

Agnieszka Dryjańska

Uniwersytet Warszawski

ORCID: 0000-0003-1649-8408

a.dryjanska@uw.edu.pl

Wykorzystanie korpusów tekstowych w nauczaniu języka francuskiego na przykładzie analizy kontrastywnej polsko-francuskiej

Use of text corpora in French Language Teaching – an example of French-Polish contrastive analysis

This paper outlines the theoretical background of data-driven learning (DDL) and its hands-on application in French Language Teaching, as well as the didactic potential of using text corpora functionalities to extract collocations. The second part focuses on the analysis of a project conducted among students of Romance Philology. Its objective was a corpus analysis of selected French terms and their Polish equivalents on the theme of social issues. The study reported a good understanding of the principles of corpus analysis by the students and an attempt to draw conclusions based on word frequency and extracted collocations. It was found that DDL could effectively deepen the development of lexical competence, especially as regards lexical combinations. Another conclusion points out that this approach allows the acquisition of extralinguistic competences – the student's autonomy in learning a foreign language, critical thinking and heuristic skills.

Keywords: text corpus, collocation, frequency, contrastive analysis, French language teaching

Słowa kluczowe: korpus tekstowy, kolokacja, frekwencja, analiza kontrastywna, dydaktyka języka francuskiego



Artykuł jest udostępniany na licencji Creative Commons – Uznanie autorstwa-Na tych samych warunkach 4.0 Międzynarodowe, <https://creativecommons.org/licenses/by-sa/4.0/>

1. Wstęp

*Należy wyjść od języka w kontekście,
aby dojść do refleksji językowej¹.*
(Oliveira Graça, Viviani, 2001 tł. własne)

W kształceniu językowym, szczególnie w kontekście filologicznym, zależy nam nie tylko na bezpośrednim prowadzeniu uczącego się do osiągnięcia poziomu C2, lecz także, co również często jest podkreślane w *Europejskim Systemie Opisu Kształcenia Językowego* (ESOKJ, 2003), na rozwijaniu jego wrażliwości językowej oraz interkulturowej, a także umiejętności heurystycznych i autonomii. Student, a później absolwent filologii, powinien też rozumieć, na czym polega fenomen znajomości języka obcego: nie można się go bowiem nauczyć perfekcyjnie raz na zawsze, konieczne jest ciągłe samodzielne korzystanie z różnorodnych źródeł, które dostarczają odpowiedzi na coraz bardziej szczegółowe pytania. W kontekście znajomości języka obcego koncepcja uczenia się przez całe życie wydaje się więc szczególnie istotna, dlatego celem kształcenia filologicznego powinno być przygotowanie studenta do wykorzystywania różnorodnych zasobów językowych w perspektywie *lifelong learning* (Ciekanski, 2014).

Tradycyjnym i cennym źródłem informacji o języku stosowanym w dydaktyce językowej są słowniki (Zajac, 2009), nowszym i dynamicznie rozwijającym się zasobem wiedzy są natomiast korpusy tekstowe. Korzystanie ze słowników pozostaje pierwszym i najbardziej intuicyjnym krokiem do wyjaśnienia kwestii leksykalnych. Nawet najobszerniejsze opracowania leksykograficzne nie zapewnią jednak odpowiednio dużej liczby przykładów, aby zilustrować znaczenie każdego wyrazu i jego użycie, dlatego warto sięgać do korpusów tekstowych, ich istotą jest bowiem wprowadzenie innego podejścia do znaczenia, które rodzi się na skrzyżowaniu leksyki i składni. Przydatność korpusów dla dydaktyki języków obcych dostrzeżono bardzo szybko, badania korpusowe w istocie rzeczy wyrosły na gruncie dydaktyki języka angielskiego jako obcego (Teubert, 2009).

Przedmiotem niniejszego artykułu jest przedstawienie podstaw teoretycznych wykorzystania analizy korpusowej w glottodydaktyce z uwzględnieniem definicji korpusu oraz kluczowych pojęć i narzędzi samej analizy, jakimi są: frekwencja oraz kolokacje wraz ze sposobami ich ekstrakcji. Przedstawimy w nim też przeprowadzony projekt, w ramach którego studenci filologii romańskiej wykorzystywali korpusy tekstowe francuski i polski w celu, po pierwsze, pogłębionego rozumienia omawianych zagadnień leksykalnych w wymiarze interkulturowym; po drugie – rozwijania również wspomnianych powyżej

¹ « Il faut partir de la langue en contexte pour arriver à la réflexion sur la langue ».

kompetencji ogólnych. Ostatnim etapem będzie szczegółowa analiza uzyskanych rezultatów w odniesieniu do różnorodnych kompetencji, najpierw omówionych w sposób teoretyczny, a następnie identyfikowanych w trakcie realizacji projektu. Naszym zamierzeniem jest ukazanie wielości i różnorodności kompetencji, które mogą być rozwijane, choć w różnym stopniu u każdego uczącego się, dzięki zastosowaniu analizy korpusowej. W związku z obszernością materiału analiza każdej z kompetencji z konieczności będzie zawężona.

2. Rys historyczny

Początek zainteresowania korpusami tekstowymi przypada na koniec lat pięćdziesiątych i lata sześćdziesiąte. Należy tu wspomnieć przede wszystkim o pracach językoznawców brytyjskich, na przykład Firtha (1962 [1957]) czy Sinclaira (1991). Pierwsze korpusy dla języka angielskiego powstały już w latach sześćdziesiątych (Bennett, 2010), np. Brown Corpus (Kučera, Francis, 1967).

Kolejna fala zainteresowań językoznawstwem korpusowym, po przezwyciężeniu spowodowanej sukcesem teorii Chomsky'ego, nastąpiła w latach dziewięćdziesiątych XX w., kiedy to powstał British National Corpus (Boulton, Tyne, 2014B), i trwa do chwili obecnej. Należy to tłumaczyć pojawieniem się nowych możliwości technologicznych umożliwiających zarówno tworzenie dużych korpusów, jak i ich szybkie przeszukiwanie. Francuskie językoznawstwo korpusowe przez dłuższy czas pozostawało sceptyczne wobec podejścia empirycznego częściowo ze względu na silną tradycję normatywną, dobrze wyrażoną przez pojęcie *le bon usage*, a także tendencje unifikacyjne we Francji, które utrudniały zainteresowanie różnorodnością użycia języka (Ledegen, Léglise, 2013). Jeden z pierwszych najważniejszych francuskich korpusów tekstowych Frantext jest dostępny od 1998² r.

Pierwszych zastosowań podejścia korpusowego w dydaktyce języków obcych można doszukiwać się już w latach dwudziestych i trzydziestych XX w., kiedy zaczęto tworzyć listy frekwencyjne na potrzeby nauczania języka angielskiego (Thorndike, 1921 za Charzyńska, 2016). Pisząc o wykorzystaniu korpusów tekstowych w nauczaniu języka angielskiego należy przede wszystkim wymienić prace Johnsa i Kinga (1991, za: Boulton, Tyne, 2014B) oraz McEnergy'ego i Wilsona (1997, za: Boulton, Tyne, 2014B). Wśród francuskojęzycznych badaczy ważne miejsce zajmują między innymi Boulton (2008), Ciekanski (2014) oraz Chambers (2009). W polskojęzycznej literaturze dotyczącej wykorzystania korpusów tekstowych w dydaktyce należy wspomnieć prace Zasiny (2018A, 2018B), Lewandowskiej (2014) czy Dryjańskiej (2020).

² <https://www.frantext.fr/>

3. Korpus tekstowy i jego definicja

Korpus tekstowy stanowi kluczowy element językoznawstwa korpusowego, dziedziny empirycznej, której głównym zainteresowaniem jest, tak ważne w glottodydaktyce, użycie języka, którego „nie da się wymyślić – można je jedynie zanotować” (Sinclair, 1987, za Rudolf, 2004: 21). Jedna z częściej cytowanych definicji korpusu brzmi:

The term corpus as used in modern linguistics can best be defined as a collection of sampled texts, written or spoken, in machine-readable form which may be annotated with various forms of linguistic information.

(McEnery i in., 2006: 4, za Beeching, 2014)

Językoznawcy francuskojęzyczni, np. Macaire, Boulton definiują korpus natomiast jako „duży zbiór tekstów autentycznych w formie elektronicznej, który jest reprezentatywny w odniesieniu do całego języka lub jego odmiany”³ (2014 tł. własne). Definicja ta akcentuje pojęcie reprezentatywności korpusu oznaczając, że „its contents can be generalized to the said language variety” (Xiao, 2010: 149). Xiao zwraca jednak uwagę, że reprezentatywność korpusu jest pojęciem ‘fluide’, gdyż w rzeczywistości jest trudna do osiągnięcia i zmierzenia, szczególnie w odniesieniu do korpusów języka ogólnego. Rozmiar korpusu jest jednakże jednym z parametrów mocno wpływających na reprezentatywność. Wielkość zbioru trzeba także brać pod uwagę w badaniach kontrastywnych, w celu zapewnienia porównywalności frekwencji leksemów konieczne jest bowiem jej znormalizowanie w odniesieniu rozmiaru korpusu. Dopiero wówczas możliwe jest odwoływanie się do frekwencji względnej lub, jak w przypadku Leipzig Corpora Collection (LCC), do klasy frekwencyjnej.

Zasina z kolei, powołując się na czeskiego badacza Čermáka (1995: 119), zwraca uwagę na jeszcze inne cechy korpusu podkreślając, że jest to zbiór „uporządkowany, ujednolicony, zwykle też indeksowany i zintegrowany” i „zaprojektowany ze względu na użycie w pewnym celu” (Zasina, 2018A: 111). Dzięki formie elektronicznej, odpowiedniej anotacji i narzędziom informatycznym możliwe jest automatyczne wyszukiwanie w nim jednostek leksykalnych występujących w „naturalnej strukturze gramatycznej i stylistycznej” (Zasina, 2018B), ułatwia to ich efektywniejszą analizę językową, a tym samym stanowi o potencjale dydaktycznym tak zorganizowanego zbioru.

Istnieją też podejścia bardziej liberalne do definiowania korpusu. Zwolennikiem takiego podejścia jest jeden z najwybitniejszych brytyjskich

³ « Une grande collection de textes authentiques, présentée sous forme électronique, et censée être représentative d'une langue ou d'une variété de langues »

językoznawców korpusowych, Kilgarriff. Zwraca on uwagę na to, że wielu badaczy myli to, czym jest korpus, z tym, jakie cechy charakteryzują dobry korpus. Jego definicja jest w związku z tym bardzo prosta i szeroka: „a corpus is a collection of texts when considered as an object of language or literary study” (Kilgarriff, Grefenstette, 2003).

W rozumieniu językoznawstwa korpusowego, korpus nie jest więc zbiorem dowolnych danych tekstowych (Xiao, 2010), lecz próbą języka wyekscerpowaną według ustalonych kryteriów zewnętrznych – sytuacyjnych (ang. *external 'situation' criteria*) (ibidem: 149), które pozostają w opozycji do wewnętrznych (ang. *internal criteria*). Kryteria zewnętrzne „are defined situationally irrespective of the distribution of linguistic features whereas internal criteria are defined linguistically, taking into account the distribution of such features” (ibidem). Korpusy opracowane w oparciu o kryteria sytuacyjne w połączeniu z wymogiem reprezentatywności stanowią w naszym przekonaniu podstawę do ich wykorzystania jako wiarygodnego⁴ narzędzia do badania języka, użytecznego także w dydaktyce języka obcego.

4. Podejście ilościowe w językoznawstwie korpusowym

Języka możemy nauczyć się w sposób nieformalny w wyniku naturalnej ekspozycji, co jest możliwe dzięki jego strukturze, gdyż „statystyczny rozkład struktur w języku w naturalnym środowisku jego użycia charakteryzuje się dużą koncentracją pewnych z nich, które stają się bardziej widoczne dla uczącego się”⁵ (Sockett, 2014 tł. własne). Frekwencja leksemów stanowi więc element ważny i z punktu widzenia akwizycji języka, i analiz korpusowych, dlatego też powinna być ona uwzględniana przy doborze materiału leksykalnego w nauczaniu języka obcego. Zjawisko częstości często jest jednak pomijane lub w najlepszym przypadku – marginalizowane. Uczący się w związku z tym często jest bezradny, gdyż musi opanować dużą ilość leksyki (proponowanej przez podręczniki czy inne materiały dydaktyczne), niekoniecznie tej częściej używanej. Przykładem może być słownictwo z zakresu turystyki prezentowane w «Le Bled Vocabulaire», które różni się częstością użycia, co mogłoby być zasygnalizowane, by uczący się miał świadomość tych

⁴ Wiarygodności korpusu nie należy mylić z poprawnością danych tekstowych zawartych w korpusie. „Niepoprawność” stanowi jeden z elementów charakterystycznych autentycznego języka (Cori, David, 2008) i w naszym przekonaniu, w kształceniu filologicznym także to zjawisko powinno znaleźć swoje właściwe miejsce.

⁵ « la distribution statistique des structures dans une langue en milieu naturel comporte aussi une forte concentration de certaines structures, augmentant ainsi la visibilité de ces dernières du point de vue de l'apprenant. »

różnic. Traktując *Web as a corpus*, zgodnie z tak zatytułowanym artykułem Kilgarriffa, przykładowo podajemy frekwencję istotnie różną kilku wyrażzeń, sprawdzoną w Google (24.08.2022 r.), ‘pris un coup de soleil’ (875 000), ‘parti à l’aventure’ (172 000), ‘roulé sa bosse’ (36 900), ‘parti avec armes et bagages’ (4 340). Frekwencja powinna też więc być jednym ze wskaźników relewantności dydaktycznej jednostek leksykalnych używanych w określonych sytuacjach komunikacyjnych⁶. Korpusy tekstowe, bez względu na dzielące je różnice, mają jedną cechę wspólną: zawierają dane o frekwencji użycia słów oraz o ich sekwencji. Dostarczają także narzędzi do ekstrakcji jednostek wyrazowych oraz ich kolokacji (kookurencji, sąsiedztwa) na podstawie algorytmów matematycznych typu: frekwencja, log-likelihood (Frantext, LCC), Chi² (NKJP), MI (Frantext)). Pierwsze trzy miary łączliwości⁷ między leksemami pozwalają na ekstrakcję połączeń jednostek wyrazowych, które są częste, natomiast Mutual Information (MI) stosuje się w identyfikacji połączeń rzadszych, ale jednak silnych (Dryjańska, 2020).

Kolejnym terminem, który wymaga wyjaśnienia w związku z analizą korpusową jest samo pojęcie *kolokacji*, jako że możliwe są jej różne definicje w zależności od stosowanego podejścia – *grosso modo* dwóch głównych: frazeologicznego i empirycznego (Evert, 2007). Definicje obowiązujące w tym drugim ujęciu, zapoczątkowanym przez Firtha (1962 (1957)), charakteryzują się wysokim poziomem ogólności, jak ta przedstawiona przez Rudolfa (2004): *kolokacja* to „współwystępujące słowa pojawiające się częściej niż inne połączenia”, W tym ujęciu ważnym elementem jest właśnie, jak widać, frekwencja pojawiania się określonych kombinacji wyrazów. W podejściu frazeologicznym z kolei *kolokacja* jest definiowana jako współwystępowanie dwóch jednostek leksykalnych (fr. *une cooccurrence lexicale binaire*) (Tutin, 2013: 49) częściowo utrwalone lub częściowo idiomatyczne (fr. *semi-figée ou semi-idiomatique*) (Tremblay, 2014: 74). W definicji tej uwagę zwraca element częściowej idiomatyczności, który trudno poddaje się formalizacji i w związku z tym nie pojawia się w korpusowym podejściu do kolokacji. LCC natomiast wprowadza pojęcie kookurencji (ang. *cooccurrence*), która jest definiowana jako „those words that occur noticeably often together”⁸. Wśród

⁶ Przy doborze leksykalnym należy także pamiętać o uwzględnieniu wymogów dyskursywnych i potrzeb uczącego się.

⁷ „Łączliwość wyrazu to zakres jego połączeń z innymi wyrazami”, jak wyjaśnia Kłosińska (Łączliwość normatywna - Poradnia językowa PWN) (DW 05.06.2023). Wyróżniamy łączliwość systemową lub normatywną, ale „w wielu przypadkach rozstrzygnięcie, z którym rodzajem łączliwości mamy do czynienia, jest trudne – często bowiem zależy od sposobu zdefiniowania kategorii, o której w danym wypadku myślimy” (Hącia, Kłosińska, 2014: 277). W korpusach tekstowych identyfikujemy przykłady obu typów łączliwości. W nauczaniu języka francuskiego wystarczające jest odwołanie się do terminu łączliwość, bez specyfikowania jej typu.

⁸ LCC / Deutscher Wortschatz (uni-leipzig.de) (DW 05.05.2023).

kookurencji możemy odnaleźć słowa występujące bezpośrednio obok siebie, wtedy nazywane są one sąsiedztwem, lub współwystępujące w jednym zdaniu. Konstrukty te są kluczowe w analizie korpusowej także dlatego, że to poprzez słowa towarzyszące danemu leksemowi docieramy do jego znaczenia zgodnie z koncepcją Firtha „You shall know a word by the company it keeps” (1962 [1957]: 11). Znaczenie w językoznawstwie korpusowym staje się więc pojęciem, które traci na swojej abstrakcyjności⁹, łatwiej poddaje się formalizacji, a w konsekwencji może być bardziej uchwytnie dla uczącego się. Dodatkowo, frekwencja pozwala na ujawnienie kulturowego wymiaru znaczenia (Wierzbicka, 2013).

5. Korpusy tekstowe w glottodydaktyce

Korpusy tekstowe dostarczają danych autentycznych (Boulton, Tyne, 2014B), są one bowiem zbiorem przykładów użycia języka, formułowanych przez bardzo wielu użytkowników w bardzo zróżnicowanych kontekstach semantycznych oraz składniowych. Oferują one nowe sposoby opisu języka, umożliwiające dokładniejsze obserwacje językowe (Vetulani, 2000: 321), które są wykorzystywane przez językoznawców. Znajdują także zastosowanie w nauce języków obcych (Ciekanski, 2014) w postaci tzw. *data-driven learning* – podejścia konstruktywistycznego, w którym uczący się przyjmuje rolę *language detective* (Johns, 2014 [1997]).

W dydaktyce języków obcych znane są dwa podejścia do korzystania z korpusów tekstowych: *hands-off (corpus-based, top-down)* i *hands-on (corpus-driven, bottom-up)* (Tognini- Bonelli, 2001). W pierwszym dane językowe są prezentowane uczącym się jako „dostosowane” (Frankenberg-Garcia, 2014), czyli poddane dydaktyzacji, np. poprzez wybór przez nauczyciela linii konkordancyjnych, które najlepiej ilustrują omawiane zagadnienie i naprowadzają studentów na właściwe wnioski, co stanowi zysk czasowy (Di Vito, 2013). W podejściu *hands-on* natomiast następuje odwrócenie tradycyjnego paradygmatu edukacyjnego: I(nitiation)-R(esponse)-F(eedback) (Sinclair, Coulthard, 1975, za: Johns, 1991), w którym nauczyciel jako bardziej aktywny aktor procesu dydaktycznego stymuluje procesy kognitywne uczącego się poprzez np. zadawanie pytania i, po odpowiedzi ucznia, udziela mu informacji zwrotnej. W pracy z korpusem tekstowym, np. w *concordance-based learning*, obowiązuje

⁹ Przykładem może być kognitywne rozumienie znaczenia, według którego jest ono utożsamiane z konceptualizacją jakiegoś fragmentu rzeczywistości (Langacker, 2021). Złożoność pojęcia, jakim jest znaczenie, czy, jak to ujmuje Wierzbicka, jego *fuzziness* (1996: 16), jest przyczyną trudności formalizacyjnych.

mechanizm: I(dentify) – C(lassify) – G(eneralise) (Johns, 1991), w którym uczący się samodzielnie formułuje zapytania do korpusu (ang. *learner-generated questions*), identyfikuje problem, a następnie dokonuje kategoryzacji uzyskanych rezultatów. Etap kategoryzacji jest szczególnie istotny w pracy z dużą ilością danych (Dryjańska, 2020), gdyż umożliwia znalezienie wśród nich prawidłowości, które następnie stają się podstawą do sformułowania wniosków. Korpus pełni wówczas rolę, którą Johns określa jako *informant*, ale jest to informator bierny, którego bierność wyzwala aktywność u uczącego się. Praca w podejściu *hands-on*, co stanowi jej dużą zaletę, pozwala uczniom rozwinąć wiele kompetencji nie tylko językowych, gdyż polega na konfrontowaniu ich i z pełnym korpusem, i jego narzędziami.

Reasumując: wykorzystanie korpusów tekstowych wpisuje się w tradycję stosowania materiałów autentycznych oraz osiągnięć technologicznych w glottodydaktyce (Ciekanski, 2014), dzięki czemu można sprawniej doskonalić kompetencje językowe, uczących się pozostawiając im jednocześnie przestrzeń na samodzielną pracę i odkrywanie ‘tajemnic’ języka.

5.1 Potrzeby współczesnego uczącego się / użytkownika Internetu

Korzystanie z korpusów tekstowych opiera się na tym samym mechanizmie, który jest stosowany obecnie przez nas wszystkich i polega na formułowaniu zapytań w wyszukiwarkach internetowych w celu dotarcia do określonych informacji (często licznych) oraz ich analizy, na przykład na podstawie kryterium częstości występowania. Dzisiejszy uczący się-użytkownik Internetu wie, że dzięki frekwencji możliwe jest uzyskanie odpowiedzi na pytanie: *Czy tak się mówi?* Nauczyciel, który nie ma dostępu do danych o frekwencji słowa czy wyrażenia, udziela odpowiedzi jedynie na podstawie własnej intuicji językowej – *armchair intuition*¹⁰ (Johns, 1991), a ta może być zawodna.

6. Kompetencje rozwijane w projekcie

6.1. Kompetencje językowe

Kompetencja leksykalna jest definiowana w ESOKJ jako „znajomość i umiejętność użycia słownictwa” (2003: 101) i łączy się ze świadomością, że może być ono wieloznaczne, a także, że może pojawiać się w związkach wyrazo-

¹⁰Vetulani także zwraca uwagę na to, że analiza korpusowa pozwala na weryfikację hipotez językowych opartych na indywidualnej intuicji lub obserwacji pojedynczych przypadków (2000: 321).

wych i kolokacjach o różnym stopniu zespolenia (ibidem). Tak rozumiane ‘poznanie’ słowa jest stosunkowo łatwo dostępne dzięki analizie korpusowej, gdyż wśród funkcjonalności korpusów znajdujemy narzędzia do ekstrakcji kolokacji lub kookurencji. W wykorzystanym w projekcie korpusie Leipzig Corpora Collection (LCC) występuje także graficzna reprezentacja łączliwości leksemów, która dodatkowo ułatwia identyfikację zjawisk polisemicznych. Dzięki funkcjonalności LCC możliwe staje się także poszukiwanie konotacji, synonimii i antonimii oraz związków frazeologicznych.

6.2. Kompetencje ogólne

Poniżej zamieszczamy krótką charakterystykę kompetencji pozajęzykowych, których elementy były rozwijane w projekcie.

Autonomia to umiejętność samodzielnej kontroli swojego procesu uczenia się (Holec, 1982: 4). Potrzeba kształtowania tej umiejętności, która jest wielokrotnie wspomniana w ESOKJ, wydaje się niezwykle ważna w kształceniu filologicznym, co zostało wyjaśnione we wstępie artykułu. Dzięki DDL można rozwijać także umiejętności heurystyczne. Stanowią one składową *savoir-apprendre* i są określane jako umiejętności „wyszukiwania i rozumienia informacji”, na przykład w bazach danych (ESOKJ, 2003: 99). W wersji francuskojęzycznej ESOKJ, umiejętności heurystyczne rozumiane są jako *zdolność do odkrywania* (fr. *aptitudes à la découverte* (CECR, 2001: 85)), co koresponduje z przedstawionymi wcześniej założeniami DDL.

ESOKJ kładzie nacisk także na kształtowanie wrażliwości interkulturowej oraz umiejętności interkulturowych. Na wrażliwość interkulturową składają się „wiedza, świadomość i rozumienie relacji między „światem społeczności pochodzenia” a „światem społeczności języka docelowego” (ESOKJ, 2003: 96). Umiejętności interkulturowe to zaś, między innymi, „umiejętność dostrzegania związku między kulturą własną a obcą” (ESOKJ, 2003: 96). Analizując różnice leksykalne i semantyczne wyekscerpowanych z korpusu wyrazów i ich połączeń odzwierciedlających różne konceptualizacje świata w języku polskim i francuskim, uczący się może więc zdobyć wiedzę, dotyczącą wybranych leksemów i na tej podstawie *konstruować* swoje rozumienie relacji pomiędzy „dwoma światami”. Kształtowanie wrażliwości i umiejętności interkulturowych to z pewnością proces długi i złożony, ale w naszym przekonaniu można go rozwijać poprzez ‘małe kroki’, tj. refleksję nad pojedynczymi słowami, która prowadzi do głębszego rozumienia istoty zróżnicowania kulturowego i językowego świata.

7. Charakterystyka projektu

W projekcie wzięło udział 20 studentów (jedna grupa zajęciowa) pierwszego roku studiów drugiego stopnia w ramach zajęć profilowanych: *Praktyczna nauka języka francuskiego: Przemiany społeczne w krajach francuskojęzycznych* (poziom językowy C1). Tematem projektu była analiza kontrastywna wybranych leksemów polskich i francuskich z zakresu zagadnień społecznych we Francji, na podstawie korpusów Leipzig Corpora Collection FR i PL. Projekt był wykonany indywidualnie w trybie asynchronicznym po zakończeniu omawiania problematyki zajęć w formie zdalnej.

W realizacji projektu można wyodrębnić cztery fazy:

1. Etap 0 – polegający na realizacji tematyki zajęć
2. Etap I – obejmujący krótkie wprowadzenie do problematyki korpusów tekstowych z koncentracją na następujących zagadnieniach:
 - Co to jest korpus? (na przykładzie LCC)
 - Jak korzystać z korpusu?
 - Jakie funkcjonalności będą potrzebne do realizacji projektu? (frekwencja, konkordancja, klasa frekwencyjna, ekstrakcja współwystępujących jednostek wyrazowych, kolokacja, kookurencja)?
 - Prezentacja przykładowej analizy korpusowej o charakterze kontrastywnym.
3. Etap II – to samodzielna praca studentów polegająca na wyborze dowolnego leksemu w języku francuskim i jego ekwiwalentu (ekwiwalentów) w języku polskim pozostającego w zakresie tematyki społecznej, a następnie poddaniu analizie kontrastywnej wyselekcjonowanej pary na podstawie danych z LCC FR i PL.
4. Etap III – ocena projektu przez nauczyciela oraz indywidualna informacja zwrotna w formie pisemnej.

7.1. Analiza rezultatów

W ramach projektu studenci samodzielnie dokonywali doboru słów, które mieli poddać analizie, jednakże musiały łączyć się one z problematyką zajęć, aby znany był szerszy kontekst społeczny ułatwiający analizę. Poniższa tabela przedstawia 14 leksemów wybranych przez studentów.

Działania zmierzające do wprowadzania równych praw społecznych były najczęściej wybieraną problematyką (6 osób). Po nich następują dwie kategorie: sytuacja kobiet, interesująca dla czterech osób oraz ideały republikańskie, także wybrane przez cztery osoby. Po trzy osoby wybrały problematykę religijną oraz kwestie związane z imigracją i tożsamością.

Tabela 1. Analizowane leksemy

Słowa analizowane	Zakresy tematyczne
(im)migration vs (i)migracja (A11) (A20) identité vs tożsamość (A2)	Wyzwania zewnętrzne
fraternité vs braterstwo (A17) liberté vs wolność (A18) prawa/o vs droit/s (A15) (A19)	Ideale republikańskie
activisme vs aktywizm (A4) grève vs strajk (A3) (A14) rebelle vs buntownik (A5) révolution vs rewolucja (A10) (A13)	Walka o prawa społeczne
laïcité vs świeckość (A1) religion vs religia (A7) (A16)	Miejsce religii w państwie
mode vs moda (A9) soumission vs uległość (A6) violence vs przemoc (A8) (A12)	Sytuacja kobiet

7.1.1. Pierwszy etap analizy – analiza ilościowa

Na tym etapie zajęliśmy się badaniem ogólnym całego korpusu zawierającego wszystkie prace studentów, z wyłączeniem cytowanych przykładów koncordancyjnych. Korpus obejmował 8 458 tokenów. Do jego wstępnej analizy wykorzystany został program do automatycznej analizy tekstu NooJ, który umożliwił analizę frekwencyjną jednostek jedno- i dwuwyrazowych. Wszyscy studenci, co niezwykle istotne, zwrócili uwagę na rolę kontekstu dla znaczenia wyrazu – leksem *contexte(s)* (pl. kontekst(y)) użyty był 61 razy, w tym w połączeniu *contexte(s) commun(s)* (16) (pl. wspólny(e) kontekst(y)) oraz frekwencji – sam leksem *fréquence* (pl. frekwencja, częstość) użyty został 31 razy, a *fréquent* (pl. częsty) i *souvent* (pl. często) pojawiły się 58 razy, w tym w połączeniu dwuwyrazowym *souvent utilisé* (18). Studenci pisali też o frekwencji leksemów w kontekście kontrastywnym używając 8 razy syntagm przysłówkowych *moins / plus fréquemment* (pl. rzadziej / częściej) lub *plus souvent* (24). Nie ograniczali się przy tym do podejścia ilościowego, ale w różnorodny sposób analizowali uzyskane kolokacje – leksem *collocation/s* został użyty 71 razy, w tym 8 razy, gdy pisano o *forte collocation* (pl. silna kolokacja).

Etap wstępny pokazał, że studenci opanowali podstawy mechanizmu pracy z danymi korpusowymi i potrafili wykorzystać je w praktyce.

7.1.2. Analiza jakościowa

Na tym etapie analizy wyodrębnionych zostało 11 kategorii obejmujących zróżnicowane umiejętności, których aspekty udało się dostrzec w wypowiedziach studenckich. Zostały one przedstawione na poniższym wykresie w połączeniu z liczbą prac, w których je dostrzeżono.

Tabela 2. Kategorie rozwiniętych kompetencji

Nr	Kategoria	Liczba prac, w których zaobserwowano daną kategorię	
1	Analiza frekwencyjna badanego leksemu	20	
2	Analiza kolokacyjna	20	
3	Autonomia we wnioskowaniu i/lub myślenie krytyczne szczególnie w odniesieniu do funkcjonalności korpusu	20	
4	Podobieństwa kontekstowe – dostrzeganie niuansów znaczeniowych	5	
5	Różnice kontekstowe/kolokacyjne	– komentarze o charakterze politycznym	7
6		– komentarze o charakterze społeczno-religijnym	5
7		– komentarze o charakterze historyczno-socjologicznym	5
8	Różnice gramatyczne	2	
9	Synonimia i polisemia	2	
10	Brak ekwiwalencji	1	
11	Prozodia semantyczna	1	

Zgodnie z wnioskami wyływającymi z etapu wstępnego, wszyscy studenci dokonali analizy frekwencyjnej i kolokacyjnej badanych leksemów. Ich analiza, co ważne, nie ograniczała się do stwierdzenia, że frekwencja danego leksemu jest wyższa lub niższa. Zwrócili uwagę, że fakt ten w połączeniu z niższą wartością klasy implikuje, że, na przykład, pojęcie *laïcité* jest bardziej zakorzenione w kulturze francuskiej niż *świeckość* w polskiej (A1).

We wszystkich pracach zaobserwowano samodzielne, bardzo zindywidualizowane refleksje, w niektórych pojawiła się dodatkowa składowa emocyjna – zaskoczenie¹¹ – wyrażona za pomocą fraz: *c'est surprenant* (pl. to zaskakujące), *ce n'est pas étonnant* (pl. nie jest to zdumiewające), *c'est intéressant* (pl. jest interesujące). Ten aspekt zwraca uwagę, gdyż wykracza poza zakładane w projekcie cele. Autor pracy badającej leksemu *activisme*

¹¹ Jedno z uczuć prymarnych według klasyfikacji Ekmana (Dortier, 2020).

i *aktywizm* (A4) wyraził zdumienie, że wśród uzyskanych kolokacji brak jest przymiotników, takich jak *ekologiczny*, *lokalny* czy *humanitarny*, a przecież zarówno we Francji jak i w Polsce bardzo często je słyszymy, gdy mowa jest o aktywizmie. Osoba analizująca leksemy *révolution* i *rewolucja* (A10) także była zaskoczona, że wśród przymiotników, które występują często ze słowem *révolution* zabrakło przymiotnika *sexuelle*. Według niej, Polacy kojarzą Francję, między innymi, z rewolucją seksualną ze względu np. na Simone de Beauvoir i legalizację aborcji. Zabrakło tu wprawdzie obserwacji, że w języku francuskim częściej używamy kolokacji *libération sexuelle* lub *liberté sexuelle* (pl. wyzwolenie, wolność seksualna)¹², co wynikało zapewne w zbyt małym jeszcze doświadczeniu w analizach korpusowych. W przypadku braku spodziewanego ekwiwalentu kolokacji, np. *rewolucja seksualna*, warto zastanowić się nad samym przymiotnikiem *seksualna* i, poszukując jego kolokacji, można ją odnaleźć właśnie w *libération sexuelle*, która częściowo jest ekwiwalentna z rewolucją seksualną.

Tego typu dojrzała refleksja miała natomiast miejsce w przypadku analizy kolokacji: *violence verbale* (częsta w korpusie fr.) i *przemoc psychiczna* (częsta w korpusie pl.) (A8). Autor zwrócił uwagę, że kolokacje te, mimo że zbliżone semantycznie, nie są w pełni ekwiwalentne. Wskazał również na interesującą różnicę: *violence* jest bardzo często używana w języku francuskim ze słowem *femme*, natomiast *przemoc* w języku polskim z leksemem *dziecko*. Oba typy przemocy są zbliżone konceptualnie, gdyż dotyczą środowiska rodzinnego (*violence domestique*), jednakże ofiara przemocy jest inna. Analiza poprowadziła autora do słusznej konkluzji, że problemy społeczne i podejście do nich znajdują odzwierciedlenie w języku. Wniosek, choć oczywisty dla językoznawcy, jest cenny ze względu na to, że został samodzielnie sformułowany przez studenta.

To, co wydaje się najistotniejsze w przedstawionych powyżej kilku przykładowych konkluzjach, to widoczny konflikt kognitywny – autorzy prac spodziewali się czegoś innego, uzyskali odmienny rezultat i próbowali go zrozumieć. Dokonywali też klasyfikacji uzyskanych danych i samodzielnego wnioskowania prowadzącego, co ważne, do pewnej, acz ostrożnej, generalizacji. Mimo że korpus tekstowy jest nowym narzędziem technicznym analizy językowej, z którym większość studentów zetknęła się po raz pierwszy, to jednak udało im się odnieść krytycznie także do pewnych kwestii technicznych, np. do zróżnicowanej reprezentatywności, wiążącej się z niekompatybilnością rozmiarową korpusów polskiego i francuskiego, która mogła wpływać na analizę (A10) (zob. 3).

¹² W korpusach Web 2012, Web 2020 w Sketch Engine, w kategorii rzeczowników modyfikowanych przez dany przymiotnik pojawiają się *liberté* i *libération*, natomiast *révolution* się nie pojawia.

W analizowanych pracach widoczna jest także różnorodność dziedzinowa wybranych przykładów z konkordancji oraz przedstawianych komentarzy i wniosków, reprezentowana przez kategorie 5–7. Najczęstsze są przykłady i komentarze o charakterze:

– politycznym:

(A6) przytaczając zdanie: „Jakoś dziwną uległość Tuska wobec Putina zauważają WSZYSCY! (www.radiozet.pl, *rampé* 20.04.2012), zauważa, że leksem *uległość* w korpusie polskim jest silnie związany z polityką oraz z postawą wobec innych krajów; jest to widoczne na podstawie sąsiedztwa tego leksemu ze słowami *Moskwa*, *Rosja*, czego nie widać w przypadku analizy leksemu *soumission*; w jego przypadku znajdujemy wręcz przykład o zupełnie innym charakterze: „Nous refusons la soumission et la crainte, et continuerons à pratiquer l'autodétermination, dans une perspective de libération collective (pl. Odrzucamy uległość oraz strach i nadal będziemy praktykować samostanowienie, dążąc do zbiorowego wyzwolenia. tł. własne) (www.cmaq.net, *rampé* 16.11.2007);

– społeczno-religijnym:

osoba (A9), analizując leksem *moda* zwraca uwagę na częste sąsiedztwo leksemu *krzyż*, co stara się wytłumaczyć obserwacją, że w Polsce od pewnego czasu nosi się na szyi krzyżyk nie jako symbol religijny, ale ze względu na modę;

– historycznym:

osoba analizująca leksem *religion* (A17) zwraca uwagę na sąsiedztwo leksemów *religion* i *guerre*, które początkowo ją dziwi, lecz zaraz przywołuje przyczyny historyczne tej sytuacji.

Dwie kategorie 8 i 9, które zostały dostrzeżone w czterech pracach, dotyczyły kwestii gramatycznych lub polisemicznych. Studenci A5 i A9 zaobserwowali, że leksem *rebelle*, a nawet *mode* mogą być w języku francuskim użyte jako przymiotniki, w przeciwieństwie do ich ekwiwalentów, jakimi są *buntownik* i *moda*. W analizie leksemu *mode*, zwrócono też uwagę, zgodnie z definicją ze słownika Le Robert, na jego polisemiczność, z którą wiąże się zmiana rodzaju.

Ciekawą kategorią, która została wydzielona, choć pojawiła się tylko w jednej pracy była: prozodia semantyczna (11). Osoba analizująca leksem *fraternité* i *braterstwo* (A18) zwróciła uwagę, że wśród słów, które często występują w ich sąsiedztwie w języku polskim, pojawiają się, między innymi: *wojennymi*, *bronii*, *oszusta*, które mają zabarwienie negatywne. Nie koreluje to z konotacją wyrazów *liberté*, *solidarité*, *égalité*, *paix*, *valeurs*, *tolérance* najczęściej występujących z leksemem *fraternité*. Zaobserwowane zjawisko

jest zdefiniowane jako *semantic prosody* oznaczające „aura of meaning with which a form is imbued by its collocates” (Louw, 1993).

8. Konkluzja

Celem powyższej analizy, z konieczności selektywnej, było pokazanie, że wprowadzenie DDL w dydaktyce języka francuskiego może przyczynić się do rozwijania kompetencji leksykalnej i semantycznej uczących się w oparciu o łączliwość wyrazów. Ponadto, analiza prac studentów, ich komentarzy, wniosków, przykładów z konkordancji ujawniła przejawy ich autonomii w zakresie doboru problematyki, kategoryzacji linii konkordancji i kolokacji oraz wnioskowania na temat języka na podstawie generalizacji¹³ (Dubois i in., 1973) zgodnie z modelem ICG Johnsa. Komponent, który pojawił się nieoczekiwanie, to sformułowania świadczące o zaangażowaniu emocjonalnym studentów. Jest to szczególnie cenne, gdyż w dobie zalewu informacyjnego, którego celem jest zwrócenie uwagi odbiorcy poprzez oddziaływanie na jego emocje, wzbudzenie zaangażowania emocjonalnego w procesie dydaktycznym stanowi wyzwanie, a jednocześnie jest pożądane, jako jeden z czynników warunkujących jego skuteczność.

BIBLIOGRAFIA

- Beeching K. (2014), *Corpora in language teaching and learning*. „Les Cahiers de l'Alcedle”, nr 11–1. Online: <https://doi.org/10.4000/rdlc.1672> [DW 24.08.2022].
- Bennett G.R. (2010), *Using Corpora in the Language Learning Classroom: Corpus Linguistics for Teachers*. Ann Arbor: The University of Michigan Press. Online: [9780472033850-part1.pdf](https://www.umich.edu/~lsl/9780472033850-part1.pdf) (umich.edu) [DW 24.08.2022].
- Berlio D., Lequeux R., Chat A.-L. (2014), *Le Bled Vocabulaire*. Paris: Hachette.
- Boulton A. (2008), *Esprit de corpus: Promouvoir l'exploitation de corpus en apprentissage des langues*. „Texte et corpus”, nr 3, s. 37–46. Online: https://www.researchgate.net/publication/29598957_Esprit_de_corpus_Promouvoir_l'exploitation_de_corpus_en_apprentissage_des_langues [DW 24.08.2022].
- Boulton A., Tyne H. (2014A), *Des documents authentiques aux corpus : démarches pour l'apprentissage des langues*. Paris: Didier.
- Boulton A., Tyne H. (2014B), Corpus-based study of language and teacher education, (w:) Bigelow M., Ennser-Kananen J. (red.), *The Routledge Handbook of Educational Linguistics*. New York: Routledge, s. 301–312. Online: https://www.researchgate.net/publication/281987589_Corpus-based_study_of_language_and_teacher_education [DW 24.08.2022].

¹³ « On infère la langue du corpus par généralisation » (Dubois i in., 1973: 165).

- Ciekanski M. (2014), *Les corpus : de nouvelles perspectives pour l'apprentissage des langues en autonomie ?* „Les Cahiers d'Acledle”, nr 11–1. Online: <https://journals.openedition.org/rdlc/1710> [DW 24.08.2022].
- Chambers A. (2009), *Les corpus oraux en français langue étrangère : authenticité et pédagogie.* „Mélanges CRAPEL”, nr 31, s. 15–34. Online: <http://fil.univ-bouira.dz/wp-content/uploads/2020/08/LES-CORPUS-ORAUX-EN-FRAN%C3%87AIS-LANGUE-ETRANGERE-.pdf> [DW 24.08.2022].
- Charzyńska E. i in. (2016), *Historia badań nad zrozumiałością tekstu*, (w:) Gruszczynski W., Ogrodniczuk M. (red.), *Jasnopis, czyli mierzenie zrozumiałości polskich tekstów użytkowych*. ASPRA-JREditors, s. 11–37. Online: https://www.researchgate.net/publication/303875967_Historia_badan_nad_zrozumialoscia_tekstu [DW 24.08.2022].
- Conseil de l'Europe (2001), *Cadre européen commun de référence pour les langues : apprendre, enseigner, évaluer*. Online: <https://www.coe.int/fr/web/common-european-framework-reference-languages> [DW 24.08.2022].
- Cori M., David S. (2008), *Les corpus fondent-ils une nouvelle linguistique ?* „Langages”, nr 171, s. 111–129. Online: <https://www.cairn.info/revue-langages-2008-3-page-111.htm> [DW 24.08.2022].
- Di Vito S. (2013), *L'utilisation des corpus dans l'analyse linguistique et dans l'apprentissage du FLE.* „Linx”, nr 68–69, s. 159–176. Online: <https://journals.openedition.org/linx/1519> [DW 24.08.2022].
- Dortier J.-F. (2014), *Les émotions en questions*, (w:) Dortier J.-F. (red.), *Le cerveau et la pensée. Le nouvel âge des sciences cognitives*. Auxerre: Éditions Sciences Humaines, « Synthèse », s. 271–277. Online: <https://www.cairn.info/-9782361060466-page-271.htm> [DW 24.08.2022].
- Dryjańska A. (2020), *Patrimoine vs. dziedzictwo – interculturality in French language teaching.* „text und diskurs”, nr 13, s. 175–193.
- Dubois J. (1993), *Dictionnaire de linguistique*. Paris: Librairie Larousse.
- Evert S. (2009), *58. Corpora and collocations*. (w:) Lüdeling A., Kytö M. (red.), *An International Handbook*. Berlin, New York: De Gruyter Mouton, s. 1212–1248.
- Firth J.R. i in. 1962 (1957), *Studies of linguistic analysis*, Oxford: Basil Blackwell.
- Frankenberg-Garcia A. (2014), *How language learners can benefit from corpora, or not.* „Les Cahiers de l'Acledle”, nr 11–1. Online: <https://journals.openedition.org/rdlc/1710> [DW 24.08.2022].
- Hącia A., Kłosińska K. (2014), *Poprawność leksykalna – łączliwość*, (w:) Będkowski M. i in. (red.), *Nowe formy i normy czyli poprawna polszczyzna w praktyce*. Warszawa: Wydawnictwo Naukowe PWN SA, s. 276–281.
- Holec H. (1982), *Autonomie et apprentissage des langues étrangères*. Paris: Hatier.
- Johns T. (1991), *Should you be persuaded: Two samples of data-driven learning materials*, (w:) Johns T., King P. (red.), *Classroom Concordancing*. ELR Journal, nr 4, s. 1–16. Online: https://lexically.net/wordsmith/corpus_linguistics_links/Tim%20Johns%20and%20DDL.pdf [DW 24.08.2022].
- Kilgariff A., Grefenstette G. (2003), *Web as Corpus.* „Computational Linguistics”, nr 29. Online: <https://www.kilgariff.co.uk/Publications/2003-KilGrefenstette-WACIntro.pdf> [DW 24.08.2022]

- Zajac J. (2009), *La compétence lexicale au service des représentations culturelles des apprenants en langues étrangères*, (w:) Actes du Congrès international, Université Aristote de Thessaloniki du 12 au 14 décembre 2008, s. 588–596.
- Zasina A. (2018A), *Korpusy językowe w nauczaniu języków obcych – metoda, narzędzia, praktyka*, (w:) Jedynak M. (red.), *Specyficzne potrzeby studentów szkół wyższych a nauczanie języków obcych*, t. II: *Praktyczne narzędzia*, s. 110–123.
- Zasina A. (2018B), *Językoznawstwo korpusowe. Empiryczne podejście w badaniach humanistycznych*. „Dziennikarstwo i Media”, nr 9, s. 169–178.

Received:13.09.2022

Revised: 18.04.2023