands-on Concordancing? W: *System*, nr 25, 301-315.
- Cobb, T. (1998) Breadth and Depth of Lexical Acquisition With Hands-on Concordancing. W: *Computer Assisted Language Learning*, nr 12(4), 345-360.
- Cobb T., Greaves C., Horst M. (2001) Can the Rate of Lexical Acquisition from Reading Be Increased? An Experiment in Reading French with a Suite of Online Resources. W: P. Raymond, C. Cornaire (red.) *Regards sur la Didactique des Langues Secondes*. Montréal: Éditions logique, 133-153.
- Cobb, T. (1999) Applying Constructivism: A Test for the Learner as Scientist. W: *Educational Technology Research & Development*, nr 47(3), 15-31.
- de Schryver, G. M. (2002) Web for/as Corpus: a Perspective for the African Languages. W: *Nordic Journal of African Studies*, nr 11(2), 266-282.
- Gabrielatos, C. (2005) Corpora and Language Teaching: Just a Fling or Wedding Bells? W: *TESL-EJ*, nr 8(4). [online] [dostęp 18.07.2016] <<http://www.tesl-ej.org/ej32/a1.html>>.
- Gaskell, D., Cobb, T. (2004) Can Learners Use Concordance Feedback for Writing Errors? W: *System*, nr 32(3), 301-319.
- Godwin-Jones, B. (2001) Tools and Trends in Corpora Use for Teaching and Learning. W: *Language Learning & Technology*, nr 5(3), 7-12. [online] [dostęp 18.07.2016] <<http://llt.msu.edu/vol5num3/emerging/>>.
- Hadley, G. (2002) Sensing the Winds of Change: An Introduction to Data-Driven Learning. W: *RELC Journal*, nr 33(2), 99-124.
- Hasselgård, H. (2001) *Corpora and Their Use in Research and Teaching* [online] [dostęp 18.07.2010] <<http://folk.uio.no/hasselg/UV-corpus.htm>>.
- Horst, M., Cobb, T., Nicolae, I. (2005) Expanding Academic Vocabulary with a Collaborative Online Database. W: *Language Learning & Technology*, nr 9(2), 90-110.
- Johns, T. (1988) Whence and Whither Classroom Concordancing? W: T. Bongaerts, P. de Haan, S. Lobbe, H. Wekker (red.) *Computer Applications in Language Learning*. Dordrecht: Foris Publications, 9-33.
- Kennedy, C., Miceli, T. (2001) An Evaluation of Intermediate Students' Approaches to Corpus Investigation. W: *Language Learning & Technology*, nr 5(3), 77-90.
- Krieger, D. (2003) Corpus Linguistics: What It Is and How It Can Be Applied to Teaching. W: *The Internet TESL Journal*, nr 9(3) [online] [dostęp 18.07.2016] <<http://iteslj.org/Articles/Krieger-Corpus.html>>.
- Lee, D., Swales, J. (2006). A Corpus-based EAP Course for NNS Doctoral Students: Moving from Available Specialized Corpora to Self-compiled Corpora. W: *English for Specific Purposes*, nr 25, 56-75.
- Leech, G. (1997) Teaching and Language Corpora – A Convergence. W: A. Wichmann, S. Fligelstone, T. McEnery, G. Knowles (red.) *Teaching and Language Corpora*. New York: Longman, 1-23.
- Leńko-Szymańska, A. (2001) Passive and Active Vocabulary Knowledge in Advanced Learners of English. W: B. Lewandowska-Tomaszczyk, P. J. Melia (red.) *PALC'99. Practical Applications in Language Corpora*. Frankfurt: Peter Lang, 287-302.
- Leńko-Szymańska, A. (2005) Korpusy w nauczaniu języków obcych. W: B. Lewandowska-Tomaszczyk (red.) *Podstawy językoznawstwa korpusowego*. Łódź: Wydawnictwo Uniwersytetu Łódzkiego, 221-239.
- Lewandowska-Tomaszczyk, B. (1998) Language Corpora, the Internet, and the Foreign Language Teacher. W: B. Lewandowska-Tomaszczyk (red.) *Perspectives on Foreign Language Teaching*. Piotrków Trybunalski: WSP Kielce, Filia w Piotrkowie Trybunalskim, 133-154.
- Lewandowska-Tomaszczyk, B. (2000) Computers, Language Corpora and Their Applications. W: B. Lewandowska-Tomaszczyk, J. Osborne, F. Schulte (red.) *Computers in Language Studies*. Piotrków Trybunalski: WSP Kielce, Filia w Piotrkowie Trybunalskim, 96-164.

- Lewandowska-Tomaszczyk, B. (2001) Acquisition of Lexis, Language Corpora and Foreign Language Teaching. W: B. Lewandowska-Tomaszczyk, I. Czwenar (red.) *A New Curriculum for English Studies*. Piotrków: College Press.
- Lewandowska-Tomaszczyk, B. (2005) *Podstawy językoznawstwa korpusowego*. Łódź: Wydawnictwo Uniwersytetu Łódzkiego.
- St. John, E. (2001) A Case for Using a Parallel Corpus and Concordancer for Beginners of a Foreign Language. W: *Language Learning & Technology*, nr 5(3), 185-203.
- Stevens, V. (1995) Concordancing with Language Learners: Why? When? What? W: *CAELL Journal*, nr 6(2), 2-10. [online] [dostęp 18.07.2016] <<http://www.vancestevens.com/papers/archive/concordance1996.htm>>.
- Thomas, J. (2003) Extending Vocabulary Knowledge with Computers. W: *Teaching English with Technology*, nr 3(2). [online] [dostęp 18.07.2016] <<http://www.tewtjournal.org/issues/past-issue-2003/past-issue-2003-issue-2/>>.
- Tribble, C. (1997) Improvising Corpora for ELT: Quick-and-dirty Ways of Developing Corpora for Language Teaching. W: J. Melia, B. Lewandowska-Tomaszczyk (red.) *PALC'97 Proceedings*. Łódź: Wydawnictwo Uniwersytetu Łódzkiego, 106-117.
- Tribble, C., Jones, G. (1990) *Concordances in the Classroom: A Resource Book for Teachers*. London: Longman.
- Yoon, H., Hirvela, A. (2004) ESL Student Attitudes towards Corpus Use in L2 Writing. W: *Journal of Second Language Writing*, nr 13, 257-283.

DR HAB. JAROSŁAW KRAJKA Pracuje na Uniwersytecie Marii Curie-Skłodowskiej w Lublinie oraz Uniwersytecie SWPS w Warszawie. Zajmuje się badaniami nad nauczaniem języków obcych wspomaganym komputerowo, kształceniem nauczycieli w e-learningu, wielojęzycznością w dobie Internetu oraz socjokulturowymi uwarunkowaniami procesu glottodydaktycznego.