

ELŻBIETA KACZMARSKA

Instytut Sławistyki Zachodniej i Południowej
Uniwersytet Warszawski, Warszawa
e-mail: e.h.kaczmarska@uw.edu.pl
ORCID 0000-0002-8838-1404

ADRIAN JAN ZASINA

Ústav Českého národního korpusu
Filozofická fakulta, Univerzita Karlova, Praha
e-mail: adrian.zasina@ff.cuni.cz
ORCID 0000-0001-9348-5833

BŁĘDY WALENCYJNE W TEKSTACH OBCOKRAJOWCÓW UCZĄCYCH SIĘ JĘZYKA POLSKIEGO W ŚWIETLE KORPUSU PoLKo

VALENCY ERRORS IN THE TEXTS OF NON-NATIVE SPEAKERS
LEARNING POLISH AS A FOREIGN LANGUAGE IN THE LIGHT
OF CORPUS PoLKo

ABSTRACT: The present article addresses valency errors in writings of non-native speakers learning Polish as a foreign language. Valency is a key element in a foreign language acquisition, and yet there are no studies on valency errors based on empirical data for Polish as a foreign language. Therefore, this study presents the first attempt to examine valency errors based on data from the Polish Learner Corpus PoLKo. The pilot analysis deals with different proficiency levels (A1–C1) and nationalities (Slavic and non-Slavic) of learners. The corpus material has shown that valency errors are present across all language levels and for different nationalities. Nevertheless, valency errors are more common in prepositional phrases among learners with uninflected mother tongue. Further investigation into the subject is obviously needed, and it may be expected that it will bring new results and new conclusions, along with the development of PoLKo.

KEYWORDS: valency, Polish as a foreign language, learner corpus

SŁOWA KLUCZOWE: walencja, język polski jako obcy, korpus uczniowski

Wstęp

Błędy językowe obcokrajowców piszących po polsku są przedmiotem coraz większej liczby prac (Dąbrowska 2004; Dąbrowska i Pasięka 2008; Kita i in. 2008; Krawczuk 2009; Skura 2013; Dąbrowska i Pasięka 2014; Górska 2015; Kowalewski 2018; Skura 2018). Brakuje jednak wśród nich opracowań dotyczących błędów w walencji¹, choć to właśnie walencja jest istotna w procesie akwizycji języka obcego, dlatego badanie uchybień wobec niej w tekstach nierodzimyich użytkowników języka polskiego jest konieczne dla optymalizacji metod nauczania i materiałów dydaktycznych. Uczniowie spotykają się z walencją już na samym początku nauki języka, choćby w takich strukturach, jak np. *jechać pociągami, iść do kina, robić zakupy* itp., dlatego warto na to zagadnienie zwrócić baczniejszą uwagę.

Właśnie temu tematowi poświęcony jest niniejszy artykuł. Jest to badanie pilotażowe, w którym analizować będziemy błędy łączliwości syntaktycznej pojawiające się w pracach zebranych w pierwszym polskim korpusie uczniowskim – PoLKO².

Naszym celem jest dostarczenie nowych informacji w świetle uchybień wobec walencji w tekstach pisanych przez obcokrajowców uczących się języka polskiego jako obcego. Korpusowy materiał umożliwi identyfikację typowych błędów dzięki empirycznemu poznaniu. W takim kontekście możemy mówić o analizie korpusowej języka uczniowskiego.

¹ O niedostatku badań nad błędami składniowymi cudzoziemców pisała Anna Dąbrowska: „Od wielu lat Anna Dąbrowska i Małgorzata Pasięka prowadzą badania polszczyzny cudzoziemców, koncentrując się przede wszystkim na typach popełnianych przez nich błędów. Szeroko zakrojone badania przynoszą ciekawe rezultaty, wykorzystywane przede wszystkim w procesie glottodydaktycznym. Mają też dużą wartość poznawczą, ponieważ wyraźnie pokazują kłopotliwe dla cudzoziemców problemy gramatyczne, w tym składniowe, nie tak często będące przedmiotem badań glottodydaktyków. Ich wyniki dają empiryczną, nieintuicyjną podstawę opracowywania zagadnień gramatycznych i leksykalnych.” (2014, s. 260)

² Jak ukazała w swej rozprawie doktorskiej Marta Skura (Skura 2018), popełniane przez Niemców błędy składniowe (w tym – walencyjne) stanowiły największą grupę (40%) analizowanych przez nią błędów.

Korpusy języka uczniowskiego

Analiza korpusowa języka uczniowskiego jest stosunkowo młodą dziedziną, która zaczęła się rozwijać wraz z powstaniem pierwszych korpusów zawierających teksty nierodzimych użytkowników języka, tzw. korpusów uczniowskich (ang. *learner corpora*; Gilquin i in. 2007, s. 320). Zainteresowanie tworzeniem tychże korpusów zaczęło się pojawiać wśród nauczycieli akademickich, specjalistów do nauczania języka angielskiego jako obcego oraz wydawnictw naukowych na początku lat 90. XX wieku (Granger 1998, s. 3–4). W tym okresie rozpoczęto prace nad jednymi z pierwszych korpusów uczniowskich, np. *International Corpus of Learner English* (ICLE), *Longman Learners' Corpus* (LLC), czy *Hong Kong University of Science and Technology Learner Corpus* (HKUST). Dwa pierwsze gromadzą teksty tworzone przez wszystkich uczących się języka angielskiego bez względu na język ojczysty, natomiast HKUST zawiera teksty uczniów z językiem chińskim jako pierwszym.

Analiza korpusowa języka uczniowskiego łączy w sobie dwie niezależne dyscypliny: językoznawstwo korpusowe i analizę języka obcego / drugiego (Granger 2002, s. 4). Wykorzystanie głównych założeń, metod i narzędzi językoznawstwa korpusowego pozwala efektywniej poznawać i opisywać charakter języka uczniowskiego. Jego szczegółowy opis może być wykorzystany do wielu celów w badaniach akwizycji języka obcego / drugiego, a także może się przyczynić do ulepszenia metod nauczania. Na przykład, Nesselhauf (2004, s. 126–127) wymienia trzy sposoby zastosowania korpusu uczniowskiego:

- identyfikacja problematycznych obszarów językowych danej grupy uczniów, na które następnie należy kłaść bezpośredni nacisk w różnych materiałach dydaktycznych,
- wyciąganie wniosków na temat przyswajania drugiego języka na podstawie analizy korpusu uczniowskiego oraz wykorzystanie tych spostrzeżeń w nauczaniu,
- zastosowanie danych z korpusów uczniowskich w pracy z uczniami podczas zajęć.

Próby analizy języka uczniowskiego za pomocą metod korpusowych są również podejmowane w Polsce. Jednym z pierwszych korpusów uczniowskich jest *PELCRA Learner English Corpus* (PLEC; <http://pelcra.pl/plec>) powstały na Uniwersytecie Łódzkim. Jego celem jest analiza języka angielskiego studiowanego przez Polaków. PLEC zawiera 3 miliony słów w tekstach zarówno pisanych, jak i mówionych. Jeżeli chodzi o język polski, na przełomie XX i XXI wieku na Uniwersytecie Wrocławskim pojawiła się inicjatywa zebrania błędów obcokrajowców uczących się polszczyzny. Powstał wtedy tzw. korpus DAMA będący bazą danych błędów, która została utworzona na podstawie prac pisemnych cudzoziemców na poziomie średnio zaawansowanym i zaawansowanym (Dąbrowska

i Pasięka 2014). DAMA nie zawiera całych prac studentów, tylko ich fragmenty (zdania) z przykładami błędów językowych. Ze względów technicznych baza nie jest dostępna publicznie.

Potrzeba stworzenia pierwszego korpusu uczniowskiego języka polskiego jako obcego była już kilkakrotnie poruszana przez nauczycieli i badaczy języka polskiego jako obcego (Chachulska 2005; Przybylska 2008; Zasina 2019). Ostatecznie, w październiku 2019 roku dwójka badaczy, Elżbieta Kaczmarek z Uniwersytetu Warszawskiego i Adrian Jan Zasina z Uniwersytetu Karola, rozpoczęła prace nad zbiorem tekstów i budowaniem potrzebnej infrastruktury (Zasina i Kaczmarek 2020). Powstający Korpus Uczniowski Języka Polskiego PoLKO (*Polish Learner Corpus*) ma być odpowiedzią na współczesne potrzeby dydaktyki języka polskiego jako obcego.

Korpus PoLKO

PoLKO jest międzynarodowym niekomercyjnym akademickim projektem, którego celem jest stworzenie pierwszego obszernego zbioru tekstów nierodzimych użytkowników języka polskiego. W chwili obecnej korpus zawiera ponad 2000 słów i stale się powiększa. Gromadzi prace pisemne uczniów w różnym wieku, z różnych krajów i na wszystkich poziomach kompetencji językowej³. Materiał ma posłużyć jako źródło empirycznej analizy języka obcokrajowców uczących się języka polskiego. Dzięki temu będzie również możliwa identyfikacja najczęstszych błędów językowych, która da sposobność do dostosowania materiałów dydaktycznych do konkretnych potrzeb uczniów.

Korpus jest w chwili obecnej opracowywany w środowisku TEITOK (Janssen 2016), który jest narzędziem umożliwiającym tworzenie, anotację oraz dystrybucję korpusu. Dzięki temu nowo powstający korpus można z łatwością modyfikować. TEITOK zawiera kilka wbudowanych narzędzi do automatycznego przetwarzania języka naturalnego (np. automatyczną tokenizację) oraz pozwala na konfigurację nowych narzędzi w zależności od potrzeb danego korpusu. TEITOK jest wyposażony w korpusową wyszukiwarkę łączącą się z lokalnym serwerem CQP (*Corpus Query Processor*), która wyświetla wyniki w postaci klasycznej konkordancji wraz z kluczowym słowem w kontekście, tzw. KWIC (zob. Obrazek 1). TEITOK pełni także funkcję bazy danych, gdzie gromadzone są wszystkie teksty wchodzące w skład korpusu.

³ Zbieramy prace tylko tych osób, które wyrażą na to pisemną zgodę i wypełnią metryczkę tekstu zawierającą metadane.

context W 2019 roku została premierem i od tego czasu stara się
 context nie jest spontaniczna pewna siebie i może wątpi o sobie.
 context że jest wytrwała, celowa i odważna. Żaden polityk od
 context trudnej pozycji w swoim kraju i na scenie międzynarodowej. I W
 context na obcasie z wzorem lamparta i charakterystyczną biżuterię.
 context myję zęby, ubieram się i robię makijaż. Na śniadanie
 context szynką albo jogurt z płatkami i mocną czarną herbatę. O
 context idę Biblioteki Narodowej czytać książki i pracować na artykułach akademickich.
 context kuchniu. Zwykle pracuję długo i okolo siudmej idę robić zakupy
 context chleb, jajka, owoc i mleko. Wieczorem czasami oglądamy

Obrazek 1. Wynik wyszukiwania w korpusie PoLKO za pośrednictwem TEITOK

Korpus PoLKO posiada szczegółowe informacje o tekstach wchodzących w skład korpusu oraz o ich autorach. Przy czym tożsamość autorów jest anonimizowana. Metadane dzielą się na dwie grupy: a) informacje dotyczące respondenta (zob. Tabela 1), b) informacje dotyczące tekstu (zob. Tabela 2). Opis metadanych korpusu PoLKO został oparty na systemie metadanych korpusu uczniowskiego języka czeskiego CZEsl (Rosen i in. 2020; Rosen 2015; Šebesta 2012). Szeroki wachlarz parametrów umożliwi wszechstronne naukowe wykorzystanie korpusu przez różnych badaczy, którzy na podstawie metadanych mogą sortować materiał w zależności od celów swojej analizy. Dzięki temu korpus nie ogranicza się tylko do potrzeb ich twórców, ale jest też cennym źródłem do badań akwizycji języka polskiego jako obcego.

Wielowartościowość	Atrybut	Wartość	Komentarz
–	s_id	<ciąg znaków>	Unikatowy identyfikator studenta, np. AJZ_M
–	s_sex	m; f; nb	Płeć studenta
–	s_age	<liczba całkowita>	Wiek w latach
–	s_age_cat	0-12 dzieci 13-18 młodzież 19-25 młodzi dorośli 26-60 dorośli 60- seniorzy	Wiekowa kategoria
–	s_L1	<kod ISO>	Pierwszy język studenta
–	s_L_group	S; IE; NIE	Rodzina językowa L1

Wielowartościowość	Atrybut	Wartość	Komentarz
+	s_foreign_langs	<kod ISO>	Inne języki
-	s_pl_CERF	A1; A2; B1; B2; C1; C2	Poziom języka polskiego
+	s_pl_in_family	matka; ojciec; rodzeństwo; partner / partnerka; inne; żadne	Związki językowe z rodziną
-	s_years_in_PL	-1; 1; -2; 2; -3; 3; 3-	Długość stałego pobytu w Polsce
+	s_study_pl	prywatnie; samouk; uczelnia; za granicą; SP; SŚ; kurs	Sposób nauki języka polskiego
-	s_study_pl_months	-3; 4-6; 7-12; 13-24; 25-36; 37-48; 49-60; 61-	Okres nauki języka polskiego w miesiącach
-	s_study_pl_hrs_week	-3; 3-14; 15-	Ilość godzin nauki w tygodniu
+	s_textbook	Hurra; KpK; ...	Nazwa podręcznika, z którego uczył się student
-	s_bilingual	tak; nie	Dwujęzyczność studenta

Tabela 1. Metadane respondenta korpusu PoLko

Wielowartościowość	Atrybut	Wartość	Komentarz
-	t_id	<ciąg znaków>	Unikatowy identyfikator tekstu, np. AZ_SOUK_TEST_JK_F_01
-	t_date	<data>	Data napisania pracy, RRRR-MM-DD
-	t_medium	ręcznie; komputer	Sposób pisania pracy
-	t_limit_minutes	10; 15; 20; 30; 40; 45; 60; inny; nie	Czasowy limit na pisanie

Wielowartościowość	Atrybut	Wartość	Komentarz
+	t_aid	tak; słownik; podręcznik; inne; nie	Możliwość korzystania z pomocy naukowych
+	t_exam	tak; końcowy; semestralny; częstkowy	Praca pisana podczas egzaminu
-	t_limit_ words	<ul style="list-style-type: none"> • bez limitu • 0–30 • 31–70 • 71–100 • 101–170 • 171–250 • 250– 	Limit słów
-	t_title	<ciąg znaków>	Nazwa pracy
-	t_topic	narzucony; wolny	Specyfikacja zadania
-	t_topic_as- signed	do wyboru; określony; wolny; inny	Zadanie tematu do wyboru, konkretne, wolne
-	t_genre_as- signed	określony; wolny	Sposób opracowania zadania był dowolny czy ściśle określony
-	t_genre_ predominant	informacja; opis; list; argumentacja; narracja	Gatunek pracy
-	t_words_ count	<liczba całkowita>	Liczba słów
-	t_words_ range	-50; 50–99; 100–149; 150–199; 200–	Przedział liczby słów

Tabela 2. Metadane tekstów korpusu PoLKO

W korpusie PoLKO oznaczane są korekty autorskie (dotyczy to prac pisanych ręcznie), jak np. skreślenie, przesunięcie słowa, rozdzielenie słów, połączenie dwóch słów itd. (zob. Obrazek 2). W przyszłości planujemy opatrzyć wszystkie teksty podwójną anotacją błędów, która do tekstów zostanie wprowadzona ręcznie. Pierwszą implementacją anotacji błędów będzie wprowadzenie różnych ich poziomów zgodnie z wypracowanym przez Państwową Komisję do spraw Poświadczania Znajomości Języka Polskiego systemem, który

dzieli błędy na: gramatyczne, leksykalne, stylistyczne, ortograficzne i interpunkcyjne. Drugą implementacją, która będzie wymagała dogłębnej analizy typów błędów, będzie szczegółowa anotacja błędów z opisem konkretnego problemu (np. brakujący diakryt, błędny rodzaj gramatyczny itp.).

AZ_SOUK_TEST_JK_F_03

Title

Tereza May jest brytyjską politykiem. W 2019 roku została premierem i od tego czasu stara się, żeby Wielka Brytania wychodziła z Unii Europejskiej. Jej pozycja w Partii Konserwatywnej nie jest za **bardzo silna między innymi też** z powodu pewnych cech jej charakteru.

Logicznie, May jest bardzo ambitną kobietą. Została tylko drugim preimrem – kobietą w historii Brytanii po Margaret Thatcher. Jak polityk musi być bezkompromisowa, no też **musi często** być często w stanie znaleźć kompromisów. Wydaję mi się, że jednym z największych problem Therezy May jest jej nie towarzyskość. Media ją czasami obwiniają z tego, że nie umie zbliżyć się do ludzi. Innymi słowy, May nie jest spontaniczna pewna **sobie** siebie i może wpaść o sobie. Nietety też nie jest dobrym oratorem.

Z **du** drugiej strony w krzesle premiera w tym roku będzie już trzy lata. To znaczy, że jest wytrwała, celowa i odważna. Żaden polityk od czasów drugiej wojny światowej w tak trudnej pozycji w swoim kraju i na scenie międzynarodowej.

W końcu **Theresu May** też można opisać **Theresu May**, jeśli chodzi o jej wygląd. Pani premier uwielbia elegancki ale też dość ekstrawagancki styl. Lubie nosić buty na obcasie z wzorem lamparta i charakterystyczną biżuterię.

Obrazek 2. Widok tekstu w korpusie PoLKO

W korpusie planowana jest również lematyzacja i anotacja morfosyntaktyczna, która identyfikowałaby kategorie gramatyczne leksemów i przydzielała im odpowiednie znaczniki (tagi). Tagowanie korpusu będzie przebiegało automatycznie za pomocą wybranego taggera (w obecnej chwili rozważamy wykorzystanie narzędzia MorphoDiTa z polskim modelem). Następnie konieczna będzie ich ręczna kontrola, ponieważ w tekstach obcokrajowców znajdują się formy niestandardowe, które nie zostaną automatycznie rozpoznane. Wprowadzenie automatycznej analizy morfosyntaktycznej wiąże się również z potrzebą implementacji dodatkowych narzędzi w interfejsie TEITOK, co jest kwestią techniczną wymagającą odpowiedniej konfiguracji.

Analiza błędów walencyjnych

Błędy w walencji⁴, czyli uchybienia wobec zasad łączliwości składniowej, są klasyfikowane jako błędy gramatyczne⁵ i można je podzielić na kilka grup⁶.

⁴ Walencję w tym opracowaniu rozumiemy jako wymagania składniowe różnych części mowy, najczęściej czasowników. Szerzej walencję komentuje Kaczmarska (2019, s. 128–133). Por. też: Čermáková 2009; Daneš 1971; Daneš i in. 1973; Panevová 1975; Panevová 1980; Panevová 1994; Panevová 1998; Przepiórkowski 2017; Ružička 1968; Rytel-Kuc 1989; Skwarska, Kaczmarska 2016; Urbańczyk-Adach 2011; Zaron 2012; Zaron, Skwarska 2017; Zaron, Skwarska 2018.

⁵ W oparciu o klasyfikację Andrzeja Markowskiego (2008) można wyróżnić kilka typów błędów: ortograficzne, leksykalne, interpunkcyjne, stylistyczne i gramatyczne. Por. Skura 2018 i Dąbrowska 2014.

⁶ W przytaczanych przykładach zaznaczamy i omawiamy tylko błędy walencyjne.

Są to m.in. zakłócenia łączliwości czasownika występujące zarówno w konstrukcjach bezprzymikowych, jak i przymikowych, a także uchybienia wobec wymagań składniowych rzeczowników.

Analizowane przykłady pochodzą z zasobów korpusu PoLKO. W samym korpusie, jak było wspomniane wcześniej, umieszczone są metadane, które pozwalają ustalić parametry (m.in. narodowość, wiek, wykształcenie, znajomość języków obcych, poziom znajomości języka polskiego, itd.) autorów popełniających błędy w pracach pisemnych. W tym artykule będziemy pracować z materiałem, w którym podajemy jedynie symbol pierwszego języka autorów tekstów⁷ (L1) i poziom ich znajomości języka polskiego.

1. Jednym z częstszych problemów jest stosowanie biernika zamiast dopełniacza po czasowniku z negacją. Klasycznym przykładem jest zdanie:

*Wczoraj **nie zobaczyłam** Ewę, było za ciemno. (A2, cs)

Szczególnie widoczne jest to w zdaniach, w których czasownik i łączący się z nim rzeczownik są rozdzielone zaimkiem, liczebnikiem, imiesłowem czy przymiotnikiem. Np.

*Moim zdaniem, jeżeli **nie wkładamy całą duszę** w naszą pracę, to marnujemy się w niej, nie rozwijamy swoich talentów. (C1, uk)

*Ale rzadko chodzi do kina, bo **nie lubi tej atmosfery**. (A1, cs)

Natomiast w poniższym przykładzie użyty został rzeczownik w bierniku lub mianowniku (nie znamy intencji autorki / autora) zamiast rzeczownika w dopełniaczu.

*Gdybym **nie dostała prezent**, byłabym bardzo rozczarowana. (B2, be)

2. Wśród błędów reakcji werbalnej zdarzają się też frazy, w których na miejscu wymaganego biernika pojawia się dopełniacz.

*Nie wolno wypożyczać pieniędzy w dni świąt lub w niedziele, ponieważ **stracisz pieniądze** bezzwrotne. (B2, be)⁸

3. Uchybienia dotyczące biernika (często mieszanie biernika i mianownika) pojawiają się również w konstrukcjach nieosobowych.

***Darmowa energia słoneczna wykorzystuje się** w nieograniczonej ilości. (B2, be)

⁷ Symbole języków zgodne z listą kodów ISO, zob. https://iso639-3.sil.org/code_tables/639/data.

⁸ W przykładzie tym jest więcej błędów, których autorzy już nie omawiają, np. łączliwość semantyczna we frazie **wypożyczać pieniądze** (zamiast *pożyczać / pożyczyć pieniądze*).

Błąd w tym zdaniu był prawdopodobnie spowodowany podobieństwem fraz, które mogłyby być użyte alternatywnie⁹: *wykorzystuje się / jest wykorzystywana*; w pierwszym przypadku wymagany jest biernik, w drugim – mianownik.

4. Innym typem błędu w tekstach z poziomów A1–A2 jest pojawianie się frazy w dopełniaczu zamiast w narzędniku.

*Interesuję się **polskich filmów**. (A2, hi-IN)

Poprawność ortograficzna zapisu zdania sugeruje, iż powodem błędu w tym przypadku może być błędny wybór przypadku gramatycznego wynikający z braku automatyzacji użycia dopełniacza i narzędnika.

5. W analizowanych tekstach spotykamy się też z niestandardową realizacją łączliwości czasowników z przyimkami, zdarza się, że użyty przyimek jest zbędny, np.

*W tym samym dniu Robert **odpowiedział do Marka**, jednego z swoich przyjaciół od dzieciństwa. (B1, ko)

W przytoczonym powyżej przykładzie występuje czasownik *odpowzieć*, który w tym kontekście wymaga frazy nominalnej w celowniku; zamiast niej pojawiła się fraza przyimkowa (do + NP_{GEN}). Podobny problem obserwujemy w zdaniu poniżej:

*Robert **odwiedził u babci** z okazji jej urodziny. (B1, ko)

Zamiast frazy bezprzyimkowej z biernikiem została błędnie użyta fraza przyimkowa (u + NP_{GEN}): *odwiedził babcię* – **odwiedził u babci*. Powodem zaistnienia tego błędu mogło być też utrwalenie przez autorkę / autora całej frazy (ang. *lexical bundles*; zob. Allen 2010), którą powtarza się na początkowym etapie nauki (A1), np. *u babci, u dentysty, u lekarza*. Uczący się mogą przenieść całą frazę i zaaplikować ją w innej strukturze; tu – niepoprawnie.

Inny problem z przyimkiem pojawia się w zdaniu poniżej:

*Piątek jest pierwszy **dzień do weekendu**. (A2, hi-IN)

Zamiast dopełniaczowej konstrukcji bezprzyimkowej (*pierwszy dzień weekendu*) użyta została fraza przyimkowa **do** + NP_{Gen}¹⁰.

6. W tekstach uczniowskich problemem jest również nieodpowiedni dobór przyimków (często towarzyszącym czasownikom ruchu), np.

*W piątek oba **poszli do koncertu**. (B1, ko)

⁹ Mieszanie rekcji biernika i dopełniacza może być też wynikiem słabego opanowania reguł ich repartycji.

¹⁰ W zadaniu tym niepoprawna jest również jego pierwsza część – zamiast konstrukcji mianownikowej (**Piątek jest pierwszy dzień do weekendu*.) powinna się pojawić fraza narzędnikowa (*Piątek jest pierwszym dniem weekendu*.), ewentualnie fraza z „to” – *Piątek to pierwszy dzień weekendu*.

Poza błędną formą liczebnika (*oba* zamiast *obaj* / *obydwaj*) pojawia się niepoprawny w tym kontekście przyimek łączący się z czasownikiem *pójść* (*do* zamiast *na*), co skutkuje również błędną formą rzeczownika (**do koncertu* zamiast *na koncert*). Przyczyną może też być nieutrwalenie różnicy semantyczno-funkcyjnej pomiędzy przyimkami *do* i *na*¹¹.

Analogiczny przypadek prezentowany jest w poniższym przykładzie:

***Ukryłem pieniądze do banku.** (B1, ko)

W zdaniu zamiast frazy przyimkowej (*w* + NP_{Loc}) pojawił się błędnie dobrany przyimek *do* (*do* + NP_{Gen}). Gdyby nie kontekst (opisywana kradzież), moglibyśmy przypuszczać, iż problemem jest tutaj dobór czasownika, por. *Wpłacić pieniądze do banku*.

Podobną sytuację przedstawia również kolejny przykład:

***Wcześniej pani Krysia poprosiła ją za opiekowanie się** swoim psem przez pięć dni. (B1, ko)

W zdaniu tym został użyty niewłaściwy przyimek (**za** zamiast **o**); mogło na to wpłynąć przeniesienie wzorca walencyjnego czasownika *przeprosić za*. Przyczyną błędu mógł być też przypadkowy rozdzielny zapis rzeczownika *zaopiekowanie się* (**za opiekowanie się*) i brak przyimka **o** (*poprosiła ją o zaopiekowanie się*).

W kolejnym przypadku problemem jest przyimek **na**, który pojawił się na miejscu oczekiwanego przyimka **w**:

***Lubię spacerować na parku.** (A2, pes)

Forma rzeczownika (miejscownik) jest poprawna i choć nie znamy intencji osoby piszącej, zakładamy, że nie chodziło tu o identycznie brzmiącą formę dopełniacza (*parku*).

Jeszcze bardziej interesujący jest dalszy przykład:

***Lubię piątki, bo wiem, że za 2 dni** będę miała wolny czas. (A2, tr)

Nie możemy być pewni, czy został tu popełniony błąd i autorce / autorowi chodziło o wolny czas w weekend (*przez 2 dni będę miała wolny czas / będę miała dwa dni wolnego*), czy może jednak ów wolny czas ma nastąpić rzeczywiście za dwa dni (wówczas zdanie byłoby poprawne).

¹¹ Przyimek **do** jest między innymi powiązany z miejscem, do którego wchodzimy (pomieszczenie, budynek). Przyimek **na** często odnosi się natomiast do wydarzenia, czynności czy celów związanych z jakimś obiektem (por. *Idę do kina na film. / Idę na letnie kino obejrzeć film na świeżym powietrzu.*) oraz do pójścia do miejsca zajmującego jakąś większą powierzchnię (*na stadion, na uniwersytet, ale też na pocztę*).

7. Korpusowe poświadczenia zawierają też błędne formy przypadka po właściwie użytym przyimku, jak w przykładzie:

*Kiedy coś zapomniał i wracasz do domu trzeba **się obejrzeć w lustro**. (B2, be)

Zamiast wymaganego miejscownika (*w lustrze*) został użyty biernik (albo mianownik; nie znamy intencji autorki czy autora)¹².

8. W analizowanych tekstach znajdują się też przykłady pominięcia przyimka:

*Czasami w piątek wieczorem **spotykam się kolegami**. (A2, hi-IN)

Ponieważ poza brakiem przyimka z nie stwierdzamy w tym zdaniu innych uchybień, zakładamy, iż osoba pisząca najprawdopodobniej zapomniała o tym przyimku (*spotykam się z kolegami*). W tym przypadku warto jednak wspomnieć o myleniu dwóch podobnie brzmiących (ale różnych pod względem znaczenia) konstrukcji *spotykać się z kimś* i *spotkać kogoś*:

***Spotykam z koleżankami** na obiad. (A2, be)

zamiast *spotykam się z koleżankami* lub

*Wczoraj **spotkałem się go** na przystanku 175. (A1, be)

gdzie został użyty czasownik *spotkać się* zamiast *spotkać*, przy czym forma zaimka (biernik) jest poprawna.

9. Błędy walencyjne występujące w tekstach uczniowskich polegają również na niewłaściwym wyborze formy łączącej się z danym czasownikiem, np.

*Wielojęzyczność **pomaga poznamy** bogactwo całego świata! (B2, be)

Czasownik *pomagać* umożliwił zastosowanie dwóch uzupełnień – *pomagać + inf* (*pomaga poznać*) i *pomagać + w + NP_{Loc}* (*pomaga w poznawaniu*); została wybrana jednak konstrukcja błędna i z punktu widzenia języka polskiego nielogiczna: **pomaga poznamy*.

10. Przykłady ekscerpowane z korpusu PoLKO zawierają również błędy łączliwości rzeczowników, np.:

*Tracimy **możliwości do spełnienia** siebie, istniejemy, a nie żyjemy. (C1, uk)

Rzeczownik *możliwość* wymaga pojawienia się frazy nominalnej w dopełniaczu. W prezentowanym przykładzie została użyta jednak fraza przyimkowa *do + NP_{Gen}*, która jest błędna.

Kolejny przykład dotyczy łączliwości tego samego rzeczownika:

¹² W tym zdaniu są także inne błędy gramatyczne: *Kiedy czegoś zapomniałeś i wracasz do domu, musisz przejrzeć się w lustrze*.

*Moja idealna praca, to praca, w której czuję się bezpieczna pod względem finansowym, która daje mi **możliwość zaspokojeniu potrzeb** codziennych i tych niezbędnych do normalnego funkcjonowania. (C1, be)

W tym zdaniu zamiast frazy nominalnej w dopełniaczu obok rzeczownika *możliwość* pojawiła się fraza nominalna w celowniku lub miejscowniku (obydwa przypadki przybierają tę samą formę, a intencja autorki / autora nie jest znana).

Podsumowanie

Niniejsza analiza przedstawia sondę otwierającą dalsze możliwości badania walencji w tekstach obcokrajowców za pomocą metod korpusowych. Zaprezentowanych zostało 10 najczęstszych typów tych uchybień. Nie są to jednak jedyne czy wszystkie błędy popełniane przez obcokrajowców¹³. Błędy i ich typy mogą być obszerniej opisane w zależności od ich klasyfikacji. Szczegółowy opis błędów zastosowany w korpusie PoLKO będzie przedmiotem odrębnej publikacji. Kolejnym planowanym przedsięwzięciem jest badanie, które typy błędów są skorelowane z natywnym językiem piszącego (L1)¹⁴.

Na podstawie naszej małej próbki tekstów zauważamy, iż częściej z walencją fraz przyimkowych problem mają osoby, których rodzimym językiem nie jest język fleksyjny (por. punkt 5 i 6)¹⁵. Powodem błędów w walencji wśród osób z różnymi pierwszymi językami bywa też słabe opanowanie form fleksji imiennej (syntetycznych i analitycznych) oraz reguł rekcji.

Błędy w walencji zdarzają się także w pracach, których autorki i autorzy są na wyższych poziomach znajomości języka (B1, B2, C1); w takich przypadkach może to być pomyłka lub błąd, który nie jest jeszcze utrwalony; należy też sprawdzić, czy dane uchybienie powtarza się. Będzie to możliwe dzięki odpowiedniemu kodowaniu prac w korpusie PoLKO.

Badanie zostało przeprowadzone na próbce powstającego korpusu uczniowskiego. Wraz z jego rozwojem będzie można przeprowadzać wiele dalszych badań w zakresie języka polskiego jako obcego. Obszerny materiał korpusowy i narzędzia umożliwiają wszechstronne przeszukiwanie danych: od uzyskiwania wyników konkretnych słów (czy części mowy), po informacje dotyczące poszczególnych grup językowych czy wiekowych. Dzięki korpusowemu podejściu będziemy mieli możliwość wyjść poza ramy naszej grupy badawczej (grupy naszych

¹³ Błędy językowe popełniają także lektorzy języka polskiego jako obcego, zob. Iwona Dembowska-Wosik (2015).

¹⁴ Warto wziąć w przyszłości pod uwagę wszystkie znane uczniom języki, ponieważ negatywny transfer może być związany z drugim lub trzecim znanym językiem.

¹⁵ U Słowian błędy te mogą być spowodowane przez negatywny transfer z ich pierwszego języka.

uczniów, których najczęściej badamy) i przeprowadzić analizę na materiale, który ze względu na jego objętość zapewnia obiektywne wyciągnięcie wniosków. Analiza poświęcona konkretnym grupom uczniów może też przynieść wymierne korzyści w dalszym rozwoju dydaktyki języka polskiego jako obcego.

Bibliografia

- Allen, D. (2010). Lexical Bundles in Learner Writing: An Analysis of Formulaic Language in the ALESS Learner Corpus, *Komaba Journal of English Education*, 1, 105–127.
- Čermáková, A. (2009). *Valence českých substantiv*. Praha: Nakladatelství Lidové noviny.
- Chachulska, B. (2005). Korpus języka polskiego w nauczaniu języka polskiego jako obcego. W: P. Garncarek (red.), *Nauczanie języka polskiego jako obcego i polskiej kultury w nowej rzeczywistości europejskiej* (344–350). Warszawa: Uniwersytet Warszawski.
- Dąbrowska, A. (2004). Najczęstsze błędy popełniane przez cudzoziemców uczących się języka polskiego jako obcego. W: A. Seretny i in. (red.), *Opisywanie, rozwijanie i testowanie znajomości języka polskiego jako obcego* (105–136). Kraków: Universitas.
- Dąbrowska, A. (2014). Glottodydaktyka polonistyczna w SJPiK. Kierunki badań naukowych. W: A. Dąbrowska, U. Dobesz (red.), *40 lat wrocławskiej glottodydaktyki polonistycznej: teoria i praktyka* (255–276). Wrocław: Oficyna Wydawnicza ATUT.
- Dąbrowska, A., Pasięka, M. (2008). Nowa typologia błędów popełnianych przez cudzoziemców w języku polskim. W: M. Kita i in. (red.), *Błąd językowy w perspektywie komunikacyjnej* (73–102). Katowice: Wyższa Szkoła Zarządzania Marketingowego i Języków Obcych.
- Dąbrowska, A., Pasięka, M. (2014). Badania błędów cudzoziemców prowadzone w Szkole Języka Polskiego i Kultury dla Cudzoziemców UW. W: A. Dąbrowska, U. Dobesz (red.), *40 lat wrocławskiej glottodydaktyki polonistycznej: Teoria i praktyka* (331–342). Wrocław: Oficyna Wydawnicza ATUT.
- Daneš, F. (1971). Větné členy obligatorní, potenciální a fakultativní. W: M. Komárek (red.), *Miscellanea Linguistica* (131–138). Ostrava: Profil.
- Daneš, F. i in. (1987). *Větné vzorce v češtině*. Praha: Academia.
- Daneš, F., Hlavsa, Z., Kořenský, J. (1973). Postavení slovesa v struktuře české věty. W: B. Havránek (red.), *Československé přednášky pro VII. Mezinárodní sjezd slavistů ve Varšavě* (129–139). Praha: Academia.
- Dembowska-Wosik, I. (2015). Błędy językowe w pracach pisemnych przyszłych lektorów języka polskiego jako obcego, *ACTA UNIVERSITATIS LODZIENSIS, Kształcenie Polonistyczne Cudzoziemców*, 22, 223–236. DOI: 10.18778/0860-6587.22.15
- Gilquin, G. i in. (2007). Learner corpora: The missing link in EAP pedagogy, *Journal of English for Academic Purposes*, 6(4), 319–335. DOI: 10.1016/j.jeap.2007.09.007
- Górska, A. (2015). Błędy studentów z Ukrainy – zapobieganie i eliminacja w grupach o zróżnicowanych możliwościach (na podstawie doświadczeń Centrum Partnerstwa

- Wschodniego Uniwersytetu Opolskiego), *ACTA UNIVERSITATIS LODZIENSIS, Kształcenie Polonistyczne Cudzoziemców*, 22, 357–370. DOI: 10.18778/0860-6587.22.24
- Granger, S. (1998). The computer learner corpus: a versatile new source of data for SLA research. W: S. Granger (red.), *Learner English on Computer* (3–18). London – New York: Addison Wesley Longman.
- Granger, S. (2002). A birds-eye view of learner corpus research. W: S. Granger i in. (red.), *Computer learner corpora, second language acquisition, and foreign language teaching* (3–33). Amsterdam – Philadelphia: John Benjamins. DOI: 10.1075/llt.6.04gra
- Janssen, M. (2016). TEITOK: Text-faithful annotated corpora. W: N. Calzolari i in. (red.), *Proceedings of the Tenth International Conference on Language Resources and Evaluation* (4037–4043). Portoroż: ELRA.
- Kaczmarek, E. (2019). *Metody ustalania ekwiwalentów czasowników wyrażających stany emocjonalne w przekładzie czesko-polskim na materiale z korpusu równoległego InterCorp*. Warszawa: Wydział Polonistyki Uniwersytetu Warszawskiego.
- Kita, M. i in. (red.). (2008). *Błąd językowy w perspektywie komunikacyjnej*. Katowice: Wyższa Szkoła Zarządzania Marketingowego i Języków Obcych.
- Kowalewski, J. (2018). Badania statystyczne nad błędami językowymi uczących się języka polskiego na Ukrainie. W: M. Maciołek (red.), *Polonistyka na początku XXI wieku. Diagnozy. Koncepcje. Perspektywy. Tom III: Współczesne aspekty badań nad językiem polskim – teoria i praktyka* (277–298). Katowice: Wydawnictwo Uniwersytetu Śląskiego.
- Krawczuk, A. (2009). Błędy leksykalne i leksykalno-stylistyczne w polszczyźnie Ukraińców, *Postscriptum Polonistyczne*, 1(3), 167–183.
- Markowski, A. (2008). *Kultura języka polskiego. Teoria. Zagadnienia leksykalne*. Warszawa: Wydawnictwo Naukowe PWN.
- Nesselhauf, N. (2004). Learner corpora and their potential for language teaching. W: J. Sinclair (red.), *How to use corpora in language teaching* (125–156). Amsterdam – Philadelphia: John Benjamins. DOI: 10.1075/scl.12.11nes
- Panevová, J. (1975). On verbal Frames in Functional Generative Description, *Prague Bulletin of Mathematical Linguistics*, 23, 17–52.
- Panevová, J. (1980). *Formy a funkce ve stavbě české věty*. Praha: Academia.
- Panevová, J. (1994). Valency frames and the meaning of the sentence. W: P. A. Luelsdorff (red.), *The Prague School of structural and functional linguistics. A short introduction* (223–243). Amsterdam – Philadelphia: John Benjamins.
- Panevová, J. (1998). Ještě k teorii valence, *Slovo a slovesnost*, 59, 1–14.
- Przepiórkowski, A. (2017). *Argumenty i modyfikatory w gramatyce i słowniku*. Warszawa: Wydawnictwa Uniwersytetu Warszawskiego.
- Przybylska, R. (2008). O potrzebie tworzenia korpusów tekstów do nauczania języka polskiego jako obcego. W: W.T. Miodunka, A. Seretny (red.), *W poszukiwaniu nowych*

- rozwiązań. *Dydaktyka języka polskiego jako obcego u progu XXI wieku* (31–34). Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.
- Rosen, A. (2015). *CzeSL-SGT – korpus češtiny nerodilých mluvčích s automaticky provedenou anotací*. https://www.researchgate.net/publication/282643850_CzeSL-SGT_-_korpus_cestiny_nerodilych_mluvchich_s_automaticky_provedenou_anotaci.
- Rosen, A. et al. (2020). *Compiling and annotating a learner corpus for a morphologically rich language: CzeSL, a corpus of non-native Czech*. Prague: Karolinum.
- Ružička, J. (1968). Valencia sloviess a intencia slovesného deja, *Jazykovedný časopis*, 19, 50–56.
- Rytel-Kuc, D. (1989). Wybrane problemy opisu walencyjnego języka, *Studia z Filologii Polskiej i Słowiańskiej*, XXVI, 237–247.
- Skura, M. (2013). Błędy wynikające z interferencji kulturowej popełniane przez Niemców uczących się języka polskiego jako obcego, *ACTA UNIVERSITATIS LODZIENSIS, Kształcenie Polonistyczne Cudzoziemców*, 20, 149–158.
- Skura, M. (2018). *Błędy popełniane przez Niemców uczących się języka polskiego jako obcego – implikacje glottodydaktyczne*, (niepublikowana rozprawa doktorska). Wydział Polonistyki Uniwersytetu Warszawskiego, Warszawa.
- Skwarska, K., Kaczmarska, E. (red.). (2016). *Výzkum slovesné valence ve slovanských zemích*. Praha: Slovanský ústav Akademie věd České republiky.
- Šebesta, K. (2012). Parametry žákovských korpusů a CzeSL. W: K. Šebesta, S. Škodová (red.), *Čeština – cílový jazyk a korpusy* (13–33). Liberec: Technická univerzita v Liberci.
- Urbańczyk-Adach, N. (2011). *Wariantywność walencji czeskiego czasownika*. Warszawa: Sławistyczny Ośrodek Wydawniczy.
- Zaron, Z. (2012). Konotacja nie jedno ma imię. Wymagania składniowe nazw osobowych. W: J.D. Apresjan i in. (red.), *Meaning, Texts and other Exciting Things. A Festschrift to Commemorate the 80th Anniversary of Professor Igor Alexandrovič Mel'čuk* (672–681). Moskva: Languages of Slavic Culture.
- Zaron, Z., Skwarska, K. (red.). (2017). *Prace Filologiczne (Tom LXX)*. Warszawa: Wydziału Polonistyki Uniwersytetu Warszawskiego.
- Zaron, Z., Skwarska, K. (red.). (2018). *Prace Filologiczne (Tom LXXII)*. Warszawa: Wydziału Polonistyki Uniwersytetu Warszawskiego.
- Zasina, A.J. (2019). Podejście korpusowe w nauczaniu języka polskiego jako obcego na przykładzie rzeczownikowych alternacji ó:o. W: K. Ziolo-Pużuk (red.), *Panorama glottodydaktyki polonistycznej. Wyzwania, pytania, kierunki* (181–199). Warszawa: Wydawnictwo Naukowe Uniwersytetu Kardynała Stefana Wyszyńskiego.
- Zasina, A.J., Kaczmarska, E. (2020). *Infrastructure of the Polish Learner Corpus PoLKo*. https://www.researchgate.net/publication/342888260_Infrastructure_of_the_Polish_Learner_Corpus_PoLKo. DOI: 10.13140/RG.2.2.23874.40648

ABSTRAKT: Niniejszy artykuł ma na celu omówienie błędów walencyjnych w pracach pisemnych obcokrajowców uczących się języka polskiego jako obcego. Walencja jest kluczowym elementem przyswajania języka obcego, jednakże w przypadku języka polskiego jako obcego nadal brakuje badań nad błędami walencji, które byłyby oparte na danych empirycznych. Prezentowane opracowanie stanowi pierwszą próbę analizy błędów walencyjnych na podstawie danych z Korpusu Uczniowskiego Języka Polskiego PoLKO. Niniejsze badanie jest pilotażowe i obejmuje różne poziomy zaawansowania (A1–C1) i narodowości (słowiańskie i niesłowiańskie) autorów tekstów. Materiał korpusowy wykazał, iż błędy walencyjne występują na wszystkich poziomach językowych i dla różnych narodowości. Niemniej jednak błędy walencyjne są częstsze we frazach przyimkowych wśród uczniów z niefleksyjnym językiem ojczystym. Potrzebne są jednak dalsze badania w tej dziedzinie, które z rozwojem korpusu uczniowskiego PoLKO przyniosą nowe wyniki i wnioski.