

Forum Lingwistyczne

Forum

nr 7
2020



UNIWERSYTET ŚLĄSKI
WYDAWNICTWO



F Forum Lingwistyczne

nr 7
2020

STUDIA

ARCHIWALIA

POLEMIKI

VARIA

KRONIKA



**Wydawca: Instytut Językoznawstwa
Uniwersytetu Śląskiego w Katowicach**

Redakcja:

Redaktor naczelna – Magdalena Pastuch
Zastępca redaktor naczelnej – Mirosława Siuciak
Sekretarz naukowa – Katarzyna Sujkowska-Sobisz

Komitet Redakcyjny:

Aleksandra Janowska, Karolina Lisczyk, Iwona Loewe, Jacek Warchała, Katarzyna Węsierska

Redakcja naukowa numeru:

Beata Duda, Karolina Lisczyk, Barbara Mitrenga

Recenzenci:

Lista wszystkich recenzentów jest publikowana raz w roku na stronie internetowej: <https://www.journals.us.edu.pl/index.php/FL>

Rada naukowa:

Nam-Shin Cho (Yonsei University, Seul), Mario Enrietti (Università degli studi di Torino), Jewgenija Karpilowska (Ukraińska Akademia Nauk, Kijów), Stanisław Koziara (Uniwersytet Pedagogiczny w Krakowie), Aleksander Łukaszaniec (Białoruska Akademia Nauk, Mińsk), Ewa Malinowska (Uniwersytet Opolski), Swietłana Mengel (Uniwersytet Marcina Lutra w Halle), Galina Nieszczymienko (Rosyjska Akademia Nauk, Moskwa), Radosław Pawelec (Uniwersytet Warszawski), Ewa Sławek (Uniwersytet Śląski w Katowicach), Jan Sokołowski (Uniwersytet Wrocławski), Irena Stramljicz Breznik (Uniwersytet w Mariborze, Słowenia), Ewa Teodorowicz-Hellman (Uniwersytet w Sztokholmie), Martine Vanryckeghem (University of Central Florida, USA), Piotr Wierzchoń (Uniwersytet im. Adama Mickiewicza w Poznaniu), Maria Wojtak (Uniwersytet Marii Curie-Skłodowskiej w Lublinie)

Redakcja:

Instytut Językoznawstwa Uniwersytetu Śląskiego w Katowicach
pl. Sejmu Śląskiego 1, 40-032 Katowice
tel. +48 32 200 95 15, ij-k.wh@us.edu.pl
forumlingwistyczne@us.edu.pl, <https://www.journals.us.edu.pl/index.php/FL>

Wersją referencyjną czasopisma od 2019 roku jest wydanie elektroniczne.
Wcześniej czasopismo ukazywało się w wersji drukowanej z numerem ISSN 2449-9587.
W formie elektronicznej publikacja dostępna jest w zasobach:

Central and Eastern European Online Library
(www.ceeol.com)

Open Journal Systems
(<http://www.journals.us.edu.pl>)

The Central European Journal of Social Sciences and Humanities
(<http://cejsh.icm.edu.pl>)

Gold open access

SPIS TREŚCI

STUDIA 1

- Od Redakcji (*Beata Duda, Karolina Lisczyk, Barbara Mitrenga*) 11
- Joanna Przyklenk** 17
Cyberprzestrzeń w polskim dyskursie parlamentarnym
- Лариса Кислюк** 33
Функционирование неологизмов и окказионализмов из сферы информатики и интернет-коммуникации в современном украинском языке
- Björn Wiemer, Joanna Wrzesień-Kwiatkowska, Marek Łaziński** 45
Badania aspektu w językach polskim, czeskim i rosyjskim za pomocą korpusów i baz danych (pierwsze podsumowanie tematu)
- Mariusz Leńczuk** 59
Projektowanie metadanych w korpusie tekstów polskich do 1500 roku – wielopoziomowa struktura informacji
- Anna Majewska-Tworek, Monika Zaśko-Zielińska, Piotr Pęzik** 71
„Polszczyzna mówiona miast” – kontynuacja badań z lat 80. XX wieku z wykorzystaniem narzędzi lingwistyki cyfrowej
- Magdalena Makowska** 89
#naukanatwitterze. O multimodalnym designie informacji w dyskursie cyfrowym

ARCHIWALIA
2

Agnieszka A. Niekrewicz 105
Funkcja politycznych słów kluczy w memach internetowych

Jarosław Pacuła 119
Było, (nie) minęło. Przypomnienie tekstu Tony'ego McEnergy'ego i Nicka Ostlera *A New Agenda for Corpus Linguistics – Working with All of the World's Languages*

POLEMIKI
3

Iliana Genew-Puhalewa 127
Język poprawności politycznej *vs.* mowy nienawiści w bułgarskich mediach. Андреана Ефтимова: *Двойственият език в медиите: езикът на политическата коректност vs езика на омразата.* (София 2016, 335 с.)

VARIA
4

Piotr Pęzik 133
Budowa i zastosowania korpusu monitorującego MoncoPL

Kinga Wąsińska 151
Cyfrowe narzędzia w dydaktyce akademickiej. Badania diachroniczne polskiej fleksji z użyciem systemu Chronofleks

5
KRONIKA

Szymon Witczak	171
Sprawozdanie z konferencji naukowej „Wyróżnij się lub zgiń. Dialog marki z konsumentem” (Katowice, 9 maja 2019)	
Katarzyna Romaniuk	175
Sprawozdanie z konferencji naukowej „Dialog bez barier. Skuteczna komunikacja w kontekście EBP – praktyki opartej na dowodach” (Katowice, 8 czerwca 2019)	
Katarzyna Wyrwas	179
Sprawozdanie z konferencji naukowej „Perspektywy rozwoju badań lingwistycznych” (Katowice, 27–28 czerwca 2019)	
Dorota Hamerlok	185
Sprawozdanie z konferencji naukowej „Nie/porozumienie, nie/tolerancja, w(y)kluczenie w języku i kulturze” (Wisła, 18–20 września 2019)	
Noty o Autorach	189



TABLE OF CONTENTS

STUDIES **1**

From the Editors (<i>Beata Duda, Karolina Lisczyk, Barbara Mitrenga</i>)	11
Joanna Przyklenk <i>Cyberspace</i> in the Polish Parliamentary Discourse	17
Larysa Kysliuk The Functioning of Neologisms and Occasionalisms from the Field of Computer Science and Internet Communication in Modern Ukrainian	33
Björn Wiemer, Joanna Wrzesień-Kwiatkowska, Marek Łaziński Investigations of Polish, Czech and Russian Aspect on the Basis of Corpora and Databases (a First Summary on a Topic)	45
Mariusz Leńczuk Metadata Creation in the Language Corpus of Polish Texts until 1500 – a Multi-level Data Structure	59
Anna Majewska-Tworek, Monika Zaśko-Zielińska, Piotr Pęzik “Spoken Polish of cities” – a Continuation of Research from the 1980s with the Use of Digital Linguistics Tools	71
Magdalena Makowska #scienceviatwitter. On the Multimodal Design of Information in Digital Discourse	89

ARCHIVAL
2

Agnieszka A. Niekrewicz 105
The Function of Political Key Words in Internet Memes

Jarosław Pacuła 119
It's (not) All just Water under the Bridge. The Reminder of the Text of Tony McEnery and Nick Ostler
A New Agenda for Corpus Linguistics – Working with All of the World's Languages

POLEMICS
3

Iliana Genew-Puhalewa 127
The Language of Political Correctness *vs.* Hate Speech in the Bulgarian Media. Андреана Ефтимова: *Двойственият език в медиите: езикът на политическата коректност vs езика на омразата* (София 2016, 335 с.)

VARIA
4

Piotr Pęzik 133
Design and Applications of MoncoPL as a Monitor Corpus of Polish

Kinga Wąsińska 151
Digital Tools in Academic Didactics. Diachronic Description of Polish Inflection Using the Chronofleks System

CHRONICLE
5

Szymon Witczak	171
A Report from the Academic Conference „Wyróżniaj się lub zgiń. Dialog marki z konsumentem” (Katowice, 9 May 2019)	
Katarzyna Romaniuk	175
A Report from the Academic Conference „Dialog bez barier. Skuteczna komunikacja w kontekście EBP – praktyki opartej na dowodach” (Katowice, 8 June 2019)	
Katarzyna Wyrwas	179
A Report from the Academic Conference „Perspektywy rozwoju badań lingwistycznych” (Katowice, 27–28 June 2019)	
Dorota Hamerlok	185
A Report from the Academic Conference, „Nie/porozumienie, nie/tolerancja, w(y)kluczenie w języku i kulturze” (Wisła, 18–20 September 2019)	
Notes on Contributors	189





Od Redakcji

Zwrot cyfrowy w światowej i polskiej nauce jest niezaprzeczalnym faktem. Na przestrzeni kilku ostatnich dekad pod wpływem szybkiego rozwoju technologii informatycznych i Internetu powstały zupełnie nowe możliwości tworzenia, wyszukiwania, przetwarzania, wykorzystywania, wizualizowania czy prezentowania licznych i różnorodnych danych, a także nowatorskie formy upowszechniania wyników badań naukowych. Dotyczą one również językoznawstwa czy też – szerzej – humanistyki. W nowej rzeczywistości ogromne możliwości pozyskania danych daje językoznawcom prowadzona na szeroką skalę przez wiele ośrodków i instytucji (nie tylko naukowych) dygitalizacja źródeł pisanych, w tym unikatowych zabytków językowych, publikacja elektronicznych wersji książek i czasopism w bibliotekach cyfrowych oraz repozytoriach naukowych, tworzenie elektronicznych reedycji słowników wydanych pierwotnie w wersji drukowanej lub redagowanie od podstaw słowników w wersji online.

Możliwość dotarcia do bogatych zasobów danych językowych i ich wykorzystania w pracy naukowej stała się dla wielu badaczy impulsem do działań mających na celu zaimplementowanie osiągnięć informatycznych do językoznawstwa, w konsekwencji czego na styku dwóch pierwotnie odrębnych dyscyplin – informatyki i językoznawstwa – powstała lingwistyka cyfrowa (ang. *digital linguistics*). W jej obszarze realizowane są zarówno prace dotyczące tworzenia konkretnych narzędzi czy programów informatycznych mających zastosowanie w badaniach lingwistycznych, przykładowo współczesnych i diachronicznych korpusów językowych, jak i badania naukowe prowadzone z ich wykorzystaniem. Ponadto w zakresie lingwistyki cyfrowej mieszczą się prace, których autorzy podejmują zagadnienia związane z cyberprzestrzenią, językiem, strategiami dyskursywnymi i rolami nadawczo-odbiorczymi w świecie wirtualnym czy gatunkami i dyskursami modelującymi cyfrowe światy. Tak zorientowana lingwistyka stwarza szerokie i różnorodne możliwości prowadzenia inter-, multi- i transdyscyplinarnych badań naukowych.

Potwierdzeniem żywego zainteresowania językoznawców badaniami w opisanym zakresie jest siódmy numer czasopisma „Forum Lingwistyczne”. Ma on charakter monograficzny, a jego tematyka dotyczy zagadnień związanych z językoznawstwem cyfrowym. Zainspirowała ona naukowców z Polski, Niemiec i Ukrainy do podzielenia się z Czytelnikami naszego czasopisma wynikami prowadzonych przez siebie prac, zarówno indywidualnych, jak i zespołowych, będących między innymi rezultatem językoznawczego namysłu nad cyberprzestrzenią, przedstawiających aktualnie realizowane projekty z zakresu lingwistyki cyfrowej czy ukazujących możliwości wykorzystania narzędzi cyfrowych w badaniach lingwistycznych i dydaktyce akademickiej.

Teksty zamieszczone w dziale *Studia* cechuje różnorodność badawczo i metodologicznie podejście do tematyki numeru. Dla Joanny Przyklenk cyfrowość jest nie tylko przedmiotem analiz, ale i źródłem materiału badawczego – autorka artykułu „*Cyberprzestrzeń w polskim*

dyskursie parlamentarnym omawia bowiem znaczenia i konteksty, w jakich parlamentarzyści w latach 2001–2018 używali rzeczownika *cyberprzestrzeń*, a przy tym korzysta z materiałów zgromadzonych w elektronicznym *Korpusie Dyskursu Parlamentarnego*. Słowa, które przeniknęły z terminologii informatycznej do innych obszarów komunikacyjnych, interesują również Larysę Kysliuk, autorkę artykułu *Функционирование неологизмов и окказионализмов из сферы информатики и интернет-коммуникации в современном украинском языке (Funkcjonowanie neologizmów i okazjonalizmów z zakresu informatyki i komunikacji internetowej we współczesnym języku ukraińskim)*. Badaczka na podstawie materiałów wyekscerpowanych z zasobów elektronicznych analizuje potencjał słowotwórczy leksyki informatycznej funkcjonującej w ukraińskim języku ogólnym.

Lingwistyka cyfrowa inspiruje językoznawców do podejmowania nie tylko nowej problematyki, ale i otwiera możliwości naukowego opracowania dostrzeganych wcześniej zagadnień lub zgromadzonych materiałów badawczych. Björn Wiemer, Joanna Wrzesień-Kwiatkowska i Marek Łaziński w artykule *Badania aspektu w językach polskim, czeskim i rosyjskim za pomocą korpusów i baz danych (pierwsze podsumowanie tematu)* nakreślają założenia projektu „DiAsPol250”, w którego ramach porównywane są systemy aspektowe trzech języków – polskiego, czeskiego i rosyjskiego. Uczestnicy projektu nie tylko korzystają z istniejących korpusów tekstowych, ale też tworzą własny, wypełniając tym samym lukę w zasobach elektronicznych (żaden z dotychczasowych korpusów nie był anotowany w taki sposób, by odnosić do siebie człony par aspektowych). Także Mariusz Leńczuk wskazuje na problemy dotyczące tworzenia korpusów tekstowych. W artykule *Metadane w korpusie tekstów polskich do 1500 roku – wielopoziomowa struktura informacji* opisuje wybrane metadane, jakimi powinny być opatrywane teksty składające się na korpus najdawniejszych zabytków języka polskiego. Autor omawia przy tym trudności, na jakie natrafiają jego twórcy, wynikające między innymi z braku podstawowych informacji na temat zachowanych tekstów czy różnych form ich transliteracji i transkrypcji. Z kolei Anna Majewska-Tworek, Monika Zaško-Zielińska i Piotr Pęzik w artykule *„Polszczyzna mówiona miast” – kontynuacja badań z lat 80. XX wieku z wykorzystaniem narzędzi lingwistyki cyfrowej* opisują, jak dzięki rozwojowi technologii informatycznych można odzyskiwać materiał badawczy zgromadzony w ubiegłym wieku podczas realizacji ogólnopolskiego projektu dotyczącego mówionej odmiany języka polskiego. Narzędzia cyfrowe dają bowiem nowe możliwości badania polszczyzny mówionej z wykorzystaniem odpowiednio opracowanego korpusu tekstów.

Dział *Studia* wieńczę dwa artykuły, w których cyfrowość rozumiana jest jako przestrzeń komunikacyjna, stanowiąca źródło zagadnień do analiz. Magdalena Makowska w tekście *#naukanatwitterze. O multimodalnym designie informacji w dyskursie cyfrowym* bada informacje zamieszczane w serwisie Twitter, pełniące funkcję upowszechniania wiedzy naukowej i popularnonaukowej, w których komunikacja odbywa się za pomocą różnych systemów znaków. Agnieszkę A. Niekrewicz w artykule *Funkcja politycznych słów kluczy w memach internetowych* również interesują krótkie komunikaty werbalno-wizualne. Badając wybrane memy internetowe, autorka wskazuje na sposoby aktualizowania treści, których nośnikiem są słowa klucze zaczerpnięte z wypowiedzi polityków, a przywoływane (również w memach) w celu ich deprecjacji.

W dziale *Archiwalia* publikujemy artykuł Jarosława Pacuły *Było, (nie) minęło. Przypomnienie tekstu Tony’ego McEnery’ego i Nicka Ostlera „A New Agenda for Corpus Linguistics –*

Working with All of the World's Languages”, zawierający omówienie przywołanego w tytule opracowania brytyjskich badaczy, powstałego przed dwiema dekadami. Tony McEnery oraz Nick Ostler pisali wówczas o potrzebie tworzenia korpusów różnych języków, a także podkreślali wyzwania czekające na ich twórców oraz korzyści płynące z rozwoju językoznawstwa korpusowego. Wiele z nich do dziś nie straciło na aktualności.

Dział *Polemiki* zawiera artykuł, którego autorka – Iliana Genew-Puhalewa – omawia książkę Andreany Eftimovej (*Двойственият език в медиите: езикът на политическата коректност vs езика на омразата*), podejmującej problematykę poprawności politycznej i mowy nienawiści w bułgarskich mediach. Autorka opracowania, podkreślając walory monografii, zwraca uwagę, że jest to pierwsze kompleksowe ujęcie wskazanej tematyki na gruncie bułgarskim.

W dziale *Varia*, zgodnie z tematyką numeru, prezentujemy dwa artykuły, które można uznać za inspirację metodologiczną w pracy zarówno naukowej, jak i dydaktycznej. Piotr Pęzik w tekście *Budowa i zastosowania korpusu monitorującego MoncoPL* omawia szczegółowo wskazany w tytule artykułu korpus oraz główne funkcje wyszukiwarki o tej samej nazwie. Kinga Wąsińska z kolei podejmuje refleksję *stricte* dydaktyczną. W opracowaniu *Cyfrowe narzędzia w dydaktyce akademickiej. Badania diachroniczne polskiej fleksji z użyciem systemu Chronofleks* autorka wskazuje na możliwości wykorzystania w kształceniu akademickim tego narzędzia, pozwalającego zestawiać na przestrzeni lat paradygmaty fleksyjne pojedynczych leksemów, ich grup, a także kategorii gramatycznych.

Ostatni dział numeru to *Kronika*, w której już tradycyjnie referowane są wybrane minione wydarzenia naukowe, organizowane przez Uniwersytet Śląski w Katowicach. Szymon Witczak omawia przebieg przygotowanej przez Koło Naukowe „Komunikator” trzeciej z cyklu konferencji naukowej „Wyróżniaj się lub zgiń. Dialog marki z konsumentem”, która odbyła się 9 maja 2019 roku. Katarzyna Romaniuk sprawozdaje wydarzenie z 8 czerwca 2019 roku – konferencję naukową „Dialog bez barier. Skuteczna komunikacja w kontekście EBP – praktyki opartej na dowodach”. W dniach 27–28 czerwca 2019 roku pracownicy Instytutu Języka Polskiego im. Ireny Bajerowej wzięli udział w konferencji naukowej „Perspektywy rozwoju badań lingwistycznych”, z której relację zdaje Katarzyna Wyrwas. *Kronikę* zamyka sprawozdanie Doroty Hamerlok ze zorganizowanej w Wiśle ogólnopolskiej konferencji naukowej „Nie/porozumienie, nie/tolerancja, w(y)kluczenie w języku i kulturze” (18–20 września 2019 r.).

Oddajemy do rąk Czytelników numer różnorodny metodologicznie i materiałowo, choć spójny tematycznie. Wierzymy, że opublikowane w nim artykuły, przybliżające zagadnienia z zakresu lingwistyki cyfrowej, stanowią nie tylko interesującą, ale także inspirującą lekturę, której pokłosiem mogą być dalsze prace prowadzące do rozwoju nurtu cyfrowego w polskim językoznawstwie. Mając nadzieję na życzliwe przyjęcie bieżącego numeru czasopisma „Forum Lingwistyczne”, zapraszamy do kolejnych naukowych prezentacji oraz dyskusji na jego łamach.

Beata Duda, Karolina Liszyk, Barbara Mitrenga



STUDIA

Cyberprzestrzeń w polskim dyskursie parlamentarnym

Функционирование неологизмов и окказионализмов из сферы информатики и интернет-коммуникации в современном украинском языке

Badania aspektu w językach polskim, czeskim i rosyjskim za pomocą korpusów i baz danych (pierwsze podsumowanie tematu)

Projektowanie metadanych w korpusie tekstów polskich do 1500 roku – wielopoziomowa struktura informacji

„Polszczyzna mówiona miast” – kontynuacja badań z lat 80. XX wieku z wykorzystaniem narzędzi lingwistyki cyfrowej

#naukanatwitterze. O multimodalnym designie informacji w dyskursie cyfrowym

Funkcja politycznych słów kluczy w memach internetowych





Cyberprzestrzeń w polskim dyskursie parlamentarnym

Cyberspace in the Polish Parliamentary Discourse

Abstract: The subject matter of the article is the conceptualization of cyberspace in the Polish parliamentary discourse. It was recognised as important to identify how representatives of legislature perceived cyberspace in 2001–2018. The source of analyses were texts excerpted from the *Polish Parliamentary Corpus*. In the research ethnolinguistic and cognitive linguistics approach (linguistic and discursive worldview) and diachronic linguistics approach were applied to establish fixed and variable elements of cyberspace conceptualization. The following components were identified as fixed: the perception of cyberspace in threat terms (legal and military contexts) and its definitional elusiveness. In turn, the transition from treating the studied phenomenon as a novelty to its empowerment may be seen as the main change here.

Key words: discourse analysis, parliamentary discourse, conceptualization, discursive worldview, cyberspace

Abstrakt: Przedmiotem opracowania jest konceptualizacja cyberprzestrzeni w polskim dyskursie parlamentarnym. Za istotne uznano rozpoznanie, jak przedstawiciele władzy ustawodawczej w latach 2001–2018 postrzegali cyberprzestrzeń. Źródłem analiz był materiał wyekscerpowany z *Korpusu Dyskursu Parlamentarnego*. W badaniach wykorzystano podejście właściwe etnolingwistyczno-kognitywnemu profilowaniu pojęć (językowy i dyskursywny obraz świata) oraz historycznojęzykowe, by ustalić elementy stałe i zmienne w konceptualizowaniu cyberprzestrzeni. Za komponenty trwałe uznano: postrzeganie cyberprzestrzeni w kategoriach zagrożenia (konteksty prawne i militarne) oraz jej definicyjną ulotność. Z kolei główną zmianę można dostrzec w przejściu od przedstawiania badanego fenomenu jako nowości do jego upodmiotowienia.

Słowa kluczowe: analiza dyskursu, dyskurs parlamentarny, konceptualizacja, dyskursywny obraz świata, cyberprzestrzeń

1. Założenia poznawcze i podstawa źródłowa badań

Cyberprzestrzeń na stałe zagościła i w polszczyźnie (jako jednostka leksykalna), i w codziennym życiu większości Polaków (jako przestrzeń wirtualna), zwłaszcza tych korzystających z cyfrowej technologii. Jako kalka angielskiego *cyberspace* definiowana jest współcześnie w słowniku polskiego języka ogólnego następująco: ‘wyobrażony świat powstały dzięki połączeniu urządzeń komputerowych w sieci umożliwiającej komunikację’ (za: WSJP PAN; ŻMIGRODZKI i in., red., 2018). Warto przy tym zaznaczyć, iż przytoczona definicja słownikowa opatrzona jest jednodziedzinowym kwalifikatorem specjalistycznym – *inform.* –

co sytuowałoby omawianą jednostkę w obszarze słownictwa specjalistycznego właściwego dziedzinie informatyki. Początki wyrazu *cyberprzestrzeń* nie wyrosły jednak z gruntu informatycznego, lecz beletrystycznego. Przyjmuje się, że po raz pierwszy użył określenia *cyberspace* William Gibson w 1982 roku w noweli pt. *Burning Chrome*. W 1984 roku w powieści zatytułowanej *Neuromancer* pisarz tak definiował cyberprzestrzeń:

To jest cyberprzestrzeń. Konsensualna halucynacja, doświadczana każdego dnia przez miliardy uprawnionych użytkowników we wszystkich krajach, przez dzieci nauczone pojęć matematycznych... Graficzne odwzorowanie danych pobieranych z banków wszystkich komputerów świata. Niewyobrażalna złożoność...

GIBSON, 2009: 59

Wyrosłe z literatury *science fiction* pojęcie dziś traktowane jako ważny termin informatyki jest obecnie zawłaszczane przez inne obszary ludzkiej komunikacji i nie ogranicza się już jedynie do terminu informatycznego, przeciwnie – staje się ważnym pojęciem licznych dyskursów, w których stale jest dyskutowane i redefiniowane. Złożona natura cyberprzestrzeni, właściwa jej konwergencja socjosfery, infosfery i technosfery (SIENKIEWICZ, 2015: 92), a nadto jej dynamicznie zmieniający się, nieostry i niematerialny charakter zarówno czyni z niej obiekt zainteresowania licznych dziedzin nauki, jak i rodzi wiele trudności definicyjnych oraz interpretacyjnych. To wszystko sprawia, iż w przypadku cyberprzestrzeni mamy do czynienia z pojęciem ważnym, bo dotyczącym istotnej sfery aktywności współczesnego człowieka, a zarazem różnoznacznym, bo cyberprzestrzeń ze względu na swą wielowymiarowość i złożoność jest różnie opisywana w zależności od przyjętej perspektywy czy punktu widzenia, różne też jej komponenty są w tych deskrypcjach podświetlane i wypuklane. Co więcej, przyjmowane na danym gruncie definicje czy sposoby konceptualizacji cyberprzestrzeni mają dalekosiężne konsekwencje, redefiniować wszak mogą także nasz status, dotąd fizykalnie dookreślony, ustalony w czasie i przestrzeni (por. ZAWOJSKI, 2002: 423).

Obszarem, w którym cyberprzestrzeń jawi się jako zjawisko szczególnie ważne, jest dyskurs parlamentarny. Waga podnoszonego w nim zagadnienia przestrzeni cyfrowej wynika zasadniczo z trzech kwestii. Po pierwsze, zintensyfikowany rozwój systemów informatycznych i telekomunikacyjnych spowodował, że wiele aktywności ludzkich „dzieje się” w cyberprzestrzeni, stymulując krajowe i światowe rynki, co przekłada się na uzależnienie i obywateli, i państw od poprawnie oraz sprawnie funkcjonującej cyberprzestrzeni. Zatem w interesie państwa leży właściwe zarządzanie tą sferą jako jedną ze sfer działalności swoich obywateli. Po drugie, dyskurs parlamentarny, rozumiany tu jako część dyskursu polityki¹, stanowi przejaw państwowotwórczych działań zmierzających do upowszechnienia wśród obywateli określonego sposobu rozumienia cyberprzestrzeni. Można by tu mówić

¹ W rozróżnieniu zaproponowanym przez Marka CZYŻEWSKIEGO, Sergiusza KOWALSKIEGO i Andrzeja PIOTROWSKIEGO (2010: 18–29) padają trzy osobne pojęcia – dyskursu publicznego, dyskursu politycznego i dyskursu polityki, przy czym ten ostatni obejmuje „wypowiedzi polityków w ramach ról przypisanych im w obrębie instytucji politycznych” (2010: 22–23). W wypadku więc wypowiedzi parlamentarzystów publicznie wygłaszanych podczas prac Sejmu i Senatu RP mielibyśmy do czynienia z dyskursem polityki, w którym z uwagi na przestrzeń komunikacji można by wyodrębnić jako jego część – dyskurs parlamentarny. O hierarchicznych zależnościach między pojęciami – dyskursu publicz-

o swoistym zawłaszczaniu pojęcia dokonywanym w duchu dyskursu władzy². Po trzecie wreszcie, dyskurs parlamentarny, na który składają się wypowiedzi padające w Parlamencie RP w ramach obowiązujących modeli komunikacyjnych (por. LASKOWSKA, 2004), ma znaczący i realny wpływ na obywateli RP, dlatego też sposób modelowania cyberprzestrzeni w dyskursie parlamentarnym rzutuje na sposób jego postrzegania w polskim społeczeństwie, ale też – podkreślimy – może odzwierciedlać to, jak społeczeństwo ową sferę odbiera. Posłowie i senatorowie są przecież przedstawicielami społeczeństwa (czy jego części – tej, która ich wybrała lub w obecnej chwili popiera), więc w debacie parlamentarnej mogą wyrażać opinie, obawy czy nadzieje, które są zbieżne z poglądami ogółu (a przynajmniej jakiejś części tego ogółu).

Interesujące byłoby zatem zbadanie sposobów konceptualizacji pojęcia cyberprzestrzeni w polskim dyskursie parlamentarnym. Punkt widzenia opisywanego tu fenomenu będzie, rzecz jasna, zależny od zajmowanego miejsca na scenie politycznej, przyjmowanej orientacji ideologicznej czy wreszcie formy tekstu, w którym mówi się o przestrzeni cyfrowej³. W przedkładanym opracowaniu skupię się na sposobach konceptualizowania cyberprzestrzeni w dyskursie parlamentarnym w ogóle, a zatem bez włączania w ogląd poznawczy rozróżnień natury partyjno-politycznej oraz genologicznej⁴. Za istotne uznaję rozpoznanie, jak przedstawiciele organu władzy ustawodawczej postrzegają cyberprzestrzeń, jakie cechy jej przypisują oraz z jakimi kontekstami najczęściej ją łączą. W celu przybliżenia konceptualizacji omawianego fenomenu przyjmuję w badaniach perspektywę właściwą etnolingwistyczno-kognitywnemu profilowaniu pojęć, za inspirujące uważając w tym zakresie dokonania przedstawicieli amerykańskiego kognitywizmu (zob. m.in.: LANGACKER, 2009; TABAKOWSKA, red., 2001) i lubelskiej szkoły etnolingwistycznej (zob. m.in.: BARTMIŃSKI, TOKARSKI, red., 1998; BARTMIŃSKI, red., 1999). Ronald Langacker zakłada, że świat postrzegany jest przez użytkownika języka na różne sposoby, co w konsekwencji prowadzi do konstruowania różnych konceptualizacji i wreszcie do tworzenia rozmaitych wyrażen językowych (TABAKOWSKA, 1998: 168). Deskrypcja językowego obrazu świata rozumianego jako „zawarta w języku, różnie zwerbalizowana interpretacja rzeczywistości dająca się ująć w postaci zespołu sądów o świecie” (BARTMIŃSKI, 2009: 12) pozwala zatem przybliżyć obraz interesującego nas wycinka rzeczywistości językowej, stanowiąc jego pewien model. Za konkretyzację nośnego poznawczo pojęcia *językowego obrazu świata* (JOŚ) uznać można *tekstowy obraz świata* (BUGAJSKI, WOJCIECHOWSKA, 1996), a za jego rozszerzenie – pojęcie *kulturowego obrazu świata* (ANUSIEWICZ, DĄBROWSKA, FLEISCHER, 2000). Z kolei rozpatrywanie tekstów na płaszczyźnie dyskursywnej zrodziło potrzebę wprowadzenia do badań pojęcia *dyskursywnego obrazu świata* (DOŚ) (CZACHUR, 2011; BINIEWICZ, 2015).

nego, dyskursu politycznego, dyskursu polityki a pojęciem dyskursu parlamentarnego zob. rozważania Anity FILIPczAK-BIAŁKOWSKIEJ (2018: 28–29).

² Na państwo jako siłę próbującą przypisać sobie możliwość mówienia o cyberprzestrzeni i przez to wpływającą na nasze jej rozumienie wskazywał Radosław BOMBA (2007: 121).

³ Możemy przypuszczać, że z uwagi na specyfikę gatunków wypowiedzi parlamentarnych i ich uwarunkowań pragmatycznych (tu zwłaszcza rola audytorium) nieco inaczej swoje sądy będą wyrażać politycy na posiedzeniach plenarnych, a inaczej w interpelacji czy podczas prac komisji.

⁴ Zagadnienia te uznaję jednak za ważne dla pełnego opisu zjawiska, o którym tu mowa, i jako takie powinny stać się przedmiotem odrębnych opracowań.

Ponieważ interesuje mnie cyberprzestrzeń w określonej sferze dyskursywnej, zdeterminowanej lokatywno-temporalnie, za obiecującą uznaję koncepcję dyskursywnego obrazu świata. Jak podkreśla Waldemar Czachur, analiza tego typu umożliwia:

- uchwycenie konkretnych (dominujących, marginalizowanych itd.) aktorów dyskursu, a przez to ich punktów widzenia,
- uchwycenie szerszego kontekstu kulturowego i społecznego mającego bezpośredni wpływ na proces profilowania wiedzy językowej w postaci pojęć, słów sztandarowych, symboli kolektywnych, metafor, wzorców argumentacyjnych itd.,
- uchwycenie chwili dyskursywnej jako wycinka pewnej całości i ciągłości kulturowej, schematów poznawczych i aksjologicznych typowych dla danej społeczności kulturowej.

CZACHUR, 2011: 88

Oczywiście badania DOŚ nie stoją w opozycji do badań JOŚ, przeciwnie – widzę w nich komplementarność i ważkie podobieństwo. „Dyskursywny obraz świata (DOŚ), podobnie jak językowy obraz świata według Bartmińskiego, to dyskursywnie wyprofilowana interpretacja rzeczywistości, dająca się ująć w postaci zespołu sądów o świecie, o ludziach, rzeczach i zdarzeniach” (CZACHUR, 2011: 87). Analizując zbiór kontekstów, w jakich występuje jednostka *cyberprzestrzeń*, można ustalić typowe dla badanego obszaru dyskursywnego (tu: dyskursu parlamentarnego) modele postrzegania przedstawianego zjawiska. Interesuje mnie bowiem nie cyberprzestrzeń w ogóle, ale cyberprzestrzeń uobecniata w tekstach przynależnych dyskursowi parlamentarnemu.

Drugą perspektywą zakładaną w badaniach jest ujęcie historycznojęzykowe. Dążąc do przybliżenia dyskursywnego obrazu cyberprzestrzeni w zdefiniowanym wycinku czasowym (por. uwagi poniżej), wskażę elementy stabilne i zmienne w sposobach konceptualizacji omawianego pojęcia.

Podstawą źródłową prowadzonych analiz był przede wszystkim *Korpus Dyskursu Parlamentarnego* (KDP), który powstał w Zespole Inżynierii Lingwistycznej Instytutu Podstaw Informatyki PAN⁵. *Korpus* zawiera lingwistycznie anotowane teksty z posiedzeń plenarnych Sejmu i Senatu, interpelacji i zapytań poselskich (oraz odpowiedzi na nie), a także z posiedzeń komisji od roku 1919 do roku bieżącego. W celu prześledzenia przebiegu interpelacji i odpowiedzi na poselskie zapytania oraz ustalenia ich datacji posiłkowałam się również materiałami zamieszczonymi na stronie Sejmu RP⁶.

⁵ Pomysłodawcą i koordynatorem prac korpusowych jest Maciej OGRODNICZUK (2012; 2018). Więcej o *Korpusie* i o projektach, w ramach których powstał, można przeczytać na stronie KDP: <http://sejm.nlp.ipipan.waw.pl/overview> oraz we wskazanych tam hiperłączami tekstach.

⁶ Dotyczyło to głównie tych fragmentów, które pochodziły z interpelacji poselskich oraz z odpowiedzi na postawione przez posła/posłów pytanie w formie interpelacji. W KDP brak oznaczeń czasowych dla interpelacji (i odpowiedzi na nie) utrudniał sytuowanie cytatu na osi czasu. Podawana kadencja obrad Sejmu była zbyt ogólnym wskaźnikiem temporalnym, stąd konieczność sięgania do źródła, czyli materiałów zamieszczonych na stronie: <http://www.sejm.gov.pl/>, które z łatwością można było odnaleźć właśnie dzięki metadansom zawartym w KDP. Brak datacji w KDP dotyczy wyłącznie interpelacji oraz odpowiedzi na poselskie interpelacje, daty są zawsze obecne przy transkrypcjach obrad sejmowych, senackich oraz posiedzeń komisji.

Zakres czasowy analiz został wyznaczony zgromadzonym materiałem. KDP obejmuje lata od 1919 roku do chwili obecnej, ale *cyberprzestrzeń* znajduje w nim poświadczenie od roku 2001⁷ do roku 2019 włącznie. Ponieważ jednak pod datą roczną 2019 znajduje się zaledwie jedno poświadczenie w *Korpusie*, co może wynikać nie tyle z nieobecności interesującej mnie jednostki w dyskursie parlamentarnym, ile z faktu, że *Korpus* – jak zaznaczają jego twórcy – jest stale uzupełniany materiałami z kolejnych posiedzeń, to uwzględniany w opracowaniu zakres czasowy obejmuje lata 2001–2018.

2. Oswajanie cybernetycznej przestrzeni

Dla postawionego w KDP zapytania [base="cyberprzestrzeń"] uzyskujemy 1167 wyników⁸. W celu zobrazowania wzrostu zainteresowania cyberprzestrzenią w dyskursie parlamentarnym przywołuję dane liczbowe w tabeli 1.

Tabela 1

Liczba wystąpień dla zapytania: [base="cyberprzestrzeń"] oraz liczba tekstów, w których wystąpiło poświadczenie badanej jednostki, z przyporządkowaniem do daty jej wystąpienia

Data roczna	Liczba poświadczeń	Liczba tekstów, w których wystąpiła badana jednostka	Data roczna	Liczba poświadczeń	Liczba tekstów, w których wystąpiła badana jednostka
2001	1	1	2011	164	14
2002	1	1	2012	61	21
2003	0	0	2013	93	22
2004	3	2	2014	221	48
2005	0	0	2015	269	44
2006	1	1	2016	204	60
2007	5	4	2017	87	37
2008	3	3	2018	34	12
2009	8	2	2019	1	1
2010	12	6			

Źródło: Opracowanie własne na podstawie danych z KDP.

Jak wynika z tabelarycznego zestawienia, realne zainteresowanie cyberprzestrzenią w środowisku ustawodawczym pojawia się wraz z rokiem 2011. Wówczas sfera ta staje się przedmiotem ożywionej dyskusji, a jeśli spojrzeć na relację liczby poświadczeń do liczby

⁷ W *Narodowym Korpusie Języka Polskiego* pierwsze poświadczenia dla słowa *cyberprzestrzeń* odnotowane są w roku 1993 – 8 poświadczeń z materiału prasowego (zob. www.nkjp.pl).

⁸ Dane liczbowe służą tylko orientacji w badanym materiale. Traktowanie ich jako danych statystycznie istotnych jest problematyczne, ponieważ w liczbie tej znajdują się fragmenty tożsame, choć obecne w różnych tekstach, nieoznaczone jednak jako cytacje czy przytoczenia. Są to w zasadzie powtórzenia fraz, zdań, a czasem całych akapitów, np. w 2014 roku – 3 poświadczenia w 3 tekstach, czyli 1 poświadczenie na tekst (INT_07-i29154-1, INT_07-i29152-1, INT_07-i29153-1); w 2018 roku – 2 poświadczenia w 2 tekstach, czyli po 1 poświadczeniu na tekst (SEJ_08-pp-063-01, SEJ_08-pp-064-01). Ponadto w liczbie tej znajdujemy także jednostkę *cyberprzestrzeń* jako człon nazwy aktu prawnego, dokumentu politycznego czy nazwy stanowiska bądź komisji, np.: *Rządowy programu ochrony cyberprzestrzeni RP na lata 2010–2015*, *Pełnomocnik Ministra Spraw Wewnętrznych do spraw Bezpieczeństwa Cyberprzestrzeni w Ministerstwie Spraw Wewnętrznych*.

tekstów, w których te poświadczenia znajdujemy, można zauważyć postępującą od roku 2011 tematyzację interesującego nas zagadnienia. W latach 2001–2010 mamy w zgromadzonym materiale do czynienia w zasadzie ze sporadycznymi i jednostkowymi przywołaniami cyberprzestrzeni, często w kontekście zgłaszanej konieczności przystosowania prawa polskiego do obowiązujących regulacji europejskich (np. KOM_4_03678 z 2004 roku, SEJ_06-pp-016-02 z 2008 roku) bądź w kontekstach luźno związanych z samą sferą cyfrową (np.: *Wspomniane sklepy z nasionami są dziś liczne w cyberprzestrzeni*. SEJ_05-pp-029-01 z 2006 roku). Z materiału wynika jednak, że nawet w tym początkowym okresie kwestią stale podnoszoną w kontekście cyberprzestrzeni był i wciąż jest temat bezpieczeństwa w świecie cyfrowym.

3. Potencjalne zagrożenie i potrzeba ochrony

Dominującym i stabilnym w badanym czasie komponentem w sposobie prezentowania cyberprzestrzeni w polskim dyskursie parlamentarnym jest przedstawianie jej w kontekście potencjalnych zagrożeń. Współwystępowanie omawianej jednostki z takimi wyrazami, jak *atak*, *zagrożenie*, *przestępstwo* bądź *bezpieczeństwo*, *ochrona*/*obrona* implikuje dwie zbieżne drogi w modelowaniu omawianego pojęcia.

Z jednej strony cyberprzestrzeń jawi się jako miejsce możliwych zagrożeń, stosunkowo rzadko w toku wywodu precyzowanych, czy ataków i/lub przestępstw, których skutki są negatywne zarówno dla obywateli, jak i dla samego państwa, na przykład:

- **Atak w cyberprzestrzeni** spowodował wielomilionowe straty finansowe z racji zakłócenia systemów przesyłu danych. [INT_05-i8428-1; 2007]⁹;
- Chciałem zapytać o akcję informacyjną, dotyczącą **ataków w cyberprzestrzeni** na komputery użytkowników indywidualnych czy instytucjonalnych [...]. [KOM_7_02410; 2013];
- Przy rosnącym uzależnieniu od technologii teleinformatycznych **konflikty w cyberprzestrzeni mogą poważnie zakłócić funkcjonowanie społeczeństw i państw** oraz zagrozić bezpieczeństwu Rzeczypospolitej Polskiej. [INT_07-i33282-1; 2015];
- I fakt, że mamy do czynienia z dynamicznym rozwojem **różnorodnych zagrożeń w odniesieniu do użytkowników cyberprzestrzeni**, ale także **do infrastruktury kluczowej dla funkcjonowania państwa, dla funkcjonowania gospodarki**, sprawia, że działania, które powinny być podejmowane i są podejmowane, dotyczą wielu obszarów. [SEJ_08-pp-054-03; 2017].

Pojawiające się w materiale związki wyrazowe, w jakich występuje *cyberprzestrzeń* z jednostkami *atak*, *zagrożenie* i *przestępstwo* (oraz z wyrazami pokrewnymi tychże)¹⁰, to

⁹ Oznaczenia skrótowe, odsyłające do konkretnego dokumentu, podaję za KDP. Po średniku odnotowuję rok, z którego cytacja pochodzi.

¹⁰ W materiale źródłowym jednostka *cyberprzestrzeń* występuje 35 razy z wyrazem *atak* (i pokrewnymi, jak: *cyberatak*, *atakowany*), 128 razy z wyrazem *zagrożenie* (i pokrewnym: *zagrożony*), a 55 razy z wyrazem *przestępstwo* (i pokrewnymi: *przestępca*, *przestępczy*, *przestępczość*), co w sumie stanowi ok. 19% wszystkich egzemplifikacji.

na przykład: *atak(i) w cyberprzestrzeni*; *atak(i) na cyberprzestrzeń*; *ataki dotyczące cyberprzestrzeni*; *zagrożeni-e/-a (w) cyberprzestrzeni*; *zagrożenia występujące w cyberprzestrzeni*; *zagrożenia związane z cyberprzestrzenią*; *przestępstwa (kryminalne/z nienawiści) w cyberprzestrzeni*; *przestępstwa popełniane/dokonywane w cyberprzestrzeni*; *zwalczanie przestępstw w cyberprzestrzeni*. Z czasem w dyskusji parlamentarnej spotyka się też sformułowania: *ataki z cyberprzestrzeni* (od 2011 r.); *zagrożenia wychodzące z cyberprzestrzeni* (od 2011 r.); *zagrożenia płynące z cyberprzestrzeni* (od 2015 r.). Cyberprzestrzeń zaczyna tu być postrzegana nie tylko jako miejsce, które można zaatakować czy któremu można zagrozić z zewnątrz lub jako cyfrowa sfera działalności przestępców, ale też przestrzeń, z której atak przyjść może. Ów ścisły związek świata digitalnego z potencjalnym zagrożeniem, a taki znajduje potwierdzenie w tekstach dyskursu parlamentarnego, może wynikać z wielu czynników, w tym miejscu jednak chciałabym zwrócić uwagę na dwa z nich. Pierwszy ma charakter asocjacyjny i dotyczy warstwy leksykalno-semantycznej. Skoro polska jednostka leksykalna *cyberprzestrzeń* jest kalką angielskiego *cyberspace*, to za tą pożyczką idą kolejne, konotujące znaczenie zagrożenia wyrazy, jak na przykład *cyberatak* (ang. *cyberattack*), *cyberprzestępstwo* (ang. *cybercrime*), *cyberterroryzm* (ang. *cyberterrorism*) czy *zagrożenie cyfrowe/cybernetyczne* (ang. *cyberthreat*)¹¹. Owo skojarzenie członu *cyber-* z poczuciem zagrożenia czy braku bezpieczeństwa jest szczególnie w polszczyźnie wyraźne, kiedy zestawimy z sobą jednostkę *cyberprzestrzeń* z takimi jej synonimami, jak: *przestrzeń wirtualna* czy *świat wirtualny*, w wypadku których konteksty ataku, zagrożenia czy przestępstwa są w badanym materiale rzadkością.

Bez wątpienia na uwypuklenie „groźnego” wymiaru cyberprzestrzeni wpływ miały także międzynarodowe regulacje prawne, które Polskę najpierw jako kandydata na członka i potem jako państwo członkowskie Unii Europejskiej i NATO obowiązywały, a w których skupiano się na niebezpieczeństwach wynikających z powstania nowej, niefizycznej i niematerialnej przestrzeni. I tak przechodzimy do czynnika drugiego – dyskursywno-retorycznego. Otóż ważnym elementem w dyskusji parlamentarnej nad cyberprzestrzenią jest uzasadnianie potrzeby działań na rzecz sprawowania kontroli państwa nad cyfrową przestrzenią czy choćby jej prawnego zdefiniowania i skategoryzowania¹². Ponieważ jed-

¹¹ W materiale zgromadzonym na bazie KDP słowa *cyberatak*, *cyberprzestępczość* oraz *cyberterroryzm* są obecne od roku 2011, choć rzadko poświadczane w bliskim sąsiedztwie jednostki *cyberprzestrzeń* (*cyberatak* – 7 poświadczeń; *cyberprzestępczość* – 12 poświadczeń; *cyberterroryzm* – 1 poświadczenie). W 2016 roku obserwujemy z kolei obecność jednostki *cyberzagrożenie* wymienionej w bliskim kontekście słowa *cyberprzestrzeń* (1 poświadczenie). Za bliski kontekst uznaję ten, który uzyskujemy w wyszukiwarce KDP, czyli 10 segmentów graficznych zarówno w lewostronnym, jak i w prawostronnym sąsiedztwie badanej jednostki.

¹² Takie uzasadnienia, wyrażone *expressis verbis*, znajdujemy w materiale źródłowym; dla przykładu: *Wszyscy zgadzają się co do tego, że istnieje potrzeba uregulowania kwestii bezpieczeństwa w cyberprzestrzeni* [KOM_6_05298; 2011]; *To bardzo ważne, niemniej jednak należy zwrócić uwagę, że wspierając czy chroniąc wolność w Internecie, musimy dążyć do ochrony obywateli przed działaniami w cyberprzestrzeni – tym bardziej, że przestrzeń tego medium jest szeroko otwarta* [KOM_7_00591; 2012]; *Przecież obserwujemy przestępstwa gospodarcze mające miejsce w cyberprzestrzeni i w związku z tym te regulacje są bardzo potrzebne* [KOM_8_00380; 2016]. Cyberprzestrzeń z trudem poddaje się definicyjnym rygorom. Kwestię złożonego prawnie statusu cyberprzestrzeni omawiała np. Joanna

nak cyberprzestrzeń ze względu na swoją specyfikę wymyka się państwowym regulacjom (np. brak granic – też państwowych, płynność, nieuchwytność), a odgórne próby jej porządkowania i kontrolowania mogą budzić opór części społeczeństwa (nie tylko w Polsce)¹³, to w tworzeniu wizerunku cyberprzestrzeni jako terytorium potencjalnie niebezpiecznego widzieć można starania polityków o uzasadnienie czy wręcz usprawiedliwienie podejmowanej nad nią debaty parlamentarnej. Temu zadaniu służą choćby zabiegi perswazyjne mające na celu wyolbrzymienie wiążanego ze światem cyfrowym poczucia zagrożenia. Spójrzmy na parę typowych strategii stosowanych w wypowiedziach polityków, które mają za zadanie potęgować skalę zjawiska poprzez:

a) prognozowanie/wskazywanie liczebności bądź wzrostu liczby przestępstw:

- Ponadto należy liczyć się: [...] ze **wzrostem liczby przestępstw w tzw. cyberprzestrzeni** [INT_05-z2418-odp-1; 2007];
- Chciałem również zwrócić uwagę, że jak wynika z ostatnich informacji, w cyberprzestrzeni w Polsce dokonano **w poprzednim roku 5800 tys. naruszeń**, można powiedzieć **przestępstw**, co powodowało **straty na kwotę 4.800.000 tys. zł.** [KOM_7_01052; 2012];

WORONA (2017), prezentując ów fenomen w świetle prawnomiędzynarodowych rozstrzygnięć. Badaczka zauważała, że „część władz krajowych stoi na stanowisku, iż współczesne zagrożenia są tak groźne, że musimy rzec się części prywatności, by zapewnić bezpieczeństwo państwa i jego obywateli. O ile duża część społeczeństwa wyraziłaby zgodę na przyznanie rozszerzonych kompetencji do monitorowania (i tym samym pewnego zakresu inwigilowania) przez własne państwo, to zgody takiej z całą pewnością nie udzieliliby innemu państwu. Tymczasem cyberprzestrzeń znajduje się poza jakimikolwiek granicami, wobec czego inne państwa są w stanie, pod byle pretekstem (choćby walki z terroryzmem), szpiegować obywateli innych krajów” (WORONA, 2017: 465). Problemem w parlamentarnym stanowieniu prawa są m.in.: dynamicznie zmieniająca się technologia (co może prowadzić do rychłej dewaluacji dotychczasowych rozstrzygnięć prawnych dotyczących cyberprzestrzeni), kwestia swobód obywatelskich (i pytania o ich granice/status w cyfrowej przestrzeni) czy postrzeganie cyberprzestrzeni oraz Internetu jako nowych dóbr wspólnych (szerokie omówienie kwestii Internetu jako dobra wspólnego znajdujemy w pracy: HOФМОКЛ, 2009; zob. również: VOGLER, 2012). Ochrona cyberprzestrzeni gwarantowana prawnie przez państwo, ale i ochrona własnych obywateli w tej tak ważnej współcześnie sferze, może rodzić pytania o prawa jednostki w tym zakresie, stąd potrzeba uzasadniania określonych działań prawnych.

¹³ Wystarczy przypomnieć falę protestów związanych z ACTA (w 2012; BIELENIA, 2013), dyskusje toczone w kontekście Dyrektywy w sprawie praw autorskich na jednolitym rynku cyfrowym (zwanej potocznie ACTA 2) czy choćby traktowanie przez część użytkowników Internetu kolejnych rozwiązań prawnych jako sposobów cenzurowania tego medium, co nie jest do końca nieuzasadnione, zważywszy na obecność cenzury np. w Chinach. Także w zgromadzonym materiale znajdujemy głosy sygnalizujące potrzebę wyważenia stanowiska między prawem wspólnoty a prawem jednostki: *Nie do zakwestionowania jest fakt, że ochrona cyberprzestrzeni jest tematem bardzo ważnym, jednakże należy pamiętać o tym, aby nie naruszyć praw i swobody obywateli* [INT_07-z01874-odp-3; 2012]; *Podkreśla się także, że istotne znaczenie, z punktu widzenia bezpieczeństwa, ma niewystarczająca wiedza użytkowników o zagrożeniach w cyberprzestrzeni oraz konieczność rozwiązania dylematu pomiędzy wolnością osobistą i ochroną praw jednostki a stosowaniem środków służących zachowaniu bezpieczeństwa państwa* [INT_08-i03036-odp-3; 2016].

- Tak jak wcześniej powiedziała koleżanka, postęp techniczny, nowe technologie informatyczne, wykorzystywanie nowych technologii informatycznych, wykorzystywanie komputerów w życiu codziennym **powodują nasilenie się przestępczości lub powstawanie nowych przestępstw w cyberprzestrzeni**. [...] Dzisiaj, szczerze mówiąc, gdyby tak na to spojrzeć, to **więcej przestępstw o charakterze rasistowskim czy ksenofobicznym dokonuje się w cyberprzestrzeni niż w rzeczywistości**. [SEJ_07-pp-074-02; 2014];
- Takie rozwiązanie jest też konieczne z tego względu, że żyjemy w czasach, gdy **ryzyko ataków na cyberprzestrzeń systematycznie rośnie**, i to jest wręcz naszym obowiązkiem. [SEJ_07-pp-077-01; 2014];
- W tym kontekście warto zwrócić uwagę, że **obserwujemy wzrost liczby przestępstw popełnianych bezpośrednio w cyberprzestrzeni**. [KST_024i_egz_2; 2016].

b) enumerację, zestawiającą przestępstwa cyfrowe z innymi zagrożeniami i przez to hiperbolizację zagrożeń w sferze cyberprzestrzeni:

- Żyjemy w czasach, w których na bezpieczeństwo Polski i Europy wpływają nie tylko klasyczne zagrożenia, ale także różnorodne wyzwania, jak **kryzysy energetyczne, zmiany klimatyczne, migracje, niebezpieczeństwo ataków terrorystycznych, konflikty regionalne, ataki w cyberprzestrzeni, łamanie praw człowieka, rozprzestrzenianie się broni masowego rażenia i inne podobne zjawiska**. [SEJ_06-pp-064-02; 2010];
- Chodzi przede wszystkim o **przestępczość branży narkotykowej, migracyjnej, przemysłowej, o oszustwa finansowe, handel ludźmi, handel bronią oraz przestępstwa w cyberprzestrzeni**. [SNT_08-pp-072-01; 2015].

Można zatem stwierdzić, że wiązanie cyberprzestrzeni z potencjalnym atakiem, zagrożeniem czy przestępstwem to jedna strona sposobu konceptualizacji świata *cyber-* w dyskursie parlamentarnym.

Z drugiej strony natomiast mamy do czynienia z komplementarną wobec wyżej opisanej drogi strategią polegającą na zgłaszaniu potrzeby ustalenia państwowoprawnej czy – jak widać w przywołanych egzemplifikacjach – militarnej ochrony/obrony cyberprzestrzeni, co ma wynikać z chęci zachowania bezpieczeństwa tegoż obszaru. Wymieńmy związki, w jakich występuje *cyberprzestrzeń* z jednostkami *ochrona/obrona* i *bezpieczeństwo*¹⁴: *ochrona (bezpieczeństwa / danych osobowych / obywateli i przedsiębiorców / dzieci i młodzieży w cyberprzestrzeni*¹⁵; *ochrona obywateli przed działaniami w cyberprzestrzeni; obrona cyberprzestrzeni; obrona (przed atakami/przestępczością/terroryzmem) w cyberprzestrzeni; bezpieczeństwo (w) cyberprzestrzeni*.

W dyskursie parlamentarnym cyberprzestrzeń konceptualizowana jest jako określone terytorium, teren, przynależny zasadniczo Polsce (przynajmniej w jakiejś mierze), nad którym

¹⁴ W KDP wyraz *cyberprzestrzeń* występuje 286 razy w kontekście wyrazu *ochrona* (i pokrewnych: *ochronny, chronić, chronienie*), 29 razy – w kontekście wyrazu *obrona* (i pokrewnych: *obronny, broń, obronność*) i aż 285 razy – w kontekście jednostki *bezpieczeństwo* (też *bezpieczny*); te konteksty stanowią ponad połowę wszystkich poświadczeń.

¹⁵ Też bardzo częste przywoływanie dokumentu pt. *Polityka ochrony cyberprzestrzeni Rzeczypospolitej Polskiej* z 25 czerwca 2013 roku i powoływanie się nań.

należy z poziomu instytucji państwowych sprawować kontrolę i ochronę, zapewniając mu bezpieczeństwo. Przekonanie o potrzebie nadzoru państwowego nad światem cyfrowym, ale też skuteczności tych działań obronnych widać w zgłaszanych raz po raz deklaracjach i postulatach:

- Mogę odpowiedzieć, że do ataku na naszą cyberprzestrzeń **jestemy przygotowani**, gdyż te dwie sieci nie są ze sobą połączone, lecz są od siebie odseparowane. [KOM_6_01730; 2009];
- Jednym słowem, **musimy w owej cyberprzestrzeni zorganizować system bezpieczeństwa w takim samym zakresie, w jakim organizujemy się w tradycyjnej geoprzestrzeni**. Musimy skorygować, uzupełnić, rozwinąć, a tam, gdzie trzeba, opracować zupełnie nowe koncepcje strategiczne, plany operacyjne, programy przygotowawcze. [KOM_6_05298; 2011];
- Czy **polskie instytucje są przygotowane na odparcie ataków w cyberprzestrzeni? Odpowiedź brzmi: tak**. [KOM_7_00234; 2012].

Dominująca¹⁶ w dyskursie parlamentarnym perspektywa patrzenia na cyberprzestrzeń jako miejsce potencjalnego niebezpieczeństwa ewokuje po stronie środowiska ustawodawczego kreację postawy aktywnej, zmierzającej do jurysdykcyjnych rozstrzygnięć, które zapewniłyby bezpieczeństwo cybernetyczne państwu i jego instytucjom oraz obywatelom Polski. Widać to szczególnie wyraźnie w tych miejscach, w których aktualizowane jest postrzeganie cyberprzestrzeni jako pola walki/wojny, stąd obecność militarnego słownictwa i retoryki wojennej:

- Polska staje się **forpocztą postępu w walce z zagrożeniami w cyberprzestrzeni**. [SE]_06-pp-099-01; 2011];
- Ważny aspekt merytoryczny, który się pojawił, to aspekt odnoszący się do nowych zagrożeń, do których zaliczane są **cyberataki** czy generalnie mówiąc **wojna w cyberprzestrzeni**. [KOM_7_00450; 2012];
- To pokazuje także, jak te zagrożenia się rozwijają. [...] To jest **ciągła walka**, ciągły proces. Jeżeli chodzi o **totalne ataki**, które możemy nazwać **wojną w cyberprzestrzeni**, to tutaj również podjęte zostały działania. [SNT_08-pp-060-02; 2014];
- Wraz z pojawieniem się nowych technologii teleinformatycznych oraz rozwojem sieci internetowej pojawiły się nowe zagrożenia, takie jak **cyberprzestępczość, cyberterroryzm, cyberszpiegostwo, cyberkonflikty** z udziałem podmiotów niepaństwowych i **cyberwojna**, rozumiana jako **konfrontacja w cyberprzestrzeni między państwami**. [INT_07-i33282-1; 2015];
- Jak słusznie wskazują coraz częściej służby obrony narodowej, **cyberprzestrzeń może być przestrzenią wojny i walki**. [KOM_8_01125; 2016];
- W zakresie tej modernizacji planuje się rozwijać Siły Zbrojne RP w oparciu o następujące priorytety: środki **cyberobrony i cyberwalki** (w tym **systemy ochrony kryptograficznej**

¹⁶ Warto podkreślić, iż współwystępowanie w egzemplifikacjach jednostki *cyberprzestrzeń* z wyrazami: *atak, zagrożenie, przestępstwo* oraz *ochrona/obrona* i *bezpieczeństwo* (wraz ze wskazanymi w przypisach 10 i 14 wyrazami pokrewnymi) stanowi ok. 70% wszystkich poświadczeń, a uwzględnwszy cytacje dokumentujące retorykę wojenną wynik ten przekracza 70%.

i bezpieczeństwa w cyberprzestrzeni), **bezzałogowe systemy uderzeniowe**, wsparcia i zabezpieczenia [...]. Ponadto, planowane reformy [...] obejmują także [...] utworzenie struktur do prowadzenia **ofensywnych i defensywnych działań w cyberprzestrzeni (wojska cybernetyczne)**, wdrożenie reformy Narodowych Sił Rezerwowych [...], a także zmiany w **szkolnictwie wojskowym**. [INT_08-i00192-odp-3; 2016];

- **W cyberprzestrzeni toczy się prawdziwa walka**. [KOM_8_02001; 2017];
- Mamy dzisiaj w rządzie, czy też będziemy mieć, trzy ośrodki związane z cyberbezpieczeństwem. [...] trzeci [dotyczy – J.P.] Ministerstwa Obrony Narodowej **szykującego ofensywną armię do walki w cyberprzestrzeni**. [KOM_8_02379; 2017];
- **Pola bitew współczesnego świata to cyberprzestrzeń**, dlatego poważne państwo musi stawiać również na cyberbezpieczeństwo. [SE]_08-pp-054-01; 2017];
- Otóż w cyberprzestrzeni te zagrożenia również stanowią kwestię **wojen dezinformacyjnych**. [KOM_8_03045; 2018].

Oprócz tej inkluzyjnej i ofensywno-defensywnej postawy względem cyberprzestrzeni obserwujemy też strategię inną, rzadszą, w której jednoznaczna aneksja świata cyfrowego traktowanego jako wirtualne terytorium Polski, wymagające nieraz i zbrojnej ochrony, jest kwestionowane lub przynajmniej podawane w wątpliwość. Dostrzec to można chociażby w poniższym przykładzie, gdzie zestawienie cybernetycznego zagrożenia z zagrożeniem terrorystycznym w wypowiedzi innego polityka zostaje potraktowane jako nadużycie:

- [...] odbyliśmy dyskusję, w której pan generał Koziej według mnie dość – szukam właściwego słowa, żeby zabrzmiało trafnie – **woluntarystycznie**, myślę, że to dobre słowo, **porównał zagrożenia w cyberprzestrzeni do zagrożeń terrorystycznych**. [KOM_6_05444; 2011].

Na koniec tej części rozważań warto zauważyć, że choć model konceptualizowania cyberprzestrzeni jako pewnego zagrożenia przeważa w dyskursie polityków¹⁷, stanowiąc tym samym uzasadnienie ich działań na rzecz tej sfery, to w materiale nieobecne są w zasadzie elementy dyskredytujące cyberprzestrzeń, podważające jej status czy też ją negujące. Znajdujemy zresztą takie wypowiedzi, które wprost dystansują cyberprzestrzeń od zagrożeń, sygnalizując jej neutralny status, na przykład:

- **Cyberprzestrzeń jest taką samą przestrzenią jak każda inna dookoła**. Przestrzeń realna też może być dobra i zła. **Cyberprzestrzeń też może być wykorzystana dobrze i niekoniecznie dobrze**. [SE]_07-pp-095-02; 2015];

¹⁷ Przykłady niewpisujące się w ów model konceptualizacji cyberprzestrzeni stanowią niespełna 1/3 poświadczeń. Trudno wśród nich wskazać dominujący trend, ale zasygnalizujemy parę powtarzających się motywów, jak np.: *obróć prawny w cyberprzestrzeni, użytkownicy cyberprzestrzeni, obywatele funkcjonujący (czy w ogóle: funkcjonowanie) w cyberprzestrzeni, działania w cyberprzestrzeni, definicja cyberprzestrzeni, informacja (i jej wymiana) w cyberprzestrzeni, agresywne treści / rasizm / przemoc w cyberprzestrzeni, stan polskiej cyberprzestrzeni, ryzyko / sytuacje kryzysowe związane z cyberprzestrzenią i inne*.

- Dodatkowo pragnę zwrócić uwagę na fakt, że **w większości przypadków** – poza atakami wymierzonymi w funkcjonowanie systemów teleinformatycznych – **cyberprzestrzeń nie stwarza nowego rodzaju przestępstw**, a jedynie dostarcza nowych środków lub metod do prowadzenia działalności przestępczej lub stanowi nową przestrzeń, w której taka działalność jest prowadzona. [INT_08-i04777-odp-2; 2016].

4. Od nowości ku podmiotowości. Zmiany w postrzeganiu przestrzeni cyber-

Zastanówmy się teraz nad tym, jak w badanym czasie zmieniało się percypowanie cyberprzestrzeni w dyskursie parlamentarnym, jakie atrybuty politycy jej przypisują. Przede wszystkim traktowana jako *novum* stanowi problem definicyjny i początkowo trudności te wyrażane są *expressis verbis*, na przykład:

- Określenie „cyberprzestrzeń” **nie występowało** dotychczas w obrocie formalnoprawnym. [SNT_07-pp-083-01; 2011];
- Definicja cyberprzestrzeni **nie jest jednoznaczna, ani w prawie międzynarodowym, ani w polskim prawie**, żeby ją pospiesznie regulować. [KOM_6_05298; 2011];
- Ale myślę, że w następnych dniach, najdalej tygodniach, będziemy mówić [...] o programie [...], który dotyka bardzo szczególnych miejsc, mianowicie **cyberprzestrzeni, która dla nas wszystkich jest miejscem tajemniczym, może nie zawsze znanym, może takim, które budzi wiele wątpliwości**. [KOM_7_04493; 2015].

Fakt jej nieprecyzyjnego dookreślenia oddaje także skrót *tzw.* (czasem przywoływany też w rozwiniętej postaci) – sygnalizujący dystans wobec nazwy i pojęcia. Przytoczmy tu parę ilustracji tekstowych:

- [...] do przeciwdziałania incydentom komputerowym, czyli zagrożeniom **w tzw. cyberprzestrzeni**. [KOM_6_01730; 2009];
- Nasilają się nowe formy przemocy, np. przemoc **w tzw. cyberprzestrzeni**. [SE]_07-pp-097-02; 2015];
- Dziś wszyscy funkcjonujemy **w tzw. cyberprzestrzeni** i to tam jesteśmy narażeni na coraz większe niebezpieczeństwa. [SE]_08-pp-063-01; 2018].

Mimo upływu czasu i wraz z nim pojawiających się kolejnych regulacji prawnych cyberprzestrzeń, już nieopisywana jako zjawisko nowe, stale budzi wiele dyskusji. Postrzegana jako *miejsce tajemnicze* [KOM_7_04493; 2015] wciąż podlega redefiniowaniu, co poświadcza poniższa egzemplifikacja z 2017 roku:

- Wraz z dynamicznym rozwojem technologii **na przestrzeni lat pojęcie cyberprzestrzeni znacznie się zmieniało**. Obecnie to przede wszystkim wirtualna przestrzeń, w której połączone siecią komputery i inne media cyfrowe komunikują się między sobą. To taki nowy rodzaj, nowy typ przestrzeni społecznej, w której spotykamy się na co dzień, korzystając z Internetu czy telefonując. [SE]_08-pp-054-03; 2017].

Politycy, mając świadomość, że *cyberprzestrzeń staje się wirtualnym odzwierciedleniem fizycznej rzeczywistości* [INT_07-i34230-1; 2015], podejmują próby dookreślenia jej prawnego statusu, w działaniu tym widzą bowiem szansę na państwową kontrolę nad światem cyfrowym, która ma zapewnić ochronę zarówno instytucjom kraju, jak i jego mieszkańcom.

Kolejny atrybut przypisywany cyberprzestrzeni to cecha terytorialnej polskości. Często w wypowiedziach polityków pojawia się precyzacja, iż w dyskusji parlamentarnej nad cyberprzestrzenią tak naprawdę chodzi o przestrzeń cyfrową Polski, stąd dookreślenia typu: *cyberprzestrzeń Rzeczypospolitej Polskiej* [np. INT_07-i01307-odp-2; 2012]; *polska cyberprzestrzeń* [np. INT_07-i05868-1; 2012; KOM_7_01870; 2013; INT_07-i29743-1; 2014; SEJ_08-pp-019-02; 2016]; *cyberprzestrzeń Polski* [np. INT_07-i20666-odp-2; 2013; KOM_7_05075; 2015]; *cyberprzestrzeń na terenie Rzeczypospolitej Polskiej* [np. SNT_08-pp-060-02; 2014]. Z czasem jednak dochodzi coraz częściej do głosu stanowisko wyrażające przekonanie o międzynarodowym statusie cyberprzestrzeni, stąd mowa między innymi o: *ustanowieniu spójnej międzynarodowej polityki w zakresie cyberprzestrzeni dla Unii Europejskiej* [KOM_7_01870; 2013]; *monitorowaniu sytuacji poza krajem, ponieważ problem ten ma charakter transgraniczny* [KST_024i_egz_2; 2016]. Próbę wpisania cyberprzestrzeni w ponadpaństwową sferę prawną widać dobitnie w poniższym cytacie, gdzie porównuje się ją do morza otwartego, mającego status przestrzeni międzynarodowej¹⁸:

- Nie wiem [...], jak się to zakończy i jakie będzie finalne rozwiązanie. Jest to jakaś próba – osoba plus sprzęt – chyba niedoskonała, ale trudno też wyobrazić sobie inną poza ogólną koncepcją otwartego morza... Czyli cyberprzestrzeń, nad którą każdy ma jurysdykcję, tak jak kiedyś na otwartych morzach miała każda marynarka tam operująca. [KST_132szue_egz_2; 2017].

Ważną zmianę w sposobie opisu cyberprzestrzeni dostrzeganą w czasie przynosi postać indywidualnego jej użytkownika. Początkowo jako postać bierna jest on z rzadka przywoływany. Z czasem jego obecność w dyskursie parlamentarnym staje się znacząca, ponieważ politycy: a) mówią wprost, że w ich działaniach prawnych chodzi o *poczucie bezpieczeństwa nie tylko instytucji rządowych, ale wszystkich użytkowników cyberprzestrzeni* [INT_07-i22248-odp-2; 2013]; b) dzielą użytkowników Internetu na świadomych i mniej (lub wcale nie) świadomych zagrożeń (za szczególnie wymagających ochrony państwa w przestrzeni cyfrowej są uznawani: dzieci i młodzież oraz seniorzy); c) włączają obywateli w dyskusję nad cyberprzestrzenią, na przykład: *Spotkania miały na celu rozpoczęcie współpracy w obszarze bezpieczeństwa cyberprzestrzeni i zebranie opinii interesariuszy co do optymalnego kształtu dalszej współpracy. A były to: spotkania w szerokiej formule, obejmujące także podmioty spoza administracji rządowej, które służą budowaniu szerokiego forum współpracy w zakresie zapewnienia bezpieczeństwa cyberprzestrzeni Polski* [INT_07-i29746-odp-2; 2014]. Podniesienie rangi działania jednostki w przestrzeni Internetu może wynikać z kilku czynników. Po pierwsze, to efekt rosnącej świadomości, że przez nierozważne zachowanie jednostki w Sieci narażona na niebezpieczeństwo cyfrowe jest większa społeczność (np.

¹⁸ Taki status mają – oprócz morza otwartego oraz dna mórz i oceanów – przestrzeń kosmiczna i Antarktyka (por. WORONA, 2017: 491–508).

całej firmy). Po drugie, zachwianiu ulega wiara w pomyślność rozwiązań systemowych na poziomie administracji państwowej (co widać w materiale zwłaszcza od roku 2014 w reakcjach na raporty NIK wykazujące słabość dotychczasowych rozwiązań systemowych). Po trzecie, pojawiają się w dyskursie: kwestia upodmiotowienia obywatela i chęć uzyskania równowagi między ograniczeniem a wolnością jednostki w cyberprzestrzeni; por. przykłady:

- Generalnie chodzi o nasze stanowisko, które idzie w kierunku takim, żeby **mając na uwadze potrzebę zwiększania bezpieczeństwa w cyberprzestrzeni**, nie tracić z horyzontu tego, co jest istotą rozwoju Internetu, czyli **pewnej wolności w Internecie, jak również kwestii udostępniania i możliwości udostępniania pewnych danych**, które mogą być objęte rygiem typowym dla danych osobowych, ale jednocześnie powinny być udostępniane do celów naukowych, społecznych, historycznych itp. [KOM_8_00051; 2015];
- **Państwo przyjazne obywatelowi dzisiaj musi ułatwiać mu funkcjonowanie w cyberprzestrzeni**. [SE]_08-pp-063-01; 2018].

5. Uwagi końcowe

Za komponenty trwałe konceptualizacji cyberprzestrzeni w polskim dyskursie parlamentarnym lat 2001–2018 można uznać jej postrzeganie w kategoriach zagrożenia/ataku i stąd konieczności ochrony ze strony państwa. Takim sposobom prezentacji sfery cyfrowej towarzyszą zwykle konteksty prawne i militarne. Za chęcią legislacyjnego uporządkowania kwestii cyberprzestrzeni stoi potrzeba jej definicyjnego dookreślenia. Zarówno zmiany w definiowaniu pojęcia, jak i sama płynność natury cyberprzestrzeni czynią z niej obiekt trudno definiowalny na gruncie prawa (por. WASILEWSKI, 2013), ale też na polu parlamentarnej dyskusji. Za właściwość trwale przypisywaną charakteryzowanemu fenomenowi należałoby uznać definicyjną nieprecyzyjność, ulotność zarówno zjawiska, jak i nazywającego je pojęcia.

Początkowo cyberprzestrzeń opisywana jest w kategoriach nowości, jednak w drugiej dekadzie widać postępujące rozpoznanie tej sfery przez parlamentarzystów, co przejawia się we wzrastającej liczbie odwołań do różnego rodzaju aktów prawnych czy w liczbie sprawozdań z działalności instytucji odpowiedzialnych za zapewnienie bezpieczeństwa przestrzeni cyfrowej. Przywoływane są konkretne tytuły dokumentów regulujących sferę *cyber-*, ale też rozrasta się nomenklatura stanowisk odpowiedzialnych za cyfrowe bezpieczeństwo (np. *koordynator krajowy do zwalczania przestępstw z nienawiści w cyberprzestrzeni* KOM_8_02665; 2017). Przejście od generalizujących prób przybliżenia cyberprzestrzeni do jej skonkretyzowanych specyfikacji można zauważyć w precyzacji podobszarów sfery cyfrowej, które stają się przedmiotem zainteresowania państwa. Przytoczmy dla przykładu parę takich ważkich zagadnień podejmowanych w kontekście cyberprzestrzeni: dane osobowe, uzależnienie, hazard, hejt, cyberprzemoc, kradzież, prawo autorskie, rasizm czy mowa nienawiści. Poza trwale obecnymi sferami – prawną i militarną – występują także: edukacyjna, naukowa i artystyczna.

Kolejną ważną zmianą, jaką możemy obserwować, jest przesunięcie dyskusji z pola *stricte* polskiego na międzynarodowe. Uwzględnienie wymiaru ponadpaństwowego łączy

się z jeszcze jednym, kluczowym, przeobrażeniem. Otóż najpierw traktując cyberprzestrzeń jako ważkie dla państwa terytorium cyfrowe, próbowano ją porządkować zapisami prawnymi i regulacjami obronnymi wpisującymi się w standardy bezpieczeństwa narodowego. Rzeczywistość jednak, a zwłaszcza dynamicznie postępująca technologizacja świata współczesnego, a także brak skutecznych rozwiązań prawnych na polu zarówno krajowym czy lokalnym, jak i międzynarodowym, sprawiły, że oprócz wciąż obecnej w dyskursie parlamentarnym postawy „twardej” – państwowcentrycznej – wobec cyfrowej przestrzeni pojawia się także postawa „miękką”, skupiająca się na użytkowniku, obywatelu i jego cyfrowym edukowaniu oraz jego wsparciu we względnie bezpiecznym istnieniu państwa i jednostki w cyberprzestrzeni. Począwszy od 2014 roku nasilające się pytania o bezpieczeństwo cyfrowe kraju oraz podważające jakość państwowych rozwiązań systemowych raporty NIK kwestionują wiarygodność wypowiedzi typu: *nie istnieją żadne zaniedbania w wykonywaniu działań na rzecz ochrony cyberprzestrzeni kraju* [INT_07-i29743-odp-2; 2014]. W zebranych materiale znajdujemy poświadczenia tekstowe dokumentujące zachwianie wiary w możliwość prawnego podporządkowania sobie cyberprzestrzeni, przekonanie o nieszczelności polskich systemów bezpieczeństwa cyfrowego itd. Obserwację tej zmiany można by pogłębić, przyglądając się wypowiedziom osobno strony rządzącej i osobno strony opozycyjnej, bo tu są zauważalne pewne różnice, ale zadanie to wykracza poza ramy artykułu.

Widoczna w dyskursie parlamentarnym reorientacja narracji dotyczącej cyberprzestrzeni z instytucji państwowych na obywateli jest zmianą znaczącą, świadczy wszak z jednej strony o upodmiotowieniu użytkownika Internetu, a z drugiej – o zdaniu sobie sprawy przez parlamentarzystów z faktu, iż omawiana przestrzeń jako ulotna (nie tylko terminologicznie) cyfrowa realność stanowi pole działania jednostek, jest zatem rzeczywistością społeczną.

Korpusy

KDP – *Korpus Dyskursu Parlamentarnego* [online: <http://sejm.nlp.ipipan.waw.pl>; data dostępu: 15.07.2019].

Słowniki

WSJP PAN – ŻMIGRODZKI P., BAŃKO M., BATKO-TOKARZ B., BOBROWSKI J., CZELAKOWSKA A., GROCHOWSKI M., PRZYBYLSKA R., WANIAKOWA J., WĘGRZYNEK K., red.: *Wielki słownik języka polskiego PAN* [online: www.wsjp.pl; data dostępu: 13.08.2019].

Literatura

ANUSIEWICZ J., DĄBROWSKA A., FLEISCHER M., 2000: *Językowy obraz świata i kultura. Projekt koncepcji badawczej*. W: DĄBROWSKA A., ANUSIEWICZ J., red.: *Język a kultura*. T. 13: *Językowy obraz świata i kultura*. Wrocław, s. 11–44.

BARTMIŃSKI J., 2009: *Językowe podstawy obrazu świata*. Lublin.

- BARTMIŃSKI J., TOKARSKI J., red., 1998: *Profilowanie w języku i w tekście*. Lublin.
- BARTMIŃSKI J., red., 1999: *Językowy obraz świata*. Lublin.
- BIELEŃ H., 2013: *Protesty przeciwko ACTA jako reakcja społeczeństw na niejawność dyplomacji*. „Studia Podlaskie” XXI, s. 279–299.
- BINIEWICZ J., 2015: *Obraz wolności w języku i dyskursach (na przykładzie tabloidu „Fakt”)*. „Znaczenia. Kultura–Komunikacja–Społeczeństwo” nr 13, s. 9–18.
- BOMBA R., 2007: *Antropologia cyberprzestrzeni*. „Kultura Współczesna” nr 1, s. 120–130.
- BUGAJSKI M., WOJCIECHOWSKA A., 1996: *Teorie językowego obrazu świata w badaniach idiolektu pisarza*. „Poradnik Językowy” nr 3, s. 17–25.
- CZACHUR W., 2011: *Dyskursywny obraz świata. Kilka refleksji*. „Tekst i Dyskurs – Text und Diskurs” nr 4, s. 79–97.
- CZYŻEWSKI M., KOWALSKI S., PIOTROWSKI A., 2010: *Wprowadzenie*. W: CZYŻEWSKI M., KOWALSKI S., PIOTROWSKI A., red.: *Rytualny chaos. Studium dyskursu publicznego*. Warszawa, s. 15–48.
- FILIPCZAK-BIAŁKOWSKA A., 2018: *Mechanizmy manifestowania orientacji ideologicznej w dyskursie politycznym*. Łódź.
- GIBSON W., 2009: *Neuromancer*. CHOLEWA P.W. tłum. Katowice.
- HOFMOKL J., 2009: *Internet jako nowe dobro wspólne*. Warszawa.
- LANGACKER R.W., 2009: *Gramatyka kognitywna. Wprowadzenie*. TABAKOWSKA E., BUCHTA M., KARDELA H., KUBIŃSKI W., ŁOZOWSKI P. tłum. Kraków.
- LASKOWSKA E., 2004: *Dyskurs parlamentarny w ujęciu komunikacyjnym*. Bydgoszcz.
- OGRODNICZUK M., 2012: *The Polish Sejm Corpus*. In: CALZOLARI N., CHOUKRI K., DECLERCK Th., UĞUR DOĞAN M., MAEGAARD B., MARIANI J., MORENO A., ODIJK J., PIPERIDIS S., eds.: *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC 2012)*. Istanbul, s. 2219–2223.
- OGRODNICZUK M., 2018: *Polish Parliamentary Corpus*. In: FIŠER D., ESKEVICH M., DE JONG F., ed.: *Proceedings of the LREC 2018 Workshop ParlaCLARIN: Creating and Using Parliamentary Corpora*. Paris, s. 15–19 [online: http://lrec-conf.org/workshops/lrec2018/W2/pdf/11_W2.pdf; data dostępu: 30.09.2019].
- SIENKIEWICZ P., 2015: *Ontologia cyberprzestrzeni*. „Zeszyty Naukowe Warszawskiej Wyższej Szkoły Informatyki” nr 13, s. 89–102.
- TABAKOWSKA E., 1998: *Profilowanie w języku i tekście – perspektywa językoznawcy, tłumacza i poety*. W: BARTMIŃSKI J., TOKARSKI J., red.: *Profilowanie w języku i w tekście*. Lublin, s. 167–184.
- TABAKOWSKA E., red., 2001: *Kognitywne podstawy języka i językoznawstwa*. Kraków.
- VOGLER J., 2012: *Global Commons Revisited*. „Global Policy” III, issue 1, s. 61–71.
- WASILEWSKI J., 2013: *Zarys definicyjny cyberprzestrzeni*. „Przegląd Bezpieczeństwa Wewnętrznego” nr 9 (5), s. 225–234.
- WORONA J., 2017 [komputeropis]: *Cyberprzestrzeń a prawo międzynarodowe. „Status quo” i perspektywy*. Białystok [online: https://repozytorium.uwb.edu.pl/jspui/bitstream/11320/5875/1/1_Worona_%20Cyberprzestrzen_a_%20prawo_miedzynarodowe_Status_quo_i_perspektywy.pdf; data dostępu: 12.07.2019].
- ZAWOJSKI P., 2002: *Monitory między nami. O byciu razem i osobno w cyberprzestrzeni*. W: GWÓŹDŹ A., ZAWOJSKI P., red.: *Wiek ekranów*. Kraków, s. 423–431.
- ŻMIGRODZKI i in., red., 2018: *Żmigrodzki P., Bańko M., Batko-Tokarz B., Bobrowski J., Czelakowska A., Grochowski M., Przybylska R., Waniakowa J., Węgrzynek K.: Wielki słownik języka polskiego PAN. Geneza, koncepcja, zasady opracowania*. Kraków.



Функционирование неологизмов и окказионализмов из сферы информатики и интернет-коммуникации в современном украинском языке

The Functioning of Neologisms and Occasionalisms from the Field of Computer Science and Internet Communication in Modern Ukrainian

Funkcjonowanie neologizmów i okazjonalizmów z dziedziny informatyki i komunikacji internetowej we współczesnym języku ukraińskim

Abstract: This paper is devoted to the analysis of term generation in the field of computer science and i-net communication using the native word-formative tools of the Ukrainian language. Particular attention is paid to the functioning of such neologisms in common usage (linguistic usus). Ukrainian electronic resources are used as a research material and verification tool. Morphemic, word-formative and comparative analysis of derivatives and the method of periphrasis allowed us to determine the reasons for the production of appropriate and inappropriate special names that can compete with English borrowings in this field.

Key words: neologism, new borrowing, neosemanticism, occasionalism

Анотация: Стаття посвящена аналізу термінотворчества в сфері інформатики і інтернет-комунікації з використанням ісконних словообразовательних средств українського мови. Особое внимание уделено функціонуванню таких неологізмів в узусі. В качестве матеріала дослідження і інструмента верифікації використовуються українські електронні ресурси. Використання морфемного, словообразовательного і сопоставительного аналізів дериватів, а также метода перифраз дозволило визначити причини створення удачних, здатних конкурувати з англійськими заїмствованиями, і неудачних спеціальних найменувань в цій сфері.

Ключевые слова: неологізм, неозаїмствование, неосемантизм, окказіоналізм

Abstrakt: Artykuł poświęcony jest analizie tworzenia terminów w dziedzinie informatyki i komunikacji internetowej z wykorzystaniem rodzimych narzędzi do tworzenia słów w języku ukraińskim. Szczególną uwagę poświęca się funkcjonowaniu takich neologizmów w uzusie. Jako materiał badawczy i narzędzie weryfikacyjne wykorzystywane są ukraińskie zasoby elektroniczne. Zastosowanie analizy morfologicznej, słowotwórczej i konfrontatywnej derywatów oraz metody peryfrazy pozwoliło nam ustalić przyczyny tworzenia udanych i nieudanych nazw specjalnych, które mogłyby konkurować z angielskimi zapożyczeniami w tej dziedzinie.

Słowa kluczowe: neologizm, nowe zapożyczenie, neosemantyzm, okazjonalizm

1. Вступление

Современные лингвистические исследования все шире используют возможности цифровых технологий для обработки языковой информации и ресурсы интернет-пространства. Новая исследовательская среда, называемая *цифровой гуманитаристикой* (англ. *digital humanities* / пол. *humanistyka cyfrowa* / укр. *цифрова гуманітаристика*), объединяет различные практики и методы гуманитарных наук и компьютерного (цифрового) инструментария: создание цифровых продуктов, т.е. представление той или иной сферы лингвистических исследований в цифровом виде, либо же использование электронных ресурсов (электронных словарей, баз данных, корпусов и т.д.). В фокусе внимания лингвистов пребывает и терминология информатики, которая осмысливает и отображает в терминах процессы сбора, хранения и передачи информации, новые средства ее распространения, а также общения между пользователями, в том числе образование новых номинаций понятий и процессов, пребывающих в постоянном развитии и обновлении (Булаховский, 2010; Jóźwiakiewicz, 2011; Клименко, 2016: 149; 2017). Формирование компьютерной терминологии, как и обновление современного лексикона в целом, происходит за счёт конкуренции различных типов новых номинаций, таких как: новообразования, т.е. новые производные (укр. *новотвори*), новые значения слов (неосемантизмы, укр. *неосемантизми*) и новые заимствования (неозаимствования, укр. *неозапозичення, новозапозичення*) (Карпловська, 2008: 12).

Среди способов номинации новых терминов в украинском языке преобладают заимствования, преимущественно интернационализмы английского происхождения (*браузер, драйвер, сервер, сайт, файл* и др.), однако исследователи отмечают использование и собственных ресурсов языка, обращая особое внимание на роль семантического словообразования в терминосистеме, метафору и метонимию как источник формирования новых значений терминов и активизации словообразовательного потенциала терминов информатики (*вікно, мережа, миша/мишка, носій, пам'ять*) (Клименко, 2017: 122–183). Остается актуальной проблема образования специальной терминологии на исконно украинской почве, особенно остро возникающая при локализации программного обеспечения, переводе интерфейса программ, поиске адекватных эквивалентов терминов и т.д.

Целью предлагаемого исследования является анализ авторского образования терминов информатики средствами украинского языка, а также производных неологизмов, функционирующих вне терминосистемы: в частности, в художественной литературе и коммуникативной практике соцсетей. Особое внимание уделено окказиональному словообразованию, роли и значению окказиональных единиц в усовершенствовании украинской компьютерной терминологии, взаимодействию и конкуренции их с узуальными наименованиями.

Окказиональное словообразование, соотношение узуального и окказионального как регулярного и нерегулярного в языке довольно глубоко освещено в работах Алексея Георгиевича Лыкова, Елены Андреевны Земской, Ольги Сергеевны Захаровой. Клара Бузашшиова и Ольга Мартинцова акцентируют внимание в своих исследованиях на важном отличии неузуального (окказионального) словообразования от словообразования узуального с точки зрения соотношения нормы словообразовательной (нестандартное образование слов) и лексической (неупотребительная, нестабилизировавшаяся в языке лексика) (Buzassyova, Martinцова, 2003: 262). В польской дериватологической традиции принято определение окказионализмов как текстовых словообразовательных дериватов, системных и внесистемных, но нестабили-

зированных относительно лексической нормы языка (СНРУЋИЊСКА, 1978: 75; ЈАДАСКА, 2001: 38), т.е. окказионализм не входит в лексическую систему языка (ВЕЛЕНТСНІКОВ 2003: 249). Исследователи отмечают также, что окказиональные дериваты отличаются от общепринятых новообразований индивидуальностью и тем, что образуются не по существующим моделям, а по аналогии с отдельными словами (образцами) и формируют особенную подсистему словообразования (ЯНКО-ТРИНИЦКАЯ, 2001: 463, 481). В энциклопедии «Українська мова» предложено следующее определение, принятое в работе:

Окказионализм – необычное, преимущественно экспрессивно окрашенное слово, образованное на базе существующего в языке слова или словосочетания, иногда с нарушением законов словообразования или языковой нормы, функционирующее только в определенном контексте, в котором оно возникло.

Пустовит, Клименко 2004: 432¹

Существенным для окказионализма признаком украинские лингвисты считают нарушение принятого образца образования слова: 1) образца когнитивной структуры номинации, 2) ономазиологической, 3) словообразовательной структуры, словообразовательной техники ее оформления (КОЛОЇЗ, 2007: 3–9; КАРПЛОВСЬКА, 2008: 52–53). Привязанность окказионализма к контексту также отличает его от узуального новообразования. Обычно окказионализмы рассматривают как результат индивидуально-авторского словотворчества в художественном тексте, публицистике или в разговорной речи с целью передачи отношения автора, его оценки происходящего. Окказиональные производные в компьютерной терминологии как результат создания нейтральных единиц, называющих новые понятия с использованием национальных ресурсов, рассматриваются впервые.

2. Материал

Для поиска материала, для анализа и верификации неологизмов или окказионализмов в текстах различных стилей и подтверждения их узуальности/неузуальности в нашем исследовании использовались прежде всего электронные ресурсы, созданные украинскими лингвистами, а также словари новой лексики, тексты, в частности художественная литература для детей.

Основным источником анализируемого материала послужил сайт «Словотвір» (Словообразование, режим доступа: <https://slovotvir.org.ua/>). Его можно рассматривать в качестве реакции на обеспокоенность ЮНЕСКО огромными рисками для многих языков в связи с глобализацией и рекомендацию сохранить языковое разнообразие в киберпространстве с помощью перевода информации в Интернете с английского на другие языки. Авторы сайта – программисты (не лингвисты!), особенно остро ощущающие «англизацию» в своей профессиональной сфере. Этот интерактивный сайт создан с целью поиска, предложения, обсуждения и оценки исконных вариантов-наименований – заменителей новых и старых

¹ Оказіоналізм – незвичне, здебільшого експресивно забарвлене слово, утворене на основі наявного в мові словосполучення, іноді з порушенням законів словотворення чи норми мовної, що існує лише в певному контексті, в якому воно виникло (Пустовит, Клименко, 2004: 432).

заимствований. Таким образом украинские программисты пытаются возродить традицию создания терминов на исконной языковой почве. Украинские эквиваленты заимствований образованы с помощью различных способов словообразования, калькирования, в том числе и семантического, и часто являются окказиональными образованиями.

Эталоном для верификации отдельных лексем как неологизмов послужил Компьютерный морфемно-словообразовательный фонд украинского языка, созданный на базе авторитетных словарей литературного языка в начале 1980-х годов сотрудниками отдела структурно-математической лингвистики под руководством профессора Нины Федоровны Клименко (сначала в составе Института языковедения имени А.А. Потебни, а с 2011 года – Института украинского языка Национальной академии наук Украины в Киеве). Фонд под руководством профессора Евгении Анатольевны Карпиловской постоянно развивается и пополняется новыми материалами и программными средствами для работы с ним. Вторым источником верификации неологизмов стал «Грамматический словарь украинского литературного языка. Словоизменение» (Киев, 2011), созданный авторским коллективом, возглавленным Валентиной Ивановной Критской. Электронная версия этого словаря, созданная совместно с программистом Виктором Михайловичем Сорокиным, свободно доступна для работы в Интернете на лингвистическом портале лаборатории компьютерной лингвистики Института филологии Киевского национального университета имени Тараса Шевченко (режим доступа: <http://www.mova.info>). Дополнительным ресурсом для верификации того или иного слова как неологизма является сайт r2u (англ. *Russian to Ukrainian*, т.е. с русского на украинский), «Русско-украинские словари», представляющий прежде всего переводные раритетные «допогромные» (до сворачивания политики «украинизации» сталинским режимом в 1930-х годах) русско-украинские словари (17 словарей), подробнее см. (СТАРКО 2017: 86–100).

Уже упомянутый лингвистический портал [mova.info](http://www.mova.info) представляет возможность верифицирования окказионализмов и неологизмов в тестовой версии корпуса текстов украинского языка, созданной в лаборатории компьютерной лингвистики Института филологии Киевского национального университета имени Тараса Шевченко (руководитель Наталья Петровна Дарчук). Корпус состоит из шести подкорпусов (по порядку убывания их размера: публицистических, художественных, научных, законодательных, поэтических и фольклорных) общим объемом около 100 млн. словоупотреблений (режим доступа: <http://www.mova.info/corpus.aspx?l1=209>). Другим корпусом, предоставленным для свободного использования в интерактивном режиме, является Генеральный регионально аннотированный корпус украинского языка (ГРАК, режим доступа: uas Corpus.org) – постоянно пополняющаяся коллекция из более чем 50 тыс. текстов различных жанров на украинском языке, охватывающая период от начала XIX до начала XXI века. Особенностью этого корпуса является его региональная разметка, дающая информацию о функционировании слова в том или ином регионе Украины.

Дополнительными ресурсами для верификации окказионализмов стали: Компьютерный фонд инноваций (КФИ) в современном украинском языке, реализованный этим же научным коллективом под руководством Е.А. Карпиловской и созданный на его основе новый, идеографический тип словаря новой лексики «Активные ресурсы современной украинской номинации» (Киев, 2013; отв. ред. Е.А. КАРПИЛОВСКАЯ); Архивная лексическая картотека украинского языка, оцифрованная в Институте украинского языка НАН Украины благодаря поддержке Украинского культурного фонда в 2018 г. (руководитель проекта Оксана

Николаевна Тищенко) и функционирующая онлайн в тестовом режиме (режим доступа: <http://mova.cf>).

3. Анализ материала

На сайте «Словотвір» представлены новые заимствования различных тематических групп, предложенные как возможность для словотворчества и поиска украинских эквивалентов. Среди них группы «ІТ» (639 слов), «Технології» (805 слов), «Медіа» (317 слов), включающие и компьютерную интернациональную терминологию: *баг, блок, браузер, демоверсія, дисплей, драйвер, інтерфейс, картридж, клавіатура, конектинг, куки, монітор, принтер, процесор, сайт, сервер, сканер, спам, файл* и др.

Описывая процессы развития терминосистемы информатики, Н.Ф. Клименко отметила ее бурное развитие в период глобализации, выражающееся в возникновении и заимствовании новых терминов, образовании фонда интернационализмов, формировании групп семантических калек (Клименко, 2017: 122). Большинство анализированных заимствований этой терминосистемы имеют небольшой словообразовательный потенциал (*вірус* – 6 производных, *процесор* – 5 производных) либо не имеют его вовсе (*блютуз, софт*) (Клименко, 2017: 158–159). Выход терминов за границы терминосистемы часто связан с процессами их метафоризации, переосмысления и словообразовательной активности в новом значении. Например, в разговорном стиле, в художественных текстах функционирует деминутив *вірусочок* (Есаулов, 2013b: 10), в значении ‘вредного файла для операционной системы компьютера’, но с оценочными признаками не только ‘вредный, опасный’, но и ‘смешной, остроумный’ и даже ‘маленький, милый’.

Исследуемая лексика находится на периферии терминосистемы информатики или вне терминосистемы, поскольку в большинстве состоит из авторских новообразований, являющихся претендентами на статус термина вместо заимствований. Это словотворчество заинтересованных пользователей сайта «Словотвір», преимущественно программистов. За типом новых номинаций анализируемую лексику предлагается классифицировать на: 1) неосемантизмы (общеупотребительная лексика, которая вследствие метонимии, расширения или сужения значения может быть использована в качестве термина; сюда же относится лексика из других терминогрупп, которая способна «мигрировать» в сферу компьютерной терминологии); 2) новообразования; 3) окказионализмы. Из общего количества около 1750 слов, список терминов, составивший ядро для словотворчества, составил около сотни слов, включающий лексику всех трех групп.

4. Неосемантизмы

Межстилевая миграция терминов вследствие метафоризации чаще касается заимствований, например, *вірус, шлюз* (Клименко, 2017: 152). На сайте автор Юрко Зеленый предлагает заменить *файл* (англ. *file* – ‘шухляда, папка’), обозначающий именованную область данных на носителе информации, сельскохозяйственным термином *сніп* (рус. *сноп*) либо другими вариантами: *зшиток, жмут* на основании внешней схожести (как и *вікно, мишка*) – все

предложенные слова обозначают целостную и одновременно собирательную единицу чего-либо (стеблей, бумажек, также, как и совокупность данных). Но в данное время термин *файл* уже прочно закрепился в терминосистеме, образует словообразовательное гнездо, и потому предложенные варианты не могут составить ему конкуренцию. Верификация функционирования предложенных слов в значении *файла* в текстовом материале корпусов не обнаружила примеров употребления.

Удачным вариантом можно считать неосемантизм *реп'яшки* вместо *куки* (англ. *cookies* – ‘пирожное, печенье’). Неосемантизм *реп'яшки* (репейнички – семена репейника, цепляющиеся за любые негладкие поверхности) в значении ‘информация в виде текстовых или бинарных данных, полученных от веб-сайта на веб-сервере, сохраняющаяся у клиента’ образован на основе функциональной метафоры: «Когда заходишь на сайт, они (репейники) к тебе цепляются, и потом они постоянно с тобой – отсылаются туда-сюда в каждом запросе» (объяснение пользователя) активизировался в узусе именно в этом значении, что подтверждают примеры и статистические данные поиска в Гугл, и даже прижилось в украинской компьютерной терминологии.

Украинское слово *равлик* (рус. *улитка*) в новом значении ‘символ в сетевых сервисах для отделения имени пользователя от названия домена’ используется как неосемантизм на основе метафорического переноса внешнего признака, подобия вместо скалькированного с русского образца *собака* для обозначения @ (англ. *at the rate*) (действительно, знак @ (коммерческое *at*) похож на улитку) и неосемантизм *равлик* становится более популярным. Параллельно (с разной активностью) функционируют и другие его названия: *собака/собачка, вухо/вушко, песик, мушля*, образованные таким же способом. Ассоциативная метафора лежит в основе названий @ и в других славянских языках (рус. *собака*, пол. *tańra*, белор. *сабака, малпа, слімак*).

Удачным вариантом можно назвать активизированную лексему *реченець* вместо *дедлайн* (англ. *deadline*). Слово *реченець* в значении ‘конечный, окончательный срок, термин’ как нейтральная единица литературного языка зафиксировано в академических словарях 1920–30-х годов (что подтверждает поиск на сайте r2u) и текстах (326 найденных примеров в корпусе ГРАК).

5. Новообразования

Авторское новообразование *вподобайка* вместо *лайк* (англ. *like* – укр. ‘до вподоби’, рус. ‘по душе (сердцу), нравится’) в значении условного выражения одобрения, предложенное на сайте «Словотвір» в 2014 году, сегодня присутствует в одном примере корпуса ГРАК и в более чем 40 000 результатов поиска в Гугл (в электронных СМИ, в соцсетях и на форумах).

Ярким примером возможности вхождения в компьютерную терминологию удачного новообразования является *застосунок* (англ. *application, app*, рус. *приложение*) – пользовательская компьютерная программа, позволяющая решать конкретные прикладные задачи пользователя. В украинских специализированных текстах его ошибочно переводили как *прикладна програма, додаток*. Русский термин *приложение* является омонимичным и переводится на украинский и как *додаток*, и как *застосування*. Но, учитывая значение английского термина, правильным будет перевод *застосовна програма* или *застосунок*. В последнее время именно этот вариант использует Microsoft. Корпус ГРАК насчитывает 149 примеров, преимущественно

в значении термина. На сайте *g2u* термин зафиксирован в «Російсько-українському народно-му словнику» с упреждением «потребує додаткового обговорення».

Среди других предложенных новообразований можно назвать *зламник* вместо *хакер* (англ. *to hack* – ‘рубать’), *переглядач* вместо *браузер* в значении ‘программа, предназначенная для просмотра сайтов, гипертекстовых документов в Интернете’, этот вариант функционирует в соцсетях, 6 примеров найдено в корпусе ГРАК; для термина *дисплей* на сайте предложено и неосемантизм *плесо*, обозначающий водную гладь в общеупотребительной лексике, и новообразование *висвітник*. Последнее слово отсутствует в корпусах и словарях, возможно, это производное образовано по аналогии с польским *wyświetlać*. Вместо термина *драйвер* в значении ‘компьютерная программа, с помощью которой операционная система получает доступ к прибору аппаратного обеспечения’ предложен вариант *розпізнавач*. Это слово встречается в словарях на сайте *g2u*, в частности, в словаре Святослава Караванского, «Російсько-українському словнику складної лексики» (2012), но с другим значением: ‘той, хто розпізнає (душу, звук)’ как перевод рус. *различающий, распознающий*. Таким образом, относительно перспектив вхождения новообразований в терминосистему информатику вырисовывается довольно пестрая картина: от принятия удачных вариантов (*застосунок*) до отсутствия их в текстах (*висвітник*).

6. Окказионализмы

Поскольку в статье рассматриваются не только новые производные, но и окказионализмы, необходим практический инструмент определения окказионального образования в массиве новой лексики. На основе приведенных выше признаков окказионализмов главным критерием их определения является нормативная маркированность – несоответствие словообразовательной норме. Дополнительным критерием служит маркер распространенности, употребительности лексической единицы в языковой практике (*узусе*), т.е. перспективы стать узуальной. Таким образом, претендент на статус окказионализма тестируется по критериям соответствия производного слова правилам словообразования и дополнительно по распространенности использования в языковой практике. Критерий распространенности в *узусе* имеет прежде всего практическую направленность оценки возможностей новообразований конкурировать с новыми английскими заимствованиями и войти в состав нормативного лексикона национального языка, что актуально в контексте процессов глобализации.

Таким образом, анализ каждого новообразования проводился по определенному алгоритму. Первый его шаг – проверка словообразовательной структуры нового слова на соответствие с существующей словообразовательной моделью (соответствует или образована с нарушением словообразовательного образца с прагматической целью (игровое, каламбурное, экспрессивно-оценочное)). Следующий шаг – верификация распространенности слова в *узусе* с помощью баз данных, корпусов, поисковых машин, предоставляющих также статистические данные об употреблении слова. Это важный фактор для понимания перспективы принятия авторского окказионализма языковым сообществом либо отсутствия его употребления вне конкретного контекста. Существенным также оказывается различие двух функциональных разновидностей окказионализмов: 1) стилистически нейтральные, созданные

с целью назвать новое явление/объект; 2) эмоционально маркированные, созданные с целью авторской оценки, выражения отношения к определенному явлению/объекту.

Авторские окказиональные варианты заимствованных компьютерных терминов на сайте «Словотвір» принадлежат к стилистически нейтральным окказионализмам.

Вместо термина *сканер* (англ. *scanner* от *scan* – ‘тщательно рассматривать, рассматривать’) предложено Юрком Зеленым *зняткувач*, а Александром Страшко производные *знятка* – сканированная страница, *зняткування* – сканирование.

Для заимствования *спам* (англ. *spam* – от *spiced ham*, то есть ‘пряная ветчина’), обозначающего ‘массовую рассылку корреспонденции рекламного или иного характера людям, которые не выразили желания ее получать’, предложен вариант *шкул*, производное от глагола *дошкуляти* – донимать. Об этом примере, возникшем в английском языке вследствие терминизации лексики, а в языках-реципиентах, как украинский, утратившем внутреннюю форму, писала Н.Ф. Клименко (Клименко, 2017: 146). Спам адаптировался в языке вместе с целым словообразовательным гнездом (*спамерський, антиспам, антиспамерський, спамити*), поэтому у других претендентов нет никаких шансов уменьшить его активность.

На замену английскому термину *драйвер* – ‘компьютерная программа, с помощью которой операционная система получает доступ к прибору аппаратного обеспечения’ – на сайте предложены варианты *стиківник, долад, керун*. Каждое из предложенных производных – мотивированное слово, образованное с учетом дефиниции термина: *долад* – ‘программа, созданная до ладу, т.е. для системы’ (в украинском языке *лад* также обозначает систему); *стиківник* – ‘программа, которая стыкуется’, *керун* – ‘программа, которая руководит’. Но тем не менее ни один из предложенных окказионализмов не обнаружен ни в одном из электронных ресурсов, т.е. не вошел в употребление.

7. Экран и его окказиональные производные

Латинизм *монітор* (лат. *monitor*) в украинском языке в разные годы XX столетия функционировал с разными значениями, в одном из них сейчас он синонимичен англицизму *дисплей* (англ. *display*) (Клименко, 2017: 158–159). Ранее заимствование из французского *екран* (фр. *écran* – ‘заслон, ширма’) сохраняет более широкую семантику (поверхность, в т.ч. различных аппаратов), носители языка обычно различают слова *екран* (телевизора) и *дисплей* (компьютера). Сайт «Словотвір» предлагает заменить галлицизм *екран* украинским *слона*, являющимся буквальным переводом фр. *écran*. Слово оформлено без приставки в женском роде с ударением на последний слог *слонá*. К достоинствам предлагаемой инновации автор причисляет ее краткость и немотивированность, позволяющие увеличить ее словообразовательный потенциал: *широкоекранний* – *широслонний*. При этом, однако, не учтена омонимия с родительным падежом единственного числа существительного *слон* (кого? *слона*), что перечеркивает вышеперечисленные достоинства и перспективы узуализации этого новообразования.

Статистика употребления заимствования *екран* в текстах корпуса ГРАК – 13 337 словоупотреблений на 1 млн также свидетельствует о его стабильном функционировании в украинской языковой практике. Его словообразовательный потенциал насчитывает десятки производных и все время возрастает, в том числе и за счет окказиональной лексики. Например, недавно изданная серия книг для детей А.Ю. Есаулова называется «Закрання»

(Заэкратье). Образован этот окказионализм со значением ‘территория/пространство, находящиеся за чем-либо, отделяющим ее/его от остального (за рекой, за Карпатами, за зеркалом)’ по словообразовательной конфикальной модели: *Заріччя, Закарпаття, Задзеркалля* («Аліса в Задзеркаллі» Л. Керрала). *Заекрання* обозначает сказочную компьютерную страну, находящуюся за дисплеем (экраном) компьютера. Автор избрал производящей основой именно *екран*, а не *дисплей* или *монитор*. Верификация в корпусах, КФИ и Гугл показала, что и слово, и его производные: *заекранський, заекранник, воїн-заекранник* функционируют только в авторских художественных текстах. Относительное прилагательное *заекранський* образовано от *Заекрання*: *долари заекранські* ‘деньги, принятые в Заэкратье’ (ЕСАУЛОВ, 2013а: 4), *заекранським ім'ям* ‘имя, принятое в Заэкратье’ (ЕСАУЛОВ, 2013а: 32). Окказиональное название лица *заекранник* (ЕСАУЛОВ, 2013а: 8, 11) обозначает не жителя Заэкратье (тогда надо было бы использовать модель *заекранець* по аналогии с *закарпатець*), а участника, «бойца» компьютерной игры, отсюда и *воїн-заекранник* (ЕСАУЛОВ, 2013а: 13). Эта модель используется писателем в окказионализме *цешанець* (житель планеты «Ц») (ЕСАУЛОВ, 2013с: 44). Таким образом окказионализмы *Заекрання, заекранський, заекранник* созданы писателем с целью номинации нового объекта в конкретном тексте и не имеют перспектив употребления вне данного контекста. Кстати, на сайте «Словотвір» есть предложение использовать слово *заекрання* вместо нового заимствования *брендвойс* – ‘голос за кадром’, но этот вариант не получил поддержки пользователей.

8. Словообразовательный потенциал неосемантизма *цифровий*

Распространение современных цифровых технологий, получение и обработка информации в *цифровом формате* (англ. *digital data*) с использованием цифровых технологий и *цифровых устройств* (англ. *digital device*) стали источником нового семантического заимствования из английского языка *цифровий*, в отличие от давнего заимствования *цифра* из арабского языка через латынь и французский (фр. *chiffre* – ‘число’). Попытки предложить украинский вариант *числина* вместо *цифра* на сайте «Словотвір» не могут увенчаться успехом, хотя слово и соответствует украинской словообразовательной модели. Но статистика употребления слова *цифра* в текстах корпуса ГРАК – 9 146 словоупотреблений на 1 миллион – свидетельствует о его стабильном функционировании в языке. Компьютерный фонд инноваций подтверждает также высокий синтагматический и словообразовательный потенциал прилагательного *цифровий* (англ. *digital*) в нескольких десятках примеров: *цифровий декодер, запис, канал, носій; цифрове обладнання, телебачення; цифрова версія, мережа, модель*; наличием новых производных *оцифрувати, оцифруватися, оцифрування, цифровик (цифровий фотоапарат)*: «На „озброєнні” юних фотографів один звичайний „цифровик”...» (Хрещатик, № 66, 2006). В словаре «Словотворчество независимой Украины. 2012–2016» А. Нелюбы, Е. Редько находим пример окказионального словообразования: *цифровізація* в значении ‘опредмеченное действие (процесс), связанное с цифровыми технологиями’ (СНУ 2017: 436). В книге А. Есаулова «Game over!» использован глагол *перещифрувати* – «...тебе ми зараз перещифруємо на цьому перетворювачі» (ЕСАУЛОВ, 2013с: 91) и название лица *цифбоєць* (ЕСАУЛОВ, 2013с: 114), т.е. *цифровий боєць* – ‘участник компьютерной игры’. Множество производных от производящей единицы в новом значении, связанном с передачей данных в цифровом формате,

свидетельствует о стабилизации слова *цифровий* именно в новом терминологическом значении (*цифровий сигнал – бітовий потік*), хотя в общеупотребительных словарях это значение отсутствует.

9. Конкуренция номинаций *смайл – усміхайлик*

Смайл как элемент массовой культуры возник благодаря распространению сети Интернет в широких слоях населения. Среди попыток найти украинский вариант замены *смайла* (англ. *smile* – ‘улыбка’, укр. *усмішка*) – ‘короткой знакограммы, отображающей эмоцию на письме’, особой популярностью пользуется *усміхайлик*, зафиксированный и в словаре «Словотворчество независимой Украины. 2012–2016» А. Нелюбы, Е. Редько, и на сайте «Словотвір», и в единичном примере в корпусе ГРАК (текст из Фейсбука; для сравнения *смайл* в ГРАК имеет 110 словоупотреблений на 1 млн), а статистические данные поиска в Гугл дают более 32 000 найденных единиц. Другие предложения на сайте «Словотвір» *емоційка, почуттівка* являются лексическими окказионализмами и более точно отображают его смысловую сущность – выражать различные эмоции, а не только улыбку, но пока остаются нераспространенными в узусе.

10. Неологизмы и окказионализмы в интернет-коммуникации

В проанализированных примерах окказионализмы исконного происхождения созданы преимущественно сознательно с целью номинации новых понятий и явлений средствами украинского языка. Однако существует большое число окказионализмов гибридного происхождения, в которых заимствованный термин (либо названия баз данных, соцсетей, поисковых систем) используется как производящая основа для образования окказиональных производных, в том числе и с целью выражения эмоциональной оценки. Например, *кібер* как сокращенная основа прилагательного *кібернетичний* во многих сложных словах-новообразованиях (в словаре «Словотворчество независимой Украины. 2012–2016» А. Нелюбы, Е. Редько насчитывается 39 таких лексем) имеет прагматический ореол, например, названия операций по блокированию мобильных телефонов террористов, организованных активистами *Кіберураган, Кіберштурм* или же ироническое название компьютера *кіберящик* по аналогии с *телеящик*. Названия соцсетей имеют также высокий словообразовательный потенциал, реализующийся и в окказиональной разговорной лексике. Это прежде всего зафиксированный в названном словаре целый спектр названий лиц – пользователей сетей, среди которых, например, Фейсбук: *фейсбукець, фейсбуківець, фейсбучанин, фейсбучник, фейсбуківка, фейсбучка*, и Инстаграмм: *інстаграмівець* и др. Значительная часть это гибридные глаголы, образованные по активным словообразовательным моделям, но с выразительным разговорным оттенком, лексические окказионализмы: *перегуглити* (избыточное действие: ‘прогуглить больше, чем кто-либо’), *оскопусити* (‘заставить писать статьи в издания, индексированные в наукометрической базе «Скопус»’), *перейсбучити* (избыточное действие: ‘писать в Фейсбуке больше, чем кто-либо’). Контексты дают ещё дополнительные эмоциональные коннотации, усиливающие окказиональность глаголов.

11. Выводы

Проанализированный материал позволяет утверждать, что хотя лексика, принадлежащая к сфере информатики, преимущественно создается с целью номинирования новых объектов и явлений, попытки создания терминов–неологизмов на основании словообразовательных ресурсов украинского языка имеют переменный успех. Частично это связано с объективными моментами: краткостью номинации, отсутствием побочных значений или коннотаций, вхождением в язык сразу целым словообразовательным гнездом, т.е. высокий словообразовательный потенциал; частично с субъективным фактором «моды» или «привычки» пользователей. Благородная миссия создавать компьютерную терминологию на базе ресурсов только украинского языка сознательно направлена на сохранение и активизацию использования украинского языка в кибер- и интернетпространстве. Такое словотворчество носителей языка приводит к успеху, когда в номинацию включаются семантическое словообразование, явление «миграции» терминов, метафоризация лексики. Но в ряде случаев авторские предложения производных украинских лексем на роль терминов терпит фиаско. Это связано с отягощенностью значения слова другими значениями, если это неосемантизм, неспособностью производного образовать нужное словообразовательное гнездо терминов. Употребление новообразований ограничивают нежелательная омонимия, низкий или нулевой словообразовательный потенциал по сравнению со стабилизировавшимися заимствованиями, привязанность к контексту либо создание окказиональных единиц с целью выразить отношение или дать оценку тому или иному явлению.

Словари

- АРСУН, 2013: Карпшовська Є.А., Кислюк Л.П., Клименко Н.Ф., Критська В.І., Пуздирева Т.К., Романюк Ю.В., ред., Карпшовська Є.А., відп. ред.: *Активні ресурси сучасної української номінації: Ідеографічний словник нової лексики*. Київ.
- ГС, 2011: Критська В.І., Недозим Т.І., Орлова Л.В., Пуздирева Т.К., Романюк Ю.В., ред., Клименко Н.Ф., відп. ред.: *Граматичний словник української літературної мови. Словозміна: Близько 140 000 слів*. Київ [Електронний ресурс, режим доступу: <http://www.mova.info/>; доступ: 19.12.2019].
- СНУ, 2017: Нелюба А., Редько Э., ред., Нелюба А., заг. ред.: *Словотворчість незалежної України. 2012–2016. Словник*. Харків.

Электронные ресурсы

- Архівна лексична картотека української мови* [Електронний ресурс, режим доступу: <http://mova.cf/>; доступ: 19.12.2019].
- Генеральний регіонально анотований корпус української мови (ГРАК)*. Шведова М., фон Вальденфельс Р., Яригин С., Крук М., Рисін А., Старко В., Возняк М., Київ, Осло, Єна, 2017–2019 [Електронний ресурс, режим доступу: uacorpus.org; доступ: 19.12.2019].
- Корпус української мови* [Електронний ресурс, <http://www.mova.info/corpus.aspx?l=209>; доступ: 19.12.2019].
- КФІ: Науково-дослідницький комп'ютерний фонд інновацій*.

Морфемно-словотвірний фонд української мови.

Російсько-українські словники [Електронний ресурс, режим доступу: <http://r2u.org.ua/>; доступ: 19.12.2019].

Словотвір [Електронний ресурс, режим доступу: <https://slovotvir.org.ua/>; доступ: 19.12.2019].

Литература


- Булаховський К.А., 2010: *Стан і проблеми розвитку україномовної локалізації комп'ютерних програм.* „Українська мова”, № 4, с. 89–94.
- ЕСАУЛОВ О.Ю., 2013а: *«Game over!»*. Харків.
- ЕСАУЛОВ О.Ю., 2013б: *Антивірус*. Харків.
- ЕСАУЛОВ О.Ю., 2013с: *Таємниця великого Сканера*. Харків.
- КАРПЛОВСЬКА Є.А., 2008: *Тенденції оновлення сучасного українського лексикону*. В: Клименко Н.Ф., відп. ред.: *Динамічні процеси в сучасному українському лексиконі*. Київ, с. 6–133.
- Клименко Н.Ф., 2017: *Процеси розвитку терміносистеми інформатики*. В: Карпловська Є.А., відп. ред.: *Вплив суспільних змін на розвиток української мови*. Київ, с. 122–183.
- КОЛОІЗ Ж.В., 2007: *Українська оказіональна деривація: монографія*. Київ.
- ПУСОВІТ Л.О., Клименко Н.Ф., 2004: *Оказіоналізм*. В: Русанівський В.М., Тараненко О.О., співгол. ред-кол.: *Українська мова: Енциклопедія*. Київ, 2-е вид., с. 432–433.
- СТАРКО В.Ф., 2017: *Комп'ютерні лінгвістичні проекти гурту r2u: стан та застосування*. „Українська мова”, № 3, с. 86–100.
- ЯНКО-ТРИНИЦКАЯ Н.А., 2001: *Оказіональное словообразование*. В: Той же: *Словообразование в современном русском языке*. Москва, с. 462–482.
- BELENTSCHIKOW R., 2003: *Неузואльное словообразование: Восточнославянские языки*. W: OHNHEISER I., red. nauk.: *Słowotwórstwo / Nominacja*. Innsbruck, Opole, s. 249–261.
- BUZASSYOVA K., MARTINCOVA O., 2003: *Neuzualni slovtovorba v zapadoslovanskykh jazycich*. W: OHNHEISER I., red. nauk.: *Słowotwórstwo / Nominacja*. Innsbruck, Opole, s. 262–275.
- CHRUŚCIŃSKA K., 1978: *O formacjach potencjalnych i okazjonalizmach*. W: SZYM CZAK M., red.: *Z zagadnień słownictwa współczesnego języka polskiego*. Wrocław–Warszawa–Kraków–Gdańsk, s. 69–79.
- JADACKA H., 2001: *System słowotwórczy polszczyzny (1945–2000)*. Warszawa.
- JÓZWIKIEWICZ P., 2011: *Ukraińska terminologia informatyczna: stan, problemy, zapożyczenia*. „Slavica Wratislaviensia”, nr 154, s. 121–132.
- JÓZWIKIEWICZ P., 2014: *Ukraińskie słowniki terminologii informatycznej (1990–2010)*. „Slavica Wratislaviensia” CLIX, Acta Universitatis Wratislaviensis, nr 3578, s. 179–188.


Björn Wiemer
Johannes Gutenberg-Universität Mainz (Niemcy)
 <https://orcid.org/0000-0001-6466-2060>
wiemberb@uni-mainz.de

Forum Lingwistyczne
nr 7, 2020
ISSN 2450-2758 (wersja elektroniczna)



DOI: <http://doi.org/10.31261/FL.2020.07.04>

Joanna Wrzesień-Kwiatkowska
Johannes Gutenberg-Universität Mainz (Niemcy)
 <https://orcid.org/0000-0001-8371-1665>
kwiatkow@uni-mainz.de

Marek Łaziński
Uniwersytet Warszawski
 <https://orcid.org/0000-0001-5718-4435>
m.lazinski@uw.edu.pl

Badania aspektu w językach polskim, czeskim i rosyjskim za pomocą korpusów i baz danych (pierwsze podsumowanie tematu)

Investigations of Polish, Czech and Russian Aspect
on the Basis of Corpora and Databases
(a First Summary on a Topic)

Abstract: The article is connected to the project “DiAsPol250” in which Polish is compared to Czech and Russian from the perspective of the evolution of their aspect systems (<http://www.diaspol.uw.edu.pl/>). We make use of existing synchronic and diachronic electronic corpora and are building a corpus of our own with annotated aspect pairs; we also create a database of aspect triplets whose role we consider as particularly important for the system. We want to assess which changes have occurred since the mid-18th century in prefixing and suffixing strategies of verb stems, both in general and by comparing particular prefixes and suffixes, especially so-called natural (*vs.* specialized) prefixes (according to Janda et al., 2013).

The article supplies a sketch of the general premises of the project, and it summarizes our experience with existing large corpora and databases which we have been employing. We also present a case study in order to demonstrate a procedure designed to compare the distribution of the Czech prefix *z-* in triplets and in the corpus. This procedure is meant to check more general tendencies; it also illustrates why electronic corpora cannot be replaced in research on distributional properties and why their role does not consist simply in providing examples for the illustration of hypotheses.

Key words: aspect, verbal prefixes, aspect triplets, electronic corpora, annotation

Abstrakt: Artykuł wiąże się z projektem „DiAsPol250”, w którym polszczyzna jest porównywana z językami czeskim i rosyjskim pod kątem rozwoju systemu aspektowego (<http://www.diaspol.uw.edu.pl/>). Do pracy wykorzystujemy istniejące elektroniczne korpusy synchroniczne i diachroniczne oraz budujemy własny korpus z anotacją par aspektowych, ale też tworzymy bazę trójek aspektowych, których rolę w systemie uważamy za szczególnie istotną. Chcemy ocenić, jakie zmiany od połowy XVIII wieku zaszły z udziałem prefiksacji i sufiksacji podstaw czasownikowych względem siebie oraz wśród poszczególnych prefiksów i sufiksów, zwłaszcza prefiksów tzw. naturalnych (*vs.* specjalne) (według JANDA i in., 2013).

W artykule krótko naszkicujemy ogólne założenia projektu i podsumujemy doświadczenia z istniejącymi dużymi korpusami i bazami danych, z których korzystamy. Zaprezentujemy również studium

przypadku, którego celem jest pokazanie, jak można porównać rozkład prefiksu *z-* w czeskich trójkach i korpusie, by sprawdzić ogólne tendencje i unaocznić, dlaczego korpusy elektroniczne są niezbędne do badań nad dystrybucją, a ich rola nie sprowadza się jedynie do ilustracji tez.

Słowa kluczowe: aspekt, prefiksy czasownikowe, trójki aspektowe, korpusy elektroniczne, anotacja

Mimo że aspekt, czyli opozycja czasowników dokonanych (DK) i niedokonanych (NDK), stanowi kategorię centralną w gramatyce dowolnego języka słowiańskiego, a jego opisowi poświęcono ogromną literaturę, to jednak stan badań nad tą kategorią wykazuje liczne białe plamy, zwłaszcza jeżeli chodzi o polszczyznę. Luki wynikają między innymi z braku rzetelnych badań diachronicznych i korpusowych. W niniejszym artykule chcemy przedstawić ogólne założenia projektu „DiAsPol250”, w którym polszczyzna jest porównywana z językami czeskim i rosyjskim pod kątem zmian w systemie aspektowym od połowy XVIII wieku. W ramach tego projektu korzystamy z różnych korpusów i innych baz elektronicznych oraz budujemy własne korpusy (punkt 1). Podsumujemy nasze dotychczasowe doświadczenia z nimi i ocenimy, w jakim stopniu nadają się one do badań aspektologicznych (punkt 2). Następnie zaprezentujemy studium przypadku czeskich prefiksów *z-* i *po-*. Pokaże ono, jak można, porównując rozkład w korpusie i w naszej bazie trójek aspektowych, sprawdzić ogólne tendencje. Poza tym unaoczni, dlaczego korpusy elektroniczne są niezbędne do badań nad dystrybucją oraz że ich rola nie sprowadza się jedynie do ilustracji tez (punkt 3). Artykuł zakończymy podsumowaniem (punkt 4).

1. Założenia i cele projektu „DiAsPol250”

Projekt *Rozwój polskiego systemu aspektowego w ostatnich 250 latach na tle sąsiednich języków słowiańskich* („DiAsPol250”, <http://www.diaspol.uw.edu.pl/>)¹ ma na celu opisanie systemu aspektowego polszczyzny i jego ewolucji od 1750 roku do dziś na tle języków rosyjskiego i czeskiego. Pod pewnymi względami polszczyzna może być traktowana jako strefa przejściowa użycia aspektu między typem wschodnim a zachodnim (DICKEY, 2000), dlatego stanowi dobry punkt wyjścia do porównań wewnątrzsłowiańskich. W naszych badaniach koncentrujemy się przeważnie na zagadnieniach pominiętych przez Stephena Dickeya albo skądinąd zanedbywanych w aspektologii: na roli tzw. trójek (zob. poniżej) i integracji nowszych zapożyczeń oraz na różnych rodzajach użycia nieaktualnych czasu teraźniejszego i wyborze aspektu w kontekstach o zawieszonej asercji. Badamy także użycie czasu teraźniejszego czasowników niedokonanych w tekstach prawnych i kodeksach oraz ich interpretację przez prawników i przeciętnych użytkowników polszczyzny. Badania te staramy się jak najbardziej opierać na korpusach.

Wychodzimy z założenia, że synchroniczne różnice między językami słowiańskimi we wzorcach derywacyjnych i w wyborze wartości aspektu w różnych kontekstach w dużym stopniu odzwierciedlają zmiany diachroniczne w systemach aspektowych poszczególnych ję-

¹ Projekt jest finansowany w ramach programu *Beethoven II* (2018-2020) przez polskie NCN i niemiecką DFG (nr projektu NCN: 2016/23/G/HS2/00922, DFG: WI 1286/19-1).

zyków. Założenie to wynika z zasad chronotopii (WIEMER, 2008; 2015). W dalszej części tego tekstu przedstawimy jedynie zagadnienia związane z wzorcami tworzenia par aspektowych oraz wynikające stąd role poszczególnych przedrostków. Nie chodzi nam o podsumowanie dotychczasowej polskiej literatury aspektologicznej, ponieważ poruszany problem konkurencji wykładników prawie w niej nie występuje. Wyjątkiem są prace historyczne Witolda Śmiecha i Jerzego Kuryłowicza (KURYŁOWICZ, 1987a [1929]; 1987b [1968]; ŚMIECH, 1986).

1.1. Ogólne założenia aspektologiczne

Ze strony formalnej trzon opozycji aspektowej w językach słowiańskich stanowi produktywna derywacja podstaw czasownikowych za pomocą prefiksów i sufiksów. Fakt ten sprawia, że aspekt jest rozróżniany nie tylko w formach finitywnych, ale w zasadzie też we wszystkich formach niefinitywnych². Derywacja podstaw nie zasada się na konkretnych wykładnikach DK i NDK, lecz na schematach derywacyjnych zmieniających aspekt; dwa najbardziej produktywne wzorce przedstawia schemat 1:

Schemat 1

Produktywne schematy derywacyjne prowadzące do zmiany aspektu

(A) podstawa bezprzedrostkowa ⇒ derywat przedrostkowy

simplex^{NDK} ⇒ przedrost^{DK} skrót: simp^{NDK}-pref^{DK}

np. *czynić* ⇒ *u-czynić*

budować ⇒ *z-budować*

(B) podstawa przedrostkowa ⇒ derywat sufigowany (= imperfektyw sekundarny)

przedrost^{DK} ⇒ przedrost+suf^{NDK} skrót: pref^{DK}-sek^{NDK}

np. *rozczynić* ⇒ *rozczygni-a-ć*

rozbudować ⇒ *rozbudow-ywa-ć*

W ten sposób tworzy się też większość par aspektowych. Ogólnie zakłada się, że człony par aspektowych łączy tożsamość znaczenia leksykalnego. Jednak nawet gdy taka tożsamość zachodzi, człony odpowiednich par podstaw nie są zwykłymi synonimami, gdyż zasadniczo różnią się co do dystrybucji w gramatycznie lub pragmatycznie relewantnych kontekstach. Konteksty te dają się opisać jako ograniczenia w różnych formatach morfoskładniowych, zaczynając od poszczególnych form (a), przez połączenia na poziomie fraz (b), do poziomu zdania prostego (c) i połączeń zdań (np. zdania warunkowe). W dodatku niektóre konteksty można ująć w postaci wymagań dyskursywnych, na przykład wiedzy presuponowanej (d), lub za pomocą różnych typów krotności, zwykle na poziomie zdania prostego (c). Siła tych ograniczeń jest różna: od kategorialnej (zob. schemat 2, przykłady 1–3) do rozkładów probabilistycznych (zob. schemat 2, przykłady 4–6). Część wymienionych tu rozróżnień wchodzi w skład tzw. funkcji kanonicznych (zob. LEHMANN, 1999: 220–223).

² Zastrzeżenia można mieć jedynie co do rzeczowników odsłownych (*nomina actionis*) w większości języków słowiańskich.

Konteksty określające użycie podstaw czasownikowych według klas (DK *vs* NDK)

Ograniczenia na poziomie

(I) form morfologicznych, np. imiesłowy nieodmienne:

(1a) DK: *-wszy vs* NDK: *-qc*

(1b) *zbudowaw-szy (*zbuduj-qc) vs buduj-qc (*budowaw-szy)*

(II) fraz, np. połączenia z czasownikami fazowymi (2), z czasownikiem pomocniczym wskazującym na czas przysłyty (3) lub o funkcji modalnej (4):

(2) *przystał pisać / *napisać*

(3) *będzie pisać / *napisać*

(4a) *Okna nie można otwierać. (Nie pozwalam.)* → zakaz (niemożliwość deontyczna)

(4b) *Okna nie można otworzyć. (Klamka jest zepsuta.)* → niemożliwość „obiektywna”
(*circumstantial modality*)

(III) zdań prostych, np.

(5a) *Codziennie pisze / ?napisze po jednym liście.*

(5b) *Czasem do nas zagłąda / zajrzy.*

(IV) wiedza presuponowana – niepresuponowana, np.

(6a) *Opowiadaj! (Bo już czekamy i pękamy od ciekawości.)*

(6b) *Opowiedz! (Jeśli chcesz.)*

Tradycyjnie do kanonicznych funkcji związanych z wyborem aspektu wchodzi też rozróżnienie zdarzenia *vs* procesu, zwłaszcza między członami par telicznych, na przykład *Rodzice budowali dom – Rodzice zbudowali dom* (KOSCHMIEDER, 1934; ŁAZIŃSKI, WIEMER, 1995). Na rozważanie związku i konkurencji między takimi heterogenicznymi funkcjami nie ma tu miejsca (omawiano je wybiórczo np. w WIEMER, 2008; 2015).

Założenie, że pary aspektowe stanowią centrum systemu aspektowego, jest słuszne w tym sensie, że człony takich par unaoczniają ścisły związek między konsekwentną i przejrzystą procedurą morfologiczną (derywacją jednych podstaw z innych, prostszych) a dystrybucją funkcjonalną dążącą do komplementarności. Wiele podstaw nie wchodzi jednak w takie pary, tak że nawet przy dość „liberalnym” ustalaniu parzystości pozostałaby ogromna liczba (chyba nawet większość) podstaw *tantum*, czyli leksemów czasownikowych bez choćby kontekstowego synonimu o komplementarnej dystrybucji według wskazanych kryteriów.

Systemu więc nie da się zrozumieć, jeżeli nie uwzględnimy również licznych podstaw stojących poza parami albo nawet „sieciami” czy „grupami akcjonalnymi” (zakładanymi na podstawie różnych przesłanek, por. JANDA, 2007; MENDE i in., 2011; TATEVOSOV, 2016), tym bardziej że właśnie produktywność przywołanych procesów morfologicznych sprawia, że derywatów używanych w mowie jest znacznie więcej, niż potrafią to oddawać nawet największe słowniki. Musimy więc uświadomić sobie, że podstawy (względnie leksemy) czasownikowe **należą do klasy** tzw. dokonanych lub niedokonanych **niezależnie od parzy-**

stości; uznając tę lub inną podstawę za dokonaną lub niedokonaną, w pierwszej kolejności odnosimy ją do jednej z dwóch klas i **tylko przez to** zawężamy jej możliwości realizowania funkcji wziętych z zasobu takiego, jak go naszkicowano (zob. schemat 1). Zważmy również, że rzadko się zdarza, iż ten sam leksem czasownikowy dysponuje wszystkimi funkcjami z zasobu funkcji przypisanych klasie DK lub NDK. Zatem, ściśle mówiąc, opozycyjne zasoby funkcji (lub ograniczeń kontekstowych) przysługują nie tyle leksemom (lub podstawom) czasownikowym w szczególności, ile ich klasom. To powoduje, że słowiańska opozycja DK–NDK ma charakter klasyfikacyjny (WIEMER, SERŻANT, 2017: 243–255).

1.2. Perfektywa naturalne i trójki

Pary typu (A) (simp^{NDK}-pref^{DK}) są możliwe tylko wtedy, kiedy przedrostek nie wnosi dodatkowego elementu semantycznego do znaczenia leksykalnego podstawy simpleksu; na przykład *budować* – *z-budować*, *pisać* – *na-pisać*, *dziękować* – *po-dziękować*, *malować się* – *po-malować się*. Przedrostki w takich parach dawniej uważano za „puste”, w rzeczy samej jednak nie brakuje im wkładu własnego do znaczenia rozszerzonej podstawy, tylko ich własny potencjał semantyczny zbiega się z jednym z elementów znaczenia podstawy prostej (= simpleks). Należy więc raczej mówić o gaszeniu pokrywających się elementów znaczeniowych, co odpowiada ogólnym zasadom łączliwości semantycznej (zob. ŁAZIŃSKI, 2011 oraz JANDA i in., 2013). Ostatni autorzy człony DK w takich parach nazywają *Natural Perfectives* (NatPerf), w odróżnieniu od *Specialized Perfectives* (SpecPerf), w których to przedrostek zmienia znaczenie simpleksu w sposób mniej lub bardziej wyrazisty. Od NatPerf mogą powstać dalsze derywaty, tym razem poprzez dodanie lub zmianę sufiksu. Tak zwana wtórna imperfektywizacja (np. *przepisać* ⇒ *przepis-ywa-ć*, *zmienić* ⇒ *zmeni-a-ć*) jest uznana za najbardziej produktywny wzorzec tworzenia par aspektowych, przynajmniej w polszczyźnie i językach wschodniosłowiańskich. Zazwyczaj pary tego typu powstają z podstaw prefigowanych należących do SpecPerf; powstają łańcuchy derywacji typu

NDK (= simpleks)	DK	NDK
(7) <i>łowić</i>	⇒	<i>wy-łowić</i> ⇒ <i>wyławi-a-ć</i> ,

w których tylko DK (*wyłowić*) i wtórny NDK (*wyławiać*) stworzony na podstawie DK mogą uchodzić za parę aspektową, a NDK simpleks (*łowić*) z wtórnym NDK o taki status nie konkuruje. Inaczej ma się sprawa, gdy DK stanowi NatPerf w stosunku do NDK simpleksu, tak że zarówno ten ostatni, jak i wtórny NDK (= NDK2) mogą uchodzić za odpowiednik NDK do DK o tym samym znaczeniu leksykalnym; por.:

NDK1 (= simpleks)	DK	NDK2
(8) <i>tworzyć</i>	⇒	<i>s-tworzyć</i> ⇒ <i>stwarz-a-ć</i>
(9) <i>szacować</i>	⇒	<i>o-szacować</i> ⇒ <i>oszacow-ywa-ć</i>

W takich to przypadkach mówimy o trójkach aspektowych. W większości prac na ich temat badacze omawiają pytanie o to, czy któregoś z dwóch NKD (zwykle NDK1) nie należy uznać za „zbędny”, oraz proponują różne sposoby, by określić, który z NDK stanowi

„lepszą” parę dla DK. Z tą ostatnią kwestią wiążą się próby podziału trójek na różne grupy wynikające z dystrybucji NDK1 i NDK2 na ten lub inny inwentarz funkcji³.

W naszym projekcie zamierzamy ustalić zasięg liczby i typów trójek w językach polskim, czeskim i rosyjskim. Należy ustalić, czy liczba trójek w okresie 1750–2020 stopniowo spadała, rosła czy podlegała wahaniom, i czy pod tym kątem badane trzy języki się nie różnią. Dla każdego z nich stworzyliśmy bazę danych o identycznej budowie zawierającą informacje o formach gramatycznych, ich frekwencji i okresach występowania oraz podstawowe testy na tożsamość leksykalną opierające się na kryteriach Jurija Masłowa (MASŁOV, 2004 [1948]). Oprócz tego dla wybranych kandydatów na trójki zaczęliśmy gromadzić dane korpusowe pozwalające prześledzić zmiany w rozkładzie form i funkcji oraz w relacji między częstotliwością typów (produktywności w zasobie leksykalnym) a częstotliwością tokenów (w dyskursie). Rzecz jasna, tego rodzaju informacje można otrzymać jedynie na podstawie korpusów i próbek losowych. Prace te trwają. W punkcie 3 zaś przedstawimy studium przypadku pokazujące, dlaczego korpus i bazy danych stanowią niezastępowalny fundament do badań nad dystrybucją i funkcjami przedrostków w trójkach. Najpierw jednak podsumujemy nasze doświadczenia z korpusami i bazami danych w Internecie, z których korzystaliśmy do celów projektu (punkt 2).

2. Praca aspektologiczna z korpusami i bazami danych

Problemy związane z leksykografią aspektu, zwłaszcza par aspektowych, Elena N. Remčukova zalicza do centralnych zadań aspektologii słowiańskiej. Jej zdaniem w tradycyjnej rosyjskiej leksykografii „видовая характеристика глагола является важнейшей составляющей грамматической информации в толковом словаре”⁴ (REMČUKOVA, 2008: 199). Taka opinia nie odpowiada jednak strukturze czy też podejściu obecnym w wielu istniejących bazach danych dostarczających gramatycznych informacji o czasownikach w różnych językach słowiańskich.

2.1. O anotacjach i parach, zwłaszcza z udziałem podstaw simplexu

Faktycznie badanie aspektu zużyciem korpusów, szczególnie relacji semantycznych i funkcjonalnych między poszczególnymi podstawami związanymi derywacyjnie, jest zajęciem nader czasochłonnym. Problemy wynikają nie tyle z tego, że dotąd w żadnym korpusie języka słowiańskiego nie zawarto anotacji umożliwiającej odnośnienie do siebie członów par aspektowych (podobnej do odsyłaczy w autorytatywnych słownikach nowszej daty), lecz także z tego, że aspekt jako taki był traktowany po macoszemu. Przedstawianie zaś stosunku między podstawami simplexu (które przeważnie należą do NDK) a ich derywatami przedrostkowymi często wynikało ze zbyt ogólnych, a przez to upraszczających, ustaleń wielu wpływowych autorów.

³ Trzeba przyznać, że dotąd takie analizy przeprowadzano niemal wyłącznie na materiale współczesnego języka rosyjskiego (zob. częściowy przegląd w WIEMER, 2017: 235–244; 2019: 51–56).

⁴ „Charakterystyka aspektualna czasownika jest jedną z ważniejszych składowych informacji gramatycznych w słowniku jednojęzycznym [objaśniającym]”.

Najlepiej to widać na przykładzie czeskiego korpusu narodowego (*Český národní korpus* – ČNK): w starszej wersji (SYN2000) aspektu nie uwzględniono w tagsecie. Widocznie przyczyną było założenie, że aspekt to sprawa słownika, a nie morfologii (Tilman Berger, informacja osobista). Dopiero od wersji SYN2005 aspekt został włączony do tagsetu na ostatniej z pozycji⁵. Jednak tag ‘aspekt’ nadal nie oddaje relacji między podstawami prostymi (zwykle NDK) a ich przedrostkowymi derywatami DK, które mogą uchodzić za NatPerf-y. Ten mankament w dużej mierze jest skutkiem wpływu wielu klasycznych prac o aspekcie w językach słowiańskich, takich jak na przykład Aleksandra Isaczenki (ISAČENKO, 2003 [1960]), w których to pary typu (A) (simp^{NDK}-pref^{DK}) nie były uznawane za „czyste” pary aspektowe. Ten wpływ wyraźnie odbija się zarówno w ČNK, jak i w słowniku walencyjnym Vallex 3.0. Ów ostatni jest pomocny, jeżeli chodzi o ustalanie znaczenia leksykalnego poszczególnych członów trójki (lub par aspektowych), ale nie w przypadku ustalania samych par lub trójek. Dla polszczyzny sytuacja do niedawna wyglądała bardzo podobnie (zob. punkt 2.2).

Opisany tu stan kontrastuje z tym, że akurat badacze języka czeskiego ogólnie mają duży wachlarz możliwości, jeśli chodzi o internetowe źródła językowe. Synchroniczna część ČNK rejestruje 5 mld tokenów. Trudności jednak napotkamy w diachronicznej części korpusu. Korpus Diakorp6 (KUČERA, ŘEHOŘKOVÁ, STLUKA, 2015) liczy 4 mln tokenów od XIV do XX wieku i nie jest anotowany morfologicznie, co utrudnia wyszukiwanie. Generalnie korpusy czeskie mogą służyć do potwierdzania, czy czasowniki wchodzące w skład (potencjalnej) trójki mają to samo znaczenie leksykalne, ale ustalenie odpowiedniości w parach lub trójkach wymaga czasochłonnej ręcznej analizy. Natomiast w celu wyszukiwania kandydatów na trójki (lub też pary) należy sięgać po inne źródła, na przykład po listę zaprezentowaną w książce François Esvana (ESVAN, 2007). Listy w tej publikacji to ekscerpcja z korpusu SYN2000, dlatego do celów naszej bazy konieczne było uzupełnienie jej o listę czasowników z korpusu syn7 (KŘEN i in., 2018) oraz ze źródeł słownikowych, na przykład ze słowników starej polszczyzny dostępnych na *Vokabuláři webovým* (zob. bibliografię)⁶. Poza pracą Esvana w żadnym ze źródeł aspekt nie stanowi centralnego zagadnienia, sama ta praca zaś nie przedstawia spójnej teorii, a zawiera głównie tabele i deskryptywną statystykę na podstawie korpusu.

Jeżeli chodzi o korpusy narodowe języków rosyjskiego (NKRJa) i polskiego (NKJP), to aspekt był w nich tagowany od samego początku. W NKRJa bogaty zestaw tagów morfologicznych legł u podstaw specjalnego interfejsu do korpusu głównego, który sięga wstecz do połowy XVIII wieku, a interfejs ułatwia zadawanie zapytań o kombinacje form i kategorii gramatycznych. Istnieje jeszcze podkorpus *starorusskij* obejmujący okres od XVI do połowy XVIII wieku, który niestety nie jest otagowany. Natomiast NKJP, który odzwierciedla stan polszczyzny od 1945 roku⁷, oferuje dwie wyszukiwarki, a tylko w jednej z nich (Poliqarp) kategorie gramatyczne, w tym aspekt, są anotowane. Składnia zapytań Poliqrpa stanowi też podstawę do zapytań w *Elektronicznym Korpusie Tekstów Polskich*

⁵ Zob. [online: <http://wiki.korpus.cz/doku.php/seznamy:tagy> i <http://wiki.korpus.cz/doku.php/en:cnk:syn2005>; data dostępu: 28.01.2020].

⁶ Pełny opis naszej procedury badawczej (wraz z wykazem wszystkich źródeł) zostanie umieszczony w innej, obszerniejszej publikacji. Tu skupiamy się jedynie na roli korpusów i baz elektronicznych.

⁷ Teksty powstałe przed 1939 rokiem weszły w skład NKJP tylko w znikomej liczbie.

z XVII i XVIII w. (do 1772 r.) (Korpus Barokowy – KorBa), powstałym pod koniec 2018 roku. Obejmuje on okres 1660–1772. Ogólny kłopot z badaniami diachronicznymi polszczyzny, nie tylko w zakresie aspektu, polega na niemożności korzystania z większych otagowanych korpusów odbijających stan języka między 1772 a 1945 rokiem. Dotąd istniał tylko *Mikro-korpus polszczyzny 1830–1918*, również oparty na systemie Poliqarpa, ale zawierający tylko ok. 1 mln tokenów. Jednak obecnie w ramach „DiAsPol250” budujemy polsko-niemiecki korpus równoległy obejmujący teksty z ostatnich 250 lat (ok. 10 mln tokenów)⁸. Powstaje też korpus XIX-wiecznej polszczyzny składający się z 12 mln tokenów w 380 tekstach.

2.2. Tagowanie par aspektowych („DiAsPol250”)

W polsko-niemieckim korpusie równoległym powstającym w ramach projektu „DiAsPol250” czasowniki otagowane są jako człony par aspektowych, dwuaspektowe, tylko niedokonane lub tylko dokonane. Taka procedura, zastosowana po raz pierwszy w korpusie jakiegokolwiek języka słowiańskiego, wymaga pewnych uproszczeń, ale daje wiele możliwości badawczych. Uproszczenia wynikają z faktu, że relacje par aspektowych, tak jak przyporządkowanie czasownika do klasy akcjonalnej, dotyczą całego konkretnego predykatu wraz z argumentami, a nie abstrakcyjnego znaczenia słownikowego, dzielonego na podznaczenia. W zdaniu *Długopis cienko pisze* mamy *imperfectivum tantum*, ale w znaczeniu procesywnym, *Jan pisze właśnie list*, lub iteratywnym, *Jan często pisze listy*, podstawa *писаć* wchodzi w opozycję z *napisać*. Przypadków niejednoznacznego przyporządkowania do pary jest znacznie więcej.

W korpusie każde wystąpienie *писаć* jest opatrzone odnośnikiem do pary *писаć* – *napisać*. Leksem, który może wchodzić w opozycję aspektową, jest traktowany tak, jakby wchodził w nią zawsze. Podobnie jest w słownikach, w których pary aspektowe są łączone na poziomie leksemów, choć bardzo niekonsekwentnie. Większość słowników łączy w jednym haśle odpowiedniki sufiksalne (z różnymi sufiksami NDK, jak *zdob-ywa-ć* – *zdobyć*, lub z sufiksem DK *-ną-*, jak *stuk-ać* – *stuk-ną-ć*), a pary prefiksalne pozostawia w oddzielnych hasłach nawet bez odnośników. Nie ma więc żadnej informacji, że *писаć* i *napisać* coś łączy. Taki jest standard SJP Dor i kolejnych słowników opracowywanych przez wydawnictwo PWN, nawet najnowszego WSJP PWN. Dopiero słowniki przełomu tysiącleci, takie jak SWJP czy ISJP, stosują odnośniki między członami par prefiksalnych. Systematyczny przegląd traktowania par w leksykografii polskiej podaje Marek ŁAZIŃSKI (2020).

Wiemy już, że czasowniki *писаć* i *napisać* tworzą parę aspektową, ale nie wiemy, w których użyciach. Tego nie dowiemy się ani ze słowników, ani z korpusu tagowanego aspektowo według słownika. Jednak nawet godząc się na uproszczenia, możemy w otagowanym korpusie zadawać pytania o cechy składniowe czy fleksyjne czasowników, pomijając czasowniki *imperfectiva tantum*, które mają największą frekwencję, na przykład *być*, *mieć*, *znaczyć*. Możemy tak jak w pracy Laury Jandy i Olgi Lyashevskiej (JANDA, LYASHEVSKAYA, 2011) dla języka rosyjskiego ustalić profil gramatyczny aspektu. Aby to wszystko zrobić z pełną dokładnością, musielibyśmy czasowniki w każdym zdaniu w korpusie tagować ręcznie według znaczeń szczegółowych aspektu i dopiero wtedy decydować, czy czasownik

⁸ O założeniach prac nad tym korpusem zob. MEGER, WOŹNIAK, VON WALDENFELS, 2016.

w danym znaczeniu tworzy parę aspektową. Takiej procedurze zostanie poddany niewielki podkorpus naszego korpusu (zob. ŁAZIŃSKI, 2020).

3. Studium przypadku: czeskie przedrostki *z-* i *po-* w bazie trójek i w korpusie

W naszej bazie trójek znajdują się *simplicia*, czasowniki dokonane prefigowane oraz czasowniki wtórnie sufigowane, na przykład czes. *tratit – ztratit – ztrácet* '(z)gubić'. Czasowniki wtórnie sufigowane (NDK2) znajdujące się w naszej bazie różni od innych wtórnie sufigowanych czasowników to, że wchodzi w skład trójki, tj. są członem dodatkowym do pary aspektowej, którą stanowią simpleks (NDK1) oraz „naturalny dokonany” (*Natural Perfective*). Oznacza to, że każdy człon ma to samo znaczenie leksykalne, tak jak człony pary. Para, w której skład wchodzi „wyspecjalizowany dokonany” (*Specialized Perfective*), na przykład czes. *zpracovat – zpracovávat* 'przetworzyć – przetwarzać', jest wprawdzie de-rywatem od *pracovat* 'pracować', ale ma inne znaczenie leksykalne, dlatego nie tworzy trójki (zob. punkt 1).

Celem badania pilotażowego jest porównanie rozkładu prefiksów *z-* i *po-* wśród czeskich czasowników NDK2 (czasowniki typu *ztrácet*) w korpusie ČNK i w bazie czeskich trójek.

W czeskiej literaturze aspektologicznej znajdziemy dokładne opisy znaczeń poszczególnych prefiksów, ale według Ivany Kolařovej (KOLAŘOVÁ, 2013) tylko niektóre z nich mają funkcję czysto perfektywizującą (czyli w naszej wykładni: służą do tworzenia naturalnych DK): *na-*, *o-*, *po-*, *s-*, *u-*, *vy-*, *za-*, *z-*. Dla pozostałych 11 prefiksów autorzy nie znaleźli przykładów par aspektowych, gdzie prefigowany czasownik byłby naturalnym perfektywem. W tabeli trójek z kolei występuje 16 prefiksów (jeśli uznamy *o-* i *ob-* za warianty tego samego prefiksu): *z-*, *do-*, *na-*, *o-*, *od-*, *po-*, *pře-*, *při-*, *pro-*, *roz-*, *s-*, *u-*, *v-*, *vy-*, *vz-*, *za-*. Poza tabelą znalazły się prefiksy *nad-*, *pod-* oraz *před-*. Nasza analiza będzie się więc odnosić do wymienionych 16 prefiksów. Ogólny udział przedrostków w bazie trójek na podstawie danych ČNK jest zróżnicowany: od 3 dla *v-* do 116 dla *z-*.

Wybór analizowanych przez nas prefiksów jest nieprzypadkowy, ponieważ prefiks *z-* w czeskiej literaturze uznawany jest za prefiks perfektywizujący:

Z- je vyznamově téměř jednotité. Vyjadřuje většinou změnu stavu, a to i ve funkci prostě perfektyvizující. Funkce lexikální i prostě zdokonaující jsou tu těsně spjatý [...]⁹.

KOPĚČNÝ, 1962: 126

Většina sloves s tímto prefixem má primárně význam vidový, tj. vyjadřuje zejména změnu nedokonavého vidu na vid dokonavý [...]¹⁰.

KOLAŘOVÁ, 2013: 271

⁹ „Z- jest pod względem znaczenia niemal jednolite. I w funkcji czysto perfektywizującej wyraża zmianę stanu. Funkcja leksykalna i perfektywizująca są z sobą blisko związane”.

¹⁰ „Większość czasowników z tym prefiksem ma głównie znaczenie czysto aspektowe, tj. wyraża głównie zmianę z aspektu niedokonanego na dokonany”.

Prefiksowi *po-* przypisuje się więcej funkcji:

U této předpony nelze mluvit o jednotném základním východisku, a to pro významovou rozběžnost (předložky i předpony)¹¹.

KOPĚČNÝ, 1962: 129

Zřídka má prefix *po-* význam pouze vidový [...]. Slovesa s tímto prefixem nejčastěji vyjadřují: pohyb určitým směrem; pokrývání plochy; upravování či poškozování objektu; oslabení intenzity procesu; distribuci děje¹².

KOLAŘOVÁ, 2013: 253

Nasze badanie nie tylko ilustruje tezy przedstawione przez czeskich badaczy, ale również pozwala wskazać, czy czasowniki w bazie czeskich trójek preferują któryś z prefiksów. Mówiąc inaczej: porównując naszą bazę z danymi korpusowymi, można sprawdzić, czy pewne prefiksy przeważają nad innymi w tworzeniu *Natural Perfectives*. Trzeba bowiem pamiętać, że istnienie *Natural Perfectives* jest warunkiem koniecznym do wytworzenia się trójek (zob. punkt 1).

W bazie czeskich trójek znajdziemy 625 trójek (100%), które mają niezerową frekwencję w korpusie¹³, w tym 116 sekundarnych imperfektywów z prefiksem *z-* (19%) oraz 36 – z prefiksem *po-* (6%). Już te liczby sugerują, że rozkład prefiksów w bazie jest nieprzypadkowy. Porównajmy to z danymi korpusowymi. W ČNK syn7 znajdziemy 8 647 typów wtórnie sufigowanych imperfektywów (100%), z czego 733 (8%) ma prefiks *z-*, a 599 – prefiks *po-* (7%). Liczby te sugerują, że w korpusie prefiksy *z-* i *po-* mają praktycznie równy udział wśród czasowników wtórnie niedokonanych. Rozkład prefiksów porównamy za pomocą testu na proporcje (ang. *proportion test*)¹⁴ (tab. 1 i 2).

Tabela 1

Rozkład czasowników z prefiksem *z-*

Liczba typów czasowników	Prefiksy		Suma prefiksów
	prefiks <i>z-</i>	reszta prefiksów	
Liczba typów czasowników (podstawy) w tabeli czeskich trójek	116	509 (= 625 – 116)	625
Liczba typów czasowników (podstaw) w korpusie	733 (z nich 116 w bazie)	7 914 (= 8 647 – 733)	8 647

¹¹ „Oдноśnie do tego prefiksu nie można mówić o spójnej podstawie z powodu znaczeniowej rozbieżności (przyminka i prefiksu)”.

¹² „Prefiks *po-* rzadko ma znaczenie wyłącznie aspektowe [...]. Czasowniki z tym prefiksem najczęściej wyrażają ruch w jakimś kierunku, pokrywanie powierzchni, naprawianie lub uszkodzenie obiektu, osłabienie intensywności procesu, dystrybucję czynności”.

¹³ W skład badania weszły również czasowniki podwójnie prefigowane, np. *poposkakovat* ‘podskakiwać trochę, kilka razy’, oraz czasowniki z wariantami tego samego sufiksu, np. *pomazovat* i *po-mazávat* ‘posmarować’.

¹⁴ Zob. [online: <https://stat.ethz.ch/R-manual/R-patched/library/stats/html/prop.test.html>; data dostępu: 28.01.2020].

2-sample test for equality of proportions with continuity correction

data: rbind(c(116, 625 - 116), c(733, 8647 - 733))

X-squared = 70.034, df = 1, p-value < e-16

Wnioskujemy zatem, że rozkład prefiksu *z-* w dwóch zbiorach jest nieprzypadkowy, prefiks *z-* preferują czasowniki wtórnie sufigowane wchodzące w skład trójki.

Analiza prefiksu *po-* przynosi inne wyniki.

Tabela 2

Rozkład czasowników z prefiksem *po-*

Liczba typów czasowników	Prefiksy		Suma prefiksów
	prefiks <i>po-</i>	reszta prefiksów	
Liczba typów czasowników w tabeli czeskich trójek	36	589 (= 625-36)	625
Liczba typów czasowników w korpusie	599 (z nich 36 w bazie)	8 048 (= 8 647 - 599)	8 647

Największą różnicę widać między obserwowaną a oczekiwaną liczbą typów czasowników z prefiksem *po-* w bazie czeskich trójek. Rezultat testu na proporcje nie pozwala nam odrzucić hipotezy zerowej, w której zakłada się, że rozkład w bazie i w korpusie jest przypadkowy. Wartość *p* jest dużo wyższa niż 0,05:

2-sample test for equality of proportions with continuity correction

data: rbind(c(36, 625 - 36), c(599, 8647 - 599))

X-squared = 1.0686, df = 1, p-value = 0.3013

Wyniki (wartości *p*) należy jeszcze skorygować, ponieważ oba prefiksy pochodzą z tego samego zbioru. Wynik dla prefiksu *z-* jednak pozostaje na poziomie bardzo sygnifikatywnym ($p < 0.0001$), dla prefiksu *po-* zaś $p > 0.6$ ¹⁵.

Badanie dotyczące typów czasowników, jakie znajdujemy w bazie czeskich trójek i czeskim korpusie *syn7*, potwierdza, że rozkład prefiksu *z-* jest nieprzypadkowy, mianowicie że czasowniki wtórnie sufigowane stanowiące część trójek preferują prefiks *z-*, z kolei dystrybucja prefiksu *po-* w bazie i w korpusie jest przypadkowa. Inaczej mówiąc: wśród czeskich trójek aspektowych udział prefiksu *z-* jest ponadproporcjonalnie wysoki na poziomie sygnifikatywnym, natomiast udział prefiksu *po-* wśród trójek jest podobny do udziału prefiksu *po-* wśród wtórnych imperfektywów w korpusie. Zatem prefiks *z-* podlega szczególnemu „przyciąganiu” wśród trójek, a *po-* na odwrót jest unikany.

Można zadać kolejne pytanie: Czy badanie dotyczące liczby tokenów zmieniłoby wyniki? Zważmy bowiem, że w bazie czeskich trójek dla okresu 1990–2017 łączna frekwencja czasowników wtórnie sufigowanych z prefiksem *z-* wynosi 1 222 690 i jest niewiele wyższa od łącznej frekwencji czasowników wtórnie sufigowanych z prefiksem *po-*, która wynosi 975 868. Może to świadczyć o produktywności prefiksu *z-* (dużo typów, mało tokenów na każdy typ)

¹⁵ *p.adjust* (0.3013, *n* = 2)

oraz leksykalizacji wtórnych imperfektywów z prefiksem *po-* (mało typów, dużo tokenów na typ). Podjęcie próby odpowiedzi na to pytanie jest jednak sprawą dalszych etapów pracy.

4. Podsumowanie

W artykule przedstawiliśmy podwaliny projektu „DiAsPol250” dotyczące badania trójek aspektowych w językach polskim, czeskim i rosyjskim. Dokonałiśmy też przeglądu korpusów i podobnych baz internetowych służących jako podstawy do tych badań. Skonstatowaliśmy, że na przedstawienie aspektu, w szczególności relacji między podstawami prostymi (simplex) a ich derywatami przedrostkowymi ze statusem perfektywów naturalnych, bardzo rzutował wpływ prac strukturalistycznych, których autorzy nie uznawali par aspektowych o składzie (simp^{NDK}-pref^{DK}). Wpływ ten nadal wyraźnie widać w bazach czeszczyzny. Mówiąc zaś ogólniej, w żadnym z istniejących korpusów dotychczas nie ma anotacji, która by pozwalała odnosić do siebie członów par aspektowych o dowolnej budowie (początek tu stanowi polsko-niemiecki korpus tworzony obecnie w ramach „DiAsPol250”). Okoliczności te bardzo utrudniają ustalenie roli i udziału trójek aspektowych w systemie. Ponadto, chcąc zbadać ich dynamikę w diachronii, spotykamy się z bardzo nierównomiernym odbijaniem rzeczywistości językowej w korpusach sprzed 1945 roku.

W ostatnim punkcie artykułu przeanalizowaliśmy rozkład dwóch przedrostków czasownikowych we współczesnym języku czeskim (*z-* i *po-*), które wykazują bardzo różną dystrybucję w ČNK i w naszej bazie trójek. Pokazaliśmy, jak za pomocą prostej analizy statystycznej można ustalić różną wagę tych przedrostków w tworzeniu trójek (*z-* jest „przyciągane”, a *po-* – „odpychane”) i jakie znaczenie taki wniosek może mieć dla oceny roli przedrostków w systemie aspektowym współczesnej czeszczyzny, jeżeli taką analizę przeprowadzi się systematycznie. Jednocześnie to studium przypadku unaocznia, że odpowiednio anotowane korpusy elektroniczne stanowią nie tylko dogodny inwentarz tekstów do ilustracji tez, ale mogą stać się narzędziem do generowania hipotez i ich sprawdzania na podstawie cech dystrybutywnych. A w tym zakresie nie sposób zastąpić ich żadnym innym narzędziem.

Podziękowania

Za pomoc i cenne wskazówki dziękujemy Tilmanowi Bergerowi, Václavowi Cvrčkowi, Janie Kockovej, Natalii Leвшinej, Vladimírowi Petkevičowi. Za uwagi dziękujemy również dwóm anonimowym recenzentom. Oczywiście odpowiedzialność za ewentualne błędy lub niedopatrzania ponosimy tylko my.

Słowniki

ISJP – BAŃKO M., red., 2000: *Inny słownik języka polskiego*. Warszawa.

SJPDor – DOROSZEWSKI W., red., 1958–1969: *Słownik języka polskiego*. Warszawa.

SWJP – DUNAJ B., red., 1996: *Słownik współczesnego języka polskiego*. Warszawa.

WSJP PWN – DUBISZ S., red., 2018: *Wielki słownik języka polskiego PWN*. Warszawa.

Korpusy i źródła internetowe

ČNK – *Český národní korpus* [online: www.korpus.cz; data dostępu: 14.07.2019].

KorBa – *Elektroniczny korpus tekstów polskich z XVII i XVIII wieku (do 1772 r.)* [online: https://korba.edu.pl/query_corpus/; data dostępu: 14.07.2019].

KŘEN i in., 2018: KŘEN M., CVRČEK V., ČAPKA T., ČERMÁKOVÁ A., HNÁTKOVÁ M., CHLUMSKÁ L., JELÍNEK T., KOVÁŘÍKOVÁ D., PETKEVIČ V., PROCHÁZKA P., SKOUMALOVÁ H., ŠKRABAL M., TRUNEČEK P., VONDŘÍČKA P., ZASINA A.: *Korpus SYN (wersja 7 z 29.11.2018)*. Ústav Českého národního korpusu FF UK. Praha [online: <http://www.korpus.cz>; data dostępu: 14.07.2019].

KUČERA K., ŘEHOŘKOVÁ A., STLUKA M., 2015: *DIAKORP: Diachronní korpus, wersja 6 z 18.12.2015*. Ústav Českého národního korpusu FF UK. Praha [online: <http://www.korpus.cz>; data dostępu: 10.05.2019].

Mikrokorpus polszczyzny 1830–1918 [online: <https://szukajwslownikach.uw.edu.pl/f19/>] (Do pobrania też pod <http://www.f19.uw.edu.pl/>; data dostępu: 14.07.2019).

NKJP – *Narodowy Korpus Języka Polskiego* [online: <http://nkjp.pl/>; data dostępu: 10.05.2019].

NKRJa – *Nacional'nyj korpus russkogo jazyka* [online: <http://www.ruscorpora.ru>; data dostępu: 10.05.2019].

Vallex 3.0 – LOPATKOVÁ M., KETTNEROVÁ V., BEJČEK E., VERNEROVÁ A., ŽABOKRTSKÝ Z., 2016: *Valenční slovník českých sloves VALLEX*. Nakladatelství Karolinum. Praha [online: <http://ufal.mff.cuni.cz/vallex/3.0/>; data dostępu: 10.05.2019].

Vokabulář webovůj: webové hnízdo pramenů k poznání historické češtiny [online]. Ústav pro jazyk český AV ČR, v. v. i., oddělení vývoje jazyka. © 2006–2018. Verze dat 1.1.4 [cit. 1.3.2018]. Praha [online: <http://vokabular.ujc.cas.cz>; data dostępu: 10.05.2019].

Literatura

DICKEY S.M., 2000: *Parameters of Slavic Aspect (A Cognitive Approach)*. Stanford, CA.

ESVAN F., 2007: *Vidová morfologie českého slovesa*. Praha.

ISAČENKO A.V., 1960: *Slovesný vid, slovesná akce a obecný charakter slovesného děje*. „Slovo a slovesnost“ XXI, z. 1.

ISAČENKO A.V., 2003 [1960]: *Grammaticeskij stroj russkogo jazyka v sopostavlenii s slovackim. Morfologija část' vtoraja*. Moskva. [Reprint za: Vyd-vo SAV. Bratislava.]

JANDA L., 2007: *Aspectual Clusters of Russian Verbs*. „Studies in Language“ XXXI, z. 3, s. 607–648.

JANDA L.A., LYASHEVSKAYA O., 2011: *Grammatical Profiles and the Interaction of the Lexicon with Aspect, Tense, and Mood in Russian*. „Cognitive Linguistics“ XX, z. 4, s. 719–763.

JANDA i in., 2013: JANDA L., ENDRESEN A., KUZNETSOVA J., LYASHEVSKAYA O., MAKAROVA O., NESSET T., SOKOLOVA S.: *Why Russian Aspectual Prefixes Aren't Empty: Prefixes as Verb Classifiers*. Bloomington.

KOLAŘOVÁ I., 2013: *Tvoření sloves: prefixální derivaty*. V: ŠTÍCHA F., red.: *Akademická gramatika spisovné češtiny*. Praha, s. 241–271.

KOPĚČNÝ F., 1962: *Slovesný vid v češtině*. Praha.

KOSCHMIEDER E., 1934: *Nauka o aspektach czasownika polskiego w zarysie. Próba syntezy*. Wilno.

- KURYŁOWICZ J., 1987a [1929]: *Pochodzenie słowiańskich aspektów czasownikowych*. W: IDEM: *Studia językoznawcze*. Warszawa, s. 434–438.
- KURYŁOWICZ J., 1987b [1968]: *O rozwoju kategorii gramatycznych*. W: IDEM: *Studia językoznawcze*. Warszawa, s. 116–144.
- LEHMANN V., 1999: *Aspekt*. In: JACHNOW H., Hrsg.: *Handbuch der sprachwissenschaftlichen Russistik und ihrer Grenzdisziplinen*. Wiesbaden, s. 214–242.
- ŁAZIŃSKI M., 2011: *Polish Aspectual Prefixes, Their Order and Functions: A Study Based on the National Corpus of Polish*. "Word Structure" IV, issue 2, s. 231–243.
- ŁAZIŃSKI M., 2020: *Pary aspektowe w polskich słownikach*. W: BAŃKO M., RUDNICKA E., DECYK-ZIĘBA W., red.: *Leksykografia w różnych kontekstach*. Warszawa, s. 159–176.
- ŁAZIŃSKI M., WIEMER B., 1995: *Terminatywność jako kategoria stopniowalna*. „Prace Filologiczne” XL, s. 99–126.
- MASLOV Ju.S., 2004 [1948]: *Vid i leksičeskoe značenie glagola v sovremennom russkom literaturnom ŗyke*. V: IDEM: *Izbrannye trudy. Aspektologija. Obščee ŗykoznanie*. Moskva, s. 71–90. [Reprint za: *Izvestija AN SSSR. Serija jazyka i literatury*, t. 7, vyp. 4, s. 303–316].
- MEGER A., WOŹNIAK M., VON WALDENFELS R., 2016: *Jak stworzyć korpus równoległy „dla wszystkich”? O pracy nad Polsko-Niemieckim i Niemiecko-Polskim Korpusem Równoległym*. W: GRUSZCZYŃSKA E., LEŃKO-SZYMAŃSKA A., red.: *Polskojęzyczne korpusy równoległe. Polish-language Parallel Corpora*. Warszawa, s. 97–118.
- MENDE i in., 2011: MENDE J., BORN-RAUCHENECKER J., BRÜGGEMANN N., DIPPONG H., KUKLA J., LEHMANN V.: *Vid i akcional’nost’ russkogo glagola. Opyt slovarâ*. München–Berlin.
- REMČUKOVA E.N., 2008: *Problemy prefiksальной vidovoj pary v sovremennom tolkovom slovare*. V: KUDRÂVCEV J.S., KJU’MOJA I.P., red.: *Humaniora: Lingua Russica (Trudy po russkoj i slavjanskoj filologii. Lingvistika XI*. Tartu, s. 199–217.
- ŚMIECH W., 1986: *Derywacja prefiksalna czasowników polskich*. Wrocław–Warszawa.
- TATEVOSOV S.G., 2016: *Glagol’nye klasy i tipologija akcional’nosti*. Moskva.
- WIEMER B., 2008: *Zur inneroslavischen Variation bei der Aspektwahl und der Gewichtung ihrer Faktoren*. In: GUTSCHMIDT K., JEKUTSCH U., KEMPGEN S., UDOLPH L., Hrsg.: *Deutsche Beiträge zum 14. Internationalen Slavistenkongreß, Ohrid 2008*. München, s. 383–409.
- WIEMER B., 2015: *O roli vida v oblasti kratnosti i pragmatičeskiŗ funkcij (čskiz s točki zrenija krontopii)*. In: BENACCHIO [BENNAK’O] R., red.: *Glagol’nyj vid: grammatičeskoe značenie i kontekst / Verbal aspect: Grammatical meaning and context*. München, s. 585–609.
- WIEMER B., 2017: *O roli pristavok i suffiksov na ranniŗ i pozdniŗ etapax istorii slavjanskogo vida*. In: BENACCHIO R., MURO A., SLAVKOVA S., eds.: *The Role of Prefixes in the Formation of Aspectuality (Issues of Grammaticalization)*. Firenze, s. 219–253.
- WIEMER B., 2019: *O semantičeski invariantnom i grammatičeski trivial’nom v russkom vide*. V: GERASIMOV D.V., DMITRENKO S.Ū., ZAIKA N.M., red.: *Sbornik statej k 85-letiju V.S. Xrakovskogo*. Moskva, s. 43–66.
- WIEMER B., SERŹANT I.A., 2017: *Diachrony and Typology of Slavic Aspect: What Does Morphology Tell Us?* In: BISANG W., MALCHUKOV A., eds.: *Unity and Diversity in Grammaticalization Scenarios*. Berlin, s. 230–307.
- ŹMIGRODZKI i in., 2018: ŹMIGRODZKI P., BAŃKO M., BATKO-TOKARZ B., BOBROWSKI J., CZELAKOWSKA A., GROCHOWSKI M., PRZYBYLSKA R., WANIAKOWA J., WĘGRZYNEK K., red.: *Wielki słownik języka polskiego PAN. Geneza, koncepcja, zasady opracowania*. Kraków.



Projektowanie metadanych w korpusie tekstów polskich do 1500 roku – wielopoziomowa struktura informacji*

Metadata Creation in the Language Corpus of Polish Texts until 1500 – a Multi-level Data Structure

Abstract: The subject of research are selected metadata that should characterize the texts collected in the corpus of the oldest attestations of the Polish language. The author of the article compares and analyses the factors affecting the development of the basic data structure used in synchronic and diachronic corpora (author, title, date of the text, text channel, text classification, source of citation). Without those factors taken into account the disambiguation of the object in the database becomes impossible, and the use of grammatical information is unreliable and impractical. The result of the presented analysis is a proposal to extend the level of description for individual markers.

Key words: language corpus, metadata, text, glosses, 13th–15th century

Abstrakt: Przedmiotem badań są wybrane metadane, które powinny charakteryzować teksty zgromadzone w korpusie najdawniejszych zabytków języka polskiego. Autor artykułu porównuje i analizuje czynniki wpływające na rozbudowywanie podstawowej struktury danych stosowanej w korpusach synchronicznych i diachronicznych (autor, tytuł, datacja tekstu, kanał, klasyfikacja tekstu, źródło cytowania), bez których uwzględnienia ujednoznaczenie obiektu w bazie danych staje się niemożliwe, a wykorzystanie informacji gramatycznych – mało wiarygodne i niepraktyczne. Rezultatem przedstawionych analiz jest propozycja rozszerzenia poziomu opisu poszczególnych znaczników.

Słowa kluczowe: korpus językowy, metadane, tekst, glosy, XIII–XV wiek

Pierwsze prace nad strukturą elektronicznego korpusu języka polskiego rozpoczęto ponad ćwierć wieku temu (PRZEPIÓRKOWSKI i in., 2012: 5). Od tamtego momentu oprócz udoskonalonego już synchronicznego korpusu języka polskiego powstały również korpusy diachroniczne: *Elektroniczny Korpus Tekstów Staropolskich do 1500 roku*, *PolDi – a Polish Diachronic Online Corpus*, *Korpus polszczyzny XVI wieku*, *the Middle Polish Diachrone Lemmatised Corpus (16th–18th c.)*, *Elektroniczny korpus tekstów polskich z XVII i XVIII wieku (do 1772 r.)* (KorBa) oraz *Korpus polszczyzny 1830–1918*. Nie wszystkie z nich można nazwać

* Artykuł powstał w ramach projektu badawczego „Baza leksykalna średniowiecznej polszczyzny (do 1500 roku). Fleksja”, przeznaczonego do realizacji w latach 2018–2023 pod kierownictwem dr hab. Ewy Deptuchowej, prof. IJP PAN w Krakowie, finansowanego ze środków Narodowego Programu Rozwoju Humanistyki (nr projektu: 0201/NPRH/H11/85/2018).

korpusami w znaczeniu, w którym termin ten używany jest obecnie. Dzieje się tak, ponieważ prace nad nimi wciąż trwają. Część z nich złożona jest wyłącznie z nieoznakowanych tekstów w różnej postaci (transliteracji lub transkrypcji) i w różnym formacie (pdf, xml), inne bazy składają się z tekstów otagowanych podstawowymi znacznikami. W końcu są i takie, które spełniają wymogi stawiane współczesnym korpusom językowym. Niezależnie od stanu bieżących prac wszystkie zespoły tworzące korpusy diachroniczne odwołują się bezpośrednio do jednego wzorca – *Narodowego Korpusu Języka Polskiego*. Jest to podyktowane pragmatyzmem i koherencją. Z jednej strony łatwiej jest budować korpus, mogąc wzorować się na istniejącym i sprawdzonym korpusie, z drugiej strony standardy budowania tego typu baz danych są ściśle określone, co pozwoli w przyszłości połączyć mniejsze korpusy w ujednolicony Narodowy Korpus Diachroniczny Polszczyzny (KRÓL i in., 2019: 92–101).

Nawet wstępna analiza struktury metadanych NKJP uwiadcza, że bezpośrednie zastosowanie podobnych rozwiązań w korpusie diachronicznym jest wykluczone. Wynika to głównie z wpływających na tę strukturę historycznych czynników, które nie zawsze są brane pod uwagę podczas opisywania obiektów zamieszczanych w korpusie współczesnym. Porównanie materiału badawczego oraz analiza wspomnianych czynników pozwoli uzasadnić potrzebę modyfikacji gotowej już struktury metadanych, użytej w korpusie synchronicznym, do potrzeb korpusu tekstów polskich do 1500 roku.

Celem niniejszego artykułu jest przedstawienie zróżnicowanej struktury metadanych charakteryzujących teksty pochodzące ze średniowiecznych rękopisów i inkunabułów oraz analiza czynników wpływających na strukturę metadanych. Zanim jednak zostaną omówione szczegółowe zagadnienia związane z porządkowaniem informacji na temat samych tekstów oraz danych źródłowych, przybliżę podstawowe terminy, którymi będę posługiwał się w dalszym ciągu wywodu.

Zgodnie z informacją umieszczoną na stronie *Narodowego Korpusu Języka Polskiego*¹ oraz definicjami słownikowymi (WSJP PAN, SJP PWN) korpusem językowym nazywamy zbiór tekstów uporządkowanych i opracowanych według określonych zasad, który służy do celów naukowych i edukacyjnych. Podstawowymi kryteriami określającymi kompletność korpusu językowego są jego zróżnicowanie, zrównoważenie oraz reprezentatywność. W przypadku korpusu tekstów polskich do 1500 roku, podobnie jak we wszystkich korpusach diachronicznych, kryteria te mogą być spełnione tylko częściowo z uwagi na stan rozpoznania i zachowania zabytków językowych (GRUSZCZYŃSKI, ADAMIEC, OGRODNICZUK, 2013; KOŁODZIEJ, KLAPPER, 2014; KLAPPER, KOŁODZIEJ, 2015). Pierwszy z czynników ma bezpośredni wpływ na przygotowanie metadanych, drugi – na zasób bazy danych.

Kolejnym pojęciem, które wymaga doprecyzowania, jest tekst. Pierwotnie duże zbiory zabytków językowych (WYDRA, RZEPKA, 2004) dzielono na teksty rękopiśmienne i drukowane, ale i ta klasyfikacja miała bardziej złożoną hierarchię, w której wyróżniano zabytki prymarne i sekundarne². Zgodnie ze współczesną nomenklaturą możemy powiedzieć, że dawne zbiory składały się z tekstów ciągłych (np. kazań, listów, modlitw, rot przysięg sądowych, statutów) oraz tekstów nieciągłych (np. glos do tekstów łacińskich, słowników

¹ NKJP: www.nkjp.pl.

² Teksty sekundarne nazywane były najczęściej „tekstami drobnymi”.

dwujęzycznych, fragmentów zdań czy rymowanek polsko-łacińskich). Podczas komponowania optymalnie zróżnicowanego i zrównoważonego korpusu językowego nie można pominąć żadnej z grup tekstów, ponieważ zaburzyłoby to obraz XV-wiecznego języka polskiego, baza nie odzwierciedlałaby częstotliwości występowania słów, liczby połączeń wyrazowych czy wreszcie struktury gramatycznej języka. Zadaniem artykułu nie jest próba stworzenia definicji tekstu ciągłego i nieciągłego, jednak w kontekście przygotowywania metadanych, zawierających informacje o autorstwie, dacie czy tytule podział na teksty ciągłe (na potrzeby tegoż tekstu określimy je jako samodzielne) i nieciągłe (niesamodzielne)³ jest niezbędny, ponieważ liczba danych, którymi powinny być opatrzone tak odmienne obiekty, będzie zróżnicowana.

Tytułowe metadane to ustrukturalizowane informacje, które służą do opisu zebranych zasobów, ułatwiają ich identyfikację i znalezienie oraz pozwalają nimi zarządzać. W każdym korpusie językowym, czy to współczesnym, czy historycznym, ich podstawowym zadaniem jest charakterystyka i klasyfikacja tekstu. Dzięki nim użytkownik jest w stanie dowolnie ograniczać zakres poszukiwań oraz tworzyć własne podkorpusy. Można zaryzykować twierdzenie, że w korpusie diachronicznym, którego zasób jest ograniczony liczbą zachowanych obiektów, wynik badań zależy nie tylko od umieszczonych w bazie tekstów, ale również od wiarygodności i przejrzystości zebranych metadanych. Przykładowo, dzięki uściślonym identyfikatorom czasowym będzie można śledzić ilościowe zmiany form gramatycznych w wybranych przedziałach XV wieku (GRUSZCZYŃSKI, BRONIKOWSKA, 2015), jeśli jednak dane pozostaną niedoprecyzowane, na przykład znacznik czasowy będzie zawierał ogólną informację pochodzącą od wydawcy, że tekst pochodzi z XV wieku, to otrzymany wynik będzie nieprecyzyjny i z tego względu praktycznie nieprzydatny, ponieważ większość tekstów w tym korpusie będzie pochodziła z tego okresu.

Jakie metadane są niezbędne do jednoznacznej identyfikacji najstarszych zabytków języka polskiego? Do czego mogą być przydatne informacje o tekstach łacińskich, z którymi powiązane są teksty polskie? W jaki sposób można weryfikować i uściślać dotychczasowe ustalenia na temat poszczególnych rękopisów i inkunabułów? Odpowiedzi na te pytania są kluczowe w procesie formułowania zapytań korpusowych. Z uwagi na różnorodność staropolskich źródeł, brak szczegółowych badań oraz ograniczone ramy artykułu zadanie to nie może być w pełni wykonane. Mam jednak nadzieję, że podjęta tu próba przeanalizowania wybranych zagadnień związanych z ustalaniem metadanych w korpusie tekstów polskich do 1500 roku pobudzi do dalszej dyskusji na ten temat, a w konsekwencji przyniesie wymierny efekt w postaci odpowiednio opisanego korpusu.

³ Zaproponowany doraźny podział tekstów staropolskich służy tu jedynie do zobrazowania odmiennych typów zabytków językowych, wśród których można wyróżnić m.in.: teksty polskie spójne treściowo i gramatycznie; teksty polsko-łacińskie spójne treściowo i gramatycznie; glosy polskie do tekstów łacińskich spójne treściowo i gramatycznie, będące przekładami fragmentów tekstu łacińskiego, bez którego nie można ich zrozumieć; pojedyncze glosy polskie do tekstów łacińskich, które są niezrozumiałe bez łacińskiego kontekstu; słowniczkii dwujęzyczne; pojedyncze zdania i zapiski na marginesach i okładkach rękopisów niepowiązane z łacińskimi tekstami. Niezależnie od liczby wydzielonych typów tekstów, można stwierdzić, że jedne są zrozumiałe bez podstawy łacińskiej, drugie zaś funkcjonują wyłącznie w kontekście tekstu podstawowego, jakim jest dla nich tekst łaciński.

Podstawowe metadane, którymi opatrywane są teksty korpusowe, to: autor, tytuł, data, kanał (forma przekazu), lokalizacja cytatu (źródło cytowania) oraz klasyfikacja tekstu. Ponadto zasoby opisywane są również innymi identyfikatorami i odsyłaczami, na przykład: wydawca, data publikacji, drukarnia, miejsce wydania, redaktor, tłumacz, wersja cyfrowa tekstu, podobizna źródła w bibliotece cyfrowej. Nie wszystkie metadane są widoczne dla zewnętrznego użytkownika, część znaczników jest istotna wyłącznie dla autorów korpusu, na przykład kto opracował źródło, kto zrobił korektę, kto był konsultantem. Należy pamiętać, że najważniejszym celem ustrukturalizowania udostępnianych informacji jest ujednoznacznianie obiektów oraz ich typologia. Poniżej opisano czynniki wpływające na strukturę metadanych w korpusie języka polskiego do 1500 roku, które nie zawsze są brane pod uwagę w korpusach synchronicznych.

1. Autor

W korpusach językowych to podstawowy znacznik, dzięki któremu pojedynczy tekst można przypisać konkretnemu autorowi lub autorom. W korpusie najstarszych zabytków języka polskiego taka sytuacja jest rzadkością. Znamy utwory Władysława z Gielniowa, na przykład *Jezusa Judasz sprzedał (Żołtarz Jezusow)*, *Anna niewiasta niepłodna (De nativitate Marie ista cancio)*, *Już się anjeli wiesielą*, czy Andrzeja Gałki z Dobczyna *Pieśń o Wiklefie*, ale są to przypadki odosobnione zarówno z uwagi na ówczesne zwyczaje pisarskie, jak i na fakt, że znakomita większość zachowanych utworów to kopie i odpisy niewiadomego pochodzenia. Z tego też względu takie teksty opatrzone będą informacją – „anonim”.

Istnieją również zabytki, które mają autora domniemanego. Za przykład niech posłużą tzw. *Kazania gnieźnieńskie*, których twórcą, według Jerzego Wolnego, mógł być Łukasz z Wielkiego Koźmina (WOLNY, 1961). W tym wypadku imię, nazwisko oraz miejsce pochodzenia autora powinny być opatrzone znakiem zapytania lub innym grafemem sygnalizującym ten fakt.

Co jednak zrobić, gdy w łacińskim zbiorze kazań, przypisywanym na przykład Hieronimowi z Pragi, występują polskie modlitwy *Zdrowaś Maria*, *Wierzę* (rękopis CBPP sygn. Lat.F.ch.I.49) lub wezwanie modlitewne *Moc Boga Ojca* (rękopis CBPP sygn. Lat.F.ch.I.240), których Hieronim nie był autorem? W dotychczasowych rozwiązaniach korpusowych metadane autora polskiego tekstu ograniczyłyby się do opisu „anonim”, a w korpusie nie zostałaby uwzględniona informacja o autorze kazań. Potrzebny jest zatem dodatkowy poziom opisu: „autor tekstu podstawowego” (np. łacińskiego, niemieckiego, czeskiego, w którym umieszczono tekst polski). Tego rodzaju znacznik będzie miał zastosowanie nie tylko w wypadku powszechnie występujących modlitw codziennych, wezwań do osób świętych oraz tekstów katechizmowych wplecionych w teksty łacińskie, ale przede wszystkim w opisie polskich tekstów nieciągłych, które prawie wyłącznie występują pośród zapisów obcojęzycznych. Zarówno dodatkowy znacznik określający autora tekstu podstawowego, jak i dodatkowy znacznik tytułu tekstu podstawowego umożliwią wydzielenie w korpusie podkorpusu, na przykład kazań określonego autora lub kazań na dany dzień roku liturgicznego czy traktatu o *Pater noster*, co pozwoli przeprowadzić badania porównawcze polskich tekstów w identycznych tematycznie podstawowych tekstach łacińskich.

Charakterystyczną cechą piśmiennictwa średniowiecznego było powielanie pojedynczych utworów (pieśni, modlitw, kazań, traktatów, słowników), a także całych rękopisów. Wśród często kopiowanych dzieł teologicznych można wymienić między innymi kolekcję kazań Piotra z Miłosławia, które były bogato glosowane przez kaznodziejów. Pośród kazań należących do tej kolekcji rozpoznane zostały między innymi łacińskie *sermones* autorstwa Łukasza z Wielkiego Koźmina, Jana Sylwana, Jana Szczekny, Jana Milicza, Mateusza z Krakowa (BRACHA, 2007: 65), co sugeruje, że Piotr mógł być kopistą lub kompilatorem zbioru. Proces adaptacji i modyfikacji tekstów popularnych wówczas autorów do indywidualnych potrzeb kaznodziejów był tak powszechny i złożony, że trudno dziś jednoznacznie stwierdzić, czy zachowane w kolekcji Miłosławczyka kazania można jeszcze przypisać wcześniejszym autorom, na przykład Łukaszowi z Wielkiego Koźmina, czy późniejszemu kompilatorowi. Warto się zastanowić, w jakim stopniu przydatne byłoby umieszczenie informacji o pierwotnym autorze łacińskiego tekstu, ewentualnie określenie go jako współ-autora. Taki zabieg pozwala skolacjonować zachowane odpisy pojedynczych kazań oraz zbadać zachowane w nich polskie glosy i przekłady.

W celu ujednoznaczenia kilku obiektów przypisanych nieznanemu autorowi (anonimowi), posiadających ten sam tytuł (np. *Sermones dominicales*), zachowanych w wielu wariantywnych odpisach, należy dodać jeszcze jeden znacznik, którego nie mają istniejące korpusy diachroniczne, a który wydaje się niezbędny w opisie tekstów średniowiecznych – „pisarz”. Warto w tym miejscu zadać pytanie: Czy pisarz mógł być autorem polskich tekstów w zbiorach łacińskich kazań? Pewne jest, że nie każdy. Znaczna część tych tekstów została uwieczniona ręką twórcy manuskryptu, a inne rękami późniejszych właścicieli czy użytkowników. Zdarza się, że w kolofonie skryba pozostawił informację o sobie oraz dniu i roku zakończenia pracy, dzięki czemu możemy z całą pewnością stwierdzić, że wszelkie uwagi sporządziła konkretna osoba, jednak nie wiemy, czy był to kopista, który przepisał polszczyznę z innego egzemplarza, czy był to autor, który samodzielnie wprowadził polskie glosy, komentarze czy teksty ciągłe. Dodatkowa informacja zamieszczona w korpusie pozwoli wyodrębnić zbiór tekstów sygnowanych ręką konkretnego pisarza. Fakt wprowadzenia takiego znacznika nie przesądza o autorstwie, jednak przy braku jakichkolwiek danych znacznik ten umożliwi rozróżnienie kilku odpisów tego samego tekstu oraz pomoże umieścić tekst w określonych ramach czasowych.

Osobnym zagadnieniem, które należy brać pod uwagę, ustalając metadane autora, pozostaje funkcjonowanie w obiegu naukowym wariantywnego nazywania autorów. Przywołany już Hieronim z Pragi znany jest również jako Jan Sylwan, a w katalogach bibliotecznych oraz opracowaniach językowo-literackich i historycznych wymieniany bywa jako Hieronimus (Hieronymus) de Praga, Joannes (Ioannes) Silvanus, a nawet jako Hieronymus Ioannes Silvanus de Praga. Wszystkie możliwe formy jego imienia, nazwiska oraz miejsca pochodzenia powinny trafić do metadanych w znaczniku „autor”. Umożliwi to przypisanie tekstów tylko jednej osobie, a nie czterem czy pięciu.

Podsumowując dotychczasowe rozważania na temat pierwszego znacznika, należy stwierdzić, iż niezbędne jest rozgraniczenie: autora tekstu polskiego, autora tekstu podstawowego (obcego) i pisarza. Znacznik autora tekstu powinien mieć rozszerzoną opcję podawania informacji uwzględniającej kilka form zapisu imienia, nazwiska, przydomka oraz miejsca pochodzenia autora.

2. Tytuł

W ustalaniu tytułu średniowiecznego tekstu należy uwzględnić funkcjonujące w literaturze przedmiotu jego warianty. Podobnie, jak w uprzednio omówionym znaczniku, tak też w tym wypadku można mówić o nadmiarze, który wymaga korpusowego ujednoznacznienia. W zależności od środowiska, w którym powstawały edycja lub opracowanie średniowiecznego tekstu, przyjmowano odmienne sposoby identyfikacji obiektu. Przykładowo, polski tekst statutów zapisany w rękopisie nr 1418 należącym do Biblioteki Książąt Czartoryskich w Krakowie określany był jako: *Kodeks Świętosława z Wojcieszyna*, *Kodeks Macieja z Rożana*, *Kodeks (Mikołaja) Suleda*, *Statuty Kazimierza Wielkiego, Władysława Jagiełły i książąt mazowieckich* czy wreszcie *Statuta Regni Poloniae*. Tytuł katalogowy ustalony przez Bibliotekę to *Kodeks Świętosława z Wojcieszyna*. Należy go wiązać z autorem tłumaczenia początkowego fragmentu kodeksu (k. 2r–41v) zawierającym statuty Kazimierza Wielkiego i Władysława Jagiełły. Tłumaczem kolejnej części (k. 43r–57r), zawierającej statuty książąt mazowieckich, był Maciej z Rożana, stąd dodatkowy wariant tytułu. Zwyczajowe określenie funkcjonujące w środowisku twórców *Słownika staropolskiego* to *Kodeks Suleda*. Tytuł ten występuje również w wielu opracowaniach historycznojęzykowych i nawiązuje do osoby pisarza kodeksu. Historycy stosują najczęściej tytuł łaciński, który zawiera informację o zawartości kodeksu.

Przytoczony przykład nie jest odosobniony, większość średniowiecznych tekstów ciągłych posiada przynajmniej dwa tytuły (najczęściej polski i łaciński), które powinny być odnotowane w korpusie. Podczas ustalania tego znacznika warto zadać dodatkowe pytania: Co z tekstami nieciągłymi? Jak je identyfikować? Czy powinny mieć jakiś umowny tytuł? Czy wiązać je z pojedynczymi tekstami łacińskimi, w których wystąpiły, czy raczej z całym kodeksem?

Odpowiedź na pytanie o tytuł może być tylko jedna. Teksty nieciągłe nie mogą mieć polskiego tytułu. Potrzebny jest im jednak punkt odniesienia do łacińskiej podstawy. Glosy w korpusie nie powinny być pozbawione kontekstu, w którym wystąpiły, ani tytułu dzieła, z którego zostały zaczerpnięte, ponieważ uniemożliwiłoby to ich zrozumienie, a co za tym idzie – prawidłowe oznakowanie w korpusie. Niezbędny jest zatem znacznik „tytułu tekstu podstawowego”, o którym była mowa już wcześniej. Dzięki niemu możliwa będzie również jakakolwiek klasyfikacja tematyczna tekstów nieciągłych zależna od klasyfikacji tematycznej łacińskiego dzieła, w którym zostały zapisane. Teksty nieciągłe w postaci zdań, powiedzeń, fragmentów czy komentarzy rozpoznawane są tradycyjnie za pomocą incipitów, tzn. początkowych słów utworu, co stosowane jest również w tekstach ciągłych, w sytuacji, gdy utworowi nie przypisano zwyczajowego tytułu.

Ostatnie pytanie, dotyczące wydzielenia tekstów nieciągłych w pojedynczych utworach łacińskich oraz rozgraniczenia tytułów tychże utworów (np. *Sermo in prima dominica quadragesima*) od tytułów całych kodeksów (np. *Sermones*), wydaje się czysto akademickie, ponieważ obecny stan rozpoznania i opracowania średniowiecznych manuskryptów nie pozwala na tak ścisłe ustalenia metadanych, a tym samym skazuje twórców korpusu albo na podjęcie samodzielnych badań tekstologicznych, albo na pozostawienie znacznika „tytuł” – nieuzupełnionego⁴.

⁴ Osobnym pytaniem pozostaje, jak potraktować w korpusie zgromadzone w kodeksie teksty, które dotychczas w środowisku filologicznym funkcjonowały jako całość (np. źródła *Słownika staropolskiego*

Podsumowując powyższe rozważania na temat drugiego znacznika, niezbędne jest rozgraniczenie: tytułu tekstu polskiego i tytułu tekstu podstawowego (obcego) oraz uwzględnienie kilku funkcjonujących w literaturze przedmiotu wariantów tytułów polskich i obcych, na przykład *Ojciec nasz*, *Modlitwa Pańska*, *Oratio Dominica*.

3. Datacja tekstu

Dotychczasowa praktyka datowania tekstów polskich opierała się na tradycyjnym pozyskiwaniu informacji z rękopisów, wydań tekstów, opisów katalogowych i bibliotecznych, a w nielicznych tylko wypadkach, gdy brakowało tego typu informacji, na współczesnych badaniach kodykologicznych. Najaktualniejszy stan rozpoznania tekstów staropolskich, zawarty w *Opisie źródeł Słownika staropolskiego* (OżSstp, 2005), nie gwarantuje jednak stworzenia wiarygodnej bazy danych, ponieważ duża część scharakteryzowanych tam zabytków języka polskiego zawiera informacje chronologiczne, choć przypisane do poszczególnych tekstów, to jednak odnoszące się do całych kodeksów, bo pochodzące z opisów katalogowych rękopisów, bez szczegółowego datowania fragmentów zawierających polskie zapisy. Istnieje oczywiście grupa tekstów ciągłych, która od dawna znajdowała się w centrum zainteresowania mediewistów, co sprawiło, że informacje na temat ich datacji są aktualne, ale nie dotyczy to niestety tekstów nieciągłych. Większość z nich została zespolona w sztucznie wyodrębnione przez wydawców całości, na przykład glosy w kazaniach, glosy w rękopisie, i opatrzona datami ogólnymi: *XV in.*, *XV med.*, *XV p. post.*, *XV ex.* lub przybliżonymi, na przykład *ca 1450*, *ca 1500*. Tak szerokie ramy czasowe są niejednoznaczne, a różnice zakresów dat dla poszczególnych tekstów oznaczonych takimi separatorami czasowymi mogą wynosić pięć, dziesięć, a nawet piętnaście lat (w przód i wstecz od wskazanej daty). Wprowadzenie tego typu separatorów do korpusu wiąże się z każdorazową decyzją o wyznaczeniu określonego przedziału czasowego dla pojedynczego tekstu. Z punktu widzenia źródłoznawczego jest to sytuacja idealna, ponieważ powstaje możliwość uporządkowania informacji na temat najstarszych zabytków języka polskiego. Natomiast z punktu widzenia korpusowego praca taka wymagałaby zaangażowania dużego zespołu badawczego oraz wielu lat poszukiwań, co realnie uniemożliwiłoby szybkie opracowanie korpusu. Z uwagi

takie jak: *GLDom*, *GLKazB 1–IV*, *GLLek*, *GLPozn*, *GLWroc*). Czy w związku z faktem, że są to historycznie odrębne teksty (powstałe w różnych latach, ale spisane w jednym czasie), należy je wyodrębnić? W wypadku opisanego powyżej rękopisu Biblioteki Książąt Czartoryskich w Krakowie sprawa jest prosta i rozdzielenie tekstów wydaje się uzasadnione. Istnieją jednak manuskrypty zawierające kilkadziesiąt odrębnych tekstów nieciągłych, które do tej pory funkcjonowały jako całość – glosy w kodeksie. W rękopisie średniowiecznym mogły znajdować się np. traktaty teologiczne, prawnicze i medyczne, wyciągi z autorów średniowiecznych czy komentarze do ksiąg biblijnych. Niejednokrotnie w takim zbiorze tekstów łacińskich zapisywano fragmenty po polsku datowane na różne lata XV wieku, wpisane przez różnych skrybów, a także różniące się tematyką. Podobnie jest z wydaniem manuskryptów zaginionych, zniszczonych i nierozpoznanych (np. źródła *Słownika staropolskiego* takie jak: *PFV*, *R XXIV*, *XXV*, *Zab*). Wydania zawierają jedynie skromne wyciągi polskich tekstów nieciągłych z różnych dzieł łacińskich, których nie jesteśmy w stanie zidentyfikować. Jak w takim wypadku opisywać metadane tekstów nieciągłych?

na ograniczenia techniczne nie można pozostawić w korpusie ogólnych i przybliżonych separatorów czasowych bez ich dookreślenia – czy to w odniesieniu do całości tekstów z określonym separatorem, na przykład *XV in.*, czy to dla każdego tekstu osobno. Decyzja ta wpłynie na strukturę chronologiczną tekstów całego korpusu.

Innym zagadnieniem wymagającym przemyślenia jest sposób traktowania kopii dokumentów dawnych, które były oznaczane podwójnymi datami, na przykład: (1490) 1636, co należy rozumieć, że tekst z 1490 roku zaczerpnięto z kopii sporządzonej w 1636 roku. Czasem wydawca pozostawiał jedną datę, na przykład: (1500), i nie informował, z jakiego okresu pochodził dokument, z którego korzystał. Jak traktować tego typu teksty? Czy powinny trafić do korpusu? Jeżeli przyjmujemy, że z uwagi na językową wartość powinny, to należałoby przyjąć takie rozwiązania techniczne, które umożliwią wydzielenie w korpusie tego rodzaju tekstów.

Podsumowując rozważania na temat trzeciego znacznika „data”, niezbędne jest rozgraniczenie daty powstania tekstu polskiego i daty powstania kodeksu zawierającego tekst polski, ustalenie przejrzystych ram czasowych dla separatorów przybliżonych i ogólnych typu *circa* (np. ± 10 lat) oraz ustalenie sposobu postępowania z kopiami dokumentów średniowiecznych.

4. Kanał i źródło cytowania (lokalizacja cytatu)

W korpusach językowych „kanał” i „źródło cytowania” to dwa znaczniki, które są od siebie zależne. Kanał to, inaczej mówiąc, miejsce zapisania tekstu, które nie zawsze jest tożsame ze źródłem cytowania. Wynika to z faktu, że teksty w korpusie diachronicznym powinny być podawane w dwóch postaciach ortograficznych: transliteracji i transkrypcji.

W korpusie tekstów do 1500 roku możemy wyróżnić trzy podstawowe kanały cytowania tekstów w transliteracji: rękopis, inkunabuł i książkę⁵. Z uwagi na czas powstania zapisów dwa pierwsze kanały są zrozumiałe i nie wymagają objaśnienia, trzeci obowiązuje jednak do skomentowania. Pod pojęciem *książki* kryje się takie współczesne wydanie (XIX- lub XX-wieczne) tekstu dawnego, które nie zawiera informacji o tym, skąd wydawca cytował zabytek staropolski (z rękopisu czy inkunabułu). Sytuacja taka mogła zaistnieć zarówno podczas wydawania różnego rodzaju wypisów z zasobów dawnych bibliotek, jak i w szczegółowych omówieniach konkretnego zagadnienia (np. prawnego), gdzie w przypisach, jako ilustracje materiałowe, cytowano fragmenty średniowiecznych dokumentów bez podawania ich lokalizacji (por. np.: *BiblWarsz*, *RocznHist*, *StPPP*). Podobną praktyką wydawniczą było cytowanie w przypisach artykułów naukowych tekstów porównawczych lub wariantów tekstowych z innych odpisów niż omawiany w artykule, bez podawania szczegółowych informacji na ich temat (por. np.: *DłKlejn*, *SalveReg*).

Teoretycznie źródło cytowania powinno być tożsame z kanałem cytowania, gdy polski tekst podawany jest w transliteracji. Wtedy będzie to rękopis, inkunabuł lub książka. Twórcy korpusu mogą zdecydować jednak inaczej: transliteracja tekstu polskiego zostanie podana za wydaniem elektronicznym, które już zweryfikowano, na przykład *Bibliotekę za-*

⁵ O jeszcze jednym kanale mowa będzie w dalszej części artykułu.

bytków polskiego piśmiennictwa średniowiecznego (płyta DVD) – w takim wypadku należy wydzielić dodatkowy kanał – publikacja elektroniczna.

Źródłem cytowania tekstów podanych w transkrypcji będą najnowsze lub najlepsze (zdaniem twórców korpusu) wydania poszczególnych tekstów. Sytuacja wydaje się prosta i schematyczna, gdy bierzemy pod uwagę teksty ciągłe, które wpisane są w określoną tradycję wydawniczą. Co jednak zrobić z tekstami nieciągłymi, które stanowią większość zasobów staropolskich? Ograniczyć się wyłącznie do tych, które są wydane, czy opracować wszystkie dostępne? Należy pamiętać, że współczesne wydania to niejednokrotnie edycje anachroniczne lub fragmentaryczne, które nie spełniają filologicznych standardów. Zadaniem twórców korpusu będzie zatem opracowanie transkrypcji, a w wielu wypadkach również gruntowne zweryfikowanie dostępnych transliteracji.

Podsumowując spostrzeżenia na temat czwartego i piątego znacznika, warto podkreślić, że korpus diachroniczny powinien wyróżniać cztery kanały: rękopis, inkunabuł, książkę i wydanie elektroniczne, co umożliwi wyseparowanie z bazy danych tekstów pochodzących wyłącznie z jednego kanału. Źródło cytowania zależne będzie od współczesnego stanu edycji tekstów oraz odpowiedniego opracowania transliteracji i transkrypcji poszczególnych tekstów. Może się okazać, że tylko teksty ciągłe będą cytowane w transliteracji i transkrypcji ze znanych wydań, a teksty nieciągłe będą musiały być przygotowane przez zespół korpusowy. W takim wypadku źródłem cytowania będzie opracowanie autorskie lub zespołowe.

5. Klasyfikacja tekstu

Przyjęte w NKJP tematyczna klasyfikacja tekstów i Uniwersalna Klasyfikacja Dziesiętna tekstów zostały zmodyfikowane i dostosowane do potrzeb korpusów dawnych. Znalazło to odzwierciedlenie w KorBie. Zastosowane w nim rozwiązania będą z pewnością wykorzystane w korpusie tekstów polskich do 1500 roku, jednak z pewnymi uzupełnieniami wynikającymi ze specyfiki najdawniejszych zabytków języka polskiego. Przypomnijmy – polskie teksty nieciągłe wymagają uzupełnienia w postaci kontekstu w języku obcym, w którym zostały dopisane. Nie można ustalić klasyfikacji tematycznej polskich głos, ponieważ nie są one tematycznie spójne. Można natomiast ustalić pewne fakty, na przykład że glosy zapisano w łacińskim kazaniu, tłumaczeniu fragmentu biblijnego, traktacie filozoficznym czy bajce. Tego typu informacje pozwalają sklasyfikować tekst podstawowy, a pośrednio – przypisane do niego polskie teksty nieciągłe; można wtedy mówić o klasyfikacji „wtórnej”. Jest to propozycja tymczasowego rozwiązania, ponieważ pozostawienie tekstów nieciągłych jako całościowej grupy niesklasyfikowanej tematycznie, gatunkowo i rodzajowo nie pozwoliłoby na prawidłowe wykorzystanie korpusu do badań językowych. Można dostrzec pewne walory takiego rozstrzygnięcia. „Wtórna” klasyfikacja umożliwi zbadanie słownictwa polskiego występującego wyłącznie w jednym rodzaju tekstów, na przykład w kazaniach lub traktatach filozoficznych czy prawniczych. Tego typu badania nie miały dotąd miejsca z uwagi na duże rozproszenie tekstów oraz fragmentaryczność dostępnych wydań.

Rozważania na temat zastosowania ogólnej lub szczegółowej klasyfikacji tekstów staropolskich wymagają uprzedniego zapoznania się z całym jego zasobem. Rozwiązania przyjęte w korpusach synchronicznych mogą być wykorzystane tylko częściowo, gdy mamy

do czynienia z tekstami jednorodnymi: pieśniami, modlitwami, kazaniami, statutami czy listami, jednak większość zachowanych tekstów dawnych należy do pogranicza tematycznego, rodzajowego i gatunkowego. Przykładowo, pojedyncze obiekty korpusu mogą być przypisane jednocześnie do kilku kręgów tematycznych: prawniczych, sądowych, administracyjnych i gospodarczych (np. wyroki sądów kościelnych) lub religijnych, mitologicznych i filozoficznych (np. kazania i traktaty). Podobne rozstrzygnięcia wielokrotnego wyboru trzeba będzie przyjąć przy znacznikach: „język tekstu” (np. *Psałterz floriański*), „rodzaj tekstu” (podobnie jak w KorBie trzeba będzie wydzielić *Biblię* oraz tłumaczenia fragmentów biblijnych), „gatunek tekstu” (np. notatki akademickie zawierające fragmenty kazań, traktatów oraz poradników).

Podsumowując rozważania na temat klasyfikacji tekstów staropolskich, należy podkreślić złożoną strukturę zachowanych zabytków językowych, która wymusza na twórcach korpusu odrębne podejście do tekstów ciągłych i nieciągłych. W przyjętych dotąd rozwiązaniach nie uwzględniono klasyfikacji glos, komentarzy, notatek dwujęzycznych czy fragmentów zdań zapisanych na okładkach manuskryptów. Dostępne opracowania filologiczne oraz bibliologiczne nie pozwalają też jednoznacznie określić rodzaju i gatunku średniowiecznych zabytków. Sklasyfikowanie dużej części tekstów będzie wymagało wprowadzenia pól wielokrotnego wyboru i uznania, że nie są to obiekty jednorodne.

Stworzenie korpusu najdawniejszych zabytków języka polskiego może wydawać się proste, jeżeli bierzemy pod uwagę zasób tekstów znanych z edukacji szkolnej lub akademickiej oraz istniejące narzędzia korpusowe umożliwiające opisanie i anotowanie tekstów. Dla historyków języka, literaturoznawców, historyków czy źródłoznawców zadanie to wydaje się o wiele bardziej skomplikowane, ponieważ nie istnieje aktualny spis wszystkich zabytków języka polskiego do 1500 roku (*Opis źródeł Słownika staropolskiego* zawiera spis wydań dużej części rozpoznanych zabytków, ale nie identyfikuje wszystkich tekstów w rozumieniu korpusowym), nie posiadamy elementarnych informacji (data, autor, tytuł) na temat zachowanych (w różnej postaci) tekstów, a rozproszone w wielu miejscach i w różnej formie transliteracje i transkrypcje tekstów nie spełniają wymogów edytorskich pozwalających na zamieszczenie ich w bazie danych. W tej sytuacji jedynym wyjściem jest stopniowe tworzenie korpusu polskich tekstów do 1500 roku oraz uzupełnianie i weryfikowanie dostępnych już danych na podstawie bezpośrednich badań kodykologicznych, a także najnowszych badań źródłoznawczych i bibliologicznych. Tylko wtedy będzie można przedstawić zbiór uporządkowanych i opracowanych według określonych zasad tekstów, który posłuży do celów naukowych i edukacyjnych.

Źródła

KorBa – *Elektroniczny korpus tekstów polskich z XVII i XVIII w. (do 1772 r.)* [online: https://korba.edu.pl/query_corpus/; data dostępu: 7.10.2019].

Korpus polszczyzny 1830–1918 [online: <http://www.f19.uw.edu.pl/>; data dostępu: 7.10.2019].

Korpus polszczyzny XVI wieku [online: <http://spkvi.edu.pl/korpus/>; data dostępu: 7.10.2019].

Korpus tekstów staropolskich [online: <https://ijp.pan.pl/publikacje-i-materialy/zasoby/korpus-tekstow-staropolskich/>; data dostępu: 7.10.2019].

NKJP – *Narodowy Korpus Języka Polskiego* [online: <http://nkjp.pl/>; data dostępu: 7.10.2019].

OżSstp 2005 – TWARDZIK W., red. we współpracy z E. DEPTUCHOWĄ i L. SZELACHOWSKĄ-WINIARZOWĄ.

Oprac. BELCARZOWA E., DEPTUCHOWA E., FRODYMA M., KALICKA K., LEŃCZUK M., SZELACHOWSKA-WINIARZOWA L., WÓJCIKOWA Z., 2005: *Opis źródeł Słownika staropolskiego*. Kraków.

Słowniki

SJP PWN – *Słownik języka polskiego PWN*, 2019 [online: <https://sjp.pwn.pl/>; data dostępu: 7.10.2019].

WSJP PAN – ŻMIGRODZKI P., red., 2008: *Wielki słownik języka polskiego PAN* [online: <https://wsjp.pl/index.php?pwk=0>; data dostępu: 7.10.2019].

Literatura

BRACHA K., 2007: *Nauczanie kaznodziejskie w Polsce późnego średniowiecza. Sermones dominicales et festivales z tzw. kolekcji Piotra z Miłostawia*. Kielce.

GRUSZCZYŃSKI W., ADAMIEC D., OGRODNICZUK M., 2013: *Elektroniczny korpus tekstów polskich z XVII i XVIII wieku (do 1772 r.) – prezentacja projektu badawczego*. „Polonica” XXXIII, s. 309–316.

GRUSZCZYŃSKI W., BRONIKOWSKA R., 2018: *Tworzenie korpusu tekstów dawnych a korpusu tekstów współczesnych: różnice teoretyczne i warsztatowe (na przykładzie Korpusu tekstów polskich XVII–XVIII wieku)* [online: https://korba.edu.pl/static/documents/publikacje/2015_gruszczyński_bronikowska.pdf; data dostępu: 7.10.2019].

KLAPPER M., KOŁODZIEJ D., 2015: *Elektroniczny Tezaurus Rozproszonego Słownictwa Staropolskiego do 1500 roku. Perspektywy i problemy*. „Polonica” XXXV, s. 87–101.

KOŁODZIEJ D., KLAPPER M., 2014: *Elektroniczny Korpus Tekstów Staropolskich do 1500 r. Perspektywy i problemy*. „Prace Filologiczne” LXXV, s. 203–212.


KRÓL i in., 2019: KRÓL M., DERWOJEDOWA M., GÓRSKI R.L., GRUSZCZYŃSKI W., OPALIŃSKI K.W., POTONIEC P., WOLIŃSKI M., KIERAŚ W., EDER M.: *Narodowy Korpus Diachroniczny Polszczyzny. Projekt*. „Język Polski” XCIX, z. 1, s. 92–101.

PRZEPIÓRKOWSKI i in., 2012: PRZEPIÓRKOWSKI A., BAŃKO M., GÓRSKI R.L., LEWANDOWSKA-TOMASZCZYK B., red.: *Narodowy Korpus Języka Polskiego*. Warszawa.

WOLNY J., 1961: *Łaciński zbiór kazań Peregryna z Opola i ich związki z tzw. „Kazaniami gnieźnieńskimi”*. W: LEWAŃSKI J., red.: *Średniowiecze. Studia o kulturze*. T. 1. Warszawa, s. 171–238.

WYDRA W., RZEPKA W.R., 2004: *Chrestomatia staropolska. Teksty do roku 1543*. Wyd. III. Wrocław.





Anna Majewska-Tworek
Uniwersytet Wrocławski
 <https://orcid.org/0000-0003-3997-5991>
anna.majewska-tworek@uwr.edu.pl

Forum Lingwistyczne
nr 7, 2020
ISSN 2450-2758 (wersja elektroniczna)



DOI: <http://doi.org/10.31261/FL.2020.07.06>

Monika Zaśko-Zielińska
Uniwersytet Wrocławski
 <https://orcid.org/0000-0001-6333-660X>
monika.zasko-zielinska@uwr.edu.pl

Piotr Pęzik
Uniwersytet Łódzki
 <https://orcid.org/0000-0003-0019-5840>
piotr.pezik@uni.lodz.pl

„Polszczyzna mówiona miast” – kontynuacja badań z lat 80. XX wieku z wykorzystaniem narzędzi lingwistyki cyfrowej

“Spoken Polish of cities” – a Continuation of Research from the 1980s
with the Use of Digital Linguistics Toolss

Abstract: The subject of the research are old texts recorded as part of the “Spoken Polish of cities” project, which was carried out in Katowice, Kraków, Łódź and Wrocław in the last decades of the 20th century. The main aim of the article is to present a new study of Wrocław oral data from that period. The tools of digital humanities and corpus linguistics were used in the new study. The result of the team’s work is the *Speech of Wrocław* collection in Spokes resources.

Key words: spoken language, sociolinguistics, corpus, Spokes

Abstrakt: Przedmiotem badań są stare teksty nagrane w ramach projektu „Polszczyzna mówiona miast”, który był realizowany w Katowicach, Krakowie, Łodzi i we Wrocławiu w ostatnich dekadach XX wieku. Głównym celem artykułu jest przedstawienie nowego opracowania wrocławskich danych mówionych pochodzących z tamtego okresu. Do nowego opracowania zastosowano narzędzia humanistyki cyfrowej oraz językoznawstwa korpusowego. Rezultatem pracy zespołu jest zbiór *Mowa Wrocławia* w zasobach Spokes.

Słowa kluczowe: mówiona odmiana języka, socjolingwistyka, korpus, Spokes

1. Wstęp

Głównym powodem powstania niniejszego artykułu jest odnalezienie w Instytucie Filologii Polskiej Uniwersytetu Wrocławskiego kaset z nagraniami, które były gromadzone na potrzeby dawnego projektu, bardzo znanego w całym środowisku polonistycznym, poświęconego badaniom polszczyzny mówionej miast w latach 70., 80. i 90. XX wieku. Informacja o znalezionym materiale spotkała się z zainteresowaniem pracowników polskiego zespołu CLARIN, dzięki czemu materiał został ponownie opracowany.

Nikt jednak nie szukałby kaset ani na nowo ich nie transkrybował, gdyby nie swoisty renesans zainteresowań polszczyzną mówioną w różnych nurtach współczesnego językoznawstwa teoretycznego i stosowanego, choć znawcy przedmiotu zwracają uwagę na brak odpowiednio przygotowanych, wystarczająco obszernych materiałów z niezbędnymi metadanymi. Istnieje zatem wielka potrzeba wzbogacenia istniejących korpusów wypowiedzi mówionych oraz tworzenia nowych, by można było z wykorzystaniem nowoczesnych narzędzi badawczych rzeczywiście zintensyfikować dociekanie istoty mówionej odmiany języka w różnych jej przejawach. Nie sposób tego robić bez odwołania się do początkowych ustaleń badaczy z Katowic, Krakowa, Łodzi i Wrocławia, ich doświadczeń oraz osiągnięć.

2. Polszczyzna mówiona miast w XX wieku – początki projektów

Badania nad polszczyzną mówioną mieszkańców Katowic prowadzone przez badaczy z Uniwersytetu Śląskiego w Katowicach pod kierunkiem Władysława Lubasia rozpoczęły się już w pierwszej połowie lat 70. Później objęły cały obszar ówczesnego województwa katowickiego. W czerwcu 1974 roku w Sosnowcu z inicjatywy pracowników Instytutu Filologii Polskiej Uniwersytetu Śląskiego w Katowicach odbyła się ogólnopolska konferencja naukowa poświęcona potocznej polszczyźnie mówionej miast. W obradach uczestniczyli przedstawiciele większości polonistyk polskich. Rezultatem tego spotkania było wydanie w 1976 roku publikacji zatytułowanej *Miejska polszczyzna mówiona. Metodologia badań*. We wstępie do tomu Lubaś podkreślił konieczność zespołowej pracy nad badaniem polszczyzny mówionej oraz znaczenie wspólnej metodologii wynikającej z polskich i zagranicznych doświadczeń. Zaznaczył także, że opublikowany zbiór tekstów ma szansę zainicjować dalszy rozwój zainteresowań naukowych polszczyzną mówioną (LUBAŚ, 1976: 6–8), której wcześniejsze badania zostały w tomie zrelacjonowane w artykułach Bronisława Wieczorkiewicza, Mieczysława Karasia i Alfreda Zaręby (LUBAŚ, 1976: 9–40).

Pod koniec 1976 roku z inicjatywy Karasia został zainaugurowany projekt krakowski, który wyróżniał się uwzględnieniem wielu pozajęzykowych zmiennych już na etapie jego planowania (DUNA, 1981: 7). Również w 1976 roku rozpoczęły się prowadzone pod kierunkiem Marii Kamińskiej badania polszczyzny mówionej mieszkańców Łodzi. Tutaj postanowiono, że pierwszym zadaniem będzie gromadzenie materiału, a dopiero później na tej podstawie zostaną sformułowane pewne ustalenia teoretyczne (KAMIŃSKA, 1979: 87). Dopiero w latach 80. do badania polszczyzny mówionej przystąpili wrocławianie, chociaż wcześniejsze prace Antoniego Furdala nad dialektologią miejską (FURDAL, 1966) oraz kwestionariusz do badania dialektów kulturalnych i miejskich (FURDAL, 1973) często stawały się punktem wyjścia do rozważań metodologicznych prowadzonych w innych ośrodkach naukowych.

3. Szczegółowe założenia wstępne. Obszar badań i dobór populacji

Mimo że w kontekście języka miasta problem obszaru badań mógłby się wydawać oczywisty, to zarówno na etapie konstruowania planów zbierania materiału, jak i w refleksjach

późniejszych okazało się, że jest to zagadnienie wymagające dyskusji i doprecyzowania. W opublikowanej w 2000 roku pracy poświęconej typologii współczesnej polszczyzny Aleksander Wilkoń wprowadził pojęcie miejskich języków mieszanych, które na podstawie istniejących już wtedy badań nad polszczyzną mówioną miast podzielił na kilka typów (WILKOŃ, 2000: 28–30). Przykładem jednego z tych typów były miejskie dialekty występujące w miastach górnośląskich (np. Katowicach), w których język ogólny i gwara istnieją obok siebie, tworząc charakterystyczny dwujęzyczny obszar językowy. Natomiast Kraków reprezentuje typ organizmu miejskiego, w którym dominuje język ogólny, ale na obrzeżach miasta (dawne wsie podkrakowskie) język zawiera wiele cech regionalnych. Jeszcze inny typ mogłaby reprezentować Łódź, która, podobnie jak opisany przez Wilkońa Szczecin, przedstawia obszar pewnej integracji językowej.

Hipoteza o jednorodności językowej przestrzeni miasta czasem skutkowałą wnioskiem, że podczas zbierania materiału stosowanie podziału terytorialnego w obrębie organizmu miejskiego jest zbędne. Takie podejście charakterystyczne było dla badań nad językiem miejskim Katowic. Badaniami objęto ówczesne terytorium Katowic bez podziałów przestrzennych (np. na centrum i przedmieścia), gdyż założono, że cały ten obszar łączą więzi społeczne. Jednak oprócz materiałów pochodzących z Katowic w ramach projektu zebrano także teksty z kilkunastu miast śląskich i zagłębiowskich. Wśród nich znalazły się: Gliwice, Bytom, Siemianowice Śląskie, Mikołów, Mysłowice, Sosnowiec, Czeladź, Będzin, Dąbrowa Górnicza, Siewierz, Chorzów, Ruda Śląska, Zawiercie, Rybnik, Świętochłowice, Tarnowskie Góry, Tychy, Wodzisław Śląski, Zabrze (LUBAŚ, 1978; 1980).

Na terytorium Krakowa badania były prowadzone w granicach sześciu dzielnic: Grzegórzki, Kleparz, Nowa Huta, Podgórze, Stare Miasto, Zwierzyniec. Uwzględniono zatem podział miasta obowiązujący w latach 1952–1973. Wiadomo było bowiem, że każda z dzielnic ma swoją odrębną strukturę socjologiczną, która może mieć wpływ na cechy językowe jej mieszkańców.

Z kolei w ośrodku łódzkim badania prowadzono w granicach Wielkiej Łodzi, ale z dbałością o równowagę danych podzielono miasto na pięć części (północną, wschodnią, centralną, zachodnią, południową) z dokładnym wyznaczeniem dzielnic, które wchodziły w ich skład (KAMIŃSKA, 1979: 89). Jeszcze przed przystąpieniem do badań przypuszczano, że raczej nie istnieje żadna odrębna łódzka gwara miejska. Zakładano również, że z powodu centralnego położenia miasta na mapie Polski oraz sąsiedztwa gwar łęczycko-sieradzkich, które nie wyróżniają się wyjątkowymi odrębnościami, w mieście „winna się wykształcić nadgwarowa odmianka języka mówionego – swoisty interdialekt” (KAMIŃSKA, 1979: 87).

We Wrocławiu postanowiono prowadzić badania w granicach administracyjnych miasta, zwracając jednocześnie uwagę na dzielnice o charakterze wiejskim i półwiejskim (słabo zintegrowane z centrum organizmu miejskiego) oraz intensywne uczestniczenie w życiu miasta osób dojeżdżających z niedalekich miasteczek, na przykład z Oławy, Trzebnicy, Kątów Wrocławskich, Obornik Śląskich. Podkreślano, że miasta te nigdy nie były prowincją względem wrocławskiej aglomeracji, stąd nie tak wielkie różnice socjolingwistyczne występowały między ich mieszkańcami jak w przypadku aglomeracji krakowskiej czy łódzkiej. Ponadto Franciszek NIECKULA pisał: „Nie ma [...] żadnych powodów, by w jakimkolwiek stopniu liczyć się z podziałem aglomeracji na 5 dzielnic: Stare Miasto, Śródmieście, Krzyki, Fabryczną i Psie Pole. Podział ten nie odpowiada żadnym realiom historycznym, architek-

tonicznym, komunikacyjnym czy społecznym” (1990: 24)¹. Po wojnie Wrocław wyróżniał się przede wszystkim wielojęzyczną mozaiką. Ogromne zniszczenie miasta (Wrocław w wyniku działań wojennych był zburzony w 70%) powodowała „ciągłą migrację wewnątrz organizmu miejskiego [...]. Wszystko to znakomicie sprzyjało integracji i utrudniało wyłanianie się wyraźnych nowych zróżnicowań socjolingwistycznych” (NIECKULA, 1990: 24).

Drugim ważnym problemem w badaniach polszczyzny mówionej miast był dobór populacji i ustalenie reprezentatywnej próby badawczej. W odniesieniu do Katowic – ze względu na to, że tu za najważniejsze kryterium uznano terytorium – ustalono, iż badaną populację będą tworzyli zarówno rodowici katowiczanie, osoby zamieszkałe w Katowicach, jak i ci, którzy mieszkają poza miastem, ale pracują w Katowicach przynajmniej od roku.

W Krakowie za konieczne uznano odróżnienie osób urodzonych w badanym mieście (wśród tych osób skrupulatnie odnotowywano miejsce urodzenia rodziców, sprawdzając tym samym stopień zasiedlenia rodziny w mieście) od przybyszów (za ważną granicę czasową świadczącą o zrośnięciu się ze środowiskiem krakowskim uznano okres 10 lat), a także ludności przepływowej, tzn. mieszkających w mieście okresowo (studenci) albo też dojeżdżających do szkół czy zakładów pracy.

W Łodzi ze względu na nieustanny napływ nowej ludności postanowiono, że informatorami staną się osoby, które mieszkają w mieście przynajmniej 12–15 lat, gdyż dopiero tak długi pobyt pozwala na asymilację ze środowiskiem miejskim (KAMIŃSKA, 1979: 89).

Prawie całkowita wymiana ludności w powojennym Wrocławiu spowodowała, że w doborze populacji trzeba było uwzględnić pochodzenie przybyszów oraz czas zamieszkania w mieście. W badaniach wrocławskich planowano skupić się na trzech grupach wiekowych. Pierwszą miała stanowić najstarsza generacja, czyli osoby ponadsześćdziesięcioletnie (w 1945 r. miały 20 lat i więcej); drugą grupę – osoby urodzone i wychowane we Wrocławiu, czyli wrocławianie trzydziestoletni i czterdziestoletni; trzecią grupę – dzieci w wieku szkoły podstawowej (7–15 lat). Liczono więc na przebadanie trzech generacji, dzięki czemu udałoby się zaobserwować pewne zmiany językowe w mowie najmłodszych: „[...] w ich mowie mogą się ujawniać, choć nie muszą, cechy językowe najstarszej generacji (dziadkowie, babcie), właściwości drugiej grupy (rodzice) oraz zapowiedź tej polszczyzny, która za jakiś czas będzie dominowała we Wrocławiu jako ostateczny rezultat procesów integracyjnych” (NIECKULA, 1990: 31).

Kolejnym ważnym zagadnieniem w opisywanych projektach był dobór badanych. Tu znów poszczególne ośrodki nieco różniły się w szczegółowych ustaleniach. Spowodowane to było możliwościami organizacyjnymi czy wielkością zespołu, a także specyfiką miejsca. Cechę wspólną stanowiła rezygnacja z doboru losowego mówców jako mało efektywnego zarówno organizacyjnie, jak i merytorycznie.

W badaniu mieszkańców miast śląskich początkowo w doborze populacji uwzględniano przede wszystkim czynniki socjalne: zawód, wykształcenie, etniczność regionalną, wiek, płeć oraz bilingwizm. Później zwrócono uwagę, że ten zespół czynników ma różną istotność w poszczególnych grupach społecznych. Dlatego też rozważano zastąpienie kryterium socjalnego „typem kultury”, na który składałyby się: uwarunkowana historycznie etniczność,

¹ Chociaż powszechnie wiadomo, że były dzielnice szczególnie często i chętnie zamieszkiwane przez inteligencję, np. Biskupin.

ograniczany pozycją zawodową zakres kontaktów społecznych, właściwości osobowe powiązane z predyspozycjami psychicznymi, wiek, płeć oraz wykształcenie (LUBAŚ, 1976: 44).

W Krakowie przed podjęciem badań lingwistycznych wytypowano we współpracy z socjologami czynniki socjalne i biologiczne, które mogą mieć wpływ na różnice językowe w środowisku miejskim. Były to: wiek, płeć, miejsce urodzenia, pochodzenie społeczne, wykształcenie, zawód oraz charakter wykonywanej pracy (DUNAJ i in., 1979: 80–81; por. też DUNAJ, 1987: 36). Dzięki przystąpieniu do próbných analiz fonetycznych, leksykalnych i składniowych (czyli do „wstępnej konfrontacji z materiałem” – DUNAJ, 1981: 9) sprawdzono, że opracowana siatka zróżnicowań socjalnych Krakowa ma charakter maksymalny. Wywnioskowano też, że do badań zagadnień szczegółowych wystarczy uwzględnić tylko niektóre czynniki.

W Łodzi do badań mogły być wytypowane osoby, które mieszkają w mieście dłużej niż 10 lat. Postanowiono nie brać pod uwagę repatriantów zza Buga ani obcokrajowców. Postanowiono też skupić się na polszczyźnie łódzian zatrudnionych w przemyśle włókienniczym: „[...] łódzkie tkaczki i prządki oraz robotnicy zawodów pokrewnych mogą być uznani za najbardziej reprezentatywnych przedstawicieli łódzkiej społeczności” (KAMIŃSKA, 1979: 88). Natomiast badania łódzkiej inteligencji odsunięto na dalszy plan. Takie ograniczenie wynikało z chęci urealnienia całego projektu w rozsądnym przedziale czasowym. Uznano bowiem za niemożliwe przebadanie całego miasta w kilkuletnim projekcie badawczym.

W badaniach wrocławskich Nieckula wymienił kilkanaście zmiennych, które miały być uwzględniane w opisie zgromadzonych nagrań. Wydaje się, że nie stanowiły one listy zamkniętej. Przypuszczano, że nie wszystkie muszą być istotne podczas badania nagranych wypowiedzi. Dotyczyły one charakterystyki socjolingwistycznej mówców: wiek partnerów, płeć, wykształcenie, zakład pracy i zajmowane stanowisko, pochodzenie społeczne, zamożność, rodzaj zainteresowań, aktywność społeczna (działalność w stowarzyszeniach); charakterystyki nagranej wypowiedzi: tematyka wypowiedzi, forma tekstu (rozmowa, opowiadanie itd.), rodzaj kontaktu językowego (relacje społeczne między partnerami, oficjalność lub nieoficjalność), sytuacja aktu komunikacji (miejsce, czas itp.); charakterystyki psycholingwistycznej nagrywanych: łatwość werbalizacji myśli, tempo mówienia, dykcja; a także „czynników typowo wrocławskich”: czas pobytu w mieście, pochodzenie regionalne (i ewentualnie społeczne) rodziny, utrzymywanie (lub niepodtrzymywanie) więzi z miejscem pochodzenia rodziny (NIECKULA, 1990: 21).

4. Cele i metody badawcze

Celem projektu w Katowicach był opis języka miasta uwzględniający wszystkie jego składniki oraz wariantywność uwarunkowaną społecznie i kulturowo. Uznano, że koncentrowanie się na dialektyzmach lub języku ogólnym byłoby eksponowaniem jednego elementu całego mieszanego systemu, jakim jest język miejski. Spośród trzech typów kontaktów językowych charakterystycznych dla mieszkańców miast postanowiono skoncentrować się głównie na kontaktach lokalnych oraz indywidualnych zachodzących w małych grupach w sytuacjach rodzinnych i przyjacielskich. W tego typu relacjach ujawnia się bowiem najbardziej wariantywność języka.

W Krakowie, formułując metody prowadzenia badań miejskiej polszczyzny mówionej, często odwoływano się do doświadczeń badań dialektologicznych. Jednak celem dialektologów – jak słusznie podkreślano – było zazwyczaj dotarcie do najstarszej formy mówionej języka na danym terenie. Z tego powodu badano przedstawicieli (nawet pojedynczych) najstarszego pokolenia. Z kolei zasadniczy cel badań tzw. dialektologii miejskiej miał być inny: „[...] idzie w nich bowiem o pokazanie zróżnicowania językowego jakiegoś środowiska, uwarunkowanego rozmaitymi czynnikami społecznymi i sytuacyjnymi. Z tego względu wybór reprezentatywnych informatorów staje się przedsięwzięciem znacznie bardziej skomplikowanym” (DUNAJ, 1987: 30). Dużym wyzwaniem badawczym była więc nie tylko decyzja o doborze próby badanych, ale także określenie techniki zbierania materiału oraz wyznaczenie sytuacji, w których to się będzie odbywało (DUNAJ, 1987: 30). W badaniach nad polszczyzną mówioną Krakowa zastosowano metodę celowego doboru reprezentacji². Polega ona na celowym doborze próby „ze względu na określone cechy socjalne i liczbę osób potrzebnych do danej grupy” (DUNAJ, 1987: 33). W przeciwieństwie do doboru losowego umożliwia uzyskiwanie zrównoważonych danych (proporcjonalnych liczbowo) od poszczególnych, wcześniej starannie wyznaczonych grup socjalnie jednorodnych. W tego typu badaniach kolejnym dużym wyzwaniem jest określenie liczebności wyznaczonej grupy. Dunaj, powołując się na badania Haliny Kurek (1981), podał liczbę grupy minimalnej: „[...] odpowiednie obliczenia wykorzystujące rachunek prawdopodobieństwa dowodzą, że w badaniach dużej populacji grupy jednorodne socjalnie winny liczyć powyżej 10 osób” (DUNAJ, 1987: 34).

W Łodzi głównym celem badań był opis współczesnej polszczyzny łódzkiej, w tym przeprowadzenie analiz porównawczych między pokoleniami w obrębie wybranych rodzin robotniczych: „[...] opis polszczyzny łódzkiej winien być dokonany zasadniczo w płaszczyźnie synchronicznej, nie rezygnujemy jednak także w miarę możliwości z badań sięgających w przeszłość” (KAMIŃSKA, 1979: 88). Zauważono bowiem, że w mowie najstarszego pokolenia są jeszcze cechy polszczyzny dziewiętnastowiecznej, które koniecznie chciano odnotować. Pierwszym celem było przygotowanie zbioru tekstów mówionych mieszkańców Łodzi. KAMIŃSKA (1979: 89) nazwała tak sformułowane cele „minimalistycznymi”. Miały być one podstawą kolejnych badań. Zastosowano celowy dobór badanych: „[...] podobnie jak to się zwykło czynić przy badaniach dialektologicznych stosujemy staranny dobór informatorów, nie są oni typowani mechanicznie” (KAMIŃSKA, 1979: 90). Poszukiwano mówców chętnych do współpracy, o znanym miejscu urodzenia, miejscu zamieszkania i znanym zawodzie. Do zbioru danych nie włączano form zasłyszanych. Dość ściśle określono liczbę badanych. Z każdej części miasta (z 5 wcześniej wyznaczonych) postanowiono uzyskać co najmniej 100 tekstów. „W sumie z całego miasta byłoby to 500 dłuższych wypowiedzi, które nie wyznaczają liczby przebadanych informatorów. [...] Liczba informatorów winna sięgnąć 800 osób, co dawałoby w rezultacie 1 na 1000 mieszkańców miasta” (KAMIŃSKA, 1979: 89).

We Wrocławiu już na wstępie badań odrzucono hipotezę o istnieniu dialektu wrocławskiego (wrocławskiej odmiany języka polskiego). Zasadniczym celem było opisanie stanu współczesnej polszczyzny w tym mieście oraz poznanie zależności zjawisk językowych od zjawisk socjalnych i kulturowych (najlepiej także na tle innych dużych miast). Podobnie

² Opis losowego i celowego doboru próby badanych przedstawił DUNAJ (1987: 32).

jak w innych ośrodkach – i tu szczególnym zainteresowaniem badawczym cieszyła się nieoficjalna polszczyzna mówiona. Natomiast celem wyróżniającym wrocławskie badania miało być śledzenie procesów integracji językowej: „[...] w wyniku tak strasznego zdarzenia historycznego, jakim była II wojna światowa i związane z nią migracje ludności na skalę niespotykaną w dziejach, doszło do wytworzenia się zupełnie nowych społeczności wielkomiejskich z elementów przynależnych do wszelkich możliwych regionów i grup społecznych. [...] Opis procesów integracyjnych i zebranie dla nich bogatej dokumentacji tekstowej wypada uznać za podstawowy, historyczny obowiązek polskiego językoznawstwa” (NIECKULA, 1990: 23). Nieckula był przekonany o wielkim znaczeniu tego celu badawczego w wymiarze zarówno czysto poznawczym, jak i praktycznym.

5. Technika zbierania materiału i przygotowanie materiału do badań

W badaniach prowadzonych przez zespół Lubasia za jeden z istotniejszych czynników decydujących o jakości uznano naturalność mówienia. Jej gwarantem miały stać się rozmowy nagrywane z podsłuchu, gdyż jawność nagrania i obecność naukowca jako jednego z interlokutorów niweczy naturalność kontaktu. Ostatecznie zebrany materiał stanowią głównie rozmowy swobodne, ale tylko nieliczne z nich opatrzone informacją o nagrywaniu z podsłuchu, zazwyczaj też uczestniczyli w nich studenci polonistyki lub asystenci Zakładu Dialektologii IJP UŚ. Wydaje się, że z nagraniem z podsłuchu wiążą się nie tylko dostrzegane problemy prawne, ale także organizacyjne. W wyniku projektu prowadzonego w tymże Zakładzie przygotowano dwutomowy (trzyczęściowy) wybór najbardziej typowych tekstów, obejmujący 67 transkryptów pogrupowanych według miast, w których dokonywano nagrania. Teksty były spisywane (transkrypcja półfonetyczna) przez autorów nagrań, następnie przesłuchiwane i sprawdzane przez Wilkonia, po czym jeszcze raz przez Lubasia, który we wstępie do drugiego tomu wyboru tekstów przyznał, że największą trudnością było ustalanie długości pauz (LUBAŚ, 1980: 7).

W Krakowie początkowo – wzorem badań dialektologicznych – planowano wykorzystywanie kwestionariuszy fonetycznych i leksykalnych (DUNAJ, 1981: 7). Po okresie pierwszych badań zrezygnowano z ich wykorzystywania na rzecz wywiadów lub nagrań z podsłuchu. Dzięki temu uzyskano wiarygodniejszy, różnorodny materiał: „spontaniczne teksty z różnych środowisk” (DUNAJ, 1989: 15)³. Już na dość wczesnym etapie projektu uznano bowiem, że w badaniu socjolingwistycznym najcenniejsze są spontaniczne wypowiedzi utrwalone w sytuacji nieoficjalnej: w gronie rodziny, przyjaciół czy znajomych (DUNAJ, 1987: 32–33). Nagrywanie krakowian zostało rozpoczęte pod koniec 1977 roku i było prowadzone przez cały 1978 rok. W tym czasie zgromadzono około 400 godzin nagrań. Jednocześnie rozpoczęto spisywanie wypowiedzi z taśm magnetofonowych. Teksty zapisywano sławistycznym alfabetem fonetycznym, „zaznaczając w nich, w miarę możliwości, jak najskrupulatniej, pauzy, potknięcia, dźwiękowe elementy paralingwistyczne. Prowadzono odsłuchiwanie zbiorowe oraz zespołową weryfikację zapisu. W sumie nad tekstem pracował zespół 4–5 osób. Do druku przygotowano wybór najciekawszych tekstów” (DUNAJ, 1989: 15).

³ Por. BIEŃKOWSKA, UMIŃSKA-TYTOŃ (2010).

W Łodzi nagrywano dialog z eksploratorem lub dialogi między informatorami. Kamińska podkreśla, że jest to bardzo czasochłonne i trudne technicznie. Szczegółowo opisuje rolę nagrywającego rozmowę (KAMIŃSKA, 1979: 90). Wskazówki dotyczą także etapu przygotowawczego. Badaczka zaznacza, że nie nagrywanie z ukrycia, ale umiejętne poprowadzenie rozmowy przez nagrywającego decyduje o wartości zebranego materiału: „[...] przy silnym zaangażowaniu w treść wypowiedzi informator zapomina o formie, nie upiększa jej, ale używa codziennych, potocznych form” (KAMIŃSKA, 1979: 80). Do każdego nagrania dołączano charakterystykę językową informatora: jakim wariantem polszczyzny się posługuje, czy zwraca uwagę na dobór środków językowych, czy w trakcie rozmowy zmienia rejestr polszczyzny. Nagrane teksty przepisywano odręcznie z wykorzystaniem alfabetu fonetycznego. Przerwy w wypowiedzi zaznaczano pionowymi kreskami (pauza krótka – jedna kreska, pauza długa – dwie kreski. Oznaczano też imitowanie cudzego głosu czy cudzego dialogu przez informatora. Opatrywano też uwagami „ciekawe, zdaniem eksploratora, formy wyrazowe, wyrazy i zwroty frazeologiczne” (KAMIŃSKA, 1979: 90).

W ośrodku łódzkim opracowywano również kartotekę, która miała być wykorzystana do redagowania słownika. Odnotowywano w niej zleksykalizowane zjawiska fonetyczne (*krzciny, ochwiara*), słowotwórcze (*rozpuścicy, pobudynek*), leksykalne, składniowe i słowotwórcze (opis fiszki – KAMIŃSKA, 1979: 90). W kartotece znalazły się też formy nie nagrane, lecz zasłyszane u nagrywanego. Wiek mówców odnotowywany na fiszkach wyrażano w trzech grupach: do lat 30, do 60, starsi. Całość nagranych (i zasłyszanych) wypowiedzi miała służyć do opisu polszczyzny mówionej łódzian, natomiast wybór wypowiedzi miał zostać wydany w postaci antologii łódzkich tekstów. Kamińska rozważała wydanie wyboru tekstów w pisowni półfonetycznej lub nawet w zapisie ortograficznym („w pisowni literackiej”), żeby mogli z niego też skorzystać socjologowie, etnografowie, historycy.

We Wrocławiu również uznano, że najlepszym materiałem do badań byłyby nieoficjalne wypowiedzi nagrywane z ukrycia bez udziału nagrywającego. Jednocześnie podkreślono, że jednak najważniejsza w czasie nagrania jest rola eksploratora, który może wpłynąć na swobodę wypowiedzi i zadbać o naturalną atmosferę całej sytuacji. Dopuszczano możliwość uzupełniania danych (w razie stwierdzenia takiej potrzeby). Postanowiono nie nagrywać według wcześniej ustalanego klucza. Zdecydowano jedynie, że zbieranie materiałów rozpocznie się od najmłodszej generacji – dzieci ze szkół podstawowych. Uznano bowiem, że dotarcie do rozmówców będzie najłatwiejsze w zaprzyjaźnionych szkołach. Planowano, że w miarę bogacenia się zbioru kryteriów doboru mówców będą one precyzowane. Transkrybowanie miało się odbywać z zastosowaniem alfabetu fonetycznego tylko wtedy, kiedy u rozmówcy zostaną stwierdzone wyraźne cechy regionalne czy gwarowe. Przypuszczano, że w wymowie wrocławian pojawi się obficie wariacja fonetyczna w ramach dialektu kulturalnego. Miała być ona odnotowana w ogólnym opisie tekstu, a w transkryptach – zaznaczona pisownią półfonetyczną. Kierowano się ekonomizacją wysiłku oraz faktem, że kasety będą zarchiwizowane, dzięki czemu zawsze będzie można je ponownie przeanalizować. Planowano opublikowanie antologii przetran-skrybowanych tekstów.

6. Główne kierunki badań w XX wieku

Materiał zebrany przez zespół Uniwersytetu Śląskiego w Katowicach był oczywiście szansą na kontynuację badań historycznojęzykowych zainicjowanych przez Stanisława Rosponda. Jednak tak duży i różnorodny zbiór tekstów otworzył przede wszystkim możliwość badania języka mówionego w miastach śląskich i zagłębiowskich pod względem różnicowania wiekowego, zawodowego, etnicznego i kulturowego mieszkańców. Specyfika badanego obszaru stała się także zachętą do zajęcia się zagadnieniem polityki językowej w środowisku wielkoprzemysłowym (LUBAŚ, 1975).

Nowy materiał pokazał przede wszystkim, że w kontekście wypowiedzi mówionych redefinicji domaga się wypowiedzenie. Poza tym uznano, że warto podjąć zagadnienie segmentacji tekstu, typów pauz, nakładania się replik, intonacji czy elementów paralingwistycznych w wypowiedzi (LEBDA, 1979). Jeszcze jednym nowym zainicjowanym zagadnieniem badawczym stała się analiza naturalnych wypowiedzi językowych przygotowywanych do edycji (nazwano ten proces „fryzowaniem” tekstu).

W ośrodku krakowskim szybko zarysowały się zakresy badawcze rozwijane przez poszczególnych członków zespołu projektowego (DUNAJ, 1981: 8; 1984: 7–8). Były to między innymi ogólne zagadnienia socjolingwistyczne o charakterze metodologicznym (Dunaj), fonetyczne różnicowanie środowiska wielkomiejskiego (Kurek), leksyka mówionej odmiany języka (Kowalik, Frodyma, Ożóg), składnia oraz pragmatyka lingwistyczna (Awdiejew, Labocha, Ożóg), podkrakowska gwara na wiejskich terytoriach wchłoniętych przez miasto (Kąś).

Badania łódzkie zostały zaplanowane wielokierunkowo. Po pierwsze, postanowiono prowadzić analizę w formie studium przypadku (w poszukiwaniu wariantywności językowo-stylistycznej uwarunkowanej sytuacyjnie): po badaniach pilotażowych typuje się pracowników łódzkiego zakładu przemysłowego, następnie nagrywa się próbki ich rozmów w różnych sytuacjach: na stanowisku pracy, z personelem inżynierjno-technicznym, z prowadzącym badania, a także – w środowisku domowym. Po drugie, zdecydowano się również na badania pogranicza Wielkiej Łodzi z okalającymi wsiami, w których mieszka ludność łącząca pracę na roli z pracą w przemyśle. Zaplanowano również drobniejsze tematy: formacje ekspresywne łodzian różnych pokoleń, język codzienny na targowiskach miejskich i w sklepach handlu uspołecznionego – analiza wybranych form językowych; składnia wypowiedzi nagrywanych z ukrycia w warunkach domowych; wyrazy obelżywe, wyzwiska, przewziska, zdrobnienia imion; polszczyzna dzieci w wieku przedszkolnym (generacja najmłodsza).

Wrocławskie badania były skoncentrowane głównie wokół zjawisk fonetycznych (NIECKULA, 1992). Niestety, nie zrealizowano bardzo ciekawego pomysłu na badanie idiolektów. Jak już zostało wspomniane, przypuszczano, że wśród badanych nie zaobserwuje się systemowych, wspólnych dla wszystkich mieszkańców, zjawisk wyróżniających polszczyznę wrocławian. Między innymi z tego powodu Nieckula uważał, że warto koncentrować się na pojedynczych mówcach i ich rodzinach: Pisał: „[...] we Wrocławiu mamy szczególne powody, aby brać pod uwagę wielką wariantywność na poziomie idiolektu (lub tego, co można nazwać idiolektem), czyli liczyć się ze znacznym różnicowaniem osobniczym i rodzinnym” (NIECKULA, 1990: 31).

Studia nad literaturą przedmiotu wskazują, że w drugiej połowie lat 90. ubiegłego wieku badania nad polszczyzną mówioną miały w ramach opisywanego uprzednio projektu nie były kontynuowane. Jednak znalazły trwałe miejsce w historii badań socjolingwistycznych.

7. Powrót do badań nad polszczyzną mówioną Wrocławia

Wiosną 2018 roku odnaleźliśmy materiały z projektu „Polszczyzna mówiona Wrocławia”, który był prowadzony w Instytucie Filologii Polskiej w latach 90. XX wieku. Z tamtych czasów zachowało się wiele nagrań oraz transkryptów w formie maszynopisów. Nie udało się wówczas, niestety, doprowadzić do ich publikacji. Część odnalezionych materiałów była zdekompletowana, tzn. pozostały albo tylko kasety, albo tylko transkrypty. Ostatecznie wybraliśmy do utrwalenia osiem godzin nagrań, które miały pełne metadane. Cały materiał został zdigitalizowany w Pracowni Humanistyki Cyfrowej UWr przez Piotra Malaka. Do transkrypcji nagrań zastosowaliśmy program ELAN⁴ wykorzystywany również do opracowywania tekstów gromadzonych w wyszukiwarce danych konwersacyjnych Spokes, w której ostatecznie dostępne są wrocławskie materiały z lat 90.⁵

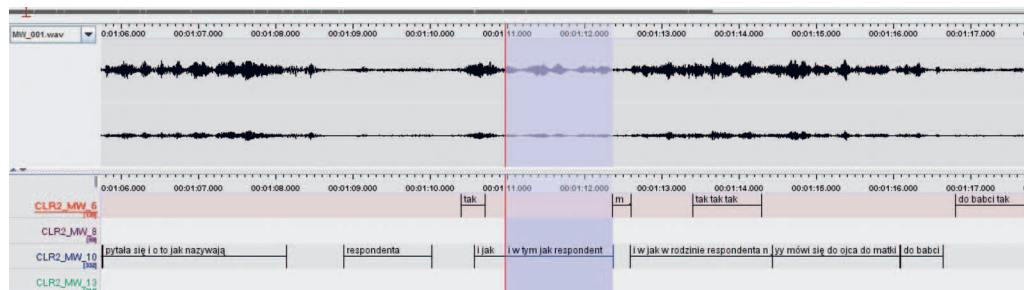
Narzędzie do transkrypcji tekstu ELAN pozwala na pracę z nagraniem tekstem, którego obraz widziany jest w postaci fali dźwiękowej z towarzyszącą jej linią czasu. Współczesny zapis transkrypcyjny utrwała się zupełnie inaczej niż w maszynopisie. Transkrypcję bowiem umieszcza się w warstwach przypisanych poszczególnym mówcom, a dodatkowa linia (*events*) pozwala na uwzględnianie zjawisk dźwiękowych towarzyszących mówieniu (np.: śmiechu – *_voice_laugh_*; kaszlu – *_voice_cough_*) oraz czasu ich trwania. Widzimy zatem nie tylko następstwo replik (jak w starych transkryptach), ale także równoległe wypowiedzi oraz sygnały z tła. Taki sposób zapisu umożliwia lepsze pokazanie dialogowości tekstu i relacji między replikami: ich następstwo, nakładanie się, pauzy między replikami mówców, wpływ zjawisk dźwiękowych towarzyszących mówieniu na jakość tekstu lub momenty rozpoczynania kolejnych replik.

Praca ze zdigitalizowanym tekstem w programie komputerowym z dodatkowym wykorzystaniem słuchawek daje, oczywiście, lepsze możliwości odsłuchu tekstu. Uzyskujemy głośniejszy, bardziej precyzyjny dźwięk, łatwość wielokrotnego odtwarzania (bez włączania i wyłączania magnetofonu), możliwość odsłuchu pojedynczych segmentów wypowiedzi zgodnie z wybranym odcinkiem fali dźwiękowej czy linii czasu, odsłuch w ramach pętli (wielokrotnego powtarzania), odtwarzanie tekstu z uwzględnieniem różnych odcinków czasowych poprzedzających fragment lub po nim następujących. Takie ułatwienia techniczne sprawiają, że osoba transkrybująca może dokładniej zapisywać słyszane dźwięki, a mniej polegać na swojej pamięci słuchowej i przewidywaniu możliwego tekstu. Przewidywany tekst (niejako odgadywany z pamięci) zwykle bardziej odpowiada wersji ogólnej języka lub też jego odmianie pisanej. W rezultacie w przygotowanych współcześnie transkrypcjach zdecydowanie mniej jest miejsc oznaczanych jako niemożliwe do transkrypcji (ELAN:

⁴ [online: <https://tla.mpi.nl/tools/tla-tools/elan>; data dostępu: 14.02.2020].

⁵ Por. też: [online: <http://ppm.uni.wroc.pl/2018/11/22/badania-nad-polszczyzna-mowiona-wroclawia-lata-80-xx-wieku/>; data dostępu: 14.02.2020].

_gap_unclear_; w maszynopisie – /xxx/) lub urwane (ELAN:_trunc_, w maszynopisie: ..., np. pu...puść). Natomiast zdecydowanie częściej są odnotowywane pauzy wypełnione (typu yy) czy inne zjawiska charakterystyczne dla wypowiedzi spontanicznej (por. rys. 1).



Rys. 1. Fragment transkrypcji w programie ELAN

Ze względu na to, że w programie komputerowym transkrypcji zawsze towarzyszy ścieżka dźwiękowa, stosuje się w nim uproszczony zapis ortograficzny⁶ (nie jest zresztą możliwe pełne oddanie idiolektalnych cech artykulacyjnych mówców w tradycyjnie stosowanej transkrypcji półfonetycznej). Dzięki automatycznemu wyszukaniu poszczególnych realizacji danego słowa można bowiem sprawdzić sposoby jego wypowiadania w całym transkrybowanym materiale. Podobnie można porównać jakość i długość wypowiadania elementów paralingwistycznych, które są odnotowywane na przykład jako yy, ii, aa, mhm. Skoordinowanie ścieżki dźwiękowej z linią czasu i dokładny zapis kolejnych fragmentów wypowiedzi umożliwia też obserwację elementarnych jednostek wypowiedzi mówionej. Nie ma pokusy segmentacji odsłuchiwanej wypowiedzi znakami interpunkcyjnymi zgodnie z zasadami obowiązującymi w tekście pisanym. Taką właśnie metodą opracowano na nowo znaczącą część wrocławskich nagrań.

Powstały korpus jest ciekawy i niejednorodny. Tworzą go prywatne i służbowe rozmowy w różnych sytuacjach komunikacyjnych prowadzone przez dorosłych i dzieci w bardzo różnym wieku. Dorosli mówcy mają różne wykształcenie, profesje i pochodzenie regionalne: są to osoby przybyłe po wojnie do Wrocławia z różnych regionów dzisiejszej Polski, a także na przykład z Wilna, Lwowa czy okolic Wołynia.

Jak zostało to już zaznaczone, uczestnikom projektu zależało przede wszystkim na nagrywaniu spontanicznych wypowiedzi nieoficjalnych. Najpewniej z tego powodu w odnalezionych materiałach stanowią one zdecydowaną większość. Znajduje to także swoje odzwierciedlenie w nowo opracowanym korpusie. Współtworzą go głównie rozmowy towarzyskie o charakterze polilogów (*Rozmowa przy obiedzie*, *Rozmowa dwóch małżeństw*, *Rozmowy u L.*, *Aktorki*, *Kłopoty z mieszkaniem*) oraz swobodne dialogi (*Wrażenia z Ameryki*). Pewną częścią korpusu są interakcje werbalne dzieci (*Rozmowa chłopców*). Utrwalono także sposób prowadzenia rozmowy dorosłych z dziećmi (*Rozmowa z córeczką*, *Rozmowa dzieci i ich mam*, *Dzieci budują statek*). Warto podkreślić, że w części nagranych rozmów nieofi-

⁶ Uproszczony zapis ortograficzny w tym projekcie oznaczał niestosowanie wielkich liter poza nazwami własnymi.

cialnych nie bierze udziału uczestnik projektu. Takie nagrania uznawano za najcenniejsze. Analiza metadanych wskazuje, że do takich należą: *Rozmowa chłopców*, *Rozmowa dzieci i ich mam*, *Rozmowy u L.*, *Aktorki*, *Rozmowa przy obiedzie*, *Dzieci budują statek*. Natomiast *Wrażenia z Ameryki*, *Rozmowa dwóch małżeństw* i *Kłopoty z mieszkaniem* to rozmowy przeprowadzone z udziałem nagrywającego. Jednak ogląd odręcznych notatek na starych, papierowych transkryptach wskazuje, że niektóre z tych rozmów eksplorator mógł nagrywać bez wiedzy pozostałych osób (na marginesie transkryptu *Kłopoty z mieszkaniem* widnieje m.in. uwaga: „z podsłuchu”). Natomiast na transkrypcie dialogu aktorki z nagrywającym *Wrażenia z Ameryki* widnieje zapisek innego rodzaju: „swobodna rozmowa z aktorką o teatrze i wrażeniach z pobytu w Stanach Zjednoczonych”. W opisie nieoficjalnej interakcji werbalnej wyeksponowano więc cechę swobodności. Uczyniono to pewnie dlatego, by nie wątpił w nią żaden badacz, który mocno przywiązywał wagę do utrwalania wypowiedzi wyłącznie tych osób, które nie wiedzą, że są nagrywane.

Dwie rozmowy, o których nie wspomiano uprzednio, zbliżają się w swej istocie do wywiadu narracyjnego (*Pan Jan z Wilna*, *Rozmowa dorosłej córki*⁷). Są próbą uzyskania informacji o historii życia ludzi zaawansowanych wiekiem przez dużo młodsze od nich osoby. Z całą pewnością nie są to rozmowy formalne, ale różnią się co do charakteru od przywołanych wcześniej nieformalnych wypowiedzi o charakterze towarzyskim.

Jak już zostało zauważone, twórcy projektu „Polszczyzna mówiona miast” bardzo dbali o zróżnicowanie sytuacyjne nagrywanych interakcji werbalnych. Znajduje to także odzwierciedlenie w materiale wrocławskim. Szczęśliwie się więc składa, że wśród odnalezionych i nowo opracowanych nagrań ujawniono także dwa polilogi utrwalone w sytuacji formalnej. Są to *Dyskusja naukowa* i *Zebranie zakładu*.

W *Dyskusji naukowej* bierze udział 7 osób. Ich wypowiedzi to zarówno wystąpienia mówione w odmianie opracowanej (powitanie uczestników konferencji, przemówienie rektora), jak i głosy członków audytorium (np. nauczycieli), odpowiedzi referentów na pytania z sali czy informacje przekazywane uczestnikom konferencji przez organizatorów. Nagrania te pokazują, jak bardzo zróżnicowane pod względem stylistycznym mogą być wypowiedzi formułowane w sytuacji oficjalnej. Bardzo ciekawa jest również obserwacja wnikania elementów dialogu do monologowych z założenia wypowiedzi konferencyjnych. Wraz z nagraniami przygotowywanymi na potrzeby projektu GeWiss (w trakcie jego realizacji utrwalano m.in. wygłaszane na konferencjach referaty z dyskusją po ich wygłoszeniu) mogą chociażby stanowić element porównawczy w kolejnych badaniach z zakresu współczesnej polszczyzny naukowej w odmianie mówionej.

Na tle całości nowo opracowanego wrocławskiego zbioru wyjątkowym nagraniem jest *Zebranie zakładu* (zachowało się z pełnymi metadany i maszynopisem starego transkryptu). Jest to rozmowa 15 naukowców podczas roboczego spotkania w IFP UW r tuż przed przystąpieniem do badań miejskiej polszczyzny mówionej. W trakcie tej rozmowy formułowane są główne cele projektu w ramach ogólnopolskiego programu węzłowego. Biorą w niej udział naukowcy różnych pokoleń: zarówno doświadczeni profesorowie, jak

⁷ *Rozmowa dorosłej córki* została ujawniona w trakcie transkrybowania ok. połowy trzydziestominutowego pliku zatytułowanego (na wzór tytułu odpowiedniej kasety magnetofonowej) *Pan Jan z Wilna*. Nie ma do niej żadnych metadanych.

i adiunkci ostrożnie formułujący tematy swoich prac habilitacyjnych (niektóre z nich zostały później faktycznie zrealizowane). Już sam odsłuch nagrania *Zebranie zakładu* przekonuje do idei utrwalania rozmów z przedstawicielami nauki choćby w formie wywiadów naracyjnych. Mają one bowiem wielowymiarową wartość, trudną do przecenienia nie tylko w kontekście badań socjolingwistycznych.

Trzeba podkreślić, że przy ponownym opracowywaniu starych nagrań intensywnie pracował duży zespół pracowników różnych ośrodków naukowych. W dalszej części artykułu zostaną wymienione wszystkie osoby, które wzięły udział w opisywanym projekcie.

Tabela 1

Projekt „Mowa_Wrocław” (maj-czerwiec 2018)

Etapy projektu	Uczestnicy projektu
Inwentaryzacja kaset i transkryptów, wstępne opracowanie metadanych	Monika Zaško-Zielińska UWr, Anna Majewska-Tworek UWr
Współpraca z CLARIN-em	Piotr Pęczik UŁ
Szkolenie w transkrypcji (ELAN)	Piotr Pęczik UŁ, Maciej Grabski UŁ, Kamil Wabnic UWr
Dygitalizacja nagrań	Piotr Malak UWr
Poprawianie jakości nagrań po dygitalizacji	Kamil Wabnic UWr
Przetworzenie metadanych dla CLARIN-u	Maciej Grabski UŁ
Anotacja morfosyntaktyczna i zindeksowanie danych w systemie Spokes	Piotr Pęczik UŁ, Paweł Kowalczyk UŁ
Anotacja czasowa narzędziami mowa.clarin.pl	Michał Kowalczyk
Przetranskrybowane nagrania (symbol, tytuł) i autorzy transkrypcji	MW_001: <i>Zebranie zakładu</i> – Anna Majewska-Tworek (IFP UWr) MW_002: <i>Rozmowy z córeczką</i> – Paulina Witkowska (IFP UWr) MW_003: <i>Rozmowa przy obiedzie</i> – Marta Śleziak (IFP UWr) MW_004: <i>Rozmowy u L.</i> – Aleksandra Płochocka (IFG UWr) MW_005: <i>Dzieci budują statek</i> – Ewa Kaczmarz (IFP UWr) MW_006: <i>Rozmowa dzieci i ich mam</i> – Kinga Sudoł (IFP UWr) MW_007: <i>Rozmowa dwóch małżeństw</i> – Kamil Wabnic (IFP UWr) MW_014: <i>Dyskusja naukowa</i> – Monika Zaško-Zielińska (IFP UWr) MW_015: <i>Kłopoty z mieszkaniem</i> – Maria Rudnicka (IFP UWr) MW_017: <i>Wrażenia z Ameryki</i> – Karolina Okurowska (IFP UWr) MW_018: <i>Pan Jan z Wilna</i> – Michalina Jaworska (IFP UWr) MW_022: <i>Aktorki</i> – Agnieszka Skwarek (IFG UWr) MW_050: <i>Rozmowa chłopców</i> – Natalia Śmiełowska (IFP UWr) MW_057: <i>Rozmowa dorosłej córki</i> – Michalina Jaworska (IFP UWr)

8. Włączenie danych do systemu Spokes

W ramach projektu CLARIN-PL od 2015 roku rozwijane i utrzymywane są dwie wersje systemu wyszukiwawczego Spokes dla korpusów polszczyzny mówionej. Pierwsza z nich pozwala na przeszukiwanie danych konwersacyjnych, które pierwotnie weszły do korpusu PELCRA (WALIŃSKI, PĘZIK, 2007) i NKJP (PĘZIK, 2012). Wyszukiwarka obsługuje specjalną składnię zapytań korpusowych (PĘZIK, 2015), a także eksportowanie wyników oraz pełnych transkrypcji zindeksowanych rozmów. Odtwarzanie fragmentów nagrań pasujących do zapytań korpusowych jest możliwe dzięki ręcznej transkrypcji rozmów oraz oznaczeniu czasu trwania poszczególnych wypowiedzi, czyli początku i końca ich wystąpienia na osi czasu. Ręczna transkrypcja rozmów jest konieczna ze względu na zaszumienie i różnorodne ekstralingwistyczne zjawiska występujące w większości nagrań, a co za tym idzie – niską skuteczność użycia systemów automatycznej transkrypcji mowy. Jednocześnie, oznaczenie granic ręcznie transkrybowanych wypowiedzi pozwala na użycie narzędzi służących przetwarzaniu mowy opracowanych w projekcie CLARIN-PL (KORŻINEK et al., 2017) do oznaczania czasu trwania słów i głosek w nagraniach. Z tej anotacji korzysta między innymi nowsza wersja wyszukiwarki o nazwie SpokesMix, która umożliwi wizualizację i eksport nie tylko wyrazów, ale ich fonemicznych segmentów w wynikach wyszukiwania (PĘZIK, 2018).

Proces włączania transkrypcji i nagrań zbioru *Mowa Wrocławia* do indeksu systemu Spokes przebiegał zgodnie z zasygnalizowaną powyżej procedurą dygitalizacji danych konwersacyjnych. Transkrypcje w formacie ELAN (.eaf) zostały najpierw podzielone na segmenty wyrazowe, a następnie poddane analizie morfosyntaktycznej za pomocą tagera APT_PL (PĘZIK, LASKOWSKI, 2017). W kolejnym kroku wyliczone zostały automatycznie iloczasy słów i głosek wewnątrz anotowanych ręcznie granic wypowiedzi. Następnie transkrypcje z anotacją czasową i lingwistyczną zostały zindeksowane w systemach Spokes i SpokesMix. Rysunek 2 ukazuje przykład zapytania o przymiotniki występujące przed rzeczownikami *sprawa*, które daje wystąpienia fraz takich, jak *ciężka sprawa*, *moja sprawa* itp. Zakres wyszukiwania jest w tym wypadku ograniczony do transkrypcji ze zbioru *Mowa Wrocławia* poprzez zdefiniowanie filtra *Source = UWR_MMW* w zaawansowanych opcjach wyszukiwania.

Transkrypcje nagrań ze zbioru *Mowa Wrocławia* włączone do indeksu SpokesMix zawierają 64 861 segmentów wyrazowych, co stanowi niespełna 2,6% wszystkich polskich danych zindeksowanych w tym korpusie. Mimo pozornie niewielkiego udziału *Mowy Wrocławia* w całej puli danych jej transkrypcje stanowią cenne uzupełnienie całości zasobów co najmniej z dwóch powodów. Po pierwsze, rejestr językowy transkrypcji oraz metodologia ich pozyskania są spójne z próbkami polszczyzny konwersacyjnej, które są głównym i unikatowym w wolnym dostępie komponentem korpusów PELCRA i NKJP udostępnianych przez system Spokes. Po drugie, utrwalone na nagraniach rozmowy pochodzą z lat 80. ubiegłego wieku, podczas gdy ogromna większość nagrań dostępnych do tej pory w zasobach Spokes pochodzi z lat 2000–2018. Dzięki wsparciu zespołu i infrastruktury CLARIN-PL nagrania sprzed około 35 lat mogą być przeszukiwane i odsłuchiwane w swobodnym dostępie przez współczesnych badaczy spontanicznej polszczyzny konwersacyjnej.

The screenshot shows the Spokes search interface. At the top, a search box contains the query '<tag=adj.*> sprawa**'. Below it, the 'Dismax' section shows 'source_corpus:UWR_MMW'. A filter section includes 'NOT', 'AND', and 'OR' options, with 'Source' set to 'equal' and 'UWR_MMW'. A 'Query name' field and a 'Mark as default' checkbox are also visible. Below the search interface, a table of concordances is displayed, showing 37 entries. The table has columns for '#', 'Left', 'Match', 'Right', 'Wav', 'Emu', and 'Audio'. The 'Match' column highlights the word 'sprawa' in various contexts from the corpus.

#	Left	Match	Right	Wav	Emu	Audio
1	no	ciężka sprawa	.			
2	. bo ja siedzę a to nie jest mieszkanie moje tylko zięcia . tu jest	ciężka sprawa	z tym ja siedzę bo siedzę .			
3		jest to konkretna spre				
4	owę więc wiesz _gap_unclear_ wiesz ja uważała m że to nie jest	moja sprawa	_gap_unclear_ wiesz			
5	sprawa nauczycieli tutaj yy była i pani o tym mówiła a ja yy	najważniejsza	: to jest w ogóle to			
6	yy to była by chyba yy ta	najważniejsza	:			
7	im się że to przedstawienie opowiada o naszych problemach o	naszycy sprawy	okazało się że jest ono bardziej uniwersalne niż nam się wyd...			
8	no to jest kwestia to sprawa scenografii ale w tekście jest po	prostu sprawa	prosta ze kilku mężczyzn buduje mur tak ?			
9	nie to są dwie	różne sprawy	odbiór za granicą i granie klasyki wydaje mi się że			

Rys. 2. Wyniki zapytania o przymiotniki przed rzeczownikiem *sprawa* w Spokes (zbiór: *Mowa Wrocławia*)

Literatura

- AWDIEJEW A., 1981: *Składnia języka mówionego jako proces bezpośredniego wyboru leksykalnego i formalnego*. W: DUNAJ B., red.: *Studia nad polszczyzną mówioną Krakowa*. [T.] 1. Kraków, s. 93–102.
- BIEŃKOWSKA D., UMIŃSKA-TYTOŃ E., 2010: *Regionalizacja czy unifikacja leksyki? Uwagi na marginesie kwestionariusza do badań regionalizmów leksykalnych*. „Studia Językoznawcze” IX, s. 23–31.
- BRUGMAN H., RUSSEL A., 2004: *Annotating Multi-media / Multi-modal Resources with ELAN*. In: LINO M., XAVIER M., FERREIRA F., COSTA R., SILVA R., eds.: *Proceedings of the 4th International Conference on Language Resources and Language Evaluation (LREC 2004)*. Paris, s. 2065–2068.
- DUNAJ B., 1981: *Badania języka mówionego mieszkańców Krakowa*. W: *Studia nad polszczyzną mówioną Krakowa*. [T.] 1. Kraków, s. 7–10.
- DUNAJ B., 1984: *Dalsze badania języka mówionego mieszkańców Krakowa*. W: *Studia nad polszczyzną mówioną Krakowa*. [T.] 2. Kraków, s. 7–11.
- DUNAJ B., 1987: *O kilku problemach metodologicznych socjolingwistycznych badań języka mówionego (w świetle dotychczasowych doświadczeń)*. „Socjolingwistyka” VI, s. 29–37.
- DUNAJ B., 1989: *Język mieszkańców Krakowa*. Cz. 1: *Zagadnienia teoretyczne, fonetyka, fleksja*. Kraków.
- DUNAJ i in., 1979: DUNAJ B., AWDIEJEW A., KOWALIK J., KUREK H., OŻÓG K.: *Ogólne założenia badań języka mówionego Krakowa i Nowej Huty*. „Socjolingwistyka” II, s. 77–85.

- FRODYMA M., 1981: *W sprawie słownika polszczyzny mówionej*. W: DUNA J. B., red.: *Studia nad polszczyzną mówioną Krakowa*. [T.] 1. Kraków, s. 89–92.
- FURDAL A., 1966: *Podział polskich dialektów miejskich*. „Rozprawy Komisji Językowej WTN” VI, s. 155–161.
- FURDAL A., 1973: *Materiały do kwestionariusza do badania dialektów kulturalnych i miejskich w Polsce*. „Rozprawy Komisji Językowej WTN” IX, s. 41–59.
- KAMIŃSKA M., 1979: *Założenia metodyczne badania polszczyzny mówionej w Łodzi*. „Socjolingwistyka” II, s. 77–94.
- KĄŚ J., 1984: *Problem wariantowości fonetycznej w gwarach ludowych w obrębie Krakowa (na przykładzie a pochylonego)*. W: DUNA J. B., red.: *Studia nad polszczyzną mówioną Krakowa*. [T.] 2. Kraków, s. 73–91.
- KORŻINEK i in., 2017: KORŻINEK D., MARASEK K., BROCKI Ł., WOŁEK K.: *Polish Read Speech Corpus for Speech Tools and Services*. In: *Selected Papers from the CLARIN Annual Conference 2016, Aix-en-Provence, 26–28 October 2016*. Linköping, s. 54–62.
- KOWALIK J., 1981: *Niektóre regionalizmy leksykalne w mowie młodzieży szkolnej*. W: DUNA J. B., red.: *Studia nad polszczyzną mówioną Krakowa*. [T.] 1. Kraków, s. 69–87.
- KUREK H., 1981: *Próba zastosowania metod statystycznych do badania fonetyki języka mówionego mieszkańców Krakowa. Na przykładzie wygłosowego -q*. W: DUNA J. B., red.: *Studia nad polszczyzną mówioną Krakowa*. [T.] 1. Kraków, s. 21–33.
- KUREK H., 1991: *Zróźnicowanie fonetyczne języka mówionego mieszkańców Krakowa (wybrane zagadnienia)*. W: DUNA J. B., OZÓG K., red.: *Studia nad polszczyzną mówioną Krakowa*. [T.] 3. Kraków, s. 147–155.
- LABOCHA J., 1981: *Przerwania jako sygnały trudności wyboru leksykalnego i formalnego w polszczyźnie mówionej*. W: DUNA J. B., red.: *Studia nad polszczyzną mówioną Krakowa*. [T.] 1. Kraków, s. 111–120.
- LEBDA R., 1979: *Dźwiękowe elementy parajęzykowe. Na przykładzie mówionego języka robotników województwa katowickiego*. „Socjolingwistyka” II, s. 149–160.
- LUBAŚ W., 1975: *Badania nad językiem jako punkt wyjścia dla ustalania i realizowania polityki językowej*. W: KANTYKA J., LUBAŚ W., red.: *Rola nauk społecznych w humanizacji środowiska wielkoprzemysłowego*. Katowice, s. 62–69.
- LUBAŚ W., 1978: *Teksty języka mówionego mieszkańców miast Górnego Śląska i Zagłębia*. T. 1. Katowice.
- LUBAŚ W., 1980: *Teksty języka mówionego mieszkańców miast Górnego Śląska i Zagłębia*. T. 2. Katowice.
- NIECKULA F., 1990: *Uwagi ogólne do badań nad polszczyzną Wrocławia*. W: NIECKULA F., red.: *Polszczyzna mówiona Wrocławia*. Cz. 1. Wrocław, s. 8–38.
- OZÓG K., 1984: *Grzecznościowe akty mowy*. W: DUNA J. B., red.: *Studia nad polszczyzną mówioną Krakowa*. [T.] 2. Kraków, s. 147–157.
- OZÓG K., 1991: *Jednostki otwierające i zamykające replikę w dialogu*. W: DUNA J. B., OZÓG K., red.: *Studia nad polszczyzną mówioną Krakowa*. [T.] 3. Kraków, s. 71–89.
- PĘZIK P., 2018: *Increasing the Accessibility of Time-Aligned Speech Corpora with Spokes Min*. European Language Resources Association (ELRA). Miyazakis, s. 4297–4300.
- PĘZIK P., 2012: *Język mówiony w NKJP*. W: PRZEPIÓRKOWSKI A., BAŃKO M., GÓRSKI R., LEWANDOWSKA-TOMASZCZYK B., red.: *Narodowy Korpus Języka Polskiego*. Warszawa, s. 37–47.
- PĘZIK P., 2015: *Spokes – a Search and Exploration Service for Conversational Corpus Data*. In: *Selected Papers from the CLARIN 2014 Conference, October 24–25, 2014, Soesterberg, The Netherlands*. Linköping, s. 99–109.

- PĘZIK P., LASKOWSKI S., 2017: *Evaluating an Averaged Perceptron Morphosyntactic Tagger for Polish*. In: *Human Language Technologies as a Challenge for Computer Science and Linguistics. 8th Language & Technology Conference November 17–19, 2017, Poznań, Poland. Proceedings*. Poznań, s. 372–376.
- WALIŃSKI J., PĘZIK P., 2007: *Web Access Interface to the PELCRA Referential Corpus of Polish*. In: WALIŃSKI J., KREDENS K., GOŹDŹ-ROSKOWSKI S., Eds.: *Corpora and ICT in Language Studies. “Łódź Studies in Language” XIII*. Ed. B. LEWANDOWSKA-TOMASZCZYK. Frankfurt am Main–Berlin–Bern–Bruxelles–New York–Oxford–Wien, s. 65–86.
- WILKOŃ A., 2000: *Typologia odmian językowych współczesnej polszczyzny*. Katowice.





#naukanatwitterze

O multimodalnym designie informacji w dyskursie cyfrowym

#scienceviatwitter

On the Multimodal Design of Information in Digital Discourse

Abstract: The subject of the article are the ways of popularizing scientific and popular science knowledge by using Twitter as a social media platform. The research corpus comprises 100 tweets which come from institutional and individual senders, including both representatives of the world of science as well as of the media that aim to popularize scientific knowledge. As a result of the mediallynguistic analysis, which focused on the structural and functional plane of the studied media texts, the high representativeness of multimodal tweets was noted. Namely, language and image contextualize each other: the image allows us to fill the semantic gap resulting from the ellipticalness of the verbal layer of a given tweet. Thus, the contribution of the multimodal information design to knowledge transfer in digital discourse has been demonstrated.

Key words: multimodality, design of information, social media

Abstrakt: Tematem artykułu są sposoby upowszechniania wiedzy naukowej i popularnonaukowej dzięki wykorzystaniu medium społecznościowego, jakim jest Twitter. Na korpus badawczy złożyło się 100 tweetów pochodzących od nadawców instytucjonalnych i indywidualnych, wśród których znaleźli się zarówno reprezentanci świata nauki, jak i mediów nastawionych na popularyzację wiedzy naukowej. W wyniku analizy mediolingwistycznej, która koncentrowała się na płaszczyźnie strukturalnej i funkcjonalnej badanych mediotekstów, zwrócono uwagę na wysoką reprezentatywność tweetów multimodalnych, w których język i obraz wzajemnie się kontekstualizują: obraz pozwala wypełnić lukę semantyczną wynikającą z eliptyczności warstwy werbalnej tweetu. Tym samym wykazano wkład multimodalnego designu informacji w transfer wiedzy w dyskursie cyfrowym.

Słowa kluczowe: multimodalność, design informacji, media społecznościowe

1. Wprowadzenie

Ponowoczesność, której współcześnie doświadczamy, przyniosła wiele transformacji w różnych sferach życia. Aby sprostać nowym wyzwaniom i dotrzymać kroku dynamicznie zmieniającej się rzeczywistości, człowiek musi wykształcać w sobie nowe kompetencje bądź też przynajmniej dokonać rewizji tych umiejętności, które już posiada, a które wymagają dostosowania do zmieniających się warunków życia, pracy czy aktywności twórczej. Sferę, w której bardzo wyraźnie widoczne jest istotne przyspieszenie, bez wątpienia stanowi komunikacja. W dobie ogromu informacji napływających bardzo różnymi kanała-

mi i przyjmujących rozmaite formy, szczególnie pożądane jest „bycie w kontakcie”, które nie tylko umożliwia „szybkie reagowanie”, ale pozwala także bez angażowania dużych sił i środków dotrzeć z przekazem do szerokiego grona odbiorców. Przestrzenią, którą cechuje bardzo wysoki poziom dynamizmu komunikacyjnego, jest przestrzeń cyfrowa, w niej zaś szczególnie miejsce przypada tzw. *nowym nowym mediom* (LEWINSON, 2010). Pojawienie się Facebooka, Instagrama czy Twittera sprawiło, że świat się zglobalizował także w tym sensie, że osoby znane dotąd z gazet, radia czy telewizji stały się dużo bardziej dostępne i osiągalne. Interakcyjność mediów społecznościowych spowodowała konieczność zredefiniowania tradycyjnych ról uczestników procesu komunikacji. Nadawca komunikatu przestał być jedyną stroną aktywnie zaangażowaną w tworzenie przekazu medialnego. Taką rolę przyjął na siebie także odbiorca wyposażony przez media społecznościowe w narzędzia pozwalające żywo reagować na informacje, recenzować je oraz – co szczególnie ważne – udostępniać kolejnym użytkownikom Sieci. Zburzony został zatem obowiązujący dotąd podział ról i zmieniła się pozycja odbiorców, których motywuje się do większej aktywności i czyni współtwórcami tekstów wypełniających przestrzeń medialną. Media społecznościowe stały się nowoczesnymi, wirtualnymi arenami wymiany myśli i sądów, ale także przestrzeniami, które coraz więcej podmiotów wykorzystuje do tego, aby konsekwentnie i zgodnie z trendami budować swój wizerunek. Z możliwości tej korzystają nie tylko osoby publiczne (politycy, dziennikarze, aktorzy), ale także instytucje, które w ten sposób popularyzują swoją działalność. Potencjał mediów społecznościowych dostrzegli i postanowili wykorzystać także przedstawiciele świata nauki.

Niniejszy artykuł poświęcony jest temu, jak w mediach społecznościowych, a dokładniej na Twitterze, prezentowane i upowszechniane są informacje o charakterze naukowym i/lub popularnonaukowym. Celem artykułu jest znalezienie odpowiedzi na pytanie o design informacji (BUCHER, 2007) ze świata nauki, czyli o to, jak reprezentanci tego środowiska radzą sobie w sytuacji, gdy chcąc podzielić się ze społecznością Twittera wiedzą o charakterze naukowym czy popularnonaukowym, do dyspozycji mają maksymalnie 280 znaków przypadających na jeden post. Właśnie to ilościowe ograniczenie znaków w jednym wpisie wyróżnia Twitter spośród innych mediów społecznościowych. Ów limit znaków może stanowić pewne utrudnienie, wymuszające na użytkownikach tego medium zwięzłość wypowiedzi, o którą zwykle trudno w przypadku treści naukowych czy popularnonaukowych, jednak z perspektywy analizy komunikacji medialnej sposób zaistnienia nauki na Twitterze stanowi interesujące zagadnienie badawcze, które może posłużyć jako kolejny głos w dyskusji nad warunkami skutecznego projektowania informacji w przestrzeni cyfrowej. Zaproponowana w niniejszym artykule analiza mediolingwistyczna ma umożliwić udzielenie odpowiedzi na pytania o to, kim są użytkownicy Twittera publikujący treści naukowe i popularnonaukowe oraz czy wykorzystują oni multimodalność jako element wspierający prezentację kwestii naukowych w sytuacji istnienia ograniczeń formalnych przekazu (maks. 280 znaków w jednym wpisie). Korpus badawczy będący przedmiotem analizy tworzy 100 tweetów powstałych między styczniem a wrześniem 2019 roku i pochodzących od dwóch grup nadawców, tj. nadawców instytucjonalnych oraz indywidualnych, których łączy to, że wykorzystują Twitter jako narzędzie porozumiewania się i wymiany informacji o charakterze naukowym i/lub popularnonaukowym.

2. Architektura informacji w dyskursie cyfrowym

Pojawienie się mediów cyfrowych bez wątpienia zrewolucjonizowało komunikację w tym sensie, że pozwoliło w niespotykanym dotąd wymiarze przełamać dystans czasowo-przestrzenny, który w wielu wypadkach poważnie utrudniał lub wręcz uniemożliwiał sprawne i wolne od barier komunikowanie się. Wraz z rozwojem cyfryzacji komunikacja z pewnością stała się szybsza i dużo bardziej zróżnicowana, jeśli chodzi o formę jej medialnego zapośredniczenia, ale jednocześnie dużo mniej otwarta na kontakt bezpośredni, który przez wieki odgrywał w niej rolę kluczową.

Media wykorzystywane przez społeczności plemienne były naturalne, oparte na biologicznym wyposażeniu człowieka, służyły do kontaktów bezpośrednich, „twarzą w twarz”, wymagających obecności i bliskości. Współczesne społeczeństwa natomiast w przeważającej mierze, choć nie wyłącznie, wykorzystują media techniczne, pokonujące barierę czasu i (lub) przestrzeni, podtrzymujące kontakty zapośredniczone. Miejsce komunikacji typu *face to face* zajmuje komunikacja typu *interface-to-interface* (komunikowanie zapośredniczone, zmediatyzowane).

GOBAN-KLAS, 2007: 43

Wprowadzenie kolejnych rozwiązań technologicznych sprawiło, że komunikacja i relacje międzyludzkie do niej wiodące nabrały zupełnie nowego wymiaru. U źródeł postępującego na przestrzeni lat rozwoju rynku medialnego leżało uświadomienie sobie faktu, że informacja może stanowić towar, zasób warunkujący pożądany rozwój różnych sfer życia i działalności człowieka. To właśnie wokół informacji zaczęła powstawać cała infrastruktura, umożliwiająca gromadzenie, przetwarzanie i udostępnianie danych, a tym samym dzielenie się wiedzą. Tak oto – na bazie *społeczeństwa informacyjnego*, a następnie *sieciowego* – wykształciło się tzw. *społeczeństwo medialne*, w którym:

- 1) międzyludzkie kontakty i stosunki są w ogromnym i przeważającym stopniu zapośredniczone przez używane media;
- 2) media masowe tworzą swoistą rzeczywistość wirtualną, kulturę medialną;
- 3) infrastruktura medialna, a w szczególności telekomunikacyjna, jest podstawą Sieci i obiegów informacyjnych o różnej skali (od lokalnej do globalnej), fundamentalnych dla wszelkich działań i kontaktów jednostkowych oraz organizacyjnych we wszystkich sferach życia;
- 4) niemal wszelkie działania ludzkie są wspomagane przez formy medialno-teleinformacyjne;
- 5) przemysły (kultury) medialne stanowią istotny i rosnący element gospodarki i zatrudnienia.

GOBAN-KLAS, 2007: 44

Zdaniem Tomasza Goban-Klasa w przełamaniu dystansu czasowo-przestrzennego warunkowanego rozwojem nowych technologii tkwi pewien paradoks. Zapśredniczenie medialne komunikacji przyczynia się bowiem do tego, że ludzie bliscy sobie w sensie fizycznym,

a uwikłani w nowe technologie, przenoszą część swoich kontaktów do sfery wirtualnej kosztem relacji bezpośrednich, ci zaś, których dotąd dzielił znaczny dystans, dzięki wsparciu technologicznemu mają szansę ów dystans pokonać. W kontekście wykorzystania nowych technologii przez sektor nauki kluczowe jest to drugie zjawisko, czyli porozumienie, do jakiego – dzięki zapośredniczeniu medialnemu – dochodzi między użytkownikami Sieci i którego przedmiotem stają się treści o charakterze naukowym czy popularnonaukowym.

Świat nauki powoli, ale konsekwentnie, odkrywa dla siebie kolejne sfery medialne, które można wykorzystać, by dzielić się wiedzą ze specjalistami, ale też popularyzować naukę wśród niespecjalistów. Nowe media stwarzają ku temu bardzo dogodne warunki, pozwalając na prezentację wiedzy naukowej w sposób przystępny, nieodbiegający od tego, jak prezentuje się wiadomości dotyczące innych sfer życia. To przekazywanie niejako „przy okazji” informacji ze świata nauki ma tę niewątpliwą zaletę, że skraca dystans, który wobec nauki może odczuwać odbiorca bezpośrednio z tą sferą niezwiązany. Fakt, że treści naukowe czy popularnonaukowe anonsowane są jak wiele innych, tj. płyną w strumieniu informacji ze świata kultury, polityki czy sportu, sprawia, że w odczuciu potencjalnego odbiorcy kontakt ze sferą nauki staje się czymś normalnym i naturalnym, czymś, wobec czego nie należy odczuwać lęku ani obaw. W erze ekranów liczne wiadomości ze świata nauki, w tym te dotyczące działalności naukowo-badawczej ludzi nauki, obecne są na ekranach tak, jak to ma miejsce w wypadku informacji z wielu innych dziedzin życia:

[...] ekran, w parze z komputerem, staje się dzisiaj podstawowym narzędziem umożliwiającym dostęp do różnorodnych informacji, czy będą to obrazy, filmy czy tekst. Używamy go do czytania codziennej prasy, do oglądania filmów, do komunikowania się ze współpracownikami, znajomymi i przyjaciółmi, i – co najważniejsze – do pracy.

MANOVICH, 2006: 177

Era ekranów to także era nowego wymiaru wizualności. Wizualność w sposób naturalny zawsze była obecna w komunikowaniu się bezpośrednim. To obraz stanowił naturalny pierwowzór pisma, a potem konsekwentnie mu towarzyszył, początkowo w formie bogato zdobionych liter inicjalnych w starodrukach, następnie zaś jako ilustracje, by współcześnie zaistnieć na przykład w postaci zróżnicowanych elementów audiowizualnych. Warto w tym miejscu podkreślić, że wizualności nie należy odnosić wyłącznie do obrazu, ale jest ona także cechą inherentną samego pisma, bez względu na to, czy jest to pismo odręczne, drukowane czy elektroniczne, które pojawia się na ekranach komputerów, laptopów, tabletów czy smartfonów, czyli zawsze wtedy, kiedy komunikacja przebiega w sposób medialnie zapośredniczony, tj. z wykorzystaniem e-maili, SMS-ów, forów internetowych, blogów czy wpisów w mediach społecznościowych. Wizualność komunikacji z pewnością nie może więc być uznana za odkrycie XX czy XXI wieku, gdyż człowiek ze swojej natury odbiera świat, kierując się swoim pierwszym wrażeniem, a ono zazwyczaj jest właśnie wrażeniem wizualnym (por. CICHĄ, 2017). Współczesność przyniosła jednak nowe formy wizualności, zwiększyła jej intensywność i pole oddziaływania, ale też oczekiwania, które się przed nią stawia także w kontekście prezentacji treści naukowych czy popularnonaukowych.

Komunikacja wizualna jest zatem z jednej strony naturalna, z drugiej wykoncy-powana, spełniająca określone zadania i oczekiwania. Jej punktem odniesienia może być zbiorowość, jak też pojedyncza osoba, może być ona bezpośrednia lub realizować się poprzez konkretne medium obrazkowe.

DĄBAŁA, 2013: 90

Utożsamianie komunikacji wizualnej wyłącznie z obrazem czy też wielością obrazów jest błędem, gdyż składa się na nią wiele relacji, których recepcja przebiega z zaangażowaniem kanału wzrokowego. Kluczowe są tu powiązania powstające na styku tego, co językowe, i tego, co obrazowe. To właśnie w kontekście tzw. *tekstów językowo-obrazowych* (STÖCKL, 2004; 2015) warto przyglądać się architekturze współczesnej komunikacji i roli, jaką odgrywają w niej słowo i obraz. W odpowiedzi na pytanie o to, czy jesteśmy świadkami końca hegemonii języka, Maryla Hopfinger podkreśla, że:

obraz przeszedł długą drogę **nobilitacji kulturowej** i dzisiaj powszechnie obecny, pełni zarazem funkcję przedstawiającą i funkcję znaczeniową. [...] W dzisiejszej przestrzeni komunikacyjnej słowo i obraz są dobrze **współobecne**, mogą pełnić zarówno samodzielne, jak i komplementarne funkcje; obraz nie tyle wypiera słowo, ile pełnić może dzisiaj zarówno swoiste funkcje przedstawiające, jak i funkcje, które kiedyś przypisywano tylko słowu.

HOPFINGER, 2013: 13, wyróżnienia w oryginale

Ta nobilitacja kulturowa obrazu, o której wspomina Hopfinger, jest kluczowa w kontekście podjętych tutaj rozważań na temat prezentacji medialnej treści naukowych i popularnonaukowych. Obraz zmienił bowiem architekturę informacji zdominowanej przez słowo, przestał być wyłącznie elementem estetyzującym, należącym do świata formy, a stał się też źródłem treści i znaczenia. Połączenie elementów językowych i form wizualnych pozwoliło stworzyć nowy, bo semiotycznie zróżnicowany, język komunikacji. Obraz unaocznia, dokumentuje, a dzięki specjalistycznym technologiom pozwala dotrzeć także tam, gdzie nie sięga ludzki wzrok, i przedstawia świat widziany w wielkim zbliżeniu, wyjaśniając przy tym jego prawidłowości i zawiłości, eksponując kluczowe fragmenty jakiegoś urządzenia czy też uwieczniając moment, w którym zachodzi określone zjawisko. W tym sensie obraz jest nie tylko przekazem informacyjnym, ale także interpretacyjnym, którego zaletą jest to, że odbierany bywa w sposób natychmiastowy i nielinearny, holistyczny. „Wizualizacja pełni funkcję wyjaśniającą i emotywną, staje się strategią nadawczą, która modeluje proces komunikacji” (JANUS-KONARSKA, 2013: 109). W obszarze nauki pozwala dokumentować przebieg badań i przyczynia się do popularyzacji dokonań naukowo-badawczych. Wobec głosów, że świat, a z nim komunikacja, stały się zbyt obrazowe i że zagraża to słowu, warto zachować krytyczny dystans, pamiętając, by angażując obraz i język, czerpać z ich niewątpliwego potencjału i pozwolić im wzajemnie się uzupełniać wszędzie tam, gdzie wymaga tego dobro komunikacji, które powinno być wartością nadrzędną. „Tak zwany zwrot obrazowy pracuje na rzecz kompilacji znaczeniowej i awansu kulturowego obrazu, a nie eliminacji słowa. Dzisiaj obraz wzbogaca i równoważy scenę komunikacyjną. Nie wypiera słowa” (HOPFINGER, 2013: 14).

Patrząc na współczesną komunikację, nie sposób nie dostrzec wspomnianej kompilacji znaczeniowej i dialogu, w którym synchronicznie współistnieją język i obraz, angażowane w proces produkcji, przetwarzania, archiwizowania i udostępniania informacji. Zdaniem Bogusława Skowronka „wielość ofert medialnych jest w istocie warunkowana wielością i różnorodnością wymagań poszczególnych grup odbiorczych” (SKOWRONEK, 2018: 16), do których trzeba dotrzeć z przekazem dla nich zrozumiałym i atrakcyjnym, ale jednocześnie takim, który spełni pokładane w nim oczekiwania i nadzieje nadawcy. Warto w tym miejscu podkreślić za Hopfinger, że przez długie lata „przedstawienia obrazowe jawiły się jako ogólnie dostępne, łatwe w odbiorze, czytelne same przez się. Nikt nie widział potrzeby edukacji wizualnej” (HOPFINGER, 2003: 141). Tymczasem kompetencja multimodalna, definiowana przez Stöckla jako zespół czynności kognitywnych i praktycznych, niezbędnych na etapie tworzenia i percepcji tekstów powstałych wskutek interakcji różnych zasobów znaków, powiązanych z sobą w syntaktyczno-semantyczną i funkcjonalną całość (STÖCKL, 2015: 45), jest kluczowa w sytuacji, gdy interfejsy urządzeń przenośnych jednocześnie pełnią funkcję nie tylko *interfejsów kultury* (por. ORZEŁ, 2014), ale także wprowadzają w świat wiedzy, w tym wiedzy naukowej. Jak podkreśla Gerd Antos, „duża część naszej wiedzy nie tylko jest reprezentowana i archiwizowana w tekstach, lecz w ogóle dopiero *konstruuje* się językowo *jako* tekst” (ANTOS, 2009: 176). Ponieważ jednak współczesna komunikacja bardzo często jest zapośredniczona medialnie, ludzka wiedza konstruuje się także w tekstach funkcjonujących w przestrzeni medialnej. Iwona Loewe używa w tym kontekście pojęcia *mediotekstu*, argumentując, że mediotekstem jest ten, który:

1. Zaistnieje w medium masowym, został dla niego przygotowany, a także przez użytkowników tego medium jest spodziewany kulturowo;
2. Zostaje wytworzony z przeznaczeniem do upublicznienia za pomocą medium masowego;
3. Zostaje wytworzony zawodowo przez pracownika redakcji medium masowego, zwanego dziennikarzem;
4. Jest formą »komunikacji jednostkowej [zainicjowaną – I.L.] za pomocą mediów technicznych – są to dzisiaj przede wszystkim komunikaty, które są tworzone, przekazywane i odbierane poprzez urządzenia elektroniczne typu komputer lub telefon komórkowy« (STÖCKL, 2015: 28).
5. Jest multimodalny.

Odpowiednio do wymienionego porządku mediotekstami są:

Ad 1. Post, wpis na blogu, monolog prezenterki, list do redakcji.

Ad 2. Infografika, horoskop, krzyżówka.

Ad 3. Felieton, słuchowisko, serwis wiadomości.

Ad 4. SMS, mejl, skype’owanie (rozmowa przez komunikator Skype), czat.

Ad 5. Film, podcast, reportaż internetowy.

LOEWE, 2018: 92

W ujęciu zaproponowanym przez Loewe mediotekst jest pojęciem szerokim i obejmuje komunikaty o różnym ciężarze gatunkowym: od SMS-ów, e-maili i pojedynczych wpisów w mediach społecznościowych, przez listy i reportaże, a na filmach kończąc. Bez względu na

zróżnicowanie gatunkowe, strukturalne czy funkcjonalne we wszystkich tych mediotekstach materializuje się ludzka wiedza, w tym także wiedza naukowa, to one czynią ją przyswajalną w sensie zarówno kognitywnym, jak i kulturowym. Nowe formy komunikacji wizualnej i powiązane z nimi nowe modele „wytwarzania wiedzy” (SKOWRONEK, 2018: 18) istotnie wpłynęły na mediasferę człowieka (por. LEPA, 2010; LOEWE, 2018) także w kontekście miejsca i sposobu anonsowania w niej treści naukowych i popularnonaukowych. Ludzie świata nauki, prowadzący swoje profile w mediach społecznościowych, stopniowo dołączają do grona tych, którzy dzięki wykorzystaniu takich narzędzi budują swoją rozpoznawalność i popularność. Jednym z wyznaczników strukturalnych architektury mediotekstów obecnych w przestrzeni nowych nowych mediów jest ich multimodalność: prezentacja treści naukowych czy popularnonaukowych odbywa się z wykorzystaniem elementów zarówno językowych, jak i obrazowych, które mają przyciągnąć i zatrzymać wzrok odbiorcy, ale także zachęcić go do zgłębienia tematu. Te zadania stają się szczególnym wyzwaniem, gdy mediotekst podlega formalnym ograniczeniom, jak to ma miejsce na Twitterze.

3. Tweet jako mediotekst

Twitter to rodzaj mikroblogu, który zarejestrowanym użytkownikom umożliwia dzielenie się wiedzą i zaspokajanie potrzeb związanych z szybkim dostępem do informacji. Od momentu powstania, tj. od 2006 roku, funkcjonuje jako medium oferujące możliwość porozumiewania się za pomocą tzw. tweetów, które w wersji prototypowej składały się zaledwie ze 140 znaków, czyli stanowiły krótkie, wręcz kompaktowe teksty, przeznaczone do szybkiego dzielenia się informacją (por. HAUSENDORF, 2009). Od stycznia 2018 roku zwiększono do 280 górny limit znaków w tweecie. Z perspektywy komunikacyjnej jest to ważna zmiana, gdyż więcej znaków to także więcej słów, za którymi może kryć się więcej wartościowej treści. Zwiększenie limitu znaków otworzyło przed użytkownikami Twittera nowe możliwości zastosowania tego medium, gdyż pozwoliło na:

- szersze „zajawianie” także bardziej skomplikowanej treści, w tym między innymi linkowanie, dzięki czemu łatwiejsze stało się wyszukiwanie zarówno pojedynczych wpisów zawierających określone informacje, jak i całych profili powiązanych z wybranymi zagadnieniami, także tymi o charakterze naukowym czy popularnonaukowym;
- wzbogacanie wpisów na przykład serią hashtagów, a tym samym dokładniejsze pozycjonowanie ich pod kątem różnorodnych, również naukowych zainteresowań odbiorców.

Tweet wypełnia warunki, które Loewe wskazała, definiując medioteksty: powstaje w przestrzeni Twittera, tam jest rozpowszechniany i stanowi formę komunikacji jednostkowej, ponieważ nawet w sytuacji, gdy jest publikowany na profilu należącym do określonej instytucji i w jej imieniu, za jego wygenerowanie odpowiada konkretny nadawca. W zakresie swojej struktury często jest multimodalny, tzn. semiotycznie zróżnicowany, bo w ten sposób rekompensuje ograniczenia wynikające z liczby znaków językowych przypadających na jeden wpis. Zdaniem Marka Palczewskiego „tweet jest nowym, wyłącznie internetowym, hybrydalnym, polimorficznym gatunkiem dziennikarskim, łączącym cechy gatunków informacyjnych i publicystycznych na płaszczyźnie nowej technologii” (PALCZEWSKI, 2013: 31). Nie posiada charakterystycznej dla newsów struktury odwróconej piramidy,

ale wykazuje cechy różnych gatunków i służy do prezentowania komentarzy, polemik czy opinii. Będąc mediotekstem, nie jest prototypowym newsem, chociaż cechuje go „szybkość, bezpośredniość, natychmiastowość, prywatność i (czasami) intymność, a równocześnie masowość (dociera od jednego/niewielu do – potencjalnie – wielu)” (PALCZEWSKI, 2013: 46). Udostępnienie cudzego tweetu swoim obserwatorom, czyli tzw. retweet (RT) wystarcza, aby odbiorca wszedł w rolę nadawcy. W ten typowy dla mediów społecznościowych sposób Twitter znosi zatem klasycznie rozumiany podział ról, a w jego miejsce wprowadza względnie równorzędnych uczestników procesu komunikacji.

Tweety nie są anonimowe, lecz podkreślają autorstwo wypowiedzi. Są więc podwójnie spersonalizowane, zarówno od strony nadawcy, jak i odbiorcy, albowiem adresowane są do wybranej osoby lub grupy osób. Jednakże – zauważmy – same pojęcia nadawcy i odbiorcy tracą w tym kontekście sens, gdyż nadawca jest jednocześnie odbiorcą, a odbiorca nadawcą.

PALCZEWSKI, 2013: 44

W przypadku Twittera można wyróżnić dwa zasadnicze typy relacji nadawczo-odbiorczych:

- relacja jednostronna: X otrzymuje powiadomienia o aktywnościach Y, ale Y nie jest informowany o aktywnościach X, ponieważ nie obserwuje konta X;
- relacja dwustronna: X otrzymuje powiadomienia o aktywnościach Y, Y zaś – o aktywnościach X, co wynika z faktu, że obie strony obserwują swoje profile.

Użytkownikiem Twittera może być nadawca indywidualny, tj. osoba prywatna lub publiczna. Twitter to również przestrzeń, w której coraz aktywniejsi są nadawcy instytucjonalni: swoje profile mają tu rozgłośnie radiowe, stacje telewizyjne, redakcje gazet i czasopism. Pojawienie się na Twitterze podmiotów indywidualnych lub instytucjonalnych związanych ze światem nauki oznacza rozszerzenie dotychczasowych funkcji tego medium. Tweety publikowane przez reprezentantów świata nauki zazwyczaj nie służą realizacji funkcji prototypowych, tj. informowaniu o nagłych, niespodziewanych, często dramatycznych wydarzeniach, sprawdzają się natomiast jako teasery, które anonsują treści równolegle publikowane w innym, wolnym od ograniczeń formalnych medium i dzięki strukturze hipertekstu do tego źródła prowadzą. W przywołanym poniżej przykładzie (rys. 1) nadawcą tweetu jest



Rys. 1. Twitter [online: @uwm_olsztyn; data dostępu: 30.09.2019]



Rys. 2. Strona internetowa Uniwersytetu Warmińskiego-Mazurskiego w Olsztynie [online: @uwm_olsztyn; data dostępu: 18.02.2020]

instytucja – Uniwersytet Warmińsko-Mazurski w Olsztynie – która na Twitterze zachęca do lektury wywiadu z Aliną Naruszewicz-Duchlińską, zajmującą się badaniami takich zjawisk, jak mowa nienawiści, hejt i trolling internetowy, który to wywiad został opublikowany na stronie Uczelni (rys. 2). Zaprezentowany tweet został użyty w celu upowszechniania i popularyzowania wiedzy humanistycznej, chociaż sam fachowej wiedzy językoznawczej nie dostarcza. Pełni jednak inną, ważną funkcję: anonsuje fakt prowadzenia badań hejtu, mowy nienawiści i trollingu oraz – dzięki wykorzystaniu struktury hipertekstu – kieruje do źródła wiedzy o tych zjawiskach, jakim jest strona internetowa Uczelni. Zaufanie odbiorcy do prezentowanych tu treści budowane jest na dwa sposoby: po pierwsze, poprzez umieszczenie tweetu na oficjalnym koncie twitterowym uczelni wyższej, po drugie, poprzez dołączenie zdjęcia osoby, która tę informację firmuje. Taka struktura sprawia, że tweet wypełnia kolejny warunek bycia mediotekstem, tzn. jest multimodalny i tworzy spójną pod względem strukturalnym, funkcjonalnym i semantycznym płaszczyznę wizualną (SCHMITZ, 2011), którą współtworzą subpłaszczyzny reprezentujące różne systemy znaków. Analizując strukturę tego tweetu w układzie pionowym można wyróżnić stały element inicjalny, informujący o właścicielu danego konta (UWM w Olsztynie) i pozwalający zidentyfikować jego login (@uwm_olsztyn). Element graficzny, który umieszczony jest zawsze w lewym górnym rogu i pełni funkcję avatara, w tym sensie uzupełnia treść werbalnego elementu inicjalnego, że pozwala rozszyfrować nazwę kryjącą się za skrótem UWM. Ten przykład pokazuje, że fizyczna bliskość elementów dzielących wspólną płaszczyznę może działać na rzecz ekonomii semiotycznej i pozwala uniknąć dublowania informacji przekazywanych w formie werbalnej i wizualnej. Ciekawym zabiegiem, który w omawianym tweecie również wykorzystano jako element subpłaszczyzny języka, jest posłużenie się dwoma zdaniami pytającymi: *Jak rozróżnić konstruktywną krytykę od hejtu czy mowy nienawiści? Jak samemu nie zostać internetowym trollem?* Wykorzystanie formuły pytań otwartych pozwala realizować funkcję fatyczną: odbiorca tweetu czuje się werbalnie zaczepiony, zagadnięty i zmotywowany do pogłębienia wiedzy na dany temat. Tutaj po raz kolejny kluczowa okazuje się bliskość fizyczna różnych elementów, tym razem elementów językowych. Doświadczenie lektury linearnego tekstu sprawia, że przejście od pytań do linku przekierowującego na stronę uczelni odbywa się w sposób płynny, tj. tak jak lektura spójnego strukturalnie i semantycznie tekstu. Ważnym elementem omawianego tutaj mediotekstu jest także jego subpłaszczyzna obrazu. Stanowi ją zdjęcie kobiety. Bliskość subpłaszczyzny języka i subpłaszczyzny obrazu wskazuje na związek obu elementów struktury tego mediotekstu, jednak można być tego pewnym dopiero po przejściu do strony Uczelni, gdzie odbiorca zapoznaje się z anonsowanym na Twitterze tekstem dotyczącym mowy nienawiści oraz poznaje z imienia i nazwiska analizującą to zagadnienie badaczkę (rys. 2).

4. #naukanatwitterze – analiza mediolingwistyczna

Zaprezentowany przykład stanowi jeden spośród 100 tweetów, tworzących korpus badawczy, na który składają się wpisy opublikowane od stycznia do września 2019 roku. Łączy je to, że 1) porusza się w nich kwestie naukowe lub popularnonaukowe oraz że 2) pochodzą od naukowych lub popularnonaukowych nadawców instytucjonalnych lub

indywidualnych. W wyselekcjonowaniu tweetów i ich autorów posłużono się hashtagami: #nauka, #LudzieNauki, #PopularyzatorNauki. Mając na uwadze strukturę tweetu, dokonano podziału korpusu na cztery kategorie. Za cechę wspólną wszystkich tweetów uznano konstytutywną dla nich subpłaszczyznę języka, którą mogą uzupełniać: (kolejna) subpłaszczyzna języka (w przypadku wpisów retwitowanych), subpłaszczyzna obrazu, link lub materiał wideo. Każda z czterech kategorii miała następującą reprezentatywność w korpusie (tab. 1):

Tabela 1

Korpus badawczy w ujęciu ilościowym, z uwzględnieniem specyfiki strukturalnej

Korpus: 100 tweetów	Tweety nadawców instytucjonalnych: 48/100	Tweety nadawców indywidualnych: 52/100
Subpłaszczyzna języka + subpłaszczyzna języka: 15/100	10/15	5/15
Subpłaszczyzna języka + subpłaszczyzna obrazu: 48/100	22/48	26/48
Subpłaszczyzna języka + link: 21/100	3/21	18/21
Subpłaszczyzna języka + wideo: 16/100	13/16	3/16

Źródło: Opracowanie własne.

Analizowany korpus był zrównoważony, jeśli idzie o reprezentatywność tweetów pochodzących od nadawców instytucjonalnych (48/100) i indywidualnych (52/100). W przypadku obu grup nadawców dało się zauważyć stosunkowo wysoką reprezentatywność tweetów bazujących na połączeniu elementów językowych i obrazowych. Stwierdzono występowanie 35/64 takich tweetów pochodzących od nadawców instytucjonalnych (język + obraz: 22/48 oraz język + wideo: 13/16) i 29/64 od nadawców indywidualnych (język + obraz: 26/48 oraz język + wideo: 3/16). Jednocześnie stwierdzono dysproporcję w częstotliwości używania linków i materiałów wideo w każdej z grup nadawców. Podczas gdy treści linkowane pojawiały się częściej w tweetach pochodzących od nadawców indywidualnych (18/21), te z materiałem wideo udostępniane były częściej przez nadawców instytucjonalnych (13/16).

Na potrzeby dalszej analizy dokonano wyboru tweetów pochodzących od czterech różnych nadawców: dwóch instytucjonalnych oraz dwóch indywidualnych. Są to:

- instytucjonalny nadawca naukowy *Nauka w Polsce* (@naukawpolsce, rys. 3)
- instytucjonalny nadawca popularnonaukowy *Magazyn Focus* (@Focus_pl, rys. 4)
- indywidualny nadawca naukowy, Michael Klemm (@mkulturlinguist)
- indywidualny nadawca popularnonaukowy, Tomasz Rożek (@RozekTom, rys. 5)

Już wstępna analiza stron startowych wszystkich omawianych nadawców pokazuje, że w podobny sposób operują oni różnymi systemami znaków, tzn. na jednej płaszczyźnie łączą zarówno elementy językowe, jak i obrazy. Powstają w ten sposób *płaszczyzny wizualne*, czyli „powierzchnie, na których dzięki przemyślanemu layoutowi teksty i obrazy tworzą wspólne jednostki znaczeniowe” (SCHMITZ, 2015: 59, por. także SCHMITZ, 2017). Płaszczyzny wizualne „skupiają [...] na sobie uwagę, umożliwiają umieszczenie na najmniejszej nawet powierzchni możliwie największej ilości informacji, pozwalają na różnorodne sposoby prezentacji, strukturyzacji, ukierunkowania i recepcji oraz wspomagają szybką orientację” (SCHMITZ, 2015: 60). W przypadku stron startowych analizowanych nadawców widać, że wszystkie one przyjmują odgórny schemat struktury Twittera, ale wypełniają go



Rys. 3. Twitter [online: profil @naukawpolsce; data dostępu: 30.09.2019]

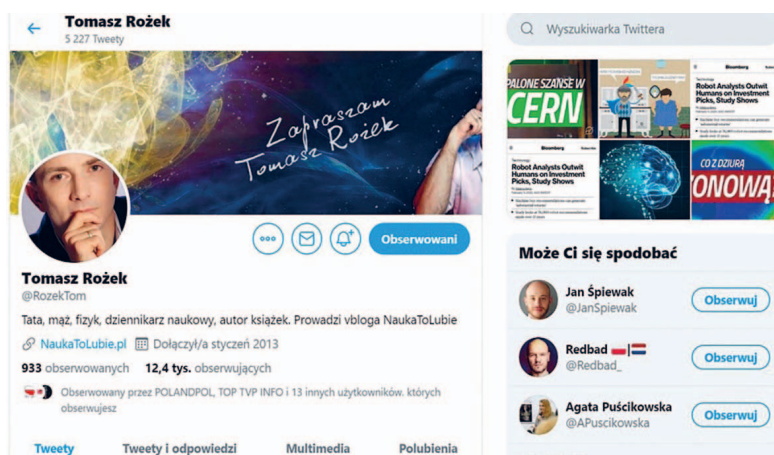


Rys. 4. Twitter [online: profil @Focus_pl; data dostępu: 30.09.2019]

różną treścią. W przypadku naukowego nadawcy instytucjonalnego (rys. 3) elementem inicjalnym profilu jest obraz przedstawiający naczynia laboratoryjne z kolorowymi cieczami, co ma zapewnić odbiorcę, że ma do czynienia z profilem o charakterze naukowym (treść obrazu może wręcz sugerować skoncentrowanie się na naukach doświadczalnych). Potwierdzeniu naukowości profilu służy także treść subpłaszczyny językowej: *Informacje o polskich badaniach naukowych oraz szkolnictwie wyższym z portalu Nauka w Polsce PAP*. W przypadku popularnonaukowego nadawcy instytucjonalnego, tj. *Magazynu Focus* (rys. 4) elementem inicjalnym analogicznego modułu uczyniono logo czasopisma „Focus”, w subpłaszczynie językowej zaś zawarto informację o tym, że jest to *najlepszy popularnonaukowy magazyn w Polsce* (po czym dołączono tłumaczenie tej informacji na język angielski). W przypadku obu profili można mówić o tendencji do ekonomii językowej, jaka przejawia się w tym, że stosowane są tu konstrukcje eliptyczne, które dla odbiorcy, ze względu na umieszczenie ich w określonym środowisku wizualnym, są zrozumiałe. Użytkownik Twittera wie, że *Nauka w Polsce* [to serwis prezentujący] *informacje o polskich badaniach naukowych oraz szkol-*

nictwie wyższym z portalu *Nauka w Polsce PAP* oraz że *Magazyn Focus* [to jest] najlepszy popularnonaukowy magazyn w Polsce.

O pewnych podobieństwach i różnicach w strukturze profili można mówić także w przypadku nadawców indywidualnych. Podczas gdy inicjalne subpłaszczyzny obrazowe zwykle są podobne (zarówno w tle, jak i na zdjęciach profilowych widoczni są zazwyczaj właściciele danego profilu), umieszczone pod nimi podpisy informują, kiedy odbiorca ma do czynienia z profilem naukowca (Michael Klemm, używający loginu @mkulturlinguist, przedstawia się jako badacz mediów, kultury i języka, specjalizujący się w badaniach nad komunikacją audiowizualną i polityczną w mediach społecznościowych), a kiedy odwiedza profil popularyzatora nauki (Tomasz Rożek, używający loginu @RozekTom, określa siebie przede wszystkim jako tatę i męża, a dopiero w dalszej kolejności jako fizyka, dziennikarza naukowego, autora książek i vlogera NaukaToLubie; rys. 5). Stałym elementem każdego



Rys. 5. Twitter [online: profil @RozekTom; data dostępu: 30.09.2019]

z profili jest także widoczny po prawej stronie moduł, w którym prezentowane są wybrane z tego profilu informacje oraz rekomendacje profili o podobnej tematyce. Analizując strukturę kont twitterowych wszystkich wymienionych nadawców, warto podkreślić, że profile te operują także pustą przestrzenią, która – jako trzeci obok subpłaszczyzny języka i subpłaszczyzny obrazu element płaszczyzny wizualnej – również odgrywa istotną rolę w tym sensie, że stanowi neutralne (białe) tło dla prezentowanych treści, pozwala dobrze te treści wyeksponować i nie rozprasza uwagi użytkownika portalu.

O tym, jak ważną i pomocną rolę w kreowaniu tweetów służących dzieleniu się wiedzą i popularyzującym naukę odgrywa multimodalność, można przekonać się, analizując konkretne wpisy indywidualnych i instytucjonalnych nadawców naukowych czy popularnonaukowych. Połączenie elementów językowych (inicjalnych w przypadku tweetów) i obrazowych sprawia, że powstaje kompaktowa jednostka komunikacyjna, spójna pod względem strukturalnym, funkcjonalnym i semiotycznym, służąca prezentowaniu, a przede wszystkim rekomendowaniu określonej wiedzy mimo ograniczeń, jakie niesie z sobą maksymalna liczba 280 znaków możliwych do użycia w jednym tweecie. Fizyczna bliskość obu subpłaszczyzn oraz ich wyraźne oddzielenie od reszty płaszczyzny wizualnej powodują, że

odbiorca z rozwiniętą kompetencją multimodalną wie, że gdy użytkownik Twittera Tomasz Rożek pisze: *Intuicja mówi, że powinien się przewrócić. A fizyka utrzymuje go na wodzie :-)*, to za podmiotem domyślnym (*powinien* – on) i zaimkiem osobowym *go* kryje się wspomniany w dalszej części tekstu i widoczny na zdjęciach największy kontenerowiec świata, MSC Gulsun (rys. 6). Obraz pozwala zatem zrozumieć tekst, zanim treść, która jest do przekazania,



Rys. 6. Twitter [online: @RozekTom; data dostępu: 30.09.2019]



Rys. 7. Twitter [online: @Focus_pl; data dostępu: 30.09.2019]

zostanie w pełni zwerbalizowana. To szczególnie ważne w sytuacji, gdy autor wpisu ma do dyspozycji tylko 280 znaków, pozwala bowiem subpłaszczyźnie obrazu przejść na siebie część zadań i odciążać tym samym subpłaszczyznę języka. Działa tu zatem mechanizm synergii semiotycznej, polegającej na wzajemnym kontekstualizowaniu się języka i obrazu. Podobny mechanizm zastosowano także w przypadku tweetu opublikowanego przez *Magazyn Focus* (rys. 7): *Jedno z ostatnich dwóch miejsc na świecie, gdzie składowano próbki wirusa powodującego ospę prawdziwą* (subpłaszczyzna języka) widoczne jest na zdjęciu (subpłaszczyzna obrazu) przedstawiającym centrum wirusologii na Syberii. Tego, co konkretnie spowodowało eksplozję butli z gazem i jakie są skutki pożaru w laboratorium Wektor, użytkownik dowie się dzięki przekierowaniu na stronę portalu. Ten szczególny przykład pokazuje jednak, że mimo ograniczeń objętościowych i dzięki zastosowaniu multimodalności w jednym tweecie można przekazać dużo informacji (takich jak: jedno z ostatnich dwóch miejsc na świecie, wirus ospy prawdziwej, wybuch, centrum wirusologii, Syberia), które mają szansę zainteresować odbiorcę i zmotywować go do dalszej lektury właściwego tekstu naukowego czy popularnonaukowego.

Multimodalność jest pomocna także wtedy, kiedy anonsowane zagadnienie naukowe jest bardziej skomplikowane. Tak jest w przypadku tweetu zaprezentowanego na rysunku 8. Ma on pogłębioną strukturę, bo oprócz 1) subpłaszczyzny języka i 2) linku zawiera jeszcze 3) subpłaszczyznę obrazu. Do zaanonsowania bardziej skomplikowanej treści użyto tu zatem rozbudowanego strukturalnie wpisu, w którym obraz staje się wsparciem dla trudnej, specjalistycznej terminologii (rys. 8). Czym są i jak wyglądają diament, grafen czy grafit, wie wielu ludzi, ale już pojęcia takie jak *fulereny* i *nanorurki* oraz ich wizualizacje będą czytelne i zrozumiałe tylko dla znawców przedmiotu. Subpłaszczyzna obrazu może



Rys. 8. Twitter [online:@naukawpolsce; data dostępu: 30.09.2019]

być jednak tym elementem, który przyciągnie wzrok użytkownika Twittera i zachęci do zgłębienia tematu zasygnalizowanego w postaci eliptycznie sformułowanego tekstu: „[Czy] podręczniki [muszą trafić] do aktualizacji? Diament, grafen, grafit, fulereny (i nanorurki) to nie [jest] wszystko! [...]” (Twitter, @naukawpolsce; data dostępu: 30.09.2019).

5. Wnioski

„Teksty ukazują wiedzę!” (ANTOS, 2009: 287). Fakt ten odnosi się do tekstów zróżnicowanych pod względem gatunkowym, strukturalnym i funkcjonalnym, obecnych zarówno w przestrzeni medialnej, jak i poza nią. Analiza mediolingwistyczna mediotekstów, którymi są tweety, wykazała, że mimo ograniczeń formalnych służą one dzieleniu się wiedzą naukową przede wszystkim w tym sensie, że przyczyniają się do jej upowszechniania, anonsując dane zagadnienie i wskazując źródła jego naukowej analizy. Ograniczenia formalne wynikające ze specyfiki Twittera nie są barierą w transferze wiedzy. Sięga się w nim bowiem po rozwiązania o charakterze multimodalnym, których potencjał tkwi w synergii semiotycznej różnych systemów znaków. Komunikaty, które są rezultatem współdziałania języka i obrazu, nie tylko przykuwają wzrok, ale także przekazują informację, dzięki której odbiorca może dotrzeć do konkretnej wiedzy naukowej lub popularnonaukowej. Możliwości, jakie niesie z sobą synergetyczne współistnienie na jednej płaszczyźnie reprezentantów różnych systemów znaków, są ogromne, pod warunkiem, że dysponuje się wiedzą i kompetencjami, które pozwolą wykorzystać efektywnie potencjał danej modalności, a jej ewentualne braki zrekompensować dzięki posłużeniu się inną modalnością. Język i obraz użyte w tweetach wzajemnie się kontekstualizują. Eliptyczność warstwy werbalnej nie stanowi problemu komunikacyjnego, jeśli w strukturze tweetu znajdzie się element, który semantycznie tę lukę wypełni. Przykład tweetów i ich wkładu w transfer wiedzy dowodzi, że multimodalny design informacji wciąż ma zatem szansę być generatorem nowej jakości w dyskursie cyfrowym.

Źródła

- Twitter, @uwm_olsztyn [online: https://twitter.com/uwm_olsztyn [data dostępu: 30.09.2019]
- Twitter, profil @naukawpolsce [online: <https://twitter.com/naukawpolsce>; data dostępu: 30.09.2019]
- Twitter, @Focus_pl [online: https://twitter.com/focus_pl?lang=pl; data dostępu: 30.09.2019]
- Twitter, profil @RozekTom [online: <https://twitter.com/rozektom>; data dostępu: 30.09.2019]

Literatura

- ANTOS G., 2009: „Teksty ukazują wiedzę!” *O prymacie medialności w pełnym napięć obszarze pomiędzy światami tekstów i (między)kulturowymi konstrukcjami rzeczywistości*. MIKOŁAJCZYK B., tłum. W: BIEUT-HOMPLEWICZ Z., CZACHUR W., SMYKAŁA M., red.: *Lingwistyka tekstu w Niemczech. Pojęcia, problemy, perspektywy. Antologia tłumaczeń*. Wrocław, s. 279–290.
- BUCHER H.-J., 2007: *Textdesign und Multimodalität. Zur Semantik und Pragmatik medialer Gestaltungsformen*. In: ROTH K.S., SPITZMÜLLER J., Hrsg.: *Textdesign und Textwirkung in der massenmedialen Kommunikation*. Konstanz, s. 49–76.
- CICHA K., 2017: *Komunikacja wizualna – humanistyczne oblicze informatyki?* „Studia Ekonomiczne. Zeszyty Naukowe Uniwersytetu Ekonomicznego w Katowicach” CXVII, s. 42–53.
- DĄBAŁA J., 2013: *Aksjologiczne konsekwencje komunikacji wizualnej a telegeniczność*. W: WOLNY-ZMORZYŃSKI K., FURMAN W., SNOPEK J., GROŃ K., red.: *Komunikacja wizualna w prasie i w mediach elektronicznych*. Warszawa, s. 89–99.
- GOBAN-KLAS T., 2007: *Nadchodzące społeczeństwo medialne*. „Chowanna” II, s. 37–51.
- HAUSENDORF H., 2009: *Kleine Texte. Über Randerscheinungen von Textualität*. [online: www.saggi-zeitschrift.unibe.ch/6_09/hausendorf.html; data dostępu: 24.07.2019].
- HOPFINGER M., 2003: *Doświadczenie audiowizualne*. Warszawa.
- HOPFINGER M., 2013: *Czy obraz wypiera słowo?* W: WOLNY-ZMORZYŃSKI K., FURMAN W., SNOPEK J., GROŃ K., red.: *Komunikacja wizualna w prasie i w mediach elektronicznych*. Warszawa, s. 11–14.
- LEPA A., 2010: *Mediosfera człowieka*. „Łódzkie Studia Teologiczne” XIX, s. 159–169.
- LEWINSON P., 2010: *Nowe nowe media*. ZAWADZKA M., tłum. Kraków.
- LOEWE I., 2018: *Od mediasfery do mediologii*. W: HOFMAN I., KĘPA-FIGURA D., red.: *Współczesne media. Zagadnienia ogólne. Multimodalność mediów drukowanych*. T. 1. Lublin, s. 21–33.
- MANOVICH L., 2006: *Język nowych mediów*. CYPRYAŃSKI P., tłum. Warszawa.
- ORZEŁ B., 2014: *Appleizacja kultury. Zmiana zachowań komunikacyjnych w kontekście nowych mediów*. Katowice.
- PALCZEWSKI M., 2013: *Tweet jako odmiana newsa czy nowy (ponowoczesny) gatunek dziennikarski?* „Nowe Media” IV, s. 31–49.
- SCHMITZ U., 2015: *Badanie płaszczyzn wizualnych. Wprowadzenie*. MAKOWSKA M., tłum. W: OPIŁOWSKI R., JAROSZ J., STANIEWSKI P., red.: *Lingwistyka mediów. Antologia tłumaczeń*. Wrocław–Dresden, s. 57–77.
- SCHMITZ U., 2017: *Grammatik und Design*. „IDS Sprachraport” Nr. 3, s. 8–17.
- SKOWRONEK B., 2018: *Technologia. Multimodalność. Komunikacja. Nowe wyzwania dla mediolingwistyki*. W: HOFMAN I., KĘPA-FIGURA D., red.: *Współczesne media. Media multimodalne*. T. 1: *Zagadnienia ogólne i teoretyczne. Multimodalność mediów drukowanych*. Lublin, s. 11–19.

- Stöckl H., 2004: *Typographie: Gewand und Körper des Textes – Linguistische Überlegungen zu typographischer Gestaltung*. „Zeitschrift für Allgemeine Linguistik“ Nr. 41 s. 5–47.
- Stöckl H., 2015: *Czytanie tekstów językowo-obrazowych? Elementy kompetencji podstawowej*. Pociask J., tłum. W: Opiłowski R., Jarosz J., Staniowski P., red.: *Linguistyka mediów. Antologia tłumaczeń*. Wrocław–Dresden, s. 114–137.



Funkcja politycznych słów kluczy w memach internetowych

The Function of Political Key Words in Internet Memes

Abstract: The subject of the study is the role of political key words used in memes. To determine their role, a semantic and stylistic analysis of Polish-language memetic statements from 2012–2019 was used, in which the following words were included: *sorrel*, *mirabelles*, *trotyl*, *octopuses*, *light bulb*, *salad*. The article shows that the use of key words even in a completely neutral context allows us to activate references to negatively evaluated events and situations, and to recall the emotions and views associated with them. In addition, quoting chronologically diverse examples of the use of key words has led to the possibility of recreating the process of extending their meanings, which is the result of releasing quotes associated with a specific situation from their function.

Key words: key word, Internet meme, political language, depreciation

Abstrakt: Przedmiotem badań jest rola politycznych słów kluczy wykorzystywanych w memach internetowych. Do jej określenia posłużyla analiza semantyczna i stylistyczna polskojęzycznych wypowiedzi memetycznych z lat 2012–2019, w których użyto jednego z następujących słów: *szczaw*, *mirabelki*, *trotyl*, *ośmiorniczki*, *żarówka*, *sałatka*. W artykule wykazano, że użycie słów kluczy nawet w całkowicie neutralnym kontekście pozwala na uruchomienie odniesień do wartościowanych negatywnie zdarzeń i sytuacji oraz przywołanie związanych z nimi emocji i poglądów. Ponadto przytoczenie zróżnicowanych chronologicznie przykładów użycia słów kluczy pozwoliło na odtworzenie procesu poszerzania ich znaczeń, będącego skutkiem uwalniania od funkcji cytatów kojarzonych z konkretną sytuacją.

Słowa kluczowe: słowo klucz, mem internetowy, język polityki, deprecjacja

1. Wprowadzenie

Narastająca polaryzacja poglądów politycznych i łatwość ich wyrażania za pośrednictwem nowych mediów przyczyniają się do intensyfikacji debaty społeczno-politycznej toczącej się w Internecie. Formułowane w jej ramach wypowiedzi – nie tylko werbalne, ale także wizualne – są publikowane przede wszystkim w mediach społecznościowych (np. Twitter, Facebook) oraz jako komentarze pod artykułami na portalach informacyjnych. Na ogół są one nacechowane emocjonalnie, wyraźnie opozycyjne wobec wybranych wypowiedzi udostępnionych wcześniej, dzięki czemu dość jednoznacznie wskazują stronę sceny politycznej reprezentowaną przez autora. Ta przejrzystość ideologiczna nie jest jednak

tożsama z formułowaniem poglądów wprost. Uczestnicy sieciowych dyskusji politycznych często posługują się ironią, aluzją, niedopowiedzeniem, metaforą etc. Jednym z popularnych środków wyrazu są słowa kluczowe, rozumiane jako jednostki leksykalne o znacznie wyższej frekwencji (ocenionej szacunkowo) w dyskusjach politycznych w porównaniu z ich częstością w innych tekstach, używane w znaczeniach odmiennych od utrwalonych w języku (ograniczonych czasowo i tematycznie) i/lub uruchamiających dodatkowe konteksty.

Odnosząc się do koncepcji „słów kluczowych” jako otwierających drzwi do zrozumienia zjawisk kulturowych (WIERZBICKA, 2007), należałoby uznać te wyekscerpowane z komentarzy politycznych za znaki objaśniające procesy toczące się we współczesnej polityce. Ich ekstensja nie jest ograniczona do tekstów jednego typu – mogą one występować w tytułach artykułów dziennikarskich, wystąpieniach i dyskusjach polityków, popularnych piosenkach, dowcipach, dyskusjach na forach internetowych itp.

Celem podjętym w artykule jest przedstawienie sposobów wykorzystywania wybranych słów kluczowych obecnych we współczesnym dyskursie politycznym (*szczaw*, *mirabelki*, *trotyl*, *ośmiorniczki*, *żarówka*, *sałatka*) do konstruowania wypowiedzi krytycznych wobec polityków w memach internetowych, a zatem w obrębie gatunku stającego się obecnie coraz popularniejszym narzędziem wyrażania opinii i poglądów oraz komentowania aktualnych zdarzeń. Główną funkcją memów nie jest jednak wyrażanie krytyki i formułowanie refleksji, ale dostarczanie rozrywki, wywoływanie rozbawienia lub przynajmniej zainteresowania, które ma skłonić do ich skopiowania i przekazania dalej, czyli – mówiąc językiem memetyki – realizacji procesu zarażania kolejnych ludzkich umysłów. Słowa kluczowe ten proces intensyfikują, uruchamiając odniesienia do wypowiedzi lub zachowań polityków, które stały się popularnym wątkiem w przekazie medialnym, a w ten sposób zyskały charakter swoistych memów informacyjnych, rozumianych w tym przypadku w klasycznym ujęciu Dawkinsowskim jako jednostki informacji o wysokim stopniu „zaraźliwości”. Przywołane w memach internetowych pozwalają na aktualizację kontekstu, który w mediach był wykorzystywany do deprecjacji polityka, a niekiedy również na tworzenie sensów dodatkowych. Ich powtarzanie w różnych, wielokrotnie kopiowanych i stale modyfikowanych wariantach, wynikające z istoty gatunku memetycznego, przyczynia się do wzmocnienia oddziaływania („zaraźliwości”) (KAMIŃSKA, 2011: 61–62; NIEKREWICZ, 2015: 14–17) – rozpoznawalność zwykle jest tak duża, a nacechowanie tak wyraziste, że zbędne staje się przytaczanie konkretnych wypowiedzi, relacjonowanie negatywnie ocenianych sytuacji czy formułowanie precyzyjnych zarzutów.

Język politycznych memów internetowych był już przedmiotem badań – analizom poddano między innymi zabiegi intertekstualne w memach politycznych (WOŁOSZYN, 2016), środki językowo-obrazowe służące etykietowaniu polityków (NIEKREWICZ, 2016), perswazji politycznej (PORYCKA, 2016), wyrażaniu agresji wobec przeciwników (NIEKREWICZ, 2019), satyrze politycznej (ZDUNKIEWICZ-JEDYNAK, 2017; BAKALARSKI, 2018) oraz formułowaniu poglądów społeczno-politycznych (NIEKREWICZ, 2014; 2015), jednak dotychczas nie wyodrębniono słów kluczowych jako istotnych elementów wypowiedzi memetycznych na tematy polityczne. Tymczasem identyfikacja słów, które w tekstach politycznych zyskują dodatkowe znaczenia odnoszące się do osób i sytuacji z tego obszaru tematycznego, może być wykorzystana w badaniach nad zmianami semantycznymi współczesnej leksyki, a także w analizach politologicznych.

W artykule scharakteryzowano werbalne i niewerbalne sposoby aktualizowania treści niesionych przez polityczne słowa klucze. Celem badania nie jest porównywanie frekwencji poszczególnych słów kluczy ani przedstawianie zmian w jej zakresie – obok siebie będą więc występować przykłady wyrazów różniących się stopniem upowszechnienia oraz mających genezę w różnych latach. Analiza służy przede wszystkim pokazaniu, jakie zabiegi słowne i obrazowe stosują twórcy memów, by wykorzystać potencjał semantyczny tkwiący w danym wyrazie kluczu, oraz jakie treści polityczne są w ten sposób wyrażane. Jedna wypowiedź memetyczna może zawierać kilka środków wyrazu – zarówno werbalnych, jak i graficznych, toteż podział artykułu na części nie jest tożsamy z rozłączną klasyfikacją wyekscerpowanego materiału. Wiele przykładów mogłoby zostać omówionych w odniesieniu do każdej z poruszonych kwestii – ich jednokrotne przywołanie w wybranym paragrafie wynika z dążenia do posługiwania się różnorodnymi przykładami i zilustrowania powszechności słów kluczy.

W analizie przywołano przykłady pochodzące z 49 memów opublikowanych na stronach internetowych o funkcji memetycznych generatorów, na portalach informacyjnych, a także w mediach społecznościowych (Asz:dziennik, Błasty, Chamsko, Demotywatory, Dziennik. Rozrywka, Fabryka Memów, Jeja, Kwejk, Mem Generator, Memy, Onet, Paczaizm, Penera, Ryjbuk). Cytowane wypowiedzi wybrano z dużo większego zbioru polskojęzycznych memów politycznych, zgromadzonego w latach 2012–2019, obejmującego około 5 tys. tekstów.

2. Słowa klucze jako element *quasi*-cytatów

W memach internetowych pierwszym etapem upowszechnienia danego słowa i zarazem nadania mu wartości symbolicznej w odniesieniu do kwestii politycznych jest umieszczanie zawierających je wypowiedzi na tle fotografii polityków – na ogół tych, którzy rzeczywiście się nim posłużyli. Przytoczenia rzadko są wierne, zwykle znacznie bardziej dosadne, ale zachowana zostaje intencja mówiącego i ogólny sens wypowiedzi, na przykład: *Jak są głodne, niech jedzą szczaw*¹; *Nie ma chleba? Żryj szczaw i mirabelki*; *Tyyyle szczawiu żeśmy zjedli z nasypu*² (podpisy na tle wizerunku Stefana Niesiołowskiego); *Trotyl, wszędzie trotyl*; *To nie była guma. To był trotyl*³ (podpisy na tle zdjęć Antoniego Macierewicza).

Im więcej czasu upływa od wygłoszenia danej wypowiedzi, tym częstsza i wyraźniejsza jest dekontekstualizacja tworzących ją słów kluczowych przywoływanych w memach

¹ W cytatach z memów internetowych zachowano oryginalną pisownię.

² Nawiązanie do wypowiedzi Stefana Niesiołowskiego w programie *Kropka nad i* w TVN24, 5.03.2013 [online: <https://www.tvn24.pl/wiadomosci-z-kraju,3/jak-chodzilem-do-szkoly-mysmy-wyjedli-caly-szczaw-z-nasypu,310126.html>; data dostępu: 22.06.2019]. Polityk swój brak wiary w istnienie w Polsce problemu niedożywionych dzieci uzasadnił niedostatkami zainteresowania przydrożnymi owocami i szczawiem, które w okresie jego dzieciństwa były powszechnie zrywane i wykorzystywane w kuchni.

³ Nawiązanie do licznych publicznych wypowiedzi Antoniego Macierewicza, w których dowodził znalezienia śladów trotylu na szczątkach polskiego samolotu rozbitego w Smoleńsku [online: <https://www.tvn24.pl/wiadomosci-z-kraju,3/macierewicz-prokurator-klamie-byl-trotyl-dowody-sa-w-polsce,286826.html>; data dostępu: 22.06.2019].

internetowych. Oznacza to, że zawierające je wypowiedzi mogą być przypisywane innym osobom lub odnoszone do odmiennych sytuacji. Ze względu na przewagę w memach ekspresji negatywnej oraz tendencji do deprecjonowania osób publicznych najczęściej tworzone są nawiązania do wypowiedzi, które wzbudziły społeczne oburzenie, były analizowane w komentarzach i dyskusjach medialnych, a nierzadko skłoniły do polemiki także przeciwników ideowych. Aluzja do tych wypowiedzi – najczęściej tworzona właśnie poprzez przywołanie słów najważniejszych pod względem semantycznym – pozwala zasugerować istnienie elementów wspólnych dla różnych z pozoru sytuacji lub osób.

Słowa *szczaw* i *mirabelki* – pierwotnie cytowane, by przypomnieć konkretną wypowiedź Niesiołowskiego – stały się uniwersalnym środkiem wyrażania dezaprobaty także wobec innych działań tego polityka, na przykład: *Szczaw i mirabelki najlepszym afrodyzjakiem*; *Stefan radzi: #NIEDLAWIAGRY. Tylko szczaw i mirabelki*; *Pokazać ci mirabelki?* (memy publikowane po oskarżeniu piosenkarza o przyjmowanie korzyści finansowych w formie usług seksualnych). Jednocześnie pojawiają się w zdaniach przypisywanych innym znanym politykom – są w nich wyraźnymi sygnałami ironii, służą zdemaskowaniu pozorów wyuczajności i skromności demonstrowanych przez polityków dla zyskania sympatii wyborców, na przykład: *Krewetki mi się przejadły. Teraz czas na szczaw i mirabelki* (wypowiedź przypisana Bronisławowi Komorowskiemu); *Pod koniec miesiąca jem tylko szczaw i mirabelki* (słowa przypisane Jarosławowi Gowinowi – nawiązanie do jego wypowiedzi nt. niskich zarobków ministrów). Demaskację uwydatnia posługiwanie się spopularyzowanym przez tabloidy kontrastem „my (zwykli, biedni ludzie) – oni (odseparowani od społeczeństwa, pozbawieni empatii politycy)”, na przykład *Donald Tusk o NASZYCH pieniądzach w ZUSie i OFE: Zawsze możecie zadowolić się szczawiem i mirabelkami* (fotografia Donalda Tuska na konferencji prasowej, na tle zdjęć przedszkolaków); *Dzieciom Szczaw, a sobie premie: Szczaw Stefana nasz budżet uratuje. Każda szkoła niech boisko szczawiove natychmiast buduje!*. *A młodzież niech się sama stołuje* (fotografia przedstawiająca rozmowę ministra finansów Jacka Rostowskiego z premierem Donaldem Tuskiem).

Dla gatunku memetycznego charakterystyczne jest mieszanie prawdy i fikcji, wyolbrzymianie, modyfikowanie cytatów – słowem: instrumentalne traktowanie faktów. Nie ma więc znaczenia to, czy słowa przypisane danej osobie rzeczywiście zostały przez nią wypowiedziane. Ważne jest natomiast potwierdzenie założonej tezy, zwykle tożsamej z przypisaniem danemu politykowi określonej etykiety, a także zwiększenie „zaraźliwości” memu. Wiele z takich tendencyjnych wypowiedzi ma formę autokrytycznych komentarzy – stanowią one realizację charakterystycznego dla memów konceptu demaskacji polegającego na konstruowaniu wypowiedzi tak, by sprawiała wrażenie wygłoszonej nieoficjalnie, w politycznych kuluarach, a więc ilustrującej prawdę nieobecną w przekazie medialnym, ujawniającej prawdziwe myśli, postawy i zamiary osoby publicznej.

Ta swoboda w traktowaniu faktów oznacza, że zmieniani są nie tylko autorzy wypowiedzi czy też sens i forma cytatów, ale także tworzone są zdania, których nikt nigdy nie wypowiedział. Pojawiają się w nich wówczas słowa klucze, którym nadano wartość symboliczną na podstawie zdarzeń, sytuacji i problemów rozpropagowanych przez media (np. *sałatka*; *ośmiorniczki*; *konstytucja*). Mogą one zarówno służyć krytyce jednostek (np. *sałatka* – Krystyna Pańkiewicz), jak i być uniwersalnym środkiem deprecjonowania

członków określonej partii, a nawet ogółu polityków (*konstytucja* – Andrzej Duda, inni politycy PiS; *ośmiorniczki* – członkowie PO, ale także inni politycy).

Przykładem słowa, które w dyskursie publicznym (nie tylko w memach) uległo wyraźnej metaforyzacji, jest *ośmiorniczka* – najczęściej w formie mnogiej (*ośmiorniczki*). Podobnie jak diada *szczaw i mirabelki* pierwotnie było ono odnoszone do wąskiej grupy polityków Platformy Obywatelskiej – bohaterów tzw. afery taśmowej. Narrację prowadzoną przez ówczesną opozycję po ujawnieniu nagrań zdominował populistyczny zarzut trwonienia publicznych pieniędzy na wystawne kolacje, dodatkowo uwydatniający kontrast między poziomem życia politycznych elit i ogółu społeczeństwa⁴. Symbolem nadmiernych zbytków stały się właśnie *ośmiorniczki* – odwołania do nich pojawiają się zarówno w artykułach dziennikarskich, wystąpieniach polityków, jak i memach internetowych oraz komentarzach publikowanych na forach dyskusyjnych, jednak to kreatywność twórców memów – nieograniczona wymogami informacyjnej rzetelności – w najwyższym stopniu przyczynia się do zmienności kontekstów, w których są one przywoływane. Mogą być znakiem nadużycia władzy (np. *Drogie wino? Ośmiorniczki? Lobbyści? A kto nam k...a zabroni?* – wypowiedź przypisana premier Ewie Kopacz), środkiem służącym uwydatnieniu przepaści między stylem życia elit i zwykłych ludzi (*Radio objadał się ośmiorniczkami. A wy?* – wypowiedź umieszczona na tle zdjęcia roześmianej Hanny Gronkiewicz-Waltz; *A to są moje ośmiorniczki jak byłem w szpitalu. Radek jadłeś?* – zdjęcie posiłku szpitalnego; *Tu mi ośmiorniczka przejedzie na motorze zanim wam się poprawi w życiu...* – wypowiedź przypisana Markowi Belce), symbolem przywilejów wywołujących tęsknotę polityków odsuniętych od władzy (*Kiedyś to było życie, ośmiorniczek się pojadło, w gałę się pocharatało... A teraz co zostało???* *Ino sznur!!!* – podpis na tle wizerunku Grzegorza Schetyny). Powyborcza zmiana partii rządzącej pozwoliła zaobserwować postępującą uniwersalizację symbolu *ośmiorniczek*, której wyrazem jest częściowe oderwanie go od sytuacji źródłowej. Znak przywilejów władzy pojawia się także w wypowiedziach krytycznych wobec członków PiS, na przykład: *A może ośmiorniczki* (zdjęcie oblizującej się Krystyny Pawłowicz). Słowo klucz *ośmiorniczki* odnoszone do szerszej grupy polityków znacznie rzadziej jest jednak wplatane do rzekomych cytatów, występuje raczej w trzecioosobowych komentarzach dotyczących apanaży polityków PiS, na przykład: *Małe pytanie do pseudoprawicy: Ile ośmiorniczek od sowy może kupić brocha⁵ na[!] za premię która się należy?* (pytanie na tle fotografii nagłówek „Faktu”: *Rozpamiętanie u Szydło, a u Tuska bieda!*); *77 lotów casq razy ponad 34 000 zł jeden lot. Wychodzi ponad 2 miliony zł. A im przeszkadzała jedna kolacja z ośmiorniczkami* (demotywator przedstawiający Beatę Szydło); *I za takie grosze chcieli go ścigać? Przecież za 650 tys. zł to za ledwie*

⁴ Np. „Zaczął się czas wakacyjny. Ponad połowa dzieci z biedy nie wyjedzie na wakacje. A ci panowie jakieś ośmiorniczki zajadają” (Mariusz Błaszczak w wywiadzie dla Tok FM, 30.06.2014 [online: <http://www.tokfm.pl/Tokfm/10,103168,16242286,polowa-dzieci-z-biedy-nie-jedzie-na-wakacje-a-ci-panowie.html?v=1&obx=16242286&order=najstarsze&pld=28381196>; data dostępu: 24.06.2019]); „Politycy Prawa i Sprawiedliwości nie bywają w takich restauracjach na koszt podatnika. Ja nigdy nie jadłem ośmiorniczek” (Jarosław Kaczyński podczas spotkania wyborczego w Bielsku-Białej, 19.10.2015 [online: <http://bielskobiala.wyborcza.pl/bielskobiala/1,88025,19048886,wybory-parlamentarne-2015-kaczynski-nie-jadam-osmiorniczek.html>; data dostępu: 24.06.2019]).

⁵ Przewisko nadawane Beacie Szydło w internetowym folklorze politycznym.

20 ton ośmiorniczek mógł kupić (fotografia Bartłomieja Misiewicza, któremu NIK w okresie rządów PiS zarzucił przekroczenie planu finansowego Ministerstwa Obrony Narodowej).

Swobodne traktowanie cytatów w memach przyczyniło się w tym wypadku do uwolnienia słowa kluczowego od funkcji cytatu kojarzonego z konkretną sytuacją oraz do uniwersalizacji jego znaczenia.

3. Gry intertekstualne

Twórcy memów wykorzystują najpopularniejsze, przyciągające uwagę elementy współczesnej kultury (np. cytaty, utarte powiedzenia, dowcipy, motywy filmowe, literackie i malarskie, stereotypowe sądy), aby w ten sposób uatrakcyjnić przekaz, a tym samym zwiększyć „zaraźliwość” kompozycji memetycznej. W tym ujęciu tradycja przestaje tworzyć zbiór oddzielnych i skończonych dzieł, stając się rezerwuarem elementów, które łatwo z sobą dowolnie łączyć, zmieniać ich znaczenie, umieszczać w innym kontekście etc. Zabiegi tego typu można uznać za ilustrację ponowoczesnej dyspozycyjności intertekstualnej, która pozwala na swobodne uruchamianie nieskończonej liczby gier elementami kultury.

Memy internetowe są zatem złożone z cytatów, aluzji rozumianych jako niebezpośrednie nawiązania do innych tekstów oraz zrekontekstualizowanych obrazów. Słowa klucze stanowią popularny element tych wielopoziomowych i wieloskładnikowych konstrukcji. Najczęściej składnikami gier intertekstualnych stają się słowa utrwalone w uzusie, o czytelnych odniesieniach sytuacyjnych i/lub ideologicznych. Dzięki ich upowszechnieniu w dyskursie politycznym zbędne staje się formułowanie uwag krytycznych wprost. Wystarczy przywołanie słowa, by zaktualizować dany kontekst, a w ten sposób zdeprecjonować określonego polityka. Jednoczesne nawiązanie do innego tekstu (utworu literackiego, filmu, reklamy, znanego cytatu historycznego lub współczesnego itp.) uatrakcyjnią wypowiedź, a tym samym pozwala na zwiększenie memetycznej „zaraźliwości”, na przykład:

- *Proszę pana, w damskiej ubikacji ktoś wkręcił żarówkę energooszczędną*⁶ (zdanie umieszczone na tle kadru z filmu *Rejs*⁷ stanowiącego popularny szablon memetyczny);
- *Nowy polski serial. Ślepnąc od żarówek. Wszystkie odcinki od października. HBO GO* (mem wzorowany na plakacie polskiego serialu *Ślepnąc od światła*⁸, twarz głównego aktora zastąpiona twarzą prezydenta Andrzeja Dudy);

⁶ Nawiązanie do wypowiedzi prezydenta Andrzeja Dudy wygłoszonej w trakcie spotkania z prezydentem Niemiec Frankiem-Walterem Steinmeierem 24 października 2018 roku. Prezydent RP stwierdził, że negatywnie ocenianą konsekwencją przynależności do Unii Europejskiej jest narzucenie korzystania z żarówek energooszczędnych i związany z tym brak w sprzedaży zwykłych żarówek, 24.10.2018 [online: <https://www.wprost.pl/kraj/10163505/andrzej-duda-w-berlinie-mowil-ozarowkach-to-sa-problemy-nad-ktorymi-zastanawiaja-sie-ludzie.html>; data dostępu: 28.06.2019].

⁷ *Rejs*, reż. M. Piłowski, prod. 1970. Scena donosu: „W damskiej ubikacji napisane jest »głupi kaowiec«”.

⁸ *Ślepnąc od światła*, reż. K. SKONIECZNY, serial TV, prod. od 2018.

- *Stefan, zjedz szczaw! bo jak jesteś głodny to strasznie gwiazdorzysz!* (nawiązanie do reklamy batonika marki Snickers⁹, fotografia Stefana Niesiołowskiego);
- – *Nie mają chleba? Niech jedzą ciastka* (nawiązanie do wypowiedzi królowej Francji Marii Antoniny; słowa przypisane Donaldowi Tuskowi). – *Albo szczaw* (odpowiedź Stefana Niesiołowskiego);
- „*A niechaj narodowie wżdy postronni znają iż Polacy nie politycy, iż ośmiorniczek nie jadają*”¹⁰ (fotografia przedstawiająca polityków PO Radosława Sikorskiego, Elżbietę Bienkowską, Jacka Rostowskiego; nad nimi talerz z ośmiorniczkami);
- *Szczaw, mirabelki i kamieni kupa*¹¹ (kontaminacja wypowiedzi Stefana Niesiołowskiego ze słowami Bartłomieja Sienkiewicza; fotografia Stefana Niesiołowskiego);
- *Tu na razie jest ściernisko, ale będzie San Francisco, bo zasadzimy szczaw!* (modyfikacja słów piosenki¹²; fotografia Stefana Niesiołowskiego na tle ścierniska);
- *Widzę ciemność! Ciemność widzę! Ciemność... #Duda #ciemność #żarówka* (nawiązanie do słów Maksa, bohatera filmu *Seksmisja*¹³; fotografia prezydenta Andrzeja Dudy umieszczona na czarnym tle).

Część nawiązań intertekstualnych nie jest przykładem gry stanowiącej cel sam w sobie, ale pozwala na wyrażenie dodatkowych sensów. Na przykład zestawienie w jednym dialogu odwołania do Marii Antoniny, będącej symbolem izolacji elit, ze słowami Stefana Niesiołowskiego sugeruje podobne podłoże obydwu zdań, a jednocześnie rozszerza diagnozę tej postawy na innych polityków sprawujących władzę.

Znaczną część nawiązań intertekstualnych stanowią te o charakterze wewnątrzgatunkowym (tj. odnoszące się do innych memów). Warunkiem ich rozpoznania jest zatem znajomość frazemów o genezie memetycznej. Większość związków wyrazowych powtarzających się w wypowiedziach memetycznych ma charakter frazematyczny ze względu na swoją odtwarzalność. Autorzy części memów, zwłaszcza tych opartych na szablonach, nie tworzą bowiem wypowiedzi z elementów dobieranych doraźnie, ale posługują się gotowymi elementami kodu memetycznego, cytując je w formie niezmienionej lub tylko w niewielkim stopniu zmodyfikowanej. Powiązanie frazemów z memami szablonowymi oznacza także

⁹ Reklama batonów marki „Snickers”, 2012 roku Właśc. „– Dobra, leszcze. Który zajmował mój dezodorant? [...] – Dan, zjedz snickersa. – Po co? – Bo jak jesteś głodny, to zacznasz strasznie gwiazdorzyć. Lepiej? – Lepiej. – Głodny nie jesteś sobą” [online: <https://www.youtube.com/watch?v=iwuoGMTEuak>; data dostępu: 28.06.2019].

¹⁰ Por. „A niechaj narodowie wżdy postronni znają, iż Polacy nie gęsi, iż swój język mają”. M. REJ: *Do tego, co czytał*. W: IDEM: *Figliki*. Kraków 1574 [online: [oai:mbc.malopolska.pl:83120](http://oai.mbc.malopolska.pl:83120); data dostępu: 28.06.2019].

¹¹ Nawiązanie do wypowiedzi ministra Bartłomieja Sienkiewicza, nagranej w restauracji i ujawnionej w ramach tzw. afery taśmowej w 2014 roku: „Chuj, dupa i kamieni kupa” (o rządowym programie Polskie Inwestycje Rozwojowe).

¹² Golec uOrkiestra: *Ściernisko*, słowa O. GOLEC, R. GOLEC, por. „Tu na razie jest ściernisko / Ale będzie San Francisco / A tam, gdzie to kretowisko / Będzie stał mój bank”.

¹³ Fraza z *Seksmisji* ma charakter kultowego cytatu, w memach internetowych odznacza się wysoką frekwencją. *Seksmisja*, reż. J. MACHULSKI, prod. 1983.

konieczność identyfikacji schematów kompozycyjnych charakterystycznych dla poszczególnych serii memów, na przykład:

- *A tam był nasyp kolejowy... Po maturze chodziliśmy na szczaw* (fotografia Stefana Niesiołowskiego); *A po marszu chodziliśmy na ośmiorniczki* (fotografia Radosława Sikorskiego z flagą) – nawiązanie do memów z różnorodnej tematycznie serii „A po maturze [ewentualnie innym zdarzeniu – A.A.N.] chodziliśmy na...” wykorzystującej zdanie wypowiedziane przez papieża Jana Pawła II w Wadowicach, w 1999 roku: „Po maturze chodziliśmy na kremówki”;
- *To je szczaw. Tego nie ogarniesz!* (obraz Jana Matejki, *Kazanie Skargi*) – naruszenie schematu charakterystycznego dla memów z wizerunkiem ks. Piotra Skargi poprzez zastąpienie powtarzalnej frazy „A idź pan w chuj z... [nazwa wydarzenia, osoby, przedmiotu, zjawiska itp.]” wypowiedzią charakterystyczną dla innej serii „To je/jest x. Tego nie ogarniesz”, służącej do krytycznej oceny zjawisk wymykających się regułom racjonalnego poznania;
- *Zdążyłem postrzelać w Afganistanie. Zdążyłem zjeść ośmiorniczki u Sowcy* (fotografia Radosława Sikorskiego); *ZDAŻYŁAM zrobić karierę. ZDAŻYŁAM zwyzywać wszystkich. ZDAŻYŁAM zjeść sałatkę¹⁴ ale... NIE ZDAŻYŁAM zostać mamą* (fotografia Krystyny Pawłowicz) – nawiązanie do memów serii opartej na przetworzeniu hasła kampanii społecznej „Zdążyłam [tu wyliczenie dokonań zawodowych i przyjemności życiowych]. Nie zdążyłam zostać mamą. Nie odkładaj macierzyństwa na później”; powyższe przykłady ilustrują różnorodność intencji wypowiedzi memetycznych zawierających takie same nawiązania intertekstualne – od częściowo zgodnej z pierwowzorem w przypadku Krystyny Pawłowicz do zupełnie z nim niezwiązanej, stanowiącej krytyczny bilans kariery politycznej Radosława Sikorskiego, należy jednak podkreślić, że w obydwu przypadkach słowa kluczowe współtworzą negatywny wizerunek polityka, są przynajmniej równie ważne co fraza tworząca szablon memetyczny.

Baumanowska intertekstualna dyspozycyjność współczesnej kultury w memach oznacza prowadzenie nieustannego dialogu z tradycją oraz nowymi tekstami, wykorzystywanie odznaczających się wysoką frekwencją frazów – często efemerycznych, stanowiących wyraz chwilowej mody. Ze względu na tę modę właśnie – łatwo rozpoznawalnych, kojarzonych z konkretnymi zdarzeniami i poglądami, ale też poddawanych modyfikacjom pozwalającym na odniesienie do innych osób lub grup politycznych.

4. Nawiązania graficzne

Współistnienie obrazu i słowa jest charakterystyczne dla większości memów internetowych. Obydwa systemy znaków uzupełniają się nawzajem lub też przeciwstawiają sobie,

¹⁴ Nawiązanie do rozpowszechnionych w mediach nagrań i fotografii przedstawiających Krystynę Pawłowicz spożywającą sałatkę w trakcie obrad sejmowych. Zob. np. J. MAKOSA: *Pawłowicz je na sali sejmowej*. Tok.fm, 18.12.2014 [online: http://www.tokfm.pl/Tokfm/1,103454,17152693,Pawlowicz_je_na_sali_sejmowej__Komisja_etyki_ma_dosc_.html; data dostępu: 6.07.2019].

pozwalając tym samym na różne sposoby odczytywania przekazu. Obraz może stanowić aluzję do słów uznanych za kluczowe dla aktualnego dyskursu politycznego – w tym przypadku ich obecność w części werbalnej memu nie jest obligatoryjna, choć często bywa uzupełnieniem elementu ikonograficznego.

Częstym zabiegiem parodystycznym stosowanym przez twórców memów jest nadawanie kształtu żarówki głowie prezydenta Andrzeja Dudy i/lub także pierwszej damy. Obraz stanowi niemy komentarz do prezydenckiej wypowiedzi na temat niedostępności w sprzedaży zwykłych żarówek jako przyczyny spadku poparcia dla integracji europejskiej, rzadziej na jego tle umieszczane są ironiczne podpisy typu *Światłość narodu; Najjaśniejsza para; Nie ma demokracji* (obok twarzy Andrzeja Dudy żarówka energooszczędna); *Jest demokracja* (obok twarzy prezydenta zwykła żarówka) itp.

Zwizualizowana żarówka, podobnie jak *ośmiorniczki* czy *szczaw i mirabelki*, może być traktowana jako symbol całokształtu poglądów politycznych członków danej partii. Taką funkcję ma na przykład w memach stanowiących graficzne modyfikacje powyborczych mapek ilustrujących różnice sympatii politycznych między wschodnią a zachodnią Polską – na tle regionów zachodnich umieszczono rysunek żarówki energooszczędnej, natomiast na tle wschodnich – zwykłej. Podpis *Oświetlenie powyborcze* stanowi wskazówkę interpretacyjną.

Z kolei coraz wyraźniejsza dekontekstualizacja *ośmiorniczek* sprawia, że w obrazach bywają one odnoszone do stylu sprawowania rządów nie jednej, lecz różnych partii. Zapewne nie bez wpływu jest tu upowszechnienie w ostatnich dziesięcioleciach jednostki znaczeniowej *ośmiornica* 'organizacja przestępcza' (ŻMIGRODZKI, 2007–2019) i popularny w popkulturze wątek związków mafii i polityki. Wynikiem skojarzeń tego typu jest mem, w którym na tle ośmiornicy umieszczono napis *Afery PiS* oraz słowa aktualizujące konkretne medialne oskarżenia wobec tej partii: *Srebrna, przetargi KGHM, Solvere, KNF, przetargi PGZ, Sensus Group, PCK, SKOK*.

Typową zabawą graficzną jest zapisywanie słów kluczy na wzór utrwalonych w uzusie hasałów lub logotypów. Formę hasła „Konstytucja” widniejącego na plakacie autorstwa Łuki Rayskiego otrzymały między innymi wyrazy *ośmiorniczki* (obok fotografia Radosława Sikorskiego); *konsumpcja* (napis na tabliczce trzymanej przez Krystynę Pawłowicz); *destylacja* (napis na koszulce Aleksandra Kwaśniewskiego). Akronim PAD (Prezydent Andrzej Duda) umieszczany jest na tle reklam lub logotypów KFC¹⁵ (często zamiast skrótowca stanowiącego nazwę firmy). Skrótowiec PiS bywa publikowany z humorystycznym objaśnieniem pseudoetymologicznym: *Pawłowicz i sałatka*.

Znaczenie słów kluczy aktualizowanych przez obrazy pozostaje niezmienione w stosunku do znaczeń słów obecnych wyłącznie w części werbalnej memu. Połączenie kodu słownego i graficznego umożliwia budowanie opozycji między niesionymi przez nie znaczeniami, a w rezultacie wywołuje efekt komizmu. Liczne są także memy, w których aluzje do słów kluczy mają charakter jedynie wizualny, pozbawiony werbalnych sugestii interpretacyjnych

¹⁵ Nawiązanie do wizyty pary prezydenckiej w restauracji KFC, w trakcie której przeciwniczka polityczna zarzuciła Andrzejowi Dudzie łamanie konstytucji. Zob. np. B. ДУБАЕА: *Kraków. Prezydent Duda w KFC i... pytanie o konstytucję*, 10.06.2018 [online: <https://gazetakrakowska.pl/krakow-prezydent-duda-w-kfc-i-pytanie-o-konstytucje/ar/13246922>; data dostępu: 6.07.2019].

(np. Andrzej Duda z żarówką zamiast głowy) – czytelność żartów tego typu można traktować jako dowód na utrwalenie w publicznym dyskursie przywoływanych słów.

5. Wnioski

Użycie słów kluczy nawet w całkowicie neutralnym kontekście przywołuje określone zdarzenia, sytuacje oraz związane z nimi emocje i poglądy. Ich rozpowszechnienie w medialnym dyskursie politycznym przyczynia się do utrwalenia słów w dodatkowych, symbolicznych znaczeniach. Należy podkreślić, że ekstensja słów kluczy nie ogranicza się do memów internetowych, ale to właśnie ten gatunek wypowiedzi – ze względu na lapidarność i łączenie kodu werbalnego z wizualnym – pozwala na potwierdzenie utrwalenia nowych znaczeń oraz weryfikację stopnia ich uniezależnienia od kontekstu. Swobodne traktowanie kategorii prawdy i fikcji przez autorów memów – polegające na mieszaniu różnych cytatów, nazwisk, wizerunków – ułatwia uwalnianie słowa klucza od funkcji cytatu kojarzonego z konkretną sytuacją i w konsekwencji przyczynia się do uniwersalizacji jego znaczenia. Ponadto typowe dla memów nadawanie politykom etykietek – tożsame z przypisywaniem im jednej wyolbrzymionej, niekoniecznie prawdziwej cechy – współgra z posługiwaniem się słowami kluczami, które w pewnych przypadkach przyjmują funkcję sygnałów etykiet (np. *szczaw* jako znak oderwania od rzeczywistości Stefana Niesiołowskiego; *sałatka* – niezgodnego z *savoir vivre*'em zachowania Krystyny Pawłowicz). Podobnie jak memetyczne etykiety słowa klucze pozwalają na deprecjację polityków, a w jej ramach – na hiperbolizację cech negatywnych oraz generalizację zarzutów.

Trudno wskazać jednoznaczne czynniki wpływające na tworzenie politycznych słów kluczy. Niewątpliwie decydująca jest tutaj wysoka frekwencja w tekstach medialnych dotyczących sytuacji, które wzbudziły społeczne oburzenie, jednak nie jest to jedyny ani też wystarczający czynnik. Funkcję słów kluczy otrzymują przede wszystkim rzeczowniki konkretne, które łatwo można przedstawić w formie obrazu (np. *sałatka*, *żarówka*, *szczaw*, *mirabelki*, *ośmiorniczki*), a jednocześnie odnoszące się do desygnatów spoza tradycyjnie ujmowanej sfery politycznej, zatem pozwalające na uruchomienie żartów opartych na mechanizmie kontrastu oraz na deprecjację kojarzonych z nimi osób.

Słowa klucze nie są zastępowane synonimami, ale zamiast dosłownych przytoczeń mogą pojawiać się aluzje werbalne i/lub wizualne do nich (np. przywołane uprzednio *Konsumpcja* [mem z Krystyną Pawłowicz]; *Światłość narodu* [mem z Andrzejem Dudą]), a także utworzone od nich derywaty (np. *Szczawian wapnia* [mem ze Stefanem Niesiołowskim]).

Wyrazistość emocjonalna słów kluczy, a nierzadko także ich asocjacje ideologiczne, czyni je jednym z najczęstszych memetycznych środków wyrazu. Ich popularność łączy się ściśle z narastającą polaryzacją sceny politycznej, a także z subiektywizacją, emocjonalizacją oraz kolokwializacją dyskursu publicznego.

Słowniki

ŻMIGRODZKI P., red., 2008: *Wielki słownik języka polskiego* [online: <https://wsjp.pl/>; data dostępu: 29.06.2019].

Źródła

„Asz:dziennik” [online: <https://aszdziennik.pl/>; data dostępu: 26.06.2019].

„Blasty” [online: <https://www.blasty.pl/>; data dostępu: 26.06.2019].

„Chamsko” [online: <https://www.chamsko.pl/>; data dostępu: 26.06.2019].

„Demotywatory” [online: <https://demotywatory.pl/>; data dostępu: 26.06.2019].

DYBAŁA B.: *Kraków. Prezydent Duda w KFC i... pytanie o konstytucję*, 10.06.2018 [online: <https://gazetakrakowska.pl/krakow-prezydent-duda-w-kfc-i-pytanie-o-konstytucje/ar/13246922>; data dostępu: 6.07.2019].

„Dziennik. Rozrywka” [online: <https://rozrywka.dziennik.pl/>; data dostępu: 26.06.2019].

„Fabryka Memów” [online: <https://fabrykamemow/>; data dostępu: 9.07.2019].

<https://www.tvn24.pl/wiadomosci-z-kraju,3/macierewicz-prokurator-klamie-byl-trotyl-dowody-sa-w-polsce,286826.html>; data dostępu: 22.06.2019.

<https://www.wprost.pl/kraj/10163505/andrzej-duda-w-berlinie-mowil-ozarowkach-to-sa-problemy-nad-ktorymi-zastanawiaja-sie-ludzie.html>; data dostępu: 28.06.2019.

Jarosław Kaczyński podczas spotkania wyborczego w Bielsku-Białej, 19.10.2015 [online: <http://bielskobiala.wyborcza.pl/bielskobiala/1,88025,19048886,wybory-parlamentarne-2015-kaczynski-nie-jadam-osmiorniczek.html>; data dostępu: 24.06.2019].

„Jeja” [online: <https://www.jeja.pl/>; data dostępu: 26.06.2019].

Kropka nad i w TVN24, 5.03.2013 [online: <https://www.tvn24.pl/wiadomosci-z-kraju,3/jak-chodzilem-do-szkoly-mysmy-wyjedli-caly-szczaw-z-nasypu,310126.html>; data dostępu: 22.06.2019].

„Kwejk” [online: <https://kwejk.pl/>; data dostępu: 4.07.2019].

Mariusz Błaszczak w wywiadzie dla Tok FM, 30.06.2014 [online: <http://www.tokfm.pl/Tokfm/10,103168,16242286,polowa-dzieci-z-biedy-nie-jedzie-na-wakacje-a-ci-panowie.html?v=1&obx=16242286&order=najstarsze&pId=28381196>; data dostępu: 24.06.2019].

MAKOŚA J.: *Pawłowicz je na sali sejmowej*. Tok.fm, 18.12.2014 [online: http://www.tokfm.pl/Tokfm/1,103454,17152693,Pawlowicz_je_na_sali_sejmowej__Komisja_etyki_ma_dosc_.html; data dostępu: 6.07.2019].

„Mem Generator” [online: <http://pl.memgenerator.pl/>; data dostępu: 20.07.2019].

„Memy” [online: <http://memy.pl/>; data dostępu: 26.06.2019].

„Onet” [online: <https://wiadomosci.onet.pl/>; data dostępu: 22.06.2019].

„Paczaizm” [online: <https://paczaizm.pl/>; data dostępu: 9.07.2019].

„Penera” [online: <https://www.penera.pl/>; data dostępu: 26.07.2019].

REJ M.: *Do tego, co czytał*. W: IDEM: *Figliki*. Kraków 1574 [online: oai:mbc.malopolska.pl:83120; data dostępu: 28.06.2019]

„Ryjbuk” [online: <https://www.facebook.com/ryjbuk/>; data dostępu: 24.06.2019].

Literatura

- BAKALARSKI K., 2018: *Memy internetowe satyrą polityczną społeczeństwa sieci*. „Studia Bobolanum” XXIX, nr 2, s. 163–186.
- KAMIŃSKA M., 2011: *Uwaga, zaraźliwe memy! Perypetie znaku w Internecie*. W: EADEM: *Niecne memy. Dwanaście wykładów o kulturze internetu*. Poznań.
- NIEKREWICZ A., 2014: *Słowo a obraz. Memetyczne sposoby wyrażania poglądów społeczno-politycznych*. W: BIZIOR R., SUSKA D., red.: *Mechanizmy ekonomizacji języka*. Częstochowa, s. 11–24.
- NIEKREWICZ A., 2015: *Od schematyzmu do kreatywności. Język memów internetowych*. Gorzów Wielkopolski.
- NIEKREWICZ A., 2016: *Środki językowo-obrazowe służące etykietowaniu polityków w memach internetowych*. „Poznańskie Studia Polonistyczne. Seria Językoznawcza” XXIII, nr 1, WITASZEK-SAMBORSKA M., red., s. 55–66.
- PORYCKA D., 2016: *Retoryczny wymiar memów politycznych*. „Res Rhetorica”, nr 4, s. 85–90.
- WIERZBICKA A., 2007: *Słowa klucze. Różne języki – różne kultury*. DURAJ-NOWOSIELSKA I. tłum. Warszawa.
- WOŁOZYN M., 2016: *Intertekstualność internetowych memów dotyczących polityki*. „Studia Filologiczne”, XXIX, s. 205–224.
- ZDUNKIEWICZ-JEDYNAK D., 2017: *Mem internetowy jako współczesny popularny gatunek satyry politycznej*. W: KOŁODZIEJEK E., SIDOROWICZ R., red.: *Internet jako przedmiot badań językoznawczych*. Szczecin, s. 134–144.

ARCHIWALIA

Było, (nie) minęło. Przypomnienie tekstu Tony'ego McEnery'ego i Nicka Ostlera *A New Agenda for Corpus Linguistics – Working with All of the World's Languages*







Było, (nie) minęło. Przypomnienie tekstu Tony’ego McEnery’ego i Nicka Ostlera *A New Agenda for Corpus Linguistics – Working with All of the World’s Languages*

It’s (Not) All Just Water under the Bridge. The Reminder of the Text of Tony McEnery and Nick Ostler *A New Agenda for Corpus Linguistics – Working with All of the World’s Languages*

Abstract: In this text I referate and comment the work from two decades ago. I recall an article important for corpus linguistics – the work of Tony McEnery and Nick Ostler from 2000. In their paper they argue that corpus linguistics needs to expand to cover a wider set of languages. The challenge of building a multilingual corps is admittedly current. The development of corpus linguistics still depends on technical capabilities, but today we see significant changes in this field. In the article I not only present the reflections of British linguists, but I also show current and outdated problems.

Key words: digital linguistics, corpus linguistics, Tony McEnery, Nick Ostler, retrospective article

Abstrakt: W tekście referuję i komentuję opracowanie sprzed dwóch dekad. Przypominam w nim artykuł ważny dla językoznawstwa korpusowego – autorstwa Tony’ego McEnery’ego i Nicka Ostlera z 2000 roku. W swoim artykule postulują oni potrzebę wyjścia poza dotychczasowe działania lingwistyki korpusowej na szerszym zestawie języków. Perspektywa czasowa pozwala na obiektywne, zdystansowane komentarze. Wyzwanie związane z budowaniem wielojęzycznego korpusu pozostaje aktualne; owszem, rozwój językoznawstwa korpusowego nadal zależy od możliwości technicznych, jednak dziś widzimy także znaczące zmiany w dziedzinie lingwistyki cyfrowej. W artykule nie tylko prezentuję refleksje brytyjskich lingwistów, ale także pokazuję aktualność/nieaktualność wymienianych przez nich problemów.

Słowa kluczowe: lingwistyka cyfrowa, językoznawstwo korpusowe, Tony McEnery, Nick Ostler, artykuł retrospektywny

Mija właśnie dwadzieścia lat od ukazania się jednego z istotniejszych dla humanistyki cyfrowej tekstu poświęconego gromadzeniu, przetwarzaniu i wykorzystaniu informacji językowej (a pośrednio – językoznawczej) – w 2000 roku Tony McEnery i Nick Ostler, lingwiści z Foundation for Endangered Languages, na łamach prestiżowego czasopisma „Article in Literary and Linguistic Computing” (No 15, s. 403–420) opublikowali artykuł *A New Agenda for Corpus Linguistics – Working with All of the World’s Languages*. Chociaż sam tytuł wspomnianego opracowania zawiera odniesienie do „lingwistyki korpusowej”, celowo wcześniej użyłem określenia „humanistyka cyfrowa”, ponieważ w świetle założeń przedstawionych w tekście brytyjskich językoznawców znajdziemy wiele uwag, które

z jednej strony dotyczą cyfryzacji humanistyki (zob. m.in. BERRY, 2012: 1–20; BOMBA, 2013: 57–71), z drugiej natomiast – wskazują możliwości wykorzystania korpusów językowych w badaniu grup kultur i społeczeństw (mniejszych bądź większych wspólnot komunikatywnych). W samej jedynie perspektywie językoznawczej artykuł trzeba potraktować nie tylko jako manifest lingwistyki korpusowej, ale także jako apel skierowany do środowisk językoznawców o zintensyfikowanie i poszerzenie działań z zakresu lingwistyki korpusowej oraz ukierunkowanie ich na budowanie korpusów dotyczących różnych języków. Słowo *język* wyróżniam, by zaznaczyć, że mam na myśli termin, który w kontekście artykułu ma sens dość szeroki – otóż McEnery’emu i Ostlerowi nie chodzi wyłącznie o język ogólny, ale również o odmiany środowiskowe i funkcjonalne języka.

Autorzy omawianego artykułu, pisząc o potrzebie poszerzenia zasobów korpusowych poszczególnych języków, podkreślają, że o ile przedsięwzięcia takie mają niepodważalnie dużą wartość dla rozwoju nauki oraz posiadają istotne zalety społeczne i historyczne, o tyle największą trudnością w ich realizacji są ograniczenia techniczne. Językoznawcy nie poprzestają jednak na przedstawieniu tej opinii (z dzisiejszej perspektywy nieco anachronicznej, przybierającej charakter truizmu), ale prezentują także potencjalne korzyści dla lingwistyki płynące z gromadzenia danych korpusowych.

Przez długi czas językoznawstwem korpusowym zajmowali się przeważnie badacze brytyjscy i – co oczywiste – skupiali się oni na języku angielskim. Niemniej jednak lingwistyka korpusowa u swoich źródeł ma charakter wielojęzyczny. Otóż jeszcze przed pojawieniem się technologii komputerowej utrwalano ginące języki w postaci baz papierowych, natomiast w drugiej połowie XX wieku, kiedy w nauce wyłoniła się informatyka humanistyczna (mechanolingwistyka), wiele danych językowych zostało przekonwertowanych z wersji papierowej na formę elektroniczną (chodzi tutaj m.in. o „dygitalizację” tekstów łacińskich, aramejskich i nabatejskich, jakiej podjął się Roberto BUSA (1980: 83–90), ale także o prace Alphonse’a JUILLANDA i Eugenia CHANG-RODRIGUEZA nad maszynowym odczytywaniem tekstów chińskich, hiszpańskich, francuskich i rumuńskich (1964; 1965)). Wspomniane tutaj kwestie McEnery i Ostler jedynie sygnalizują, nie poświęcając im większej uwagi prawdopodobnie dlatego, że mają one wyłącznie dawać pewne tło dla dalszych, kluczowych problemów artykułu i niejako przy okazji argumentować sens kontynuacji działań podjętych kilkadziesiąt lat wcześniej (zresztą szeroką prezentację historii badań korpusowych zawiera opublikowana w 1996 r. i wznowiona kilka miesięcy po ukazaniu się tekstu *A New Agenda for Corpus Linguistics* współautorska książka McEnery’ego i Adrew Wilsona *Corpus Linguistics. An Introduction*). Zwracają jednak uwagę na istotną sprawę – według nich dobry punkt oparcia dla rozwoju lingwistyki cyfrowej tworzą osiągnięcia językoznawców amerykańskich, które – choć w dużej mierze zniszczone przez rewolucję metodologiczną towarzyszącą publikacjom prac Noama Chomsky’ego – przetrwały w Europie Północnej, zwłaszcza w Wielkiej Brytanii, dzięki dużemu zainteresowaniu językoznawców przyswajaniem (akwizycją, nauczaniem) języka z wykorzystaniem danych korpusowych.

Rzecz jasna, rzetelność naukowa wymagała od McEnery’ego i Ostlera wskazania nie tylko watorów budowania korpusów językowych, ale także wyzwań związanych z tworzeniem takich zbiorów. Jeśli spojrzymy na postęp, jaki miał miejsce w ostatnich dziesięcioleciach w dziedzinie technologii informacyjno-komputerowej, to jest oczywiste, że w momencie powstawania tekstu badacze mieli powody do utyskiwań na temat technicznych/technolo-

gicznych uwarunkowań projektowanych działań językoznawczych. W zasadzie przeszkody te już nie istnieją. Aktualne pozostają natomiast inne wyzwania, natury merytorycznej, bardzo często sygnalizowane również na gruncie polskiej nauki. Otóż także dziś najważniejszym wyzwaniem dla autorów korpusów językowych pozostaje brak jednolitego sposobu budowania baz, zwłaszcza doboru tekstów i konstruowania parametrów wyszukiwania materiału językowego. Niestety, z reguły korpusy powstają w ramach innych projektów badawczych lub są ukierunkowane na konkretny cel. Tak więc nadal ścierają się teorie dotyczące gromadzenia danych: po pierwsze, dotyczą one wielkości (ilości) materiałów (Czy korpus powinien być „otwarty”, systematycznie uzupełniany, a może należy założyć skończoną liczbę słów i po osiągnięciu zaplanowanej wielkości należy zakończyć dodawanie materiału?); po drugie, wiążą się z jakością i reprezentatywnością danych (Czy korpus ma być zbiorem autentycznej, najbardziej naturalnej komunikacji ludzi, czy też powinien zawierać notacje normatywne i „standardowe”? Czy powinien zawierać dane dotyczące odmian regionalnych, terytorialnych i socjolektalnych? Jeśli tak, to w jakich proporcjach?); po trzecie, odnoszą się do formatu danych, ich przetwarzania i – co za tym idzie – narzędzi ich wyszukiwania i prezentowania (Czy, a jeśli tak, to w jakich zakresach, według jakich kryteriów, możliwe jest maszynowe przetwarzanie, selekcjonowanie danych?).

W kontekście wspomnianych zagadnień wypada osadzić ideę tworzenia korpusu wielojęzycznego. Po upływie dwóch dekad widać w tym zakresie niemały postęp, jeśli chodzi o korpusy dotyczące kilku języków (powstaje coraz więcej korpusów zarówno paralelnych, jak i porównywalnych¹). Niestety, przypadki te w mniejszym stopniu dotyczą polskiego językoznawstwa – na rodzimym gruncie pomysł brytyjskich językoznawców wciąż jest w fazie wstępnej. Inna sprawa, że w skali światowej wspomniany postęp nie jest aż tak znaczący, co zapewne wynika z faktu, że nadal nie istnieje jeden wspólny model budowy korpusów dla poszczególnych języków świata. Co więcej, dalekie od jednolitości bywają korpusy odnoszące się do konkretnych języków, bo też – jak wspomniano – koncepcje budowania baz wciąż podporządkowane są raczej wąskiemu celom, wyznaczanym przez obszar i okres badań. Tymczasem zaproponowany przez McEnergy'ego i Ostlera pomysł opracowania korpusów obejmujących większą liczbę języków świata, opiera się na założeniu, że korpusy powinny być gromadzone w taki sposób, aby umożliwiły przyjęcie zmian metodologicznych w badaniach nad poszczególnymi językami, w tłumaczeniach (zob. m.in. KREDENS, 2005: 270–279; GRABOWSKI, 2011: 89–112; SZELA, 2016: 210–226) i w nauce języków obcych. Owszem, brytyjscy językoznawcy zdają sobie sprawę z faktu, że zawsze będą istnieć dostosowania charakterystyczne dla jakiegoś języka i posługującej się nim grupy społecznej, ale według nich twórcy korpusów powinni zmierzać do minimalizowania takich ograniczeń.

¹ Wprawdzie McEnergy i Ostler w swoim artykule nie wprowadzają rozróżnienia na wspomniane rodzaje korpusów (nie posługują się więc samymi terminami *parallel corpus* i *comparable corpus*), niemniej jednak pierwszy z autorów wespół z innymi specjalistami wyjaśnia koncepcje obu typów korpusów: korpus porównywalny jest dwu- lub wielojęzyczny i składa z tekstów, które nie są własnymi translacjami (teksty są dobrane według jednakowych parametrów stylistycznych, gatunkowych, tematycznych, chronologicznych itd.), z kolei korpus paralelny składa się z tekstów oryginalnych oraz ich przekładów na jeden lub więcej języków. Zob. MCEENERY, XIAO, TONO, 2006.

Widać dzisiaj aktualność także innych utrudnień, o których wspominają McEnery i Ostler. Nadal niewielki procent języków świata to języki urzędowe; tylko w samej Europie mamy sporo takich sytuacji, bo na przykład w Norwegii językiem urzędowym jest język norweski, ale lokalnie także języki lapońskie, w Hiszpanii – w skali kraju język kastylijski, a w skali regionu także kataloński czy baskijski, w Rosji – rosyjski, natomiast lokalnie między innymi ałtajski, czeczeński, czukocki, jakucki, kałmucki, komi czy tatarski². Mało tego, nadal nie jest rozstrzygnięty status wielu dialektów (języków regionalnych), jak na przykład walijski czy kaszubski oraz śląski. Ponadto istnieją języki, którymi posługuje się duża liczba osób, ale języki te nie mają oficjalnego statusu, jak na przykład język Sylheti w Bangladeszu; i odwrotnie: istnieją języki narodowe z niewielką liczbą użytkowników, jak islandzki. Inne wyzwania: Jak potraktować między innymi odmiany języka chińskiego istniejące wyłącznie w wersji mówionej? Jaką pozycję nadać sposobom komunikacji wizualno-przestrzennej, zwłaszcza językom migowym (przecież występują między nimi spore różnice, a szacuje się, że istnieje ich na świecie nawet 300 i nie zawsze posługiwanie się nimi pokrywa się z granicami państw: mamy na przykład PJM [polski język migowy] obowiązujący w Polsce³, ale już BSL [brytyjski język migowy] jest zupełnie różny od ASL [amerykańskiego języka migowego], używanego nie tylko w Stanach Zjednoczonych, ale również w Meksyku czy Kanadzie)? Oczywiście, podniesione tu kwestie mają znaczący wpływ na sposób budowania i wykorzystywania korpusów językowych. Niestety, problemy podjęte przez brytyjskich językoznawców w znacznym zakresie nadal nie zostały rozwiązane.

McEnery i Ostler zwracają również uwagę na ekscerpcję tekstów mających stanowić korpusy językowe. Entuzjastycznie wyrazili się na temat Internetu, wskazując go jako dobre źródło, z którego można czerpać niemal nieograniczoną liczbę egzemplifikacji. Z dzisiejszej perspektywy koncepcja ta wydaje się nie dość przekonująca – współczesność nie pozostawia wątpliwości: tworząc korpusy, nierzadko do zbiorów notacji włącza się wypowiedzi użytkowników Internetu, a przecież dzisiaj w Sieci można znaleźć wszystko i zarazem nic (ogrom form językowych zadziwia i zaskakuje jednocześnie, a sam fakt ich obecności nie może konstytuować faktu językowego i pozwalać na uogólnienia; zresztą, znane są praktyki, że jeśli czegoś w Sieci nie ma, a jest komuś potrzebne na potwierdzenie jakiejś tezy, zawsze można to coś zamieścić), a więc z konstruowaniem korpusu wiązałyby się żmudna praca oddzielania ziarna od plew (zob. uwagi zamieszczone w: LOEWE, 2006: 93–103; ПОДНАЈЕКА, 2006: 338–347; PIOTROWSKI, GRABOWSKI, 2013: 59–71; ZABAWA, 2019: 211–232). Owszem, Sieć oferuje również wiele innej wartości danych – zdigitalizowane teksty literackie, naukowe czy publicystyczne, które mogą być kanwą korpusu językowego, w większym stopniu

² Zob. m.in.: *Urzędowy wykaz nazw państw i terytoriów niesamodzielnych* (Komisja Standaryzacji Nazw Geograficznych poza Granicami Rzeczypospolitej Polskiej, wyd. 4, 2017); TNS OPINION & SOCIAL: *Europeans and their Languages* [online: https://ec.europa.eu/commfrontoffice/publicopinion/archives/ebs/ebs_243_en.pdf; data dostępu: 12.11.2019]; *О языках народов Российской Федерации* (с изменениями и дополнениями) Закон РФ от 25 октября 1991 г. N 1807-1; Zob. FOOTITT, KELLY, 2012.

³ Na marginesie warto dodać, że w rodzimej przestrzeni istnieje już *Korpus Polskiego Języka Migowego* i oparty na nim *Korpusowy słownik polskiego języka migowego* – dzieła opracowane przez Pracownię Lingwistyki Migowej Uniwersytetu Warszawskiego w ramach realizowanego od 2011 roku projektu badawczego „Ikoniczność w gramatyce i leksyce polskiego języka migowego (PJM)” [online: <http://www.slovníkpjm.uw.edu.pl/>; data dostępu: 12.11.2019].

zapewniając zasadę równowagi i reprezentatywności. Dane takie mają także tę zaletę, że procedura ich zbierania jest i tania, i stosunkowo prosta, podobnie jak łatwe jest ich maszynowe odczytywanie.

W założeniu autora szkic niniejszy nie ma być „rozliczeniem” koncepcji McEnergy'ego i Ostlera ogłoszonej przed dwiema dekadami, ale jedynie przypomnieniem ich apelu, na który w wielu aspektach środowiska językoznawców odpowiedziały z entuzjazmem i – co więcej – ideę brytyjskich lingwistów wciąż realizują. Gwoli ścisłości wypada jednak zaznaczyć, że choć w swoim tekście McEnergy i Ostler przywołują przykłady istnienia w świecie zachodnim pewnych podwalin dla realizacji ich pomysłu, w 2000 roku mających już nawet kilkadziesiąt lat, to nie powinno się zapominać, że również na gruncie rodzimego językoznawstwa – choć nieco później i na mniejszą skalę – w drugiej połowie XX wieku były podejmowane działania w zakresie inżynierii językoznawczej (dość wspomnieć tutaj o zapoczątkowanej w latach 70. na Uniwersytecie Warszawskim współpracy informatyków i lingwistów) (zob. np.: LEWANDOWSKA-TOMASZCZYK, 2005; ŚWIDZIŃSKI, 2006: 23–34), a obecnie możemy się poszczycić między innymi wieloma korpusami równoległymi (zob. artykuły w: GRUSZCZYŃSKA, LEŃKO-SZYMAŃSKA, 2016) i narzędziami służącymi automatycznemu przetwarzaniu wielkich korpusów tekstów polskich (SZAFRAN, 1997: 51–64; RUDOLF, 2004; RUDOLF, ŚWIDZIŃSKI, 2006: 31–43). Niemniej jednak do osiągnięcia celu wyznaczonego przez naukowców z Foundation for Endangered Languages, a więc zrealizowania przedstawionej przez nich koncepcji budowania wielojęzycznych korpusów na podstawie jednolitych zasad, jeszcze dość daleko. Dlatego też tekst Ostlera i McEnergy'ego, mimo że ma charakter archiwalny, może i powinien być przypomniany jako zachęta i inspiracja do zintensyfikowania działań w tym zakresie.

Literatura

- BERRY D.M., 2012: *Introduction: Understanding the Digital Humanities*. In: BERRY D., ed.: *Understanding Digital Humanities*. London, s. 1–20.
- BOMBA R., 2013: *Narzędzia cyfrowe jako wyznacznik nowego paradygmatu badań humanistycznych*. W: RADOMSKI A., BOMBA R., red.: *Zwrot cyfrowy w humanistyce*. Lublin, s. 57–71.
- BUSA R., 1980: *The Annals of Humanities Computing. The Index Thomisticus*. „Computers and the Humanities” XIV, nr 2, s. 83–90.
- FOOTITT H., KELLY M., 2012: *Languages at War. Policies and Practices of Language Contacts in Conflict*. London.
- GRABOWSKI Ł., 2011: *Korpusy dwu- i wielojęzyczne w służbie tłumacza, leksykografa i badacza: poszukiwanie ekwiwalentów przekładowych w świetle hipotez dotyczących istnienia uniwersaliów tłumaczeniowych*. W: CHLEBDA W., red.: *Na tropach translatów. W poszukiwaniu odpowiedników przekładowych*. Opole, s. 89–112.
- Ikoniczność w gramatyce i leksyce polskiego języka migowego (PJM)* [online: <http://www.slownikpjm.uw.edu.pl/>; data dostępu: 12.11.2019].
- JUILLAND A., CHANG-RODRIGUEZ E., 1964: *Frequency Dictionary of Spanish Words*. The Hague.
- JUILLAND A., CHANG-RODRIGUEZ E., 1965: *Frequency Dictionary of Rumanian Words*. The Hague.
- KREDENS K., 2005: *Korpusy językowe w językoznawstwie sądowym*. W: LEWANDOWSKA-TOMASZCZYK B., red.: *Podstawy językoznawstwa korpusowego*. Łódź, s. 270–279.

- LEWANDOWSKA-TOMASZCZYK R., red., 2005: *Podstawy językoznawstwa korpusowego*. Łódź.
- LOEWE I., 2006: *Internet i jego zasoby w polskich badaniach lingwistycznych. Rekonesans*. „Biuletyn Polskiego Towarzystwa Językoznawczego” LXII, s. 93–103.
- MCENERY T., WILSON A., 1996: *Corpus Linguistics*. Edinburgh.
- MCENERY T., XIAO R., TONO Y., 2006: *Corpus-Based Language Studies. An Advanced Resource Book*. London–New York 2006.
- PIOTROWSKI T., GRABOWSKI Ł., 2013: *Interpretacja danych frekwencyjnych z korpusów językowych: opis pewnych problemów (na kilku przykładach z życia wziętych)*. W: CHLEBDA W., red.: *Na tropach korpusów. W poszukiwaniu optymalnych zbiorów tekstów*. Opole, s. 59–71.
- PODHAJECKA M., 2006: *Kilka uwag o wykorzystaniu zasobów internetowych do analiz korpusowych języka*. „Język Polski” LXXXVI, z. 5, s. 338–347.
- RUDOLF M., 2004: *Metody automatycznej analizy korpusu tekstów polskich*. Warszawa.
- SZAFRAN K., 1997: *Automatyczne hasłowanie tekstu polskiego*. „Polonica” XVIII, s. 51–64.
- SZELA M., 2016: *O wykorzystaniu Angielsko-Polskiego Korpusu Równoległego Tekstów Prawnych w badaniu cech języka tekstów tłumaczonych*. W: GRUSZCZYŃSKA E., LEŃKO-SZYMAŃSKA A., red.: *Polskojęzyczne korpusy równoległe. Polish-language Parallel Corpora*. Warszawa, s. 210–226.
- ŚWIDZIŃSKI M., 2006: *Lingwistyka korpusowa w Polsce – źródła, stan, perspektywy*. „LingVaria” I, nr 1, s. 23–34.
- ŚWIDZIŃSKI M., RUDOLF M., 2006: *Narzędzia informatyczne obsługi wielkich korpusów tekstów: wyszukiwarka Holmes*. „Biuletyn Polskiego Towarzystwa Językoznawczego” LXI, s. 31–43.
- TNS OPINION & SOCIAL: *Europeans and their Languages* [online: https://ec.europa.eu/commfrontoffice/publicopinion/archives/ebs/ebs_243_en.pdf; data dostępu: 12.11.2019].
- ZABAWA M., 2019: *O roli samodzielnie przygotowanych korpusów w badaniach językoznawczych (na przykładzie korpusu wykorzystującego zasoby internetowe)*. „Półrocznik Językoznawczy Tertium. Tertium Linguistic Journal” IV, nr 1, s. 211–232.

О языках народов Российской Федерации (с изменениями и дополнениями), Закон РФ от 25 октября 1991 г. N 1807-I.

POLEMIKI

Język poprawności politycznej vs. mowę nienawiści w bułgarskich mediach. Андреана Ефтимова: *Двойственият език в медиите: езикът на политическата коректност vs езика на омразата*. (София 2016, 335 с.)







Język poprawności politycznej *vs.* mowy nienawiści w bułgarskich mediach

Андреана Ефтимова: *Двойственият език в медиите: езикът на политическата коректност vs езика на омразата*
(София 2016, 335 с.)

Książka Andreany Eftimovej, bułgarskiej lingwistki i medioznawczyni, poświęcona jest – jak czytamy w tytule – „dwoistości języka mediów” w odniesieniu do dwóch antagonistycznych w swojej naturze zjawisk: poprawności politycznej i mowy nienawiści. Przedmiotem badań *sensu stricto* są procesy eufemizacji i tabuizacji oraz dysfemizacji i detabuizacji, w których autorka rozpoznaje podstawowe językowe mechanizmy służące realizacji tytułowych zjawisk w bułgarskich mediach. Jest to publikacja warta uwagi, dotyczy bowiem językowego wymiaru problemów trapiących aktualnie nie tylko bułgarskie, lecz także polskie życie publiczne¹. Opiera się na nowoczesnym aparacie teoretyczno-metodologicznym i dogłębnych analizach; autorka zaś wykorzystuje solidną bazę materiałową i rzetelnie zebrane dane empiryczne (w tym ankietowe).

Monografia składa się ze wstępu, pięciu rozdziałów, zakończenia, bibliografii oraz aneksu, zawierającego przykłady z analizowanych materiałów prasowych. Całość uzupełniają trzy obszernie wywiady, które Eftimova przeprowadziła z ekspertami w przedmiotowej dziedzinie. Jednym z nich jest wywiad z drem hab., prof. UJ Wojciechem Kajtochem, badaczem języka mediów i komunikacji społecznej, dzięki któremu bułgarski czytelnik dowiaduje się między innymi o poglądach polskiego środowiska lingwistyki medialnej na zagadnienia mowy/języka nienawiści.

Rozdziały I–IV monografii mają charakter teoretyczny. W I rozdziale autorka przedstawia podstawowe, z jej punktu widzenia, założenia aksjologii dziennikarskiej i szerzej – aksjologii lingwistycznej, odwołując się przede wszystkim do prac autorów rosyjskojęzycznych i anglojęzycznych, prezentujących odmienne podejście do wartościowania niż to, przyjęte na przykład w nurcie polskiej aksjolingwistyki czy też aksjologii mediów. Wprowadziła i następnie zoperacjonalizowała pojęcia *funkcjonalnej kategorii semantyczno-stylistycznej*, relewantnej dla *aksjologizacji* i *wartościowania* oraz *aksjologemu*; przyjęła *tolerancję/nietolerancję* za podstawową (uniwersalną) wartość/antywartość w odniesieniu do badanych zjawisk. Podkreślając wpływ czynników kontekstowych na interpretacje konkretnych użyć

¹ Po zabójstwie prezydenta Gdańska Pawła Adamowicza w styczniu 2019 roku „mowa nienawiści” stała się w Polsce przedmiotem nie tylko szerokiej debaty środowiskowej, lecz wręcz apelu o przyjęcie ustawy zapobiegającej mowie nienawiści w przestrzeni publicznej i penalizującej ją, tzw. Lex Adamowicz [online: <http://wiesz.com.pl/2019/01/15/lex-adamowicz-apel-o-nowe-prawo/>; data dostępu 10.09.2019].

językowych, Eftimova buduje swój model badawczy na konkretyzowaniu wartościowania za pomocą takich operatorów oceny, jak: *akceptacja, respekt, lekceważenie, ironia* itp.

W rozdziale II zostały omówione: cechy kategorialne eufemizmów i dysfemizmów, rozmaite sposoby klasyfikacji, ogólne mechanizmy ich tworzenia (zachowanie wspólnego denotatu, różnice w funkcjach pragmatycznych i in.) oraz sposoby identyfikacji. Eufemistyczność bądź dysfemistyczność, jak podkreśla badaczka wielokrotnie, nie są charakterystykami języka jako systemu, lecz generowane są na płaszczyźnie mowy, zależą od czynników pozajęzykowych i stanowią złożony fenomen lingwokulturowy oraz pragmatyczny.

Po dokonaniu wnikliwego przeglądu dostępnej literatury przedmiotu, a także na podstawie własnych obserwacji, w rozdziale III bułgarska lingwistka sytuje zjawiska *język poprawności politycznej* oraz *język nienawiści* na jednej płaszczyźnie i wyróżnia ich wspólne cechy, a mianowicie: 1) zorientowane są na ochronę/naruszenie praw pewnych grup społecznych; 2) realizowane są w treści oraz warstwie ideologicznej przekazu, ale także na płaszczyźnie wyrażania; 3) dotyczą przynależności zbiorowej osoby/jednostki; 4) przejawiają się w mowie, a nie w języku; 5) funkcjonują tylko w sferze publicznej; 6) są wysoce zależne od kontekstu; 7) związane są z naturalnymi procesami zachodzącymi w mowie – z eufemizacją i dysfemizacją; 8) realizują się przeważnie za pomocą substytucji leksykalnej; 9) mają naturę manipulacyjną; 10) stanowią kontrolowane oraz motywowane ideologicznie praktyki o powtarzalnym charakterze; 11) są elementami ukierunkowanej i kontrolowanej polityki językowej. Z kategorycznością stwierdzenia, że *język poprawności politycznej* oraz *język nienawiści* funkcjonują wyłącznie w sferze publicznej, można się z autorką nie zgodzić, ponieważ istnieją dowody na to, że pewne eufemizmy i dysfemizmy przenikają do języka poszczególnych użytkowników (odpowiednio – do dyskursu potocznego i codziennych rozmów).

Akcentując wzajemne powiązania pomiędzy *tabu językowym*, *eufemizmem* i *leksyką poprawnie polityczną*, w rozdziale IV Eftimova rozważa kwestię rekonfiguracji podstawowych pojęć składających się na system wartości Bułgarów. Badaczka stwierdza obecność tendencji do przemiany fundamentalnych wyobrażeń i poglądów grupowych wobec obszarów dotychczas tradycyjnie tabuizowanych (np. stosunków międzypłciowych, części ciała czy zdrowia). Na przykładzie eksperymentu dotyczącego uświadamiania sobie akceptowalności lub nieakceptowalności określonych nominacji, wyekscerpowanych z bułgarskich tabloidów, autorka pokazuje, że częstokroć interpretacje ich znaczeń i konotacji są rozchwiane.

Na rozdział V książki składa się właściwa część badania, w której konkretne eufemizmy/dysfemizmy pochodzące z tekstów medialnych zostały zanalizowane w obrębie następujących klas: 1) profesje i instytucje; 2) choroby i wady; 3) śmierć; 4) płeć (tożsamość płciowa) i stosunki seksualne; 5) zachowania przestępcze; 6) etnonimy; 7) wygląd i wiek; 8) eufemizmy polityczne (*doublespeak* – „język kłamstwa”). Szczególnie interesujące są wywody autorki w zakresie ostatniej z wymienionych klas – ugruntowuje je koncepcja tzw. lingwistyki kłamstwa. Eftimova wyraża opinię i udowadnia na materiale językowym, że eufemizmy polityczne są potężnym narzędziem maskowania rzeczywistości i manipulowania umysłem odbiorcy, których celem jest całkowita zmiana sposobu postrzegania pewnego wydarzenia. Jeden z podrozdziałów tej empirycznej części monografii został poświęcony eufemizmom politycznym opisującym kryzys ukraińsko-rosyjski (2014 r.) – dyskurs medial-

ny w Bułgarii, i nie tylko, dostarczył badaczce wyjątkowo obfitego materiału, który wart byłby porównania z jego odpowiednikiem polskim.

Za wartościowe osiągnięcie Eftimovej należy uznać wyodrębnienie *nowomowy brukselskiej* (autorski termin badaczki) w obrębie *profesjolektu polityków* oraz rzucenie światła na jego cechy. Autorka uważa, że „bombastyczny żargon polityczny” jest częścią schematu, dzięki któremu działania i decyzje polityków trzymane są w tajemnicy przed ich potencjalnymi oponentami/krytykami. Podobnie, *mowa górnolotna*, tzn. biurokratyczna, dąży do przedstawiania pewnych osób i sytuacji jako ważniejszych i bardziej złożonych niż są one w rzeczywistości.

Wracając do głównej kwestii, podejmowanej w książce Eftimovej, tzn. do kwestii eufemizacji i dysfemizacji w bułgarskich tekstach medialnych, należy wyprowadzić wnioski ogólne, który autorka wysnuła na podstawie zaobserwowanego jednoczesnego występowania w tekstach eufemizmów i dysfemizmów. Dzięki tej „dwoistości” w stosunku do komentowanego obiektu (jednoczesne oględne i dosadne nazywanie tej samej osoby lub rzeczy) osiąga się niejednoznaczną, rozchwianą interpretację ze strony odbiorcy, która uzasadnia negatywne i ironiczne oceny. Autorka odnotowuje ponadto wysoką frekwencję eufemizmów w nieformalnych rejestrach mowy medialnej, co istotnie zmienia stan wiedzy bułgarskiej socjolingwistyki na temat eufemizmów, dotąd postrzeganych wyłącznie jako element literackich/książkowych odmian dyskursywnych.

Na koniec wypada zaakcentować, że na gruncie bułgarskim problemy języka poprawności politycznej i mowy nienawiści do tej pory były dyskutowane fragmentarycznie i przeważnie w kontekście nielingwistycznym (etycznym, społeczno-filozoficznym, prawnym i prawniczym)². Monografia Eftimovej jest pierwszym kompleksowym ujęciem tych zagadnień i bez wątpienia stanowi ważną pozycję dla bułgarystów i slawistów, a także dla czytających po bułgarsku lingwistów medialnych i komunikologów.

² Por. С. МИНЕВА, 2018: *Езикът на омразата в контекста на истината и толерантността/политкоректността (етически прочит)*. В: КАРАГЕОРГИЕВА А. и др., ред.: *Език и познание*. София, s. 176–178 i cytowaną tam literaturę.



VARIA

Budowa i zastosowania korpusu monitorującego
MoncoPL

Cyfrowe narzędzia w dydaktyce akademickiej. Bada-
nia diachroniczne polskiej fleksji z użyciem systemu
Chronofleks







Budowa i zastosowania korpusu monitorującego MoncoPL

Design and Applications of MoncoPL as a Monitor Corpus of Polish

Abstract: This paper introduces the methodology of compiling and maintaining MoncoPL, a large monitor corpus of web-based Polish. Furthermore, an overview of the search engine of the same name is provided to show how the size and composition of the corpus, currently reaching over 5.6 billion word tokens, facilitates research on distributional properties of rare words, neologisms and phraseological units. Finally, the article exemplifies some advantages of using a densely-sampled diachronic corpus for the purposes of observing frequency trends and cycles of various constructions in online media discourse.

Key words: MoncoPL, monitor corpus, Polish, diachronic corpora

Abstrakt: Artykuł przedstawia budowę i zastosowania korpusu monitorującego polszczyzny MoncoPL oraz główne funkcje wyszukiwarki o tej samej nazwie. Omówiona zostaje rola referencyjna korpusu oraz jego użycia w identyfikacji neosemantyzmów występujących w komunikacji internetowej. Wielkość indeksu MoncoPL (5,6 mld słów w połowie 2019 r.) pozwala zaobserwować subtelne zjawiska językowe, takie jak łączliwość frazeologiczna rzadkich słów czy wariantywność frazemów. Z kolei wysoka częstość próbkowania danych umożliwia badanie trendów i cykliczności występowania różnorodnych konstrukcji językowych w internetowych rejestrach dyskursu medialnego.

Słowa kluczowe: MoncoPL, korpus monitorujący polszczyzny, korpusy diachroniczne

1. Wstęp

Korpusy językowe stanowią dziś podstawę badań lingwistycznych, leksykograficznych, a także prac informatycznych z dziedziny komputerowego przetwarzania języka naturalnego. Szczególne znaczenie dla empirycznych badań języka mają korpusy określane mianem referencyjnych, takie jak *Narodowy Korpus Języka Polskiego* (NKJP) (PRZEPIÓRKOWSKI i in., 2012). Są to zazwyczaj liczące co najmniej kilkaset milionów słów zbiory wielu odmian, rejestrów i typów funkcjonalnych tekstów, skompilowane według określonych z góry kryteriów z myślą o szerokich zastosowaniach praktycznych i teoretycznych. Korpusy monitorujące można uznać za specjalny typ korpusów referencyjnych. Ich struktura i zawartość mają w założeniu umożliwiać ciągłą obserwację najnowszych zmian i trendów w języku. W odróżnieniu od nieaktualizowanych korpusów diachronicznych, które również składają się z tekstów próbkowanych z podokresów historycznych, a także inaczej niż w przypadku niezrównoważonych diachronicznie korpusów referencyjnych, korpusy monitorujące mają strukturę otwartą i są stale uzupełniane o nowe próbki reprezentowanych w nich odmian

języka. Korpus monitorujący, który przestaje być aktualizowany, pozostaje jedynie zamkniętym korpusem diachronicznym, a jego funkcja referencyjna ulega nieuchronnemu osłabieniu w miarę upływu czasu. Podczas gdy ogólne korpusy referencyjne opisuje się stałą wartością ich wielkości, na przykład liczbą słów, zdań czy też tekstów w nich zawartych, podstawowym parametrem ilościowym korpusu monitorującego jest tempo przyrostu (ang. *rate of flow*, por. SINCLAIR, 1996), mierzone zazwyczaj w segmentach słów przypadających na jednostkę czasu. Do niedawna tempo przyrostu zawartości korpusów monitorujących wynosiło nie więcej niż kilkadziesiąt milionów słów na rok, na przykład popularny korpus monitorujący angielszczyzny amerykańskiej COCA powiększył się w latach 2010–2017 z około 400 do 560 mln segmentów wyrazowych (DAVIES, 2010). W epoce cyfrowej dostępność niektórych typów danych językowych jest tak duża, że współczesne korpusy monitorujące można powiększać w tempie milionów słów dziennie. Przykładowo, korpus monitorujący angielszczyzny globalnej NOW powiększa się o 140–160 mln słów miesięcznie¹.

Prawdopodobnie jedynym publicznie dostępnym, stale aktualizowanym korpusem monitorującym polszczyzny jest obecnie MoncoPL² i to właśnie temu zasobowi oraz narzędziom do jego przeszukiwania dostępnym w serwisie *monco.frazeo.pl* poświęcony jest ten artykuł. Pierwsza część tekstu przedstawia metodologię budowy i aktualizacji MoncoPL, główne funkcje jego wyszukiwarki, w tym składnię zapytań, funkcje agregowania wyników według metadanych oraz moduł ekstrakcji kolokacji. W dalszej części omówione są przykłady zastosowań korpusu monitorującego w badaniu neosemantyzmów oraz identyfikacji kolokacji występujących w internetowych rejestrach polszczyzny³.

2. Indeks MoncoPL

Podstawowym źródłem danych indeksowanych w korpusie MoncoPL są kanały RSS⁴ udostępniane przez ponad 1500 serwisów informacyjnych, branżowych oraz platform blogowych, z których około 600 było aktywnych w lipcu 2019 roku. Poza źródłami dynamicznymi do indeksu MoncoPL zostały jednorazowo włączone niektóre dostępne na otwartej licencji korpusy, z których największym jest *Polski Korpus Sejmovy* (OGRODNICZUK, 2017). Aktualną listę największych źródeł korpusu można znaleźć w widoku *Narzędzia -> Statystyki -> Indeks*. Średni przyrost danych wynosi aktualnie około 1,64 mln słów, czyli około 87 tys. zdań dziennie. Liczba ta waha się z odchyleniem standardowym = 409 tys. słów, w wyraźnych cyklach tygodniowych, wynikających z dużo niższej podaży artykułów w serwisach informacyjnych w soboty i niedziele w porównaniu z pozostałymi dniami tygodnia. Wykres przyrostu danych od 2010 roku w odstępach miesięcznych ukazano na rysunku 1.

¹ Zob. online: <https://www.english-corpora.org/now/>; data dostępu: 15.06.2020.

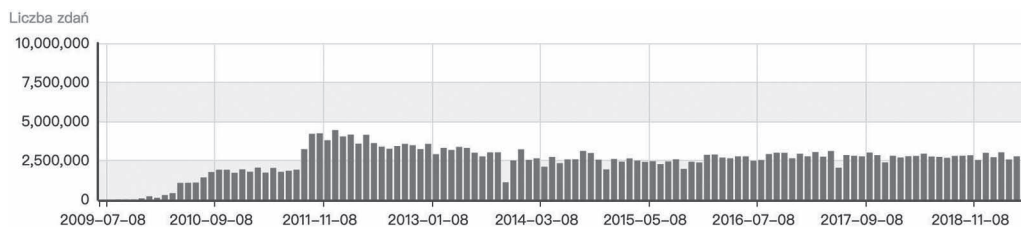
² Ponieważ jedynym narzędziem umożliwiającym korzystanie z korpusu jest dostępna pod adresem *monco.frazeo.pl* wyszukiwarka, nazwa MoncoPL jest używana w tym artykule zamiennie jako określenie witryny wyszukiwarki i samego korpusu.

³ O zastosowaniach MoncoPL w dydaktyce polonistycznej piszą Beata Duda i Karolina Liszczyk (DUDA, LISZCZYK, 2018: 148–149).

⁴ RSS (ang. *RDF Site Summary*) to oparty na języku XML format używany do udostępniania list nagłówków artykułów publikowanych w serwisach WWW.

Poza okresem rozruchu na przełomie lat 2010/2011 i dwoma epizodami awarii systemu zbierania danych indeks MoncoPL powiększał się w tempie około 2,5 mln zdań (47 mln słów) miesięcznie. W połowie 2019 roku korpus osiągnął chwilową wielkość 5,644 mld segmentów wyrazowych w ponad 300 mln jednostek zdaniowych.

Ogromną większość tekstów wchodzących w skład korpusu stanowią artykuły publikowane w internetowych serwisach informacyjnych. Ponieważ za ekstrakcję danych tekstowych ze stron WWW odpowiada generyczny algorytm, część artykułów jest indeksowanych wraz z sekcjami komentarzy, co uwidacznia się w niektórych wynikach. Takie nie do końca zamierzone zróżnicowanie można uznać za pewną zaletę, jako że styl anonimowych komentarzy znacząco się różni od rejestru artykułów, nawet jeżeli są one z nimi powiązane tematycznie. To z kolei pozwala znaleźć w korpusie przykłady form językowych, które w artykułach występują sporadycznie.



Rys. 1. Przyrost danych w indeksie MoncoPL liczony w zdaniach na miesiąc

Równomierne próbkowanie danych z ziarnistością na poziomie jednego dnia otwiera możliwości badań trendów i cykli frekwencyjnych w danych językowych. Również sam fakt zindeksowania niemal 6 mld segmentów wyrazowych z tekstów opublikowanych głównie po roku 2010, a więc po zakończeniu konstrukcji NKJP, nadaje korpusowi MoncoPL cechy uzupełniającego korpusu referencyjnego polszczyzny, choć ograniczonego tylko do kilku typów funkcjonalnych tekstów. Przykłady tych funkcji i zastosowań omówiono w dalszej części artykułu.

3. Funkcje wyszukiwarki MoncoPL

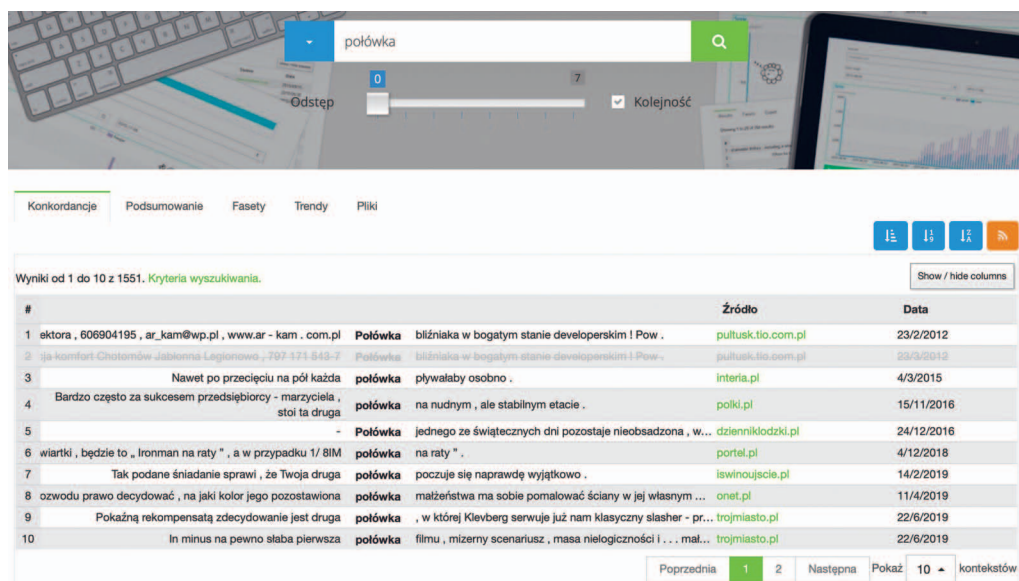
Ze względu na ograniczenia wynikające z praw autorskich, którymi objęte są teksty zindeksowane w MoncoPL, jedynym narzędziem do udostępniania tego korpusu użytkownikom zewnętrznym jest wyszukiwarka dostępna w serwisie *monco.frazeo.pl*. Funkcje tego narzędzia można podzielić na wyszukiwawcze i agregacyjne. Podstawową funkcją wyszukiwawczą jest generowanie konkordancji dla zapytań korpusowo-bibliograficznych. Do funkcji agregacyjnych można zaliczyć tworzenie tzw. faset⁵ dla wyników wyszukiwania oraz oddzielny moduł ekstrakcji kolokacji.

⁵ *Faseta* (od ang. *facet*, a pierwotnie od fr. *facette*) to w kontekście systemów wyszukiwawczych termin oznaczający widok wyników oparty na agregacji wyników po jednej metadanej lub większej ich liczbie. Fasetą może być tabela określająca liczbę wystąpień danej frazy w różnych źródłach lub przedziałach czasowych.

3.1. Konkordancje i składnia zapytań

3.1.1. Słowoformy i dokładne frazy

Aby wyszukać dokładne wystąpienia pojedynczych słowoform lub fraz składających się z dwóch lub więcej słowoform, należy wpisać w polu wyszukiwania dosłowną sekwencję wyszukiwanych słowoform i nacisnąć ikonę lupy. Wielkość liter nie ma znaczenia w zapytaniach. Inaczej niż w przypadku niektórych wyszukiwarek internetowych, fraz nie należy wpisywać w cudzysłowach⁶. Na rysunku 2 ukazano ekran wyników wyszukiwania słowoformy *połówka*. W całym korpusie MoncoPL znaleziono 1551 zdań zawierających ten wyraz.



The screenshot shows a search interface with a search bar containing 'połówka'. Below the search bar, there are navigation tabs: 'Konkordancje', 'Podsumowanie', 'Fasety', 'Trendy', and 'Pliki'. The search results are displayed in a table with the following columns: '#', 'Źródło', and 'Data'. The table contains 10 rows of results, each with a unique identifier, a source URL, and a publication date.

#	Źródło	Data
1	ektora, 606904195, ar_kam@wp.pl, www.ar - kam .com.pl	23/2/2012
2	ga komfort Chotomów Jabłonna Legionowo, 797 171 845-7	23/3/2012
3	Nawet po przecięciu na pół każda	4/3/2015
4	Bardzo często za sukcesem przedsiębiorcy - marzyciela, stoi ta druga	15/11/2016
5	-	24/12/2016
6	wiarki, będzie to „Ironman na raty”, a w przypadku 1/ BIM	4/12/2018
7	Tak podane śniadanie sprawi, że Twoja druga	14/2/2019
8	ozwodu prawo decydować, na jaki kolor jego pozostawiona	11/4/2019
9	Pokażną rekompensatą zdecydowanie jest druga	22/6/2019
10	In minus na pewno stała pierwsza	22/6/2019

Rys. 2. Konkordancje słowoformy *połówka*. Zob. <https://tinyurl.com/ykuorcsu>

Konkordancji na danej stronie może być nieco więcej, niż sugerowałaby to wybrana przez użytkownika wartość opcji limitu (w powyższym przykładzie było to 10 zdań). Dzieje się tak w sytuacji, gdy co najmniej jedno z pobranych z indeksu zdań zawiera więcej niż jeden kontekst pasujący do zapytania. Innymi słowy, gdyby wśród pierwszych dziesięciu zdań znalazło się dokładnie jedno zdanie zawierające podwójne wystąpienie wyrazu *połówka*, to na pierwszym ekranie wyników wyświetlonych byłoby 11 wierszy konkordancji. Poza samym dopasowaniem i jego kontekstem domyślnie wyświetlane są kolumny z nazwą źródła (portalu), z którego pochodzi cytat, oraz datą publikacji tekstu. Nazwa źródła jest jednocześnie bezpośrednim odnośnikiem do oryginalnej strony z pełną wersją tekstu. W niektórych przypadkach wskazywana strona może być w chwili wyświetlenia konkordancji niedostępna. Warto pamiętać, że możliwe jest wyświetlenie dodatkowych lub ukrycie aktualnie wyświetlanych kolumn z metadanymi. Do konfiguracji wyświetlanych kolumn służy przycisk znajdujący się nad tabelą konkordancji. Aktualnie, na jednej

⁶ Wyniki przykładu zapytania o dosłowne wystąpienie frazy *od wielkiego dzwonu* można zobaczyć pod adresem online: <https://tinyurl.com/y3ypz7u5>; data dostępu: 17.04.2020.

stronie można wyświetlić do 1 tys. wyników. Przechodzenie między kolejnymi stronami wyników umożliwiającą kontrolki u dołu tabeli. Wykryte w bieżących wynikach duplikaty kontekstów są przekreślane i wyszarzane, co pozwala na ich szybką identyfikację. Po dwukrotnym kliknięciu na wiersz z wynikiem ukazuje się nieco większy kontekst wystąpienia wyszukiwanego wyrazu.

Nieco mniej intuicyjne może się wydawać wyszukiwanie słowoform, które są zapisywane w tekstach łącznie, a następnie dzielone przez analizator morfologiczny Morfeusz (WOLIŃSKI, 2014) na osobne segmenty. Przykładem może być wyszukiwanie niektórych form czasownika *być*, na przykład formy *byłem*, która jest dzielona w indeksie na dwa segmenty *był* i *-em*. Zapytanie o tę formę powinno mieć postać *był em*⁷.

3.1.2. Rozszerzenia fleksyjne i ortograficzne

Niezwyczajnie istotnym elementem składni wyszukiwarki dla korpusu polszczyzny jako języka z bogatą fleksją jest możliwość wyszukania wariantów fleksyjnych zadanej formy podstawowej (tzw. lematu). Dwa przykłady takich zapytań dla słów i prostych fraz przedstawiono poniżej:

*ściemniać***
*Wielki** Brytania**⁸*

Pierwsze z tych zapytań zwraca konteksty zawierające różne formy czasownika *ściemniać*, a drugie – różne warianty frazy *Wielka Brytania*. Warto zauważyć, że w drugim wypadku formą podstawową rozszerzanego fleksyjnie terminu przymiotnika jest forma rodzaju męskiego, tj. *wielki*, mimo iż jej dosłowne wystąpienie nie pokrywa się z intencją zapytania, ze względu na stałe uzgodnienie rodzaju przymiotnika *wielki* z głową⁹ frazy *Wielka Brytania*.

Wyniki uzyskane za pomocą automatycznego rozszerzenia fleksyjnego mogą nie być pełne, jako że nie wszystkie formy podstawowe zostały poprawnie rozpoznane w indeksie. Istnieją dwa sposoby zwiększenia pokrycia form fleksyjnych w zapytaniu. Po pierwsze, możliwe jest wymienienie wszystkich wariantów fleksyjnych, które mają być uwzględnione na danej pozycji, na przykład:

fejs|fejssem|fejsa|fejsie

Po drugie, możliwe jest zastosowanie rozszerzenia ortograficznego, które jednak może dać nadmiarowe wyniki, na przykład:

*fejs.**

⁷ Zob. online: <https://tinyurl.com/yjyhbl9b>; data dostępu: 24.06.2020.

⁸ Zob. online: <https://tinyurl.com/y2skutxc>; data dostępu: 24.06.2020.

⁹ Termin *głowa* jest używany tu jako odpowiednik ang. terminu *head* (of a phrase).

Możliwe jest ograniczenie generowanych w ten sposób wariantów na danej pozycji za pomocą operatora negacji, na przykład:

fejs.|!fejsbukow.**¹⁰

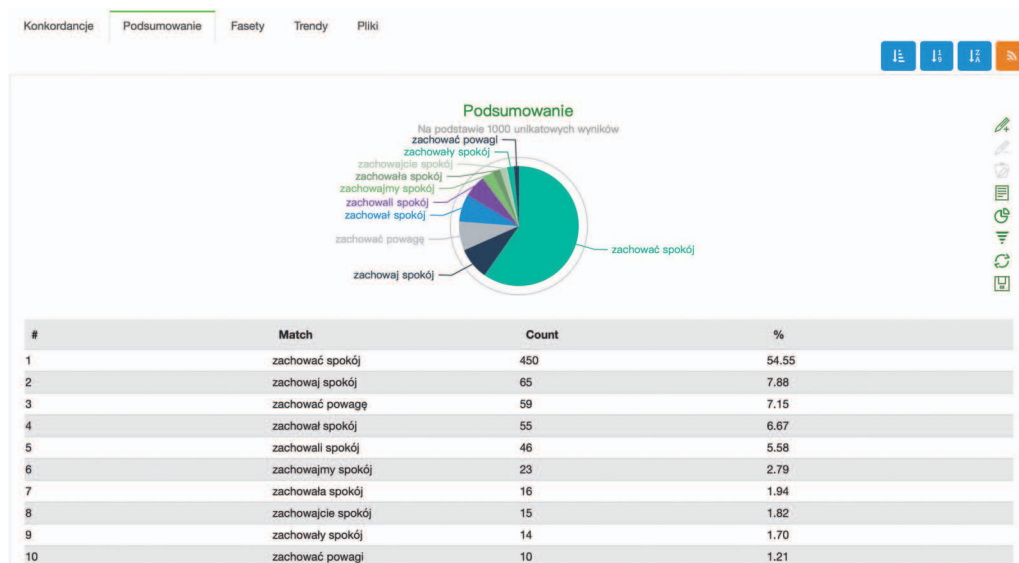
To ostatnie zapytanie zwróci różne formy wyrazowe rozpoczynające się od łańcucha znaków *fejs-* z wyjątkiem form, które zaczynają się od łańcucha *fejsbukow-*. Takie ograniczenie wykluczy więc z wyników wystąpienia przymiotników, co może być efektem pożądanym, jeżeli wyszukujemy jedynie form rzeczownikowych.

3.2. Warianty

Operator wariantu | został już częściowo wprowadzony powyżej. Warto pamiętać, że można go używać również do rozszerzeń leksykalnych. Na przykład, aby znaleźć jednocześnie wystąpienia fraz *zachować spokój* lub *zachować powagę*, można użyć następującego zapytania:

*zachować** spokój**|powaga***

W wynikach zapytań z wariantami lub operatorami rozszerzeń przydatna może być zakładka *Podsumowanie*, w której na podstawie próby o wielkości maksymalnej 1 tys. wystąpień tworzona jest lista frekwencyjna form pasujących do zapytania. Rysunek 3 ukazuje podsumowanie dopasowań do powyższego zapytania, z którego wynika, iż w losowym zbiorze 1 tys. wystąpień najczęściej pojawiała się forma *zachować spokój*, co można wytłumaczyć jej zagnieżdżeniem w dłuższych frazach z czasownikiem modalnym, wyrażającym potrzebę zachowania spokoju, na przykład *należy, powinno się, trzeba + zachować spokój*.



Rys. 3. Widok zagregowanej konkordancji

¹⁰ Zob. online: <https://tinyurl.com/yylpyf9h>; data dostępu: 12.07.2020.

3.3. Kolejność i odstęp między terminami zapytania

Domyślnie wyszukiwarka MoncoPL próbuje znaleźć dopasowania terminów występujących w indeksie korpusu w takiej kolejności, w jakiej zostały one zdefiniowane w zapytaniu. Aby wyłączyć ten warunek, należy rozwinąć panel opcji i odznaczyć w nim pole *Zachowaj szyk*. Dla przykładowego zapytania:

*kłamać** bezczelnie*

zostaną wtedy zwrócone zarówno wystąpienia frazy *kłamać bezczelnie* oraz jej wariantu pozycyjnego *bezczelnie kłamać*.

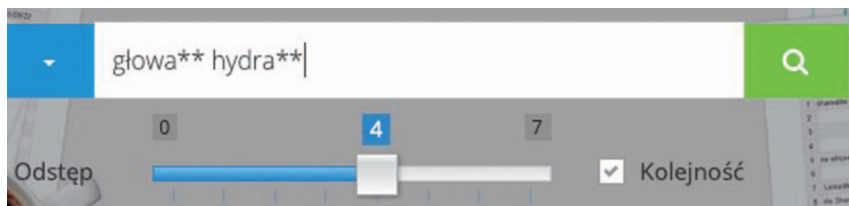
Składnia MoncoPL umożliwia stosunkowo wygodne wyszukiwanie różnego rodzaju związków wyrazowych, na przykład idiomów, formuł konwersacyjnych, kolokacji, a także nieutralonych kombinacji wielowyrazowych. Czasem trudne jest określenie z góry dokładnego zbioru, kolejności, a nawet relacji gramatycznych dla wyrazów tworzących dany związek lub frazę. Formułując zapytanie o frazy, możemy zwiększyć pokrycie zwracanego zbioru wyników poprzez odpowiednie ustawienie parametru odstępu. Domyślnie jego wartość wynosi 0, co oznacza, że między kolejnymi pozycjami dopasowania wyrażanymi terminami zapytania nie może wystąpić żaden wyraz. Dla przykładu, zapytanie:

*głowa** hydra***

z domyślnym ustawieniem odstępu zwróci jedynie konteksty, w których te terminy zapytania występują bezpośrednio obok siebie. Wyrazy te jednak tworzą często figuratywne kolokacje i warianty idiomu opartego na metaforze złożonego, powracającego problemu jako mitycznej Hydry, której głowy odrastają po odcięciu. Po zwiększeniu wartości tego parametru do 4 wśród wyników znajdują się między innymi takie dopasowania:

hydra, która podnosi głowę
hydra, która podnosi głowę
hydra nie ma akurat tyłu głów
Hydrze administracji odrastają liczne głowy

Jeszcze większe pokrycie w tym przypadku można uzyskać poprzez odznaczenie opcji *Zachowaj szyk*, jak to ukazano na rysunku 4¹¹.



Rys. 4. Zwiększenie pokrycia zapytania o frazy poprzez jednoczesne rozluźnienie parametrów kontekstu i szyku terminów

¹¹ Zob. online: <https://tinyurl.com/y2r45xc7>; data dostępu: 12.08.2020.

Poza podanymi uprzednio kontekstami w wynikach dla tego zapytania znajdują się również takie dopasowania, w których wyraz *głowa* występuje przed wyrazem *hydra*, na przykład:

głowa komunistycznej hydry
głów biurokratycznej hydry
głowę wymyślonej przez siebie faszystowskiej hydrze
głowa ohydnej hydry

3.4. Zapytania morfosyntaktyczne

Składnia MoncoPL umożliwia definiowanie sekwencyjnych wzorców leksykalno-gramatycznych. Kategorie morfosyntaktyczne można definiować, używając specjalnej składni: `<tag=XX>`, gdzie *XX* to znacznik określający część mowy i inne kategorie gramatyczne słowa zgodnie z tagsetem NKJP¹². Załóżmy, że interesują nas przymiotniki występujące przed różnymi formami rzeczownika *wiara*. Poniższe zapytanie wymusza dopasowanie wystąpienia przed dowolną formą tego słowa segmentu, który został oznakowany jako przymiotnik:

`<tag=ADJ.*> wiara**13`

Przykładowe wyniki tego zapytania ukazuje tabela 1. Pierwszy z wyników pokazuje również, że zapytania morfosyntaktyczne mają charakter czysto pozycyjny i nie uwzględniają relacji składniowych między terminami. Na przykład słowoforma przymiotnikowa *anglosaskich* modyfikuje rzeczownik *krajach* w tym zdaniu, podczas gdy powyższe zapytanie miało raczej na celu znalezienie przymiotników modyfikujących rzeczownik *wiara*.

Tabela 1
Przykładowe wyniki zapytania `<tag=ADJ.*> wiara**`

Lp.	Lewy kontekst	Dopasowanie	Prawy kontekst	Źródło
1	W krajach	anglosaskich wiarę	w Jedi zadeklarowało około 500 tysięcy osób...	gazeta.pl
2	... wyraźnie pobrzmiwa tu	archaiczna wiara	żydów, że wszystko, co dobre może...	onet.pl
3	Czym jest	autentyczna wiara	?	katolik.pl
4	Wtedy z	autentyczną wiarą	zaczęliśmy mecz, ale w miarę upływu...	weszlo.com
5	Czy można budować	autentyczną wiarę	na wzruszających piosenkach o Jezusie...	denon.pl

Składnia wyszukiwarki jest kompozycyjna, co oznacza, że jej elementy można łączyć w celu zmaksymalizowania pokrycia lub precyzji zapytania. Na przykład, po wyłączeniu opcji zachowania szyku wyrazów, zapytanie:

`<tag=adv.*> ściemniać**`

¹² Zob. online: <http://nkjp.pl/poliqarp/help/ense2.html>; data dostępu: 12.08.2020.

¹³ Zob. online: <https://tinyurl.com/y3lnqer7>; data dostępu: 12.08.2020.

zwróci zdania zawierające kombinacje przysłówków i dowolnej formy czasownika *ściemniać* w dowolnej kolejności tych terminów. Tabela 2 przedstawia przykłady zapytań o wyrazy, proste frazy i wzorce leksykalno-gramatyczne:

Tabela 2
Przykłady zapytań wykorzystujących różne aspekty składni wyszukiwarki MoncoPL

#	Zapytanie	Odstęp	Szyk	Uwagi
1	<i>brać** jak leci</i>	2	tak	Terminy mogą występować w odległości do 2 tokenów od siebie, np. 'brali wszystko jak leci'.
2	<i>stąpać** chodzić** po <tag=adj.*> <tag=subst.*></i>	2	tak	Sekwencja <i>stąpać + po + przym. + rzeczownik</i> . Do 2 nieokreślonych segmentów między terminami.
3	<i>koń** zqb** patrzeć** zaglądać**</i>	4	nie	Różne warianty idiomu.
4	<i><tag=fin.*> <tag=infin.*> <tag=praet.*> przykład**</i>	2	tak	Wybrane formy czasownikowe (zob. tagset NKJP), po których występuje dowolna forma rzeczownika 'przykład'.
5	<i>brzęk** <tag=.*gen.*></i>	2	tak	Dowolna forma wyrazu 'brzęk', po której występuje dowolny wyraz w dopełniaczu.

3.5. Sortowanie wyników

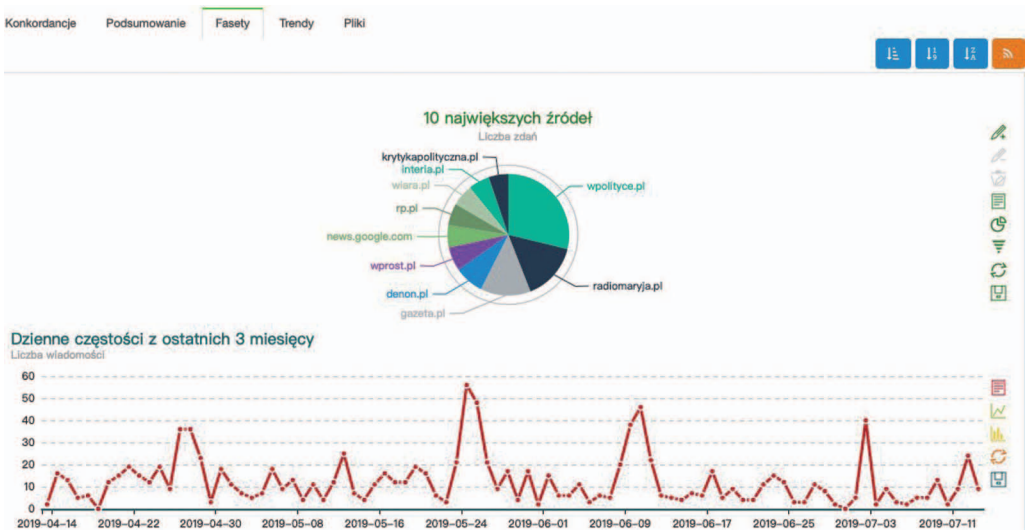
Wyszukiwarka obsługuje dwie niezależne metody sortowania wyników. *Sortowanie głębokie* to sortowanie wszystkich zindeksowanych zdań, które pasują do zapytania według metadanych takich, jak data publikacji i nazwa źródła. Opcja *Sortowanie konkordancji* pozwala z kolei posortować zbiór wydobytych w danym żądaniu konkordancji według pasujących fragmentów zdań lub ich bezpośrednich kontekstów. Obie opcje sortowania, a także opcje filtrowania wyników według źródeł tekstów dostępne są w rozwijanym pod polem zapytania panelu opcji zaawansowanych. Rysunek 5 ukazuje, w jaki sposób można ograniczyć wyniki wyszukiwania form rzeczownika *demokracja* do jednego źródła, tj. *Polskiego Korpusu Sejmowego*, i jednocześnie posortować je rosnąco według daty publikacji.

Rys. 5. Zaawansowane opcje wyszukiwania i sortowania

Nazwy źródeł są podpowiadane po wpisaniu dwóch pierwszych liter w polu *Źródła*. W formularzu można też ograniczyć zakres wyszukiwania do konkretnego przedziału czasowego.

3.6. Fasety

Dla każdego zapytania korpusowego obliczane są całkowite częstości dopasowań w różnych kategoriach metadanych. Aktualnie są to źródła oraz przedziały czasowe, w których znaleziono pasujące zdania. Wykresy z faset są dostępne w zakładce *Fasety* w widoku wyników wyszukiwania. Rysunek 6 przedstawia wykresy 10 najczęstszych źródeł dla zapytania *gender**, które zwraca wystąpienia rzeczownika *gender* oraz jego derywaty, na przykład *genderowy*, *genderyzm* itp. Najczęściej pojawiającymi się źródłami w próbie ponad 33 tys. wystąpień tego typu form okazują się portale *wpolityce.pl* oraz *radiomaryja.pl*.



Rys. 6. Największe źródła i częstości z 3 miesięcy dla zapytania *gender**

3.7. Eksportowanie wyników

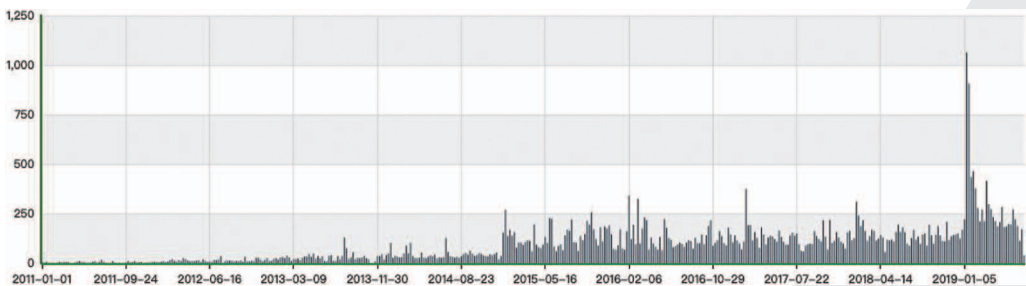
Aby umożliwić użytkownikom bardziej zaawansowaną analizę danych korpusowych, wyszukiwarka MoncoPL pozwala na pobranie w zakładce *Pliki* arkusza kalkulacyjnego zawierającego do 10 tys. wierszy konkordancji dla danego zapytania. W arkuszach wynikowych poza konkordancjami z pełnymi metadanymi znajdują się tabele z fasetami oraz dokładne informacje o rozmiarze i wersji korpusu użytego w chwili wysłania zapytania.

4. Funkcja referencyjna MoncoPL

Można zaryzykować tezę, że po zakończeniu konstrukcji NKJP w 2011 roku, MoncoPL pełni w ograniczonym stopniu funkcję dużego korpusu referencyjnego, przynajmniej w badaniach zmian językowych występujących w polszczyźnie. Zilustrujmy to prostym przykładem neosemantyzmu, jakim jest (a raczej jakiś czas temu był) czasownik *ściemniać* w znaczeniu *kłamać/kręcić* oraz jego rzeczownikowy derywat *ściema*. Dzięki dostępności w metada-

nych NKJP daty publikacji każdego tekstu możliwe jest znalezienie najwcześniejszych poświadczeń wystąpienia danego słowa lub frazy w tymże korpusie oraz sprawdzenia zmian w częstości ich występowania na przestrzeni kilkudziesięciu lat. Rzeczownik *ściema* po raz pierwszy pojawił się w danych NKJP w tekście opublikowanym w listopadzie 1998 roku na łamach „Zielonych Brygad” jako głowa dłuższej frazy rzeczownikowej *totalna ściema słupskich policjantów*. W latach 1998–2010 obserwujemy od kilku do kilkudziesięciu wystąpień tego rzeczownika rocznie. Z kolei pierwsze wystąpienie czasownika *ściemniać* w znaczeniu *kłamać/kręcić* pojawia się w NKJP w powieści beletrystycznej z 1999 roku. Można więc przypuszczać, że nowe znaczenie czasownika *ściemniać* pojawiło się na szerszą skalę najpóźniej w drugiej połowie lat 90. XX wieku. Ponieważ jednak ostatnie próbki języka włączone do NKJP pochodzą z roku 2011, w Korpusie Narodowym nie można już zaobserwować dalszych tendencji ani ilościowych, ani też jakościowych w użyciu tych dwóch słów. Z kolei w danych MoncoPL, częściowo prawdopodobnie ze względu na wysoki udział polemicznych i mocno nieformalnych komentarzy pod artykułami w ogólnej puli danych, znajdujemy ponad 10 tys. wystąpień samego rzeczownika *ściema*. Należy zaznaczyć, że sporą zaletą zrównoważonej kompozycji NKJP w tego typu badaniach jest możliwość sprawdzania stopnia przyjęcia neosemantyzmów i neologizmów we wtórnych rejestrach i typach funkcjonalnych tekstów, czyli na przykład ich poświadczeń w publicystyce i w rejestrach literackich, jeżeli ich pierwotnym rejestrem był język mówiony. Możliwości MoncoPL są w tym zakresie bardziej ograniczone.

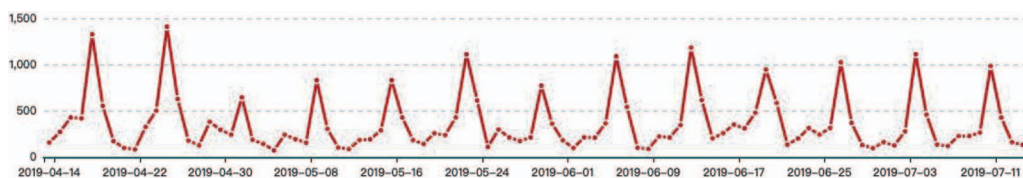
Jeszcze bardziej jaskrawym przykładem luki, jaką wypełnia MoncoPL po zamrożeniu puli tekstów NKJP, są innowacje leksykalne, które w korpusach polszczyzny powstałych przed rokiem 2011 nie były odnotowywane w ogóle lub były bardzo rzadko. Przykładem niech będzie cała rodzina wyrazów i frazów zbudowanych na bazie morfemu *hejt*, na przykład *hejt*, *hejtować*, *hejter*, *hejterski* itp. W pełnej puli danych NKJP wyrazy takie występują sporadycznie. Rysunek 7 ukazuje stopniowy przyrost częstości ich występowania po roku 2011 aż do połowy roku 2019 na podstawie 40 tys. wystąpień w korpusie MoncoPL. Ogromną popularność wyrazu *hejt* i jego derywatów można tłumaczyć faktem, iż stanowią one swoiste autodeskryptory dyskursu internetowego znamionujące agresję słowną jako jeden z wyróżników języka internautów. Świadczą o tym typowe modyfikatory rzeczownika *hejt* zidentyfikowane za pomocą opisanego w dalszej części artykułu modułu ekstrakcji kolokacji, wśród których znajdziemy przymiotniki takie, jak *internetowy* (363 wystąpienia) *wszechobecny* (26), *zmasowany* (23), *obrzydliwy* (17), *prymitywny* (14), *pisowski* (14) oraz *lewacki* (8). Kolokacje te świadczą o tym, że *hejt* ma jednoznacznie negatywny wydźwięk i jest przypisywany różnym stronom sporów politycznych i ideologicznych.



Rys. 7. Częstości tygodniowe występowania wyrazów zaczynających się od członu *hejt-* w latach 2011–2019 w MoncoPL

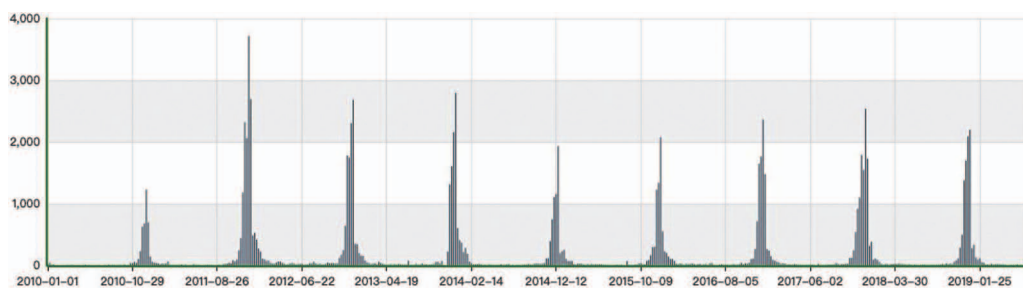
5. Cykle frekwencyjne

Jedną z unikatowych zalet dużego korpusu monitorującego z dzienną ziarnistością próbkowania danych jest możliwość badania cykli frekwencyjnych, w których zazwyczaj występują jednostki leksykalne. Jest to być może mało zbadany, ale bardzo interesujący aspekt użycia słów i skonwencjonalizowanych wyrażen wielowyrazowych. Przykładem bardzo przewidywalnego i ustalonego cyklu frekwencyjnego jest występowanie nazw dni tygodnia w prasie codziennej i internetowych tekstach informacyjnych. Dla przykładu, rysunek 8 przedstawia częstości dzienne występowania rzeczownika *czwartek* w okresie od kwietnia do lipca 2019 roku. Regularne maksima lokalne tego szeregu czasowego można wyjaśnić tendencją do opisywania daty wydarzenia w prasie codziennej za pomocą nazwy dnia tygodnia, która jest określeniem mniej względnym niż na przykład wyrazy *dziś*, *wczoraj* lub *jutro*.



Rys. 8. Częstości dzienne występowania rzeczownika *czwartek*

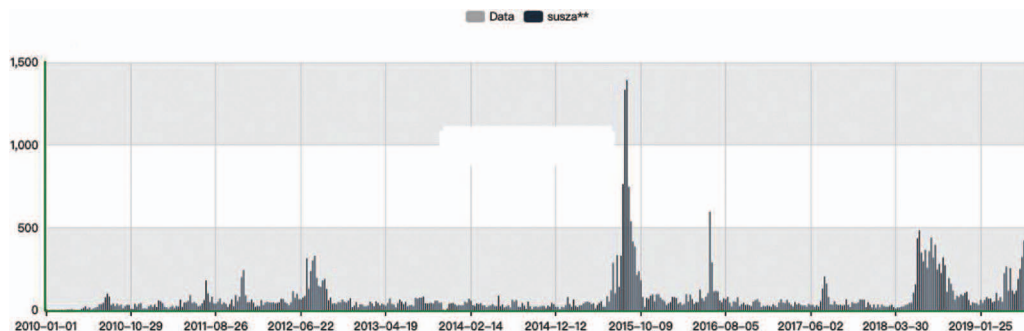
Cykle frekwencyjne wielu słów mają mniej precyzyjnie określone, ale równie niezawodnie występujące maksima. Przykładem może być słowo *choinka*, które w latach 2010–2019 osiągało maksimum częstości w okolicach świąt Bożego Narodzenia, utrzymując wyższą od przeciętnej częstość przez kilka dni. Widać to wyraźnie na rysunku 9, który został wygenerowany z konkordancji tego rzeczownika za pomocą funkcji *Trendy* wyszukiwarki MoncoPL.



Rys. 9. Częstości tygodniowe rzeczownika *choinka* w latach 2010–2019

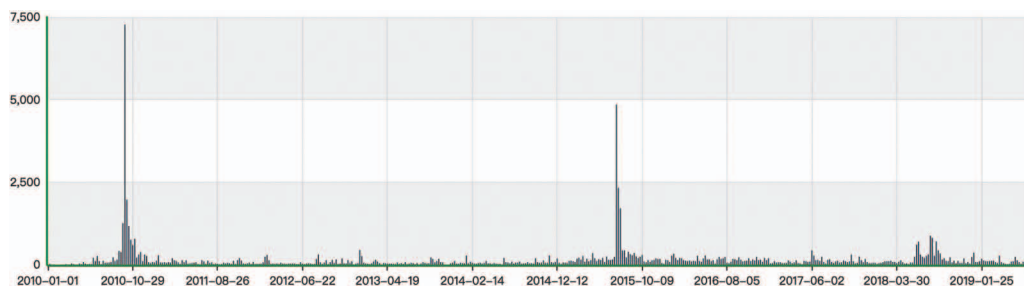
Terminy oznaczające zjawiska pogodowe lub astronomiczne mogą mieć cykle bardzo regularne z punktowym maksimum (np. *Perseidy*), ale mogą też występować z mniej przewidywalną częstością. Na przykład największe względne częstości występowania rzeczownika *susza* obserwuje się w miesiącach letnich w latach, w których stwierdza się wystąpienie suszy rolniczej na obszarze Polski. Co ciekawe, wykres z rysunku 10 w zasadzie pokrywa się z wystąpieniami tego zjawiska w ostatnich latach, a zwłaszcza z rekordową suszą w roku 2015 i jej brakiem w roku 2017¹⁴.

¹⁴ Zob. online: <http://www.susza.iung.pulawy.pl>; data dostępu: 14.07.2020.

Rys. 10. Częstotliwości tygodniowe występowania rzeczownika *susza* w latach 2010–2019

Chociaż ocena istotności cykliczności jako elementu opisu leksykograficznego nie jest tematem tego artykułu, to jednak warto zaznaczyć, że MoncoPL daje pewne możliwości badań cykli frekwencyjnych na przykład rzeczowników pospolitych w polszczyźnie.

Szeregi czasowe generowane z korpusu MoncoPL pozwalają również zweryfikować pewne intuicje z pogranicza językoznawstwa korpusowego i badań medioznawczych. Na przykład rysunek 11 przedstawia szereg czasowy częstotliwości tygodniowych występowania rzeczownika *dopalacz*. Z szeregu tego wynika, że w polskim dyskursie medialnym wyróżniamy dwa lub trzy epizody szczególnego zainteresowania tematyką *dopalaczy* w ciągu ostatniej dekady, co może stanowić punkt wyjścia do dalszej analizy korpusowo-medioznawczej.

Rys. 11. Częstotliwości tygodniowe występowania rzeczownika *dopalacz* w latach 2010–2019

6. Ekstrakcja kolokacji

Korpusowe narzędzia do ekstrakcji kolokacji, idiomów i różnych innych typów frazemów znajdują szerokie zastosowania w pracach leksykograficznych i językoznawczych. W serwisie MoncoPL moduł ekstrakcji kolokacji można znaleźć w zakładce *Narzędzia* -> *Kolokacje*. Wykorzystuje on pozycyjny mechanizm identyfikacji kolokacji z możliwością użycia filtrów morfosyntaktycznych. Formularz ekstrakcji kolokacji przedstawia rysunek 12. W głównym polu wyszukiwania należy wpisać zapytanie korpusowe definiujące tzw. ośrodek kolokacji w składni opisanej powyżej. W omawianym przykładzie jest to zapytanie o różne formy rzeczownika *hejt*. Parametr *Próbka* określa maksymalną liczbę zdań, z których wyszukiwarka ma ekstrahować potencjalne kolokaty (czyli regularnie współwystępujące wyrazy) dla

zadanego ośrodka kolokacji. Ograniczenie to wynika z ogromnej liczby danych, jakie musiałyby być przetwarzane dla niektórych ośrodków kolokacji w czasie rzeczywistym. Limit można zwiększać do 50 tys. wystąpień, ale wiąże się to z wydłużonym czasem oczekiwania na wynik zapytania. Dodatkowo, podczas ekstrakcji kolokacji identyfikowane i eliminowane są prawdopodobne duplikaty, czyli identyczne lub niemal identyczne zdania o zadanej minimalnej długości. Po zakończeniu ekstrakcji wyszukiwarka podaje liczbę unikatowych kontekstów (zdań), których użyto do wygenerowania listy kolokacji.

W polu *Części mowy* można określić klasy części mowy kolokatów, które mają być brane pod uwagę w procesie ekstrakcji. Z kolei w polu *Pozycje kolokatów* określamy maksymalną odległość pozycyjną potencjalnych kolokatów od ośrodka kolokacji. W przykładzie z rysunku 12 pod uwagę brane są tylko przymiotniki występujące na pozycjach od -3 do +1 w odniesieniu do pozycji rzeczownika *hejt*. Tego rodzaju ograniczenia zwiększają prawdopodobieństwo wystąpienia relacji składniowych między ośrodkiem kolokacji i współwystępującymi z nim wyrazami. W przykładowym zapytaniu zwiększamy szansę na zaobserwowanie modyfikatorów przymiotnikowych rzeczownika *hejt*.

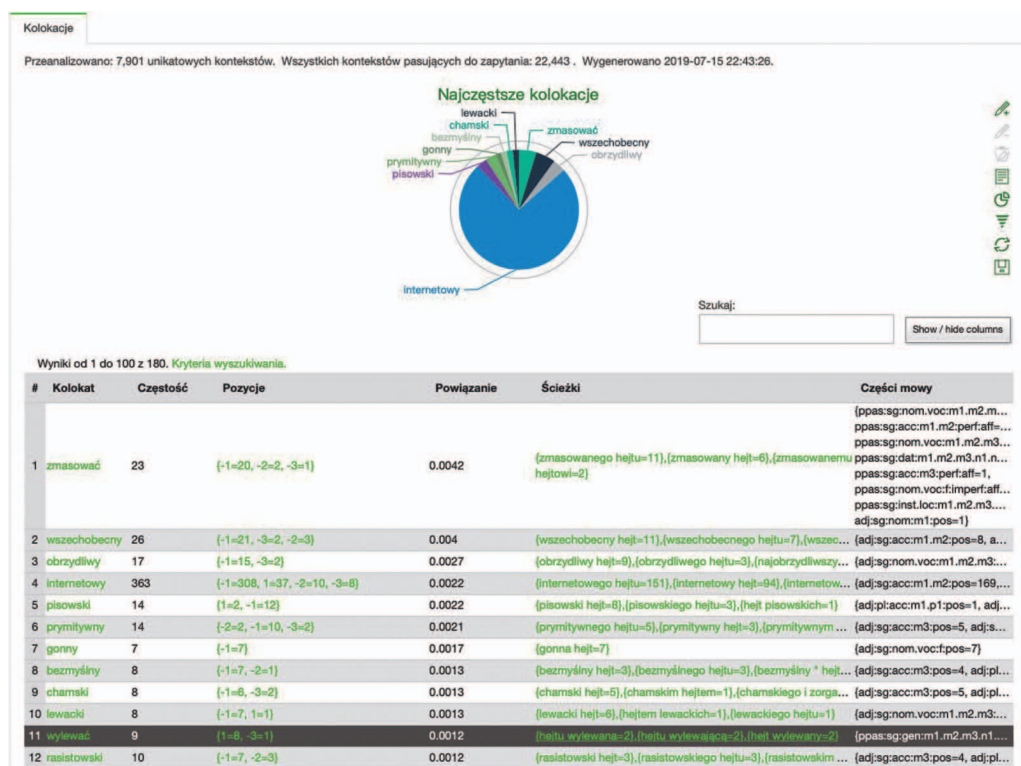
The image shows a web-based form for configuring a search query. At the top, a search bar contains the text "hejt|hejtem|hejtu|hejtowi|hejty|hejtom|hejcie" and a magnifying glass icon. Below the search bar, there are several control elements: a slider for "Odstęp" (Offset) with values 0 and 7, and a checked checkbox for "Kolejność" (Order). The main form area includes:

- "Odstęp" (Offset) input field with the value 0.
- A checked checkbox labeled "Zachowaj kolejność" (Maintain order).
- "Próbka" (Sample) input field with the value 10000.
- "Części mowy" (Parts of speech) section with a dropdown menu showing "przymiotnik" (adjective) and a "Wprowadź znacznik" (Add marker) button.
- "Pozycje kolokatów" (Collocation positions) section with a dropdown menu showing values -3, -2, -1, and 1.
- "Min. częstość" (Min. frequency) input field with the value 2.
- "Zapytanie Dismax" (Dismax query) input field.
- "Źródła" (Sources) dropdown menu.
- "Zakres dat" (Date range) section with a date range selector and a "ON" button.
- An "Odśwież cache" (Refresh cache) checkbox at the bottom.

Rys. 12. Formularz modułu ekstrakcji kolokacji

Opcja minimalnej częstości pozwala odciąć okazjonalne współwystąpienia. W polu *Źródła* można określić nazwy źródeł, z których mają być ekstrahowane kolokacje, dzięki czemu możliwe jest identyfikowanie nacechowania pewnych słów i terminów w niektórych serwisach. Dzięki tej opcji dowiadujemy się na przykład, że wśród kolokatów przymiotnikowych rzeczownika *ideologia* występujących w portalu *gazeta.pl* stosunkowo często występują przymiotniki *faszystowska*, *zbrodnicza* (głównie w odniesieniu do nazizmu), *nazistowska* czy *neonazistowska*. Z kolei w portalu *wpolityce.pl* typowymi modyfikatorami przymiotnikowymi rzeczownika *ideologia* są wyrazy: *zbrodnicza* (głównie w odniesieniu do komunizmu), *lewacka*, *marksistowska* i *genderowa*. Może to świadczyć o wyraźnej negatywnej prozodii semantycznej rzeczownika *ideologia*, który używany jest do relatywizacji i deprecjacji konkurencyjnego systemu przekonań.

Formularz wyszukiwania kolokacji umożliwia zawężenie zakresu dat publikacji tekstów, w których występują potencjalne kolokacje. I tak na przykład wśród najczęściej występujących w 2011 roku kolokatów rzeczownika *fala* znajdziemy dopełniaczowe formy rzeczowników *tsunami* i *protesty*¹⁵. To samo zapytanie z zakresem dat ograniczonym do roku 2018 jako jedno z najczęstszych współwystąpień *fali* zwraca rzeczownik *hejt(u)*¹⁶. Tego typu zestawienia mogą być przydatne w identyfikacji dominujących lub wręcz powstających w danym czasie sensów wyrazów wieloznacznych.



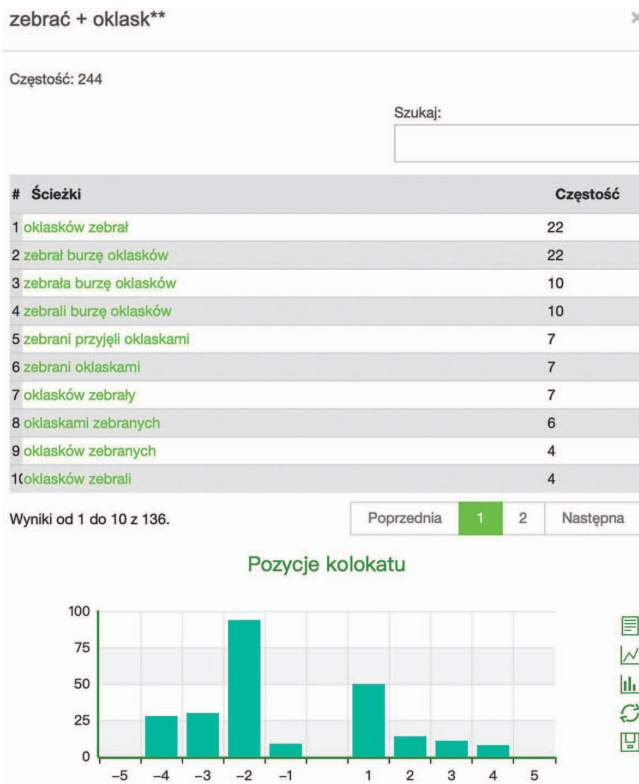
Rys. 13. Wyniki ekstrakcji kolokacji przymiotnikowych rzeczownika *hejt*

¹⁵ Zob. online: <https://tinyurl.com/y33bkoqg>; data dostępu: 14.07.2020.

¹⁶ Zob. online: <https://tinyurl.com/y4qc9tds>; data dostępu: 14.07.2020.

Rysunek 13 przedstawia główny ekran wyników ekstrakcji kolokacji. U góry tego widoku podana jest informacja o wielkości zdeduplikowanej próby użytej w zapytaniu, a także o całkowitej liczbie wystąpień ośrodka kolokacji w korpusie. Nieco niżej znajduje się wykres kołowy obrazujący rozkład frekwencji 10 najczęstszych kolokacji zaobserwowanych w przeanalizowanej próbie zdań. Potencjalne kolokacje ukazane są w zestawieniu pod wykresem. W przykładowym zestawieniu z rysunku 13 znajduje się łącznie 180 potencjalnych kolokacji przymiotnikowych rzeczownika *hejt*, które w próbie 7901 zdań wystąpiły co najmniej dwukrotnie. Domyślnie wyniki posortowane są malejąco według prostej miary powiązania opartej na współczynniku Dice'a, która promuje nadspodziewanie częste współwystąpienia wyrazów. Klikając na nagłówki kolumn *Częstość* i *Powiązanie*, można zmieniać odpowiednio kolejność wyników w tabeli.

W kolumnie *Pozycje* podane są częstości występowania pozycji, na których wystąpił dany kolokat. W omawianym przykładzie pozwala to stwierdzić, że różne formy przymiotnika *wszechobecny* wystąpiły tylko w prepozycji do rzeczownika *hejt*, w tym 21 razy bezpośrednio przed nim, dwukrotnie jako drugie i raz jako trzecie słowo przed ośrodkiem kolokacji. W kolumnie *Ścieżki* podane są częstości ciągów słowoform (tzw. n-gramów wyrazowych) tworzących kolokacje wraz z wyrazami występującymi między nimi. Takie listy n-gramów pozwalają zaobserwować w niektórych przypadkach typowe formy gramatyczne składników kolokacji, a czasami również dodatkowe wyrazy tworzące dłuższe związki



Rys. 14. Widok szczegółowy n-gramów wyrazowych tworzących kolokację *zebrać + oklaski*

frazeologiczne. Na przykład na liście czasownikowych kolokatów rzeczownika *oklaski* znajdziemy formy dokonane czasownika *zebrać*. Po kliknięciu w ten wynik otwiera się okno dialogowe podobne do tego z rysunku 14, z którego wynika, że ponad 25% (22 + 22 + 10 + 10 z 244) współwystąpień czasownika *zebrać* z rzeczownikiem *oklaski* to kombinacje uwięzione w nieco dłuższej konstrukcji idiomatycznej *zebrać burzę oklasków*. Dodatkowo w oknie dialogowym ścieżek wyświetlany jest wykres pozycji kolokatu w stosunku do ośrodka kolokacji.

Warto zauważyć, że ośrodkiem wyszukiwanych kolokacji może być wyrażenie wielowyrazowe. Przykładem ilustrującym zasadność próby ekstrakcji kolokacji wyrażenia wielowyrazowego może być zapytanie o kolokaty frazy *parasol ochronny*¹⁷, które zwraca wyniki ujęte częściowo w tabeli 3. Na uwagę zasługuje stosunkowo duża dowolność w konstrukcji idiomów opartych na metaforze *roztaczania parasola ochronnego* nad kimś lub czymś. Co ciekawe, poza czasownikiem *roztaczać* w wynikach pojawiają się bardzo często formy synonimicznych czasowników, takich jak *rozpiąć*, *rozpostrzeć*, *rozciągnąć* czy też *rozłożyć*. Tego typu listy okazują się pomocne w weryfikacji wariantywności wielowyrazowego idiomu, co ma pewnie praktyczne zastosowania na przykład w redakcji frazeostylistycznej tekstów.

Tabela 3

Kolokaty czasownikowe frazy *parasol ochronny*

#	Kolokat	Częstość występowania	Powiązanie
1	roztoczyć	138	0.077
2	roztaczać	185	0.0635
3	rozpiąć	94	0.0347
4	rozpostrzeć	61	0.0333
5	rozpinać	18	0.0103
6	rozciągnąć	40	0.0093
7	zwinąć	27	0.0065
8	rozpościerać	14	0.0065
9	zwijać	25	0.0061
10	rozłożyć	63	0.0041

7. Implementacja

Na koniec warto wspomnieć o rozwiązaniach technicznych, bez których nie byłoby możliwe wydajne aktualizowanie i serwowanie tak dużego korpusu monitorującego. Mechanizm wyszukiwania MoncoPL oparty jest na technologii Apache Solr. Format indeksu został dostosowany do potrzeb omówionej tu korpusowej składni zapytań. W szczególności, każdy segment wyrazowy jest reprezentowany w indeksie jako trójelementowa krotka złożona ze słowoformy, lematu i znacznika morfosyntaktycznego. Zapytania w składni MoncoPL są tłumaczone na wyrażenia regularne odpowiadające segmentom słów wraz z ich znac-

¹⁷ Zob. online: <https://tinyurl.com/y39ves7t/>; data dostępu: 12.06.2020.

nikami, a następnie mapowane na składnię zapytań logicznych Apache Solr. Zastosowane rozwiązanie jest skalowalne horyzontalnie, dzięki czemu możliwe jest rozproszenie indeksu wyszukiwarki na dodatkowe maszyny na odrębnych węzłach.

8. Plany utrzymania korpusu

Celem podjętym w tym artykule było omówienie budowy i funkcji, ale również wykazanie zasadności utrzymywania dużego korpusu monitorującego polszczyzny, jakim jest w pewnym zakresie MoncoPL. Aktualizacja tego korpusu jest planowana przynajmniej do czasu rozpoczęcia prac nad kolejną wersją NKJP. W przekonaniu autora, docelowo to właśnie korpus narodowy powinien pełnić funkcję monitorującą, również dla tych rejestrów i odmian polszczyzny, których odpowiednia reprezentacja wymaga znacznych nakładów i zasobów.

Literatura

- DAVIES M., 2010: *The Corpus of Contemporary American English as the First Reliable Monitor Corpus of English*. "Literary and Linguistic Computing" XXV, No. 4, s. 447–465. DOI: 10.1093/lilc/fqq018.
- DUDA B., LISZYK K., 2018: *Narzędzia cyfrowe w polonistycznej dydaktyce akademickiej – zastosowania, możliwości, perspektywy*. „Forum Lingwistyczne” nr 5, s. 143–154.
- OGRODNICZUK M., 2018: *Polish Parliamentary Corpus*. In: FIŠER D., ESKEVICH M., JONG F. DE, eds.: *Proceedings of the LREC 2018 Workshop ParlaCLARIN: Creating and Using Parliamentary Corpora*. Paris, s. 15–19.
- PRZEPIÓRKOWSKI i in., 2009: PRZEPIÓRKOWSKI A., GÓRSKI R.L., ŁAZIŃSKI M., PĘZIK P.: *Recent Developments in the National Corpus of Polish*. "NLP, Corpus Linguistics, Corpus Based Grammar Research", s. 302–309.
- SINCLAIR J., 1996: *EAGLES Guidelines*. Expert Advisory Group on Language Engineering Standards [online: <http://www.ilc.cnr.it/EAGLES96/browse.html>; data dostępu: 30.06.2020].
- WOLIŃSKI M., 2014: *Morfeusz Reloaded*. In: CALZOLARI N., CHOUKRI KH., DECLERCK TH., LOFTSSON H., MAEGAARD B., MARIANI J., MORENO A., ODIJK J., PIPERIDIS S., eds.: *Proceedings of the Ninth International Conference on Language Resources and Evaluation*, LREC. Reykjavik, s. 1106–1111.



Cyfrowe narzędzia w dydaktyce akademickiej Badania diachroniczne polskiej fleksji z użyciem systemu Chronofleks

Digital Tools in Academic Didactics
Diachronic Description of Polish Inflection
Using the Chronofleks System

Abstract: This paper describes the possibilities of the diachronic description of Polish inflection by the Chronofleks system and using it in academic didactics of historical linguistics classes. In new didactic seems necessary to use solutions based on new technologies that use students' digital skills. Therefore, the observations carried out were to check the extent to which Chronofleks is helpful both in verifying the information available in academic textbooks and in the independent acquisition of knowledge by students. Three grammar issues were selected for research: Kopczyński's rules, distribution of historic singular endings of male nouns, and ending frequency of the first person plural of verbs. The tests showed that a student can successfully learn the morphology of old Polish using the Chronofleks tool. In addition, Chronofleks allows you to implement a self-study teaching model.

Key words: academic didactics, digital tools in academic didactics, Chronofleks system, diachronic description of Polish inflection

Abstrakt: W artykule sprawdzane są możliwości wykorzystania systemu Chronofleks w dydaktyce akademickiej przedmiotów historycznojęzykowych. Autorka wybrała trzy zagadnienia fleksyjne: funkcjonowanie w uzusie tzw. reguły Kopczyńskiego, dystrybucję historycznych końcówek D. l. poj. r.m. rzeczowników oraz frekwencję zakończeń *-m*, *-my* końcówki 1. os. l. mn. czasowników. Przeprowadzone testy wykazały, że automatyczne wyszukiwanie w systemie działa sprawnie, a wyniki prac wizualizowane są w przejrzystej formie. Dzięki wykorzystaniu posiadanych sprawności cyfrowych student może zdobywać nową wiedzę. Może on zatem samodzielnie uczyć się morfologii dawnej polszczyzny z użyciem systemu, jednak powinien wspierać go nauczyciel, czuwając nad wyciąganiem prawidłowych wniosków. Realizowany w ten sposób model dydaktyczny nastawiony na samokształcenie może uatrakcyjnić ofertę dydaktyczną uczelni wyższych.

Słowa kluczowe: dydaktyka akademicka, narzędzia cyfrowe w dydaktyce akademickiej, system Chronofleks, diachroniczny opis fleksji polskiej

1. Wstęp

Konieczność wykorzystywania w dydaktyce akademickiej nowych technologii przyjmowana jest dziś bez większych zastrzeżeń. Interaktywność i multimedialność mogą podnosić

atrakcyjność prezentowanych treści, jednak nie ma pewności, że hipertekstowa i polisensoryczna forma przekazu zwiększają jego skuteczność (por. CUBAN, 2001; SPITZER, 2007; KUREK, 2013; MAYER, 2013). Użycie technologii i narzędzi cyfrowych w procesie nauczania wydaje się niezbędne z innego, bardziej fundamentalnego powodu. Obecni studenci to osoby urodzone w latach 90. ubiegłego wieku, które wzrastały otoczone nowymi technologiami. „Cyfrowi tubylcy”¹ nie znają rzeczywistości bez Internetu. Sieć komputerowa, laptop, telefon (smartfon) to dla nich naturalne elementy otaczającego świata, które służą zarówno komunikacji międzyludzkiej, jak i pozyskiwaniu informacji. Stanowią źródło wiedzy, narzędzie do nauki i pracy, a przy tym dostarczają rozrywki. Dla pokolenia Y² urządzenia cyfrowe są nieodłączną częścią życia. Za sprawą możliwości technicznych jednocześnie prowadzi się rozmowę i przegląda strony internetowe; ogląda film, słucha muzyki, wysyła wiadomości. W ten sposób łączy się w jednym urządzeniu funkcje między innymi telefonu, telewizora i odtwarzacza muzyki. Konwergencja multimediów, czyli ich wzajemne przenikanie się, spowodowane ciągłym rozwojem technologii, zmusza młode osoby do bycia wielozadaniowymi (КОРЕЦКА-ПЕЧ, 2011: 11–25). Aplikacje i oprogramowanie szybko się zmieniają, co pociąga za sobą niezwykłą różnorodność dostępnych środków wyrazu i konieczność opanowania ich funkcji. Prognozuje się, że w przyszłości hipermedia nie będą limitowane ekranem komputera. Konstruowane są właśnie aplikacje osadzone w naszym codziennym środowisku: w ubraniach, meblach, samochodach.

Aktualny profil studenta uświadamia, że:

[j]edynym sensownym i praktycznie uzasadnionym powodem wprowadzania technologii w obszar nauczania nie jest więc ich hipotetyczna rola narzędzia zwiększającego efektywność dydaktyczną, lecz rzeczywista i pragmatyczna rola narzędzia codziennej pracy, nauki i zabawy poza szkołą.

WOBALIS, 2017: 448

Cyfrowe umiejętności studentów warto wykorzystać w dydaktyce. Działania akademickie będą dzięki temu w stylu i sposobie bliższe codzienności „cyfrowych tubylców” i bardziej z nią spójne. Jednak posiadany przez studentów warsztat technologiczny należałoby również w odpowiedni sposób ukierunkować, by pomóc uczącym się odnaleźć wiarygodne i stabilne oparcie pośród dynamicznie zmieniającego się świata mediów elektronicznych.

Należy zwrócić uwagę na to, że technologie informacyjne, choć podnoszą komfort nauki i zwiększają elastyczność pracy, powodują też rozproszenie uwagi i spadek efektywności zapamiętywania (SPITZER, 2013; 2016). Postęp cywilizacyjny i związana z nim szybkość oraz wygoda dotarcia do wiedzy sprawiają, że – żartobliwie nazywanym – wieloletnim uczniom „wujka Googla i ciotki Wiki” (SZEKWO, 2012) brakuje cierpliwości. Cechuje ich potrzeba ciągłej stymulacji (bo szybko się nudzą) i praktyczne podejście do nauki. Są pewni

¹ Twórcą terminu *cyfrowi tubylcy* (ang. *digital natives*) jest amerykański badacz mediów Marc PRENSKY (2001ab).

² Nazwa pokolenia ludzi urodzonych w latach 1980–2000. W literaturze przedmiotu pokolenie Y (ang. *Generation Y*, *Gen Y*) określane jest również jako „pokolenie Milenium”, milenialsi (ang. *Millennials*). Za twórców teorii pokoleń powszechnie uznawani są William STRAUSS i Neil HOWE (1991; 2000).

siebie i oczekują natychmiastowego wynagrodzenia za swoje działania (FAZLAGIĆ, 2008). W rzeczywistości młode pokolenie ma problem z selekcją informacji, a czynności wykonuje raczej sekwencjami niż równocześnie (SAJDUK, 2014). Ponadto słabo radzi sobie z krytycznym myśleniem. Szybki dostęp do informacji nie ułatwia bowiem ich interpretacji. Studentom niełatwo jest zająć stanowisko w dyskusji, wyrazić swoją opinię czy samodzielnie przeprowadzić analizę danych. Trudność sprawia im także komunikowanie swoich uczuć i utrzymywanie prawidłowych relacji społecznych (MORBITZER, 2012: 131–153). Dostrzegając ograniczenia uczniów, Dominik Batorski postuluje redefinicję roli komputerów i Internetu w procesie edukacji. Badacz jest zdania, że:

[n]acisk nie powinien być kładziony na naukę kompetencji informatycznych, które dzieci i tak opanowują dość wcześnie same, ale przede wszystkim na naukę kompetencji informacyjnych – tj. pracy z informacją, wyszukiwania, oceny wiarygodności, weryfikacji źródeł, przetwarzania i krytycznej analizy informacji itp., a także rozwój motywacji skłaniających do lepszego korzystania z zasobów znajdujących się w sieci, w tym również ich współtworzenia.

BATORSKI, 2013: 334

Pobieżna analiza zawartości kanałów w serwisie YouTube pozwala zaobserwować, że porad w Internecie udzielają bardzo młode osoby. Zaskakuje również tematyka dostępnych wskazówek: od błahych po bardzo specjalistyczne. Błażej Sajduk interpretuje to zjawisko w kategoriach nobilitacji społecznej:

Możliwość tworzenia wiedzy może wytwarzać w nich [studentach – K.W.] doświadczenie do niedawna dostępne wyłącznie nielicznym. Teraz każdy może poczuć, że jest autorytetem i wzmacniać poczucie własnej wartości. Ma to zapewne przełożenie na proces dydaktyczny, w którym coraz częściej wśród studentów można zaobserwować tendencję, by traktować wywód wykładowcy (i samego wykładowcę) nie jako apriorycznie akceptowaną mądrość, ale opinię prawdopodobną, jak każdą inną.

SAJDUK, 2014: 24

Przegląd literatury na temat przemian sposobów uczenia się i nauczania w dobie mediów cyfrowych pozwala wyróżnić przynajmniej dwa ważne elementy w procesie dydaktycznym. Są to: sposób pozyskiwania wiedzy przez studenta oraz możliwość jej współtworzenia wraz z nauczycielem.

W dalszej części artykułu chciałabym zaprezentować nowe w swej formie i funkcji cyfrowe narzędzie badawcze, które pozwala na włączenie do dydaktyki akademickiej historycznych przedmiotów językoznawczych źródła o wysokiej wiarygodności danych. Student za pomocą programu, który wyszukuje poświadczony w tekstach formy gramatyczne, może uczyć się morfologii dawnej polszczyzny. Program ten pozwala zarówno sprawdzać wiadomości dostępne w podręcznikach, jak i samodzielnie odkrywać wiedzę. Nad prawidłową interpretacją danych językowych musi jednak czuwać wykładowca akademicki. Pod jego merytoryczną opieką student może współtworzyć wiedzę przez przygotowanie nagrań instrukcji ekranowych, tutoriali i screencastów, w których dzieli się wiadomościami z innymi.

Realizację przykładowych działań dydaktycznych przedstawię, opierając się na systemie Chronofleks i jego technicznych możliwościach.

2. System Chronofleks – opis³

System Chronofleks powstał w latach 2015–2019 w ramach grantu badawczego finansowanego przez Narodowe Centrum Nauki. Celem projektu OPUS nr 2014/15/B/HS2/03119 było opracowanie formalnego modelu diachronicznego opisu fleksji polskiej i jego komputerowej implementacji. Zadaniem wykonawców projektu było stworzenie narzędzi, które umożliwią zestawianie na przestrzeni lat paradygmatów fleksyjnych pojedynczych leksemów, ich grup, a także kategorii gramatycznych. Wykonawcy projektu to zarówno informatycy, jak i językoznawcy, którzy dzięki ścisłej współpracy przygotowali opis podsystemu fleksyjnego w formie programu komputerowego. Na stronie Instytutu Podstaw Informatyki Polskiej Akademii Nauk (IPI PAN) można przeczytać, że:

[c]elem projektu jest przedstawienie formalnego modelu polskiej fleksji, który umożliwi przedstawienie zmian w polskiej fleksji na przestrzeni dziejów Polski. Model powinien uwzględniać zmiany systemu fleksyjnego jako całości, jak również zmiany poszczególnych leksemów. [...] Ponadto model będzie na tyle ogólny, aby stanowić platformę do prezentacji polskiego systemu fleksyjnego w prospektywnym słowniku diachronicznym.

Chronofleks, online; data dostępu: 7.06.2019

W ramach prac badawczych zespół pod kierownictwem Marcina Wolińskiego stworzył narzędzia informatyczne, które pozwalają na prawidłowe działanie systemu Chronofleks. Wymienić można wśród nich tzw. transkryber, czyli program przekształcający dawny tekst z formy transliterowanej na znormalizowaną ortograficznie wersję transkrybowaną, a także systemy wspomagające przypisywanie formom wyrazowym w tekstach wartości kategorii fleksyjnych (tager, Anotatoria 2) oraz specjalny analizator morfologiczny oparty na słownikach dawnej polszczyzny (Korbeusz). Proces przypisywania wartości gramatycznej wyrazom polega na zastosowaniu automatycznego analizatora form gramatycznych na odpowiednio przygotowanym technicznie materiale językowym. W celu wyeliminowania błędów aplikacji wyniki weryfikowała grupa anotatorów.

Opracowany na potrzeby badań polszczyzny historycznej analizator Korbeusz stanowi wariant systemu Morfeusz (<http://morfeusz.sgjp.pl>), który przeznaczony jest dla współczesnej polszczyzny (WOLIŃSKI, 2014). Podstawą działania Korbeusza są sporządzone przez zespół słowniki fleksyjne, osobno dla polszczyzny XVII i XVIII wieku (słownik oparty na materiale językowym *Elektronicznego Korpusu Tekstów Polskich z XVII i XVIII wieku (do 1772 r.)*), w skrócie *Korpusu Barokowego*, dalej: KorBa) oraz wieku XIX (słownik na bazie danych *Korpusu polszczyzny 1830–1918*, dalej: F19). Podczas konstruowania Korbeusza wyko-

³ Szczegółowych danych i objaśnień na temat systemu Chronofleks udzielili mi Marcin Woliński, kierownik projektu, oraz Dorota Komosińska, programistka w zespole opracowującym opisywany system.

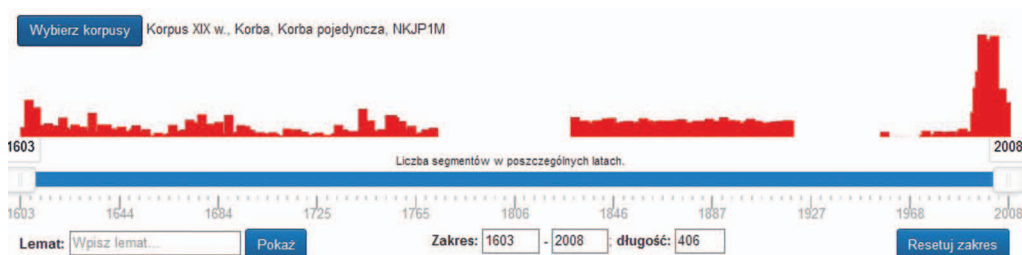
rzymano dane ze *Słownika gramatycznego języka polskiego* pod redakcją Zygmunta Saloniego (<http://sgjp.pl/leksemy>), konieczne było jednak uzupełnienie występujących w nim paradygmatów o wzory regularnie tworzonych dawnych form, które nie występują w tekstach współczesnych. Z uwagi na specyfikę dawnej polszczyzny zmodyfikować należało reguły zarówno fleksyjne, jak i segmentacyjne. Wprowadzono na przykład końcówki liczby podwójnej, formy przymiotników i imiesłówów odmiany niezłożonej, alternatywne końcówki fleksyjne rzeczowników, wzory odmiany nazw własnych, możliwość zapisu partykuły *nie* łącznie z formami czasownikowymi, reguły przyłączania aglutynantów do szerszej klasy form niż współcześnie, pisownię przyimków łącznie z następującą formą (więcej online: http://chronofleks.nlp.ipipan.waw.pl/mod_analizu/; data dostępu: 12.08.2019). Podobne procedury zmian fleksyjnych i pisownianych przeprowadzono, przystosowując system Morfeusz do znakowania słownictwa z pierwszej połowy doby nowopolskiej w korpusie F19 (por. DERWOJEDOWA i in., 2014).

System Chronofleks obejmuje również dane fleksyjne XX-wiecznej polszczyzny (na bazie *Narodowego Korpusu Języka Polskiego*, dalej: NKJP). Połączenie danych historycznych ze współczesnymi doprowadziło do skonstruowania zautomatyzowanego rejestru diachronicznych reguł fleksyjnych polszczyzny. Na stronie projektu można przeczytać, że w wyniku prowadzonych prac powstał „prototyp przekrojowego słownika fleksyjnego języka polskiego dostępnego przez Internet umożliwiający wizualizację zmian paradygmatów w czasie” (online: <http://chronofleks.nlp.ipipan.waw.pl/podsumowanie>; data dostępu: 31.07.2019).

System komputerowy Chronofleks składa się z dwóch modułów. Pierwszy z nich o nazwie „Paradygmaty” prezentuje zmienność wzorców fleksyjnych poszczególnych leksemów w czasie, drugi – „Wykresy” – pozwala na bardziej szczegółowe wyszukiwanie grup wyrazów ze względu na ich charakterystykę fleksyjną, umożliwia również przedstawienie danych w postaci graficznej. Informacje na temat szczegółowych wyników projektu, jak przebieg procesu tworzenia reguł transkrypcyjnych, sposobów implementacji modułu analizy fleksyjnej i metod ujednoznaczniania fleksyjnego, a także opis zasad związanych ze znakowaniem tekstów można znaleźć na <http://chronofleks.nlp.ipipan.waw.pl/>. Ponadto strona zawiera objaśnienie procedury wdrożenia aplikacji Anotatornia 2, przykładowe zagadnienia językowe, które badać można na podstawie segmentów dostępnych w systemie Chronofleks, oraz dane bibliograficzne prac, które powstały na temat systemu Chronofleks lub z jego wykorzystaniem.

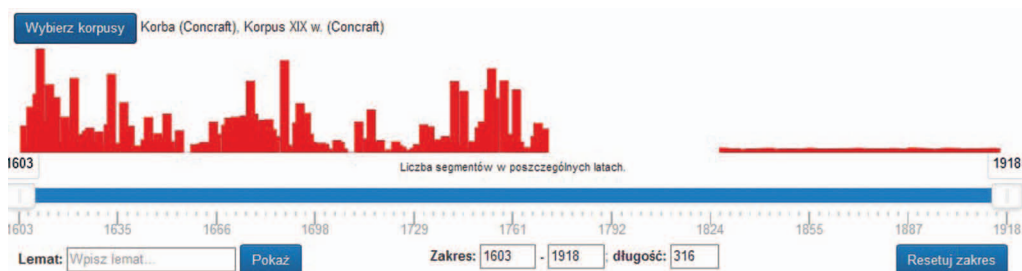
Zawartość danych w systemie Chronofleks, podobnie jak w NKJP, liczona jest w segmentach. Są to głównie słowa ortograficzne liczone ‘od spacji do spacji’, jednak w niektórych wypadkach segmenty mogą być krótsze. Sytuacja taka dotyczy końcówek aglutynacyjnych (np. [spał][eś], [często][by][śmy]), partykuł (np. [chcesz][li], [stój][że]), poprzyimkowych form zaimka *-ń* (np. [do][ń], [po][ń]), słów zawierających łącznik (np. [polsko][-][niemiecki]) oraz skrótów i inicjałów (zob. PRZEPIÓRKOWSKI, BUCZYŃSKI, WILK, 2011). Liczba danych językowych, z których korzysta Chronofleks w rozciągłości czasowej, to około 6,6 miliona segmentów dla XVII i XVIII wieku (KorBa), około 1 miliona – dla XIX wieku (F19) oraz 230 tysięcy dla polszczyzny XX-wiecznej (NKJP). Pozwala to mówić o niespełna 8-milionowym zbiorze segmentów z okresu 320 lat w wolnym dostępie. Do bazy systemu Chronofleks nie włączono wszystkich segmentów rejestrowanych przez oryginalne korpusy. Pominięto między innymi 1) segmenty odpowiadające interpunkcji, 2) segmenty oznaczone w korpusach

jako język obcy oraz 3) leksemy nieodmienne. W przypadku NKJP zrezygnowano z danych z tekstów, którym w metryczce brakuje daty ich powstania. W konsekwencji baza systemu Chronofleks zawiera mniej segmentów niż całe korpusy, co jest zrozumiałe ze względu na cel i funkcję narzędzia. Liczbę segmentów w poszczególnych latach dostępnych w systemie przedstawiają niżej zamieszczone wykresy. Ręcznie znakowane dane prezentuje rysunek 1, z kolei rysunek 2 zawiera dane znakowane automatycznie. Należy zwrócić uwagę na brak danych z lat 1772–1830 (koniec KorBy, początek F19) oraz 1918–1945 (koniec F19, początek NKJP). Zakresy te są widoczne jako białe strefy na osi czasu.



Rys. 1. Liczba segmentów znakowanych ręcznie dostępna w systemie Chronofleks w poszczególnych latach

Źródło: [online: http://chronofleks.nlp.ipipan.waw.pl/show/#c=f19_m,korba_m,korba_p_m,nkjp_m&yf=1603&yt=2008&l=^^&p=subst; data dostępu: 25.06.2020].



Rys. 2. Liczba segmentów znakowanych automatycznie dostępna w systemie Chronofleks w poszczególnych latach

Źródło: [online: http://chronofleks.nlp.ipipan.waw.pl/show/#c=f19_a,korba_a&yf=1603&yt=1918&l=^^&p=subst; data dostępu: 25.06.2020].

Korzystanie z programu Chronofleks nie wymaga znajomości składni języka zapytań. System pokazuje historyczne dane językowe na podstawie uwspółcześnionej formy hasłowej wyszukiwanego wyrazu niezależnie od jego zapisu ortograficznego oraz wielkości pierwszej litery w tekstach źródłowych⁴, na przykład *komisja* (zapis między rokiem 1650

⁴ Za pomocą systemu Chronofleks dokonuje się normalizacji zapisu transkrybowanego, z uwzględnieniem zapisu lematu. Jeśli więc postać w tekście brzmi „Bogiem”, ale przypisana jest do lematu „bóg”, to do systemu trafia znormalizowana postać „bogiem”. W tym sensie wyniki będą niezależne od wielkości pierwszej litery (bo segmenty „Bogiem” i „bogiem” dostaną taki sam zapis, pod warunkiem tylko, że oba mają przypisany lemat „bóg”). Jeśli jednak postać „Bogiem” ma przypisany lemat „Bóg” (wielką literą), to zapis znormalizowany zachowa wielką literę i będzie oddzielnie liczony (szczegółowe wyjaśnienie udzielone przez Komosińską).

a 1918): *commissiey, commissyi, komisja, komisją, komisję, komisji, komissja, komissya, komissyę, komissyi, komisya, komisją, komisye, komisję, komisyi, komisyj, komisyją, komisyję, kommisij, kommisja, kommisjów, kommissii, kommissij, kommissją, kommissje, kommissjów, kommissya, kommissyi, kommissyiey, kommissyi* (przykład za: WOLIŃSKI i in., 2018). Dane na temat frekwencji współwystępujących w konkretnym przypadku gramatycznym końcówek fleksyjnych pozyskane za pomocą systemu umożliwiają ocenę zasięgu czasowego ich użycia.

Aby sprawdzić, jakie paradygmaty fleksyjne danego wyrazu rejestrują teksty z konkretnego przedziału czasowego, na przykład z lat 1603–1700, 1772–1830, należy w systemie sprecyzować okres wyszukiwania. Można również zawęzić pole obserwacji do jednego wybranego korpusu spośród trzech dostępnych w systemie, a także porównać dane z zasobów anotowanych ręcznie lub automatycznie. Rysunek 3 przedstawia porównanie danych na temat paradygmatów leksemu KOMISJA w zasobach KorBy, F19 i NKJP. Można zaobserwować, że w korpusie XVII i XVIII wieku w dopełniaczu liczby pojedynczej nad formą *komisjej* (5) przeważa *komisji* (63), która to od wieku XIX staje się jedyną poświadczoną tekstowo formą tego przypadku, por. *komisji* (150 w F19; 34 w NKJP). Na przestrzeni wieków rywalizowały z sobą również końcówki biernika liczby pojedynczej. Zasób systemu Chronofleks dokumentuje, że w tekstach KorBy znacznie częściej rejestrowane były formy z pierwotną końcówką *-q*, por. *komisją* (46) : *komisję* (5); w F19 ich liczba ulega zrównaniu, por. *komisją* (12) : *komisję* (18); NKJP dokumentuje całkowity zanik form fleksyjnych z końcówką *-q*, por. *komisję* (4). Obserwacja danych tekstowych biernika liczby pojedynczej leksemu KOMISJA pozwala na rekonstrukcję rozwoju tego przypadku dla pozostałych rzeczowników miękkotematowych rodzaju żeńskiego obcego pochodzenia, na przykład *historia, litania, ewangelia*.

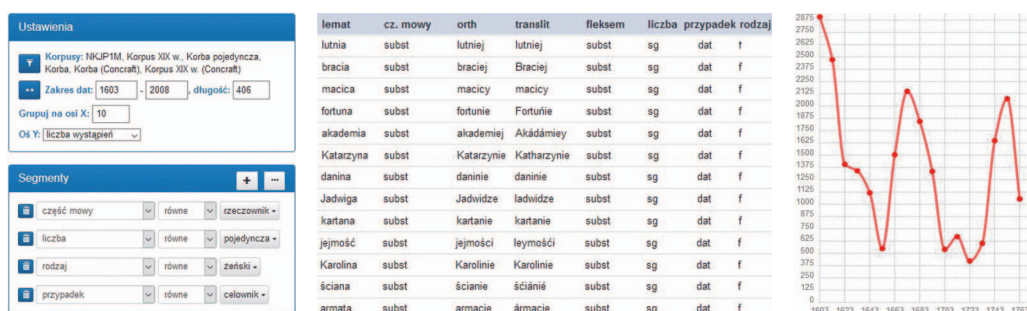
komisja <i>subst, f</i> KorBa			komisja <i>subst, f</i> F19			komisja <i>subst, f</i> NKJP		
	I. poj.	I. mn.		I. poj.	I. mn.		I. poj.	I. mn.
M.	komisja (85)	komisje (6)	M.	komisja (85)	komisje (20)	M.	komisja (27)	-
D.	komisjej (5) komisji (63)	komisjy (71) komisji (3)	D.	komisji (150)	komisjy (5) komisjów (4)	D.	komisji (37)	komisji (1)
C.	komisji (5)	-	C.	komisji (8)	-	C.	komisji (1)	-
B.	komisję (5) komisją (46)	komisje (4)	B.	komisję (18) komisją (12)	komisje (2)	B.	komisję (4)	-
N.	komisją (1)	komisjami (1)	N.	komisją (12)	-	N.	komisją (2)	-
Msc.	komisjej (1) komisji (37)	komisjach (10)	Msc.	komisji (24)	komisjach (2)	Msc.	komisji (1)	komisjach (3)
W.	-	-	W.	-	-	W.	-	-

Rys. 3. Wyniki wyszukiwania leksemu KOMISJA w zależności od korpusu (KorBa, F19, NKJP)

Źródło: [online: http://chronofleks.nlp.ipipan.waw.pl/show/#c=korba_a,korba_m,korba_p_m&yf=1603&yt=1772&l=komisja&p=subst; data dostępu: 6.07.2020; http://chronofleks.nlp.ipipan.waw.pl/show/#c=f19_a,f19_m&yf=1918&yt=1918&l=komisja&p=subst; data dostępu: 6.07.2020; http://chronofleks.nlp.ipipan.waw.pl/show/#c=nkjp_m&yf=1957&yt=1957&l=komisja&p=subst; data dostępu: 6.07.2020].

Moduł o nazwie „Wykresy” pozwala na zobrazowanie w czasie omówionych wcześniej danych tabelarycznych. Dodatkowo umożliwia tworzenia serii zapytań o różnym stopniu szczegółowości. Funkcja ta pomocna jest zwłaszcza w odniesieniu do końcówek równo-

ległych, ponieważ pozwala obserwować ich wzajemną relację na przestrzeni lat. Dzięki modułowi można sporządzić listę 100 form tekstowych spełniających zadane warunki. Jeśli wyników wyszukiwania jest mniej, wtedy pojawią się wszystkie zarejestrowane segmenty, jeśli jest ich więcej – wyświetlą się losowo wybrane przykłady. Natomiast generowany wykres ujmuje wszystkie dane dostępne systemowi. W zależności od potrzeb można grupować je według częstotliwości wystąpień, ich liczby lub typów (liczby leksemów, od których tworzona jest obserwowana forma fleksyjna). Rysunek 4 przedstawia zmienność w czasie form fleksyjnych celownika liczby pojedynczej rzeczowników rodzaju żeńskiego pogrupowanych w zależności od liczby wystąpień.



Rys. 4. Zapytanie o końcówki C. I. poj. rzeczowników r. żeńskiego wraz z wynikami tego wyszukiwania w postaci listy przykładowych rekordów i wykresu zmienności danych w czasie

Źródło: [online: http://chronofleks.nlp.ipipan.waw.pl/chart/#c=f19_m,korba_m,korba_p_m,nkjp_m&yf=1603&yt=1763&x=10&y=seg&r=1&rt=Reszta&srs=Segmenty;3_0_0;5_0_0;7_0_3;6_0_2; data dostępu: 6.07.2020].

3. System Chronofleks – badania

Poświadczenia tekstowe zjawisk językowych wykorzystywane są na zajęciach kursowych z morfologii historycznej oraz na wykładach z historii języka. Również podczas laboratoriów z leksykografii rozwijane są umiejętności związane z dokumentacją danych gramatycznych i ich analizą. Chronofleks stanowi narzędzie, które może ułatwić studentowi zdobywanie wiedzy, a także wzmocnić jego motywację do samodzielnej nauki. Twórcy systemu zaproponowali wykorzystanie stworzonego przez siebie internetowego narzędzia do obserwacji zaniku w XVII i XVIII wieku tzw. form niezłożonych przymiotnika, liczby podwójnej oraz pojawienia się wykładników fleksyjnych rodzaju męskoosobowego rzeczowników (online: <http://chronofleks.nlp.ipipan.waw.pl/chronofleks>; data dostępu: 21.08.2019). Tematów do badań jest bardzo wiele. Funkcjonowanie w uzusie tzw. reguły Kopczyńskiego, dystrybucja końcówek dopełniacza liczby pojedynczej rzeczowników rodzaju męskiego oraz frekwencja końcówki w 1. osobie liczby mnogiej czasowników to przykładowe zagadnienia, których udokumentowania można szukać w zasobach dostępnych systemowi Chronofleks. Kolejne akapity zawierają opracowanie tych zagadnień za pomocą funkcji dostępnych w systemie.

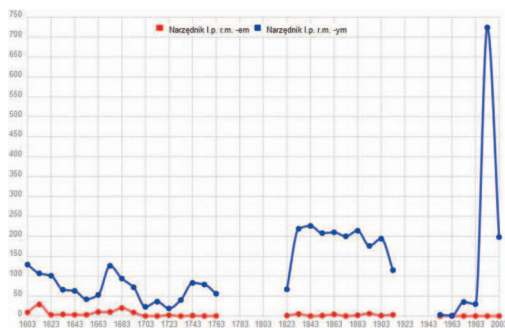
3.1. Rozwój końcówek fleksyjnych narzędnika i miejscownika liczby pojedynczej przymiotników rodzaju męskiego i nijakiego (tzw. reguła Kopczyńskiego)

Zjawiskiem doby średniopolskiej było mieszanie się końcówek narzędnika i miejscownika liczby pojedynczej rodzaju męskiego oraz nijakiego, zarówno przymiotników, jak i zaimków. Zadaniem studenta może być odtworzenie reguł fleksyjnych dla wskazanych przypadków gramatycznych z wykorzystaniem systemu Chronofleks. Ustawienie odpowiednich parametrów wyszukiwania w module „Wykresy” pozwala na generowanie wykresów przedstawiających zmienność w czasie liczby obu przymiotnikowych końcówek.



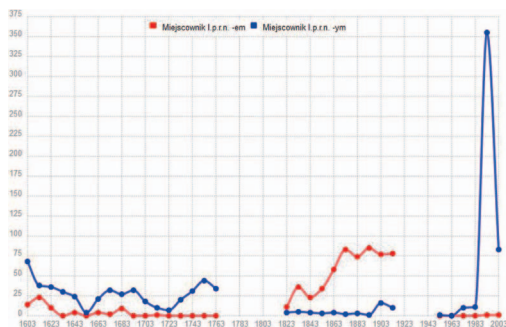
Rys. 5a. Wizualizacja rozwoju końcówki N. przymiotników rodzaju nijakiego w I. poj.

Źródło: [online: http://chronofleks.nlp.ipipan.waw.pl/chart/#c=f19_m,korba_m,korba_p_m,nkjp_m&yf=1603&yt=2008&x=10&y=seg&r=1&rt=Reszta&srs=Segmenty;3_0_1;5_0_0;7_0_4;6_0_4;0_4_em; data dostępu: 6.07.2020].



Rys. 5b. Wizualizacja rozwoju końcówki N. przymiotników rodzaju męskiego w I. poj.

Źródło: [online: http://chronofleks.nlp.ipipan.waw.pl/chart/?#c=f19_m,korba_m,korba_p_m,nkjp_m&yf=1603&yt=2008&x=10&y=seg&r=1&rt=Reszta&srs=Segmenty;3_0_1;5_0_0;7_0_0;6_0_4;0_4_ym; data dostępu: 6.07.2020].



Rys. 5c. Wizualizacja rozwoju końcówki Msc. przymiotników rodzaju nijakiego w I. poj.

Źródło: [online: http://chronofleks.nlp.ipipan.waw.pl/chart/?#c=f19_m,korba_m,korba_p_m,nkjp_m&yf=1603&yt=2008&x=10&y=seg&r=1&rt=Reszta&srs=Segmenty;3_0_1;5_0_0;7_0_0;6_0_5;0_4_em; data dostępu: 6.07.2020].



Rys. 5d. Wizualizacja rozwoju końcówki Msc. przymiotników rodzaju męskiego w I. poj.

Źródło: [online: http://chronofleks.nlp.ipipan.waw.pl/chart/?#c=f19_m,korba_m,korba_p_m,nkjp_m&yf=1603&yt=2008&x=10&y=seg&r=1&rt=Reszta&srs=Segmenty;3_0_1;5_0_0;7_0_0;6_0_5;0_4_ym; data dostępu: 6.07.2020].

Na rysunkach 5a–5d widoczna jest w latach 1603–1763 podobna odpowiedniość, która zachodzi między liczbą form przymiotnikowych z końcówką *-em* oraz *-ym*. Dotyczy to zarówno form tekstowych narzędnika, jak i miejscownika; przeważa użycie końców-

ki *-ym*. Sytuacja taka jest skutkiem zjawisk fonetycznych związanych z zanikiem iloczasu w polszczyźnie. Podwyższenie artykulacyjne samogłoski *e* spowodowało zbliżenie brzmienia końcówki *-ém* do *-ym*, a tym samym rozpoczęło mieszanie się tych końcówek. Teksty barokowe dowodzą, że użycie końcówki *-ym* (pierwotnej w obu rodzajach dla narzędnika) lub *-em* (pierwotnej w obu rodzajach dla miejscownika) zależało często od upodobań pisarza. Zgodnie z postulatami Piotra Statoriusa-Stojeńskiego, Walentego Szylarskiego i Onufrego Kopczyńskiego w drugiej połowie XVIII wieku wprowadzono normę różnicującą końcówki tych przypadków według kryterium rodzaju, w miejsce pierwotnego przypadku (DŁUGOSZ-KURCZABOWA, DUBISZ, 2001: 246). Widoczna na rysunkach 5a i 5c w latach 1823–1923 znaczna przewaga form z końcówką *-em* dla rodzaju nijakiego (kolor czerwony) oraz przewaga w tym samym okresie form z końcówką *-ym* dla rodzaju męskiego widoczna na rysunkach 5b i 5d (kolor niebieski) jest potwierdzeniem funkcjonowania reguł postulowanych przez grono gramatyków. Wspomniane arbitralnie wprowadzone przepisy fleksyjne zostały z czasem zniesione. Polska Akademia Umiejętności uznała je za sprzeczne z rozwojem historycznym oraz niezgodne z uzusem językowym. Wynik reformy ortograficznej z 1936 roku można odczytać z każdego z zaprezentowanych wykresów. Widoczny na rysunkach 5a–5d pik koloru niebieskiego wraz z jednoczesnym wypłaszczeniem linii w kolorze czerwonym to ilustracja przepisów wprowadzających jedną końcówkę *-ym* w obu przypadkach i obu rodzajach.

Student może spróbować dokonać rekonstrukcji etapów rozwoju form narzędnika i miejscownika w wyniku samodzielnej pracy z systemem. W ten sposób tradycyjna podająca metoda pracy zostanie zastąpiona podejściem problemowym do zagadnienia. Zaproponowana postawa badawcza rozwija produktywne myślenie i krytycyzm poznawczy, ponadto pozwala zwiększyć motywację uczenia się studenta i sprzyja utrwalaniu jego wiedzy. Prócz tego rozwijane są takie ogólne umiejętności, jak sztuka prowadzenia obserwacji oraz nauka wyciągania wniosków. Cyfrowe narzędzia dają możliwość połączenia kształcenia historyczno-językowego z poszerzaniem ogólnych zdolności intelektualnych studenta.

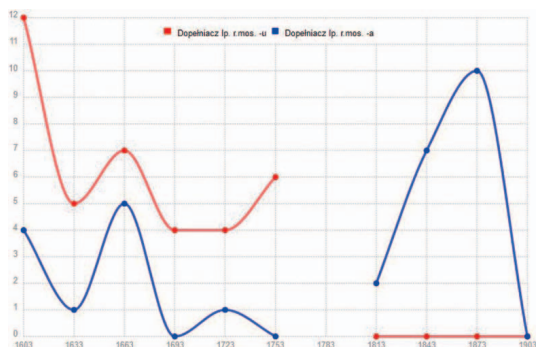
Podobne ćwiczenia można przeprowadzić na podstawie analiz form narzędnika liczby mnogiej w rodzaju męskoosobowym i niemęskoosobowym. Działanie tzw. reguły Kopczyńskiego student może sprawdzić także na materiale zaimków.

3.2. Kształtowanie się rodzaju męskożywotnego na podstawie form dopełniacza liczby pojedynczej rzeczowników

Podręczniki do gramatyki historycznej (KLEMENSIEWICZ, 1952; STIEBER, 1969; ROSPOND, 1971; DŁUGOSZ-KURCZABOWA, DUBISZ, 2001) informują o czterech praszewiańskich końcówkach dopełniacza liczby pojedynczej deklinacji męskiej. Dwie z nich (*-e*, *-i*) wyszły z użycia jeszcze w dobie przedpiśmiennej, natomiast pozostałe dwie (*-a*, *-u*) konkurują z sobą od czasów najdawniejszych. Na podstawie danych ekstrahowanych ze znakowanych fleksyjnie korpusów system Chronofleks daje możliwość porównania frekwencji rywalizujących końcówek *-u* oraz *-a*. Ekspansywność końcówki *-u* (wśród rzeczowników twar-dotematowych) odczytać można z wykresu na rysunku 6 (kolor czerwony). Znaczny spadek jej częstotliwości związany jest z rozwojem kategorii żywotności rzeczowników męskich. Pogłębiona obserwacja wyników wyszukiwania systemu Chronofleks o analizę wartości semantycznej materiału językowego, który system wyekstrahował z korpusów, pozwoli wykazać, że na-

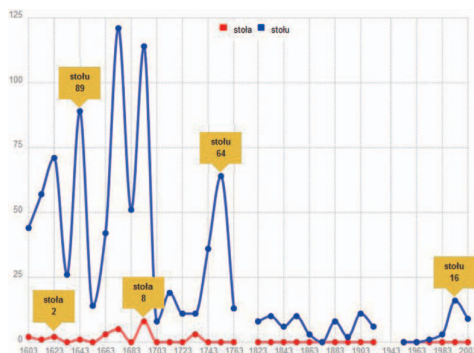
stąpiła specjalizacja końcówki *-a* do rzeczowników żywotnych (wyjątek *wołu*, *bawołu*). Wariantywność końcówki *-a* lub *-u* to zaś cecha rzeczowników nieżywotnych.

Rysunek 7 przedstawia frekwencję form *stoła* (czerwona linia) i *stołu* (linia niebieska) w rozciągłości czasowej. Zauważalne jest w latach 1603–1733 współwystępowanie form dopełniacza z pierwotną końcówką *-a* oraz wtórną *-u* (historycznie ograniczoną do nie-licznej deklinacji męskiej *ŭ*-tematowej). Z biegiem lat stabilizuje się jedna końcówka *-u* dla tego leksemu. System Chronofleks nie znajduje w dostępnych mu zasobach korpusowych poświadczeń tekstowych formy *stoła* po 1740 roku. Dokładna liczba poświadczeń wybranej formy fleksyjnej w tekstach z danego roku wyświetlana jest na żółtym polu po ustawieniu kursora w odpowiednim miejscu na krzywej wykresu (np. rys. 7).



Rys. 6. Wizualizacja liczby wystąpień końcówek *-u* oraz *-a* w D. l. poj. rzecz. mż. w latach 1603–1903

Źródło: [online: http://chronofleks.nlp.ipipan.waw.pl/chart/#c=f19_m,korba_m,korba_p_m,nkjp_m&yf=1603&yt=1903&x=10&y=seg&r=1&rt=Reszta&srs=Segmenty;0_0_!Seria%201;0_0_sto%C5%82a;0_0_sto%C5%82u; data dostępu: 6.07.2020].



Rys. 7. Wizualizacja liczby wystąpień form D. l. poj. *stoła* oraz *stołu* w latach 1603–2008

Źródło: [online: http://chronofleks.nlp.ipipan.waw.pl/chart/#c=f19_a,f19_m,korba_a,korba_m,korba_p_m,nkjp_m&yf=1603&yt=2008&x=10&y=rel&r=1&rt=Reszta&srs=Segmenty;0_0_!Seria%201;0_0_sto%C5%82u;0_0_sto%C5%82a; data dostępu: 6.07.2020].

Rozwinięciem omówionych zagadnień może być próba odnalezienia dodatkowych kryteriów semantycznych, które decydują obecnie o dystrybucji obu końcówek wśród rzeczowników męskich nieżywotnych, czyli scharakteryzowanie na bazie danych z korpusu NKJP końcówki *-a* jako dopełniacza nazw narzędzi, sprzętów, naczyń, miar, wag, liczb, nazw miesięcy, marek samochodów (np. *śrubokręta*, *telewizora*, *komputera*, *słoika*, *metra*, *kilogram*, *kwadrans*, *kwietnia*, *opła*); nazw części ciała, tańców, papierosów, gier, owoców, grzybów (np. *nosa*, *zęba*, *walca*, *camela*, *tenisa*, *banana*, *borowika*); natomiast końcówki *-u* jako dopełniacza rzeczowników pochodzenia obcego (internacjonalizmów), nazw abstrakcyjnych, a także skrótowców (np. *teatru*, *weekendu*, *bólu*, *śmiechu*, *cukru*, *PAN-u*, *LOT-u*).

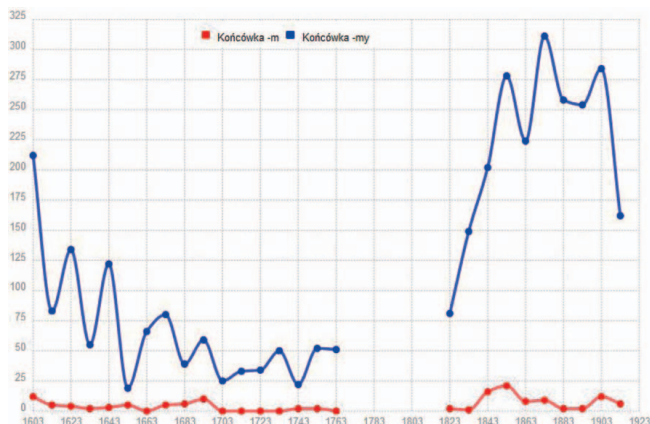
Fakultatywność występowania obu końcówek, które student rozpoznaje jako zjawisko współczesnej polszczyzny (np. *fotela*//*fotelu*, *krawata*//*krawatu*), można prześledzić w tekstach z XVII wieku, w których notuje się formy zarówno *wieka*, *obiada*, *dwora*, jak i *wieku*, *obiadu*, *dworu*. Polszczyzna najnowsza pozwala na wyzyskanie obocznych końcówek dopełniacza do sformalizowania różnic znaczeniowych, na przykład przypadku 'losu': *przypadka* 'formy gramatycznej', *strachu* 'uczucie': *stracha* 'na wróble', *zbiega* 'uciekającego': *zbiegu* 'okoliczności', *geniusza* 'człowieka genialnego': *geniuszu* 'wybitnego talentu', *postu* 'wstrze-

mięźliwość od jedzenia’ : *posta* ‘wiadomości na forum internetowym’ (JADACKA, 2005; MARKOWSKI, 2005). Proponowane połączenie wiadomości z zakresu morfologii historycznej z zagadnieniami gramatyki opisowej współczesnej polszczyzny i kultury języka polskiego może pomóc studentowi w zrozumieniu celu nauczanych treści historycznojęzykowych oraz wzmocnić w nim poczucie holistycznego programu studiów.

3.3. Rozwój końcówki fleksyjnej 1. osoby liczby mnogiej czasowników trybu oznajmującego

W 1. osobie liczby mnogiej odmiany czasownika mówimy dziś o końcówkach *-emy*, *-imy*, *-my*. Jednocześnie obecne są w polszczyźnie rzadsze formy, które kończą się na *-m*, por. *przejdziem*, *będziem*, *rzucim*. Pytanie o frekwencję tych końcówek może stanowić dla studenta zadanie do rozwiązania za pomocą systemu Chronofleks. Podręczniki do gramatyki historycznej zawierają wyjaśnienie ich historii. Zgodnie z prawami głosowymi wykształciła się w polszczyźnie końcówka fleksyjna na bazie psł. **-m̥* > pol. *-m* i psł. przyrostka tematycznego lub jego elementu samogłoskowego. Zakończenie takie dominowało w okresie staropolskim, często notowane było również w tekstach doby średniopolskiej. Widać jednak, że nowe zakończenie ukształtowane pod wpływem zaimka 1. osoby liczby mnogiej *my*, które najpierw stało się dominantą w czasownikach atematycznych (z uwagi na tożsamość formy 1. os. l. poj. i l. mn.), a następnie w pozostałych paradygmatach stosowanych w języku literackim (tj. pisanej odmianie języka ogólnopolskiego) wcześniej wyparło końcówkę pierwotną. W okresie średniopolskim końcówkami podstawowymi są już *-emy*, *-imy*, *-my*, choć przez cały ten okres, a nawet w XIX wieku, w tekstach współwystępują *-em*, *-im*, *-m* (m.in. utwory Jana Kochanowskiego, Ignacego Potockiego, Wespazjana Kochowskiego, Stanisława Staszica, Józefa Wybickiego, Juliusza Słowackiego, Marii Konopnickiej), które to końcówki otrzymują status końcówek regionalnych, następnie gwarowych oraz artystyczno-stylizacyjnych. Stosowane są dziś głównie w tekstach literackich jako przejaw archaizacji, dialektyzacji czy hieratyczności stylu. Jako gwarowe notowane są na terenie Polski północnej. Na rysunku 8a poszczególne krzywe odpowiadają częstości występowania (frekwencji) tekstowych form 1. osoby liczby mnogiej czasowników: formy koniugacyjne z zakończeniem pierwotnym *-em*, *-im*, *-m* (linia czerwona), zakończenie wtórne *-emy*, *-imy*, *-my* (linia niebieska). Z wykresu wynika, że formy koniugacyjne z zakończeniem pierwotnym już od początku XVII wieku były nieliczne. Ich bogatszą reprezentację notują teksty z okresu walk z imperium osmańskim (1683), z najazdu tureckiego na Rzeczpospolitą (1688), utwory drukowane podczas powstania krakowskiego (1843), powstania wielkopolskiego i Wiosny Ludów (1848–1849). Przykładowe rekordy form starszych zawiera lista zamieszczona na rysunku 8b.

Z dokumentacji tekstowych wynika, że największą liczbę form czasownikowych w 1. osobie liczby mnogiej rejestrowano w drugiej połowie XIX wieku. Wyniki wyszukiwania pozwalają łączyć zachowania językowe Polaków z wydarzeniami historycznymi. Czas powstania styczniowego w Królestwie Polskim w 1863 roku wymagał pisania utworów o charakterze apelów i odezw.



Rys. 8a. Wizualizacja zakończeń *-m*, *-my* 1 os. l. mn. czasowników

Źródło: [online: http://chronofleks.nlp.ipipan.waw.pl/chart/#c=f19_a,f19_m,korba_a,korba_m,korba_p_m,nkjp_m&yf=1603&yf=2008&n=10&y=rel&r=0&rt=Reszta&srs=Segmenty!Seria%201;3_0_3;5_0_2;8_0_0;0_4_m!Seria%202;3_0_3;5_0_2;8_0_0;0_4_ym; data dostępu: 6.07.2020].

Przykładowe rekordy:

orth	orth
ważym	uprzedzamy
nazwiem	dokażemy
musielim	pytamy
siedziem	widzimy
wierzym	stanęliśmy
wierzym	wspomniemy
pomieścim	żądaliśmy
stoim	zrucamy
pędzimy	traktujemy
siędziem	wróżymy
wywodzimy	zdziwimy
widzimy	zaznaczyliśmy

Rys. 8b. Przykładowe formy 1 os. l. mn. czasowników

Źródło: [online: http://chronofleks.nlp.ipipan.waw.pl/chart/#c=f19_a,f19_m,korba_a,korba_m,korba_p_m,nkjp_m&yf=1603&yf=2008&n=10&y=rel&r=0&rt=Reszta&srs=Segmenty!Seria%201;3_0_3;5_0_2;8_0_0;0_4_m!Seria%202;3_0_3;5_0_2;8_0_0;0_4_ym; data dostępu: 6.07.2020].

Cenne byłoby uzupełnienie danych systemu o okres wcześniejszy (sprzed XVII w.), co pozwoliłoby odpowiedzieć na pytanie: Od kiedy rozpoczyna się eliminacja form z zakończeniem *-m*? Możliwości takie daje włączenie do zasobów systemu Chronofleks efektu prac prowadzonych od 2018 roku w Instytucie Języka Polskiego Polskiej Akademii Nauk w Krakowie w ramach grantu badawczego *Baza leksykalna średniowiecznej polszczyzny (do 1500 roku). Fleksja*. Celem przedsięwzięcia jest interpretacja gramatyczna słownictwa tekstów doby staropolskiej (przyprządkowanie leksemu do danej części mowy, katalog form fleksyjnych poświadczonych w materiale źródłowym, przykładowe cytaty ilustrujące użycie formy w tekście), por. <https://ijp.pan.pl/nauka-i-badania/projekty/projekty-realizowane/baza-leksykalna-sredniowiecznej-polszczyzny-do-1500-roku-fleksja/> (data dostępu: 30.01.2020). Wymagałoby to zapewne kolejnych prac techniczno-inżynierskich, jednak próby zastosowania automatycznej anotacji morfosyntaktycznej tekstów do 1500 roku za pomocą tagera TaKIPi zostały już podjęte (por. EDER, KLAPPER, KOŁODZIEJ, 2015). Obecnie zainteresowany student może skorzystać z zamieszczonych na stronie *Słownika polszczyzny XVI wieku* (<http://spkvi.edu.pl/wersja-cyfrowa/>) paradygmatów czasowników. W zakładce „Dodatki” znajduje się elektroniczna wersja (pliki PDF) opracowania Krystyny Wilczewskiej *Czasowniki w „Słowniku polszczyzny XVI wieku” (odmiana i słowotwórstwo)* (data dostępu: 30.01.2020).

Aktualne funkcje systemu Chronofleks nie dają możliwości podglądu szczegółowych danych na wykresie, tzn. system nie informuje, które z form notowane są w poszczególnych latach (dekadach, przedziałach czasowych) widocznych na wykresie. Podobnie jest z charakterystyką autorów, którzy w wybranym przedziale czasu używali konkretnej formy fleksyjnej. Funkcje te mogłyby okazać się przydatne dla osób zainteresowanych studiami z zakresu

językowego obrazu świata. Jak dowodzą wyniki różnych językowych prac badawczych, wysoka frekwencja jakiejś formy może zupełnie nie odpowiadać jej popularności, a być jedynie upodobaniem językowym jednego autora lub wynikiem niezrównoważenia tekstów z danego okresu w korpusie. Wprowadzenie możliwości ustalenia informacji na temat źródła paradygmatów mogłoby zapobiec formułowaniu nie do końca trafnych wniosków.

W podsumowaniu opinii na temat użytkowania systemu Chronofleks i jego technicznych możliwości można dodać, że brakuje w nim funkcji pozwalającej na zapisanie wygenerowanego wykresu. Przydatna byłaby również możliwość wykonania kopii wyników wyszukiwania w postaci dokumentu Excel. Ponadto, jako mało intuicyjny ocenić można sposób tworzenia serii wybranych grup fleksyjnych lub form wyrazowych na bazie konkretnego segmentu w module „Wykresy”, gdzie należy zawsze ustalić nadrzędny zakres wyszukiwania jako segment, w którego zakresie możliwe jest dopiero tworzenie serii o bardziej szczegółowych parametrach zapytania. Można pomyśleć także o integracji systemu z hasłami słowników historycznych powstałych na bazie korpusów, z których Chronofleks korzysta. W przypadku braku znajomości semantyki jakiegoś wyrazu, prócz oglądu paradygmatu odmiany, możliwe byłoby poznanie jego znaczeń i przejrzanie użyć w tekstach, por. *apanaż*, *sęstwo*, *marglować*. Realizacji takich działań sprzyja ciągły rozwój elektronicznych prac korpusowych i słownikarskich.

4. Zakończenie

Przeprowadzone badania językowe z wykorzystaniem możliwości systemu Chronofleks dowodzą, że jest to narzędzie, które pozwala na samodzielną obserwację zagadnień fleksyjnych. Automatyczne wyszukiwanie działa sprawnie. System w krótkim czasie wyświetla zadane rekordy. Wyniki wyszukiwania można zwizualizować w przejrzysty sposób. Ponadto system Chronofleks wyróżnia dostęp do informacji ilościowej, co umożliwia badanie nie tylko tego, jakie formy danych leksemów są potwierdzone, ale też, ile jest tych potwierdzeń i jak są one rozłożone w czasie. Praca z systemem ułatwia przyswajanie wiedzy, a możliwość opracowania zagadnień językowych na bazie autentycznych danych sprzyja wyciąganiu prawidłowych wniosków. Chronofleks pozwala realizować model dydaktyczny nastawiony na samokształcenie. Umożliwia studentowi korzystanie z posiadanych sprawności cyfrowych, daje możliwość ich wyćwiczenia i spożytkowania do zdobycia nowych informacji.

Na temat elektronicznych zasobów językowych można przeczytać w przewodniku Pastrycji PAŁKI i Agaty KWAŚNICKIEJ-JANOWICZ (2017). W ramach przeglądu zasobów cyfrowych w publikacji przeznaczonej dla „studentów kierunków humanistycznych, magistrantów, doktorantów, pracowników naukowych, a także miłośników języka polskiego, którzy od podstaw poznają zasoby elektroniczne oraz sposoby i możliwości ich eksploatacji” (PAŁKA, KWAŚNICKA-JANOWICZ, 2017: 14) przedstawione zostały zarówno biblioteki cyfrowe, bazy skanów, słowniki, korpusy i kartoteki cyfrowe oraz kompendia elektroniczne. Zbiór ten poszerza i uzupełnia o narzędzia stworzone przez konsorcjum CLARIN-PL oraz mniejsze projekty, typu MoncoPL, Korpusomat, Wirtualne Laboratorium Transkrypcji, Jasnopis artykuł Beaty Dudy i Karoliny Liszczyk (2018). Autorki uważają, że narzędzia oraz aplikacje

internetowe „mogą okazać się cennym wsparciem w dydaktyce akademickiej i pozwolą uatrakcyjnić zajęcia kursowe skierowane do studentów polonistyki” (DUDA, LISZYK, 2018: 144).

Rozwój technologii sprzyja powstawaniu coraz nowszych narzędzi przeznaczonych do tego typu pracy dydaktycznej i naukowej. Nie wydaje się jednak, by dokonywała się technologizacja nauki, której należałoby się obawiać, jak głoszą pesymiści skoku technologicznego (por. WOBALIS, 2017). Nauczanie akademickie jest systemem, któremu nie zagraża napór nowych technologii i narzędzi cyfrowych. Rozwijające się media elektroniczne zdają się wręcz wzmacniać jeszcze rolę, jaką przypisuje się kształceniu akademickiemu. Jak podsumowuje Mirosław Wobalis:

[...] taki model edukacji jest właśnie w dzisiejszych czasach paradoksalnie najbardziej potrzebny – stanowi bowiem jedną z kilku (wraz z kulturą wysoką, dobrym wychowaniem, tradycyjnymi normami społecznymi i wartościami humanistycznymi) ważną przeciwwagę dla świata opisanego w wizjach pesymistów, ważną kotwicę gwarantującą nam kontakt z przeszłością, bez której ani terażniejszość, ani przyszłość nie mają racji bytu. [...] [t]o szkoły i uniwersytety, ze szczególnym uwzględnieniem przedmiotów szkolnych oraz kierunków akademickich związanych z kształceniem humanistycznym, powinny stanowić dla zmęczonych cyberwędrowców oazę skupienia, koncentracji, refleksji, selekcji i żmudnej analizy.

WOBALIS, 2017: 443

Zajęcia prowadzone klasycznymi metodami kształcenia uzupełnione narzędziami wykorzystującymi technologie informacyjne powinny zaspokoić oczekiwania odbiorców oferty akademickiej. Zespolecie tradycji z nowoczesnością niesie wiele korzyści zarówno dla studentów, jak i nauczycieli, dla których innowacyjne rozwiązania stanowić mogą źródło satysfakcji dydaktycznej wynikającej z zadowolenia studenta z formy przebiegu kursu. Nowe technologie powinny zatem wesprzeć dydaktykę w murach uczelni, a nie ją zastąpić. Redaktorka książki *Perspektywy zmian w praktyce kształcenia akademickiego*, Dorota Ciechanowska, zauważa we wprowadzeniu:

Współczesny uniwersytet stał się areną ścierania się tradycyjnych i nowoczesnych poglądów. [...] Studia przestały być elitarne w rozumieniu, jakie nadawano im jeszcze dwadzieścia lat temu. Jakość kształcenia akademickiego rozpatrywać dziś należy w perspektywie oczekiwań studentów. A oczekują oni dobrego, merytorycznego i praktycznego przygotowania do zawodu, wiedzy szybkiej i studiowania bardziej skoncentrowanego na odpowiedziach niż pytaniach.

CIECHANOWSKA, 2014: 7

Uczymy nowe pokolenie, spróbujmy zatem uczyć je w nowy sposób.

Słowniki

- Czasowniki w „Słowniku polszczyzny XVI wieku” (odmiana i słowotwórstwo)*. WILCZEWSKA K., oprac. [online: https://spɣvi.edu.pl/media/uploads/dodatki/03_a_1.pdf; data dostępu: 30.01.2020].
- Słownik gramatyczny języka polskiego*. SALONI Z., red. [online: <http://sgjp.pl/leksemy>; data dostępu: 21.08.2019].
- SPXVI – *Słownik polszczyzny XVI wieku* [online: <http://spɣvi.edu.pl/wersja-cyfrowa/>; data dostępu: 7.06.2019].

Korpusy

- Baza leksykalna średniowiecznej polszczyzny (do 1500 roku). Fleksja* – [online: <https://ijp.pan.pl/nauka-i-badania/projekty/projekty-realizowane/baza-leksykalna-średniowiecznej-polszczyzny-do-1500-roku-fleksja/>; data dostępu: 7.06.2019].
- Chronofleks [online: <http://chronofleks.nlp.ipipan.waw.pl/>; data dostępu: 21.08.2019].
[online: http://chronofleks.nlp.ipipan.waw.pl/mod_analazy; data dostępu: 7.06.2019].
[online: <http://chronofleks.nlp.ipipan.waw.pl/podsumowanie>; data dostępu: 31.07.2019].
[online: <http://chronofleks.nlp.ipipan.waw.pl/chronofleks>; data dostępu: 21.08.2019].
[online: <http://zil.ipipan.waw.pl/Chronofleks>; data dostępu: 7.06.2019].
- F19 – *Korpus polszczyzny 1830–1918* [online: <http://www.f19.uw.edu.pl/>; data dostępu: 21.08.2019].
- KorBa – *Elektroniczny korpus tekstów polskich z XVII i XVIII wieku (do 1772 r.)* [online: https://www.korba.edu.pl/query_corpus/; data dostępu: 21.08.2019].
- Morfeusz – [online: <http://morfeusz.sgjp.pl/>; data dostępu: 21.08.2019].
- NKJP – *Narodowy Korpus Języka Polskiego* [online: <http://nkjp.pl/>; data dostępu: 7.06.2019].

Literatura

- BATORSKI D., 2013: *Polacy wobec technologii cyfrowych – uwarunkowania dostępności i sposobów korzystania*. „Contemporary Economics” VII, s. 317–341 [online: www.ce.vizja.pl/en/+download-pdf/volume/7/issue/3.1/id/295; data dostępu: 20.08.2019].
- CIECHANOWSKA D., red., 2014: *Perspektywy zmian w praktyce kształcenia akademickiego*. Szczecin.
- CUBAN L., 2001: *Oversold and Underused. Computers in the Classroom*. Cambridge.
- DERWOJEDOWA i in., 2014: DERWOJEDOWA M., KIERAŚ W., SKOWROŃSKA D., WOŁOŚZ R.: *Współczesne narzędzia leksykograficzne a analiza tekstów dawniejszych*. „Polonica” XXXIV, s. 21–27.
- DŁUGOSZ-KURCZABOWA K., DUBISZ S., 2001: *Gramatyka historyczna języka polskiego*. Warszawa.
- DUDA B., LISZYK K., 2018: *Narzędzia cyfrowe w polonistycznej dydaktyce akademickiej – zastosowania, możliwości, perspektywy*. „Forum Lingwistyczne” nr 5, s. 143–154.
- EDER M., KLAPPER M., KOŁODZIEJ D., 2015: *Dawna polszczyzna i nowe technologie: testowanie metod przetwarzania języka naturalnego na materiale polskiego piśmiennictwa od średniowiecza po wiek XX*. „Biuletyn Polskiego Towarzystwa Językoznawczego” nr 71, s. 189–202.
- FAZLAGIĆ J., 2008: *Charakterystyka pokolenia Y* [online: <http://www.e-mentor.edu.pl/drukuj/arttykul/numer/25/id/549>; data dostępu: 20.08.2019].
- JADACKA H., 2005: *Kultura języka polskiego. Fleksja, słowotwórstwo, składnia*. Warszawa.
- KLEMENSIEWICZ Z., 1952: *Podstawowe wiadomości z gramatyki języka polskiego*. Kraków.
- KOPECKA-PIECH K., 2011: *Koncepcje konwergencji mediów*. „Studia Medioznawcze” nr 3 (46), s. 11–25.

- KUREK O., 2013: *Wykładowca w dobie aplikacji mobilnych i tabletów*. W: Rozmus A., red.: *Wykładowca doskonały. Podręcznik nauczyciela akademickiego*. Warszawa, s. 241–251.
- MAYER R., 2013: *Uczenie się z wykorzystaniem technologii*. W: Dumont H., Instance D., Benivides F., red.: *Educational Research and Innovation. The Nature of Learning: Using Research to Inspire Practice / Istota uczenia się. Wykorzystanie wyników badań w praktyce*. Warszawa, s. 277–307.
- MARKOWSKI A., 2005: *Kultura języka polskiego. Teoria. Zagadnienia leksykalne*. Warszawa.
- MORBITZER J., 2012: *O istocie medialności młodego pokolenia*. „Neodidagmata” 33/34, s. 131–153.
- PAŁKA P., KWAŚNICKA-JANOWICZ A., 2017: *Przewodnik po elektronicznych zasobach językowych dla polonistów (słowniki, kartoteki, korpusy, kompendia)*. Kraków [online: http://tmjp.pl/ima-ges/pdf/przewodnik_po_elektronicznych_zasobach.pdf; data dostępu: 29.11.2017].
- PRENSKY M., 2001a: *Digital Natives, Digital Immigrants*. Part 1. “On the Horizon” IX, no 5.
- PRENSKY M., 2001b: *Digital Natives, Digital Immigrants*. Part 2: *Do They Really “Think” Differently?* “On the Horizon” IX, no 6.
- PRZEPIÓRKOWSKI A., BUCZYŃSKI A., WILK J., 2011: *Ściągnawka do Narodowego Korpusu Języka Polskiego* [online: <http://nkjp.pl/poliqarp/help/plse1.html#x2-10001>; data dostępu: 31.07.2019].
- ROSPOND S., 1971: *Gramatyka historyczna języka polskiego*. Warszawa.
- SAJDUK B., 2014: *Pokolenie Y a metody dydaktyki akademickiej*. „Kultura i Polityka w Krakowie” nr 16, s. 12–28.
- STIEBER Z., 1969: *Zarys gramatyki porównawczej języków słowiańskich*. Warszawa.
- SPITZER M., 2007: *Jak uczy się mózg*. Warszawa.
- SPITZER M., 2013: *Cyfrowa demencja. W jaki sposób pozbawiamy rozumu siebie i swoje dzieci*. Słupsk.
- SPITZER M., 2016: *Cyberchoroby. Jak cyfrowe życie rujnuje nasze zdrowie*. Słupsk.
- STRAUSS W., HOWE N., 1991: *Generations: The History of America’s Future, 1584 to 2069*. New York–London–Toronto–Sydney.
- STRAUSS W., HOWE N., 2000: *Millennials Rising: The Next Great Generation*. New York.
- SZEWKO W., 2012: *Oto pokolenie Y* [online: <https://www.bankier.pl/wiadomosc/Oto-pokolenie-Y-2531947.html>; data dostępu: 1.09.2014].
- WOBALIS M., 2017: *W stronę cyfrowej dydaktyki akademickiej XXI wieku. Próba rozpoznania*. „Scripta Neophilologica Posnaniensia” XVII, s. 435–454.
- WOLIŃSKI i in., 2018: WOLIŃSKI M., KIERAŚ W., KOMOSIŃSKA D., MODRZEJEWSKI E.: *Narzędzia przetwarzania tekstów historycznych* [online: https://korba.edu.pl/static/documents/publikacje/2018_uw_wolinski.pdf; data dostępu: 18.08.2019].
- WOLIŃSKI M., 2014: *Morfeusz Reloaded*. In: CALZOLARI N., CHOUKRI K., DECLERCK T., LOFTSSON H., MAEGAARD B., MARIANI J., MORENO A., ODIJK J., PIPERIDIS S., eds.: *Proceedings of the Ninth International Conference on Language Resources and Evaluation*. Reykjavík, s. 1106–1111.





KRONIKA

Sprawozdanie z konferencji naukowej „Wyróżnij się lub zgiń. Dialog marki z konsumentem” (Katowice, 9 maja 2019)

Sprawozdanie z konferencji naukowej „Dialog bez barier. Skuteczna komunikacja w kontekście EBP – praktyki opartej na dowodach” (Katowice, 8 czerwca 2019)

Sprawozdanie z konferencji naukowej „Perspektywy rozwoju badań lingwistycznych” (Katowice, 27–28 czerwca 2019)

Sprawozdanie z konferencji naukowej „Nie/porozumienie, nie/tolerancja, w(y)kluczenie w języku i kulturze” (Wisła, 18–20 września 2019)







Sprawozdanie z konferencji naukowej „Wyróżnij się lub zgiń. Dialog marki z konsumentem” (Katowice, 9 maja 2019)

A Report from the Academic Conference „Wyróżnij się lub zgiń.
Dialog marki z konsumentem”
(Katowice, 9 maja 2019)

9 maja 2019 roku w katowickiej Strefie Centralnej została zorganizowana przez Koło Naukowe „Komunikator” trzecia z cyklu konferencja zatytułowana „Wyróżnij się lub zgiń”, która z uwagi na tematykę bez wątpienia wpisuje się w nurt wydarzeń naukowo-marketingowych coraz chętniej podejmowanych na uczelniach Górnego Śląska. W poprzednich latach poruszane tematy zasadniczo dotyczyły nowoczesnych form promocji, wizerunku i sposobu jego kreowania. Tegorocznemu spotkaniu przyświecało hasło „Dialog marki z konsumentem”, a zasadniczym celem uczestników konferencji było ukazanie postaw osób w sytuacjach komunikacyjnych wyraźnie osadzonych w działaniach marketingowych. I tym razem, podobnie jak we wcześniejszych edycjach, w wydarzeniu wzięli udział liczni specjaliści i naukowcy z dziedziny promocji, kreowania wizerunku, komunikacji społecznej i marketingu. Adresatami spotkania byli zaś studenci, osoby pracujące w branży marketingowej, przedsiębiorcy i hobbisci.

Zaraz po uroczystym otwarciu konferencji jako pierwsza wystąpiła Ewelina Tyc z Uniwersytetu Śląskiego w Katowicach, wiceprezes Stowarzyszenia „Via Linguae”, która w referacie *Wizerunek w krzywym zwierciadle (na przykładzie kanału Red Lipstick Monster)* omówiła strategie kreowania wizerunku wykorzystywane przez Red Lipstick Monster w mediach społecznościowych. Szczególną uwagę zwróciła na językowe aspekty autopromocji Ewy Grzegorzewskiej-Kostoglu, autorki przywołanego kanału. Następna prelegentka, Agnieszka Ganiek z Uniwersytetu Ekonomicznego w Katowicach, zawodowo związana z Urzędem Miejskim w Mikołowie oraz Fundacją „GaSzo”, w wystąpieniu zatytułowanym *Budowanie marki organizacji pozarządowej w kontekście ograniczonych zasobów finansowych* zwróciła uwagę na potrzebę działań brandingowych wśród takich organizacji, które obecnie przyczyniają się do wzrostu gospodarczego państwa. Dalszą część panelu wypełniły wystąpienia studentów. Marta Cholewa i Paulina Drobisz z Uniwersytetu Śląskiego w Katowicach, członkinie koła „Komunikator” i działaczki w Samorządzie Studenckim tejże Uczelni, wygłosiły referat pt. *Relacja marki z klientem, czyli jak Netflix wykorzystuje real time marketing*. Z kolei Joanna Blondzik i Jagoda Głowacka z Uniwersytetu Ekonomicznego w Katowicach w prelekcji zatytułowanej *Kampania reklamowa płyty „Małomiasteczkowy” D. Podsiadły, jako przykład wyróżniającej się strategii marketingowej* zwróciły szczególną uwagę na ambientowy charakter przywołanej strategii. Zgromadzona publiczność miała

również okazję zapoznać się z działaniami firmy „Socialpoint.pro”, której przedstawicielki, Magdalena Hoderny-Jeznach i Marta Bral, wyjaśniły, jak skutecznie stawiać pierwsze kroki na Instagramie i dlaczego właśnie z tego medium korzystać w celu budowania wizerunku. Pierwszą część spotkania zakończyło wystąpienie Marka Migalskiego z Uniwersytetu Śląskiego w Katowicach zatytułowane *Emocje i ich rola w polityce*. Prelegent nie tylko poruszył kwestię wykorzystania emocji w marketingu, lecz także zademonstrował zebranym słuchaczom, jakie strategie budowania historii działają na współczesnego odbiorcę i w jaki sposób tworzyć skuteczną narrację.

Drugą część spotkania otworzyła Anna Ledwoń-Błacha, która na co dzień zajmuje się mediami społecznościowymi w agencji marketingowej „More Bananas”. Prelegentka w swoim wystąpieniu omówiła sposoby skutecznej komunikacji w social mediach. Kolejny referat zatytułowany *Red Bull – Content marketing, który dodaje skrzydeł* zaprezentowała Maria Bajak z Uniwersytetu Ekonomicznego w Krakowie. W swoich badaniach skupiła się na odbiorze przez konsumentów strategii wizerunkowych marki „RedBull”. W dalszej części spotkania wystąpił Szymon Witczak z Uniwersytetu Śląskiego w Katowicach, przewodniczący koła naukowego „Komunikator”, który w referacie pt. *Kreowanie treści w internecie – webwriting* przybliżył kilka sposobów na udany przebieg sytuacji komunikacyjnych mających miejsce w nowych mediach. Ukazał też narzędzia do kreowania komunikatów pozwalających skutecznie docierać do potencjalnych użytkowników Sieci. Drugą część spotkania zakończyła prelekcja Wojciecha Orlińskiego, grafika, project menagera, który zawodowo związany jest z marketingiem politycznym. W wystąpieniu zatytułowanym *Budowanie wizerunku polityka i partii w mediach społecznościowych* poruszył między innymi kwestię doboru koloru i wpływ barwy na odczucia wyborców.

Trzecią część obrad rozpoczęła Agnieszka Nęcka z Uniwersytetu Śląskiego w Katowicach, która wygłosiła referat pt. *(Nie)skuteczne strategie komunikacji. Wydział Filologiczny Uniwersytetu Śląskiego w social mediach*. Referentka przybliżyła mechanizmy wykorzystywane w celu komunikowania się uczelni ze studentami. Zwróciła uwagę zarówno na zabiegi przynoszące pozytywny odbiór prezentowanego wydziału, jak i na działania, które nie trafiają do potencjalnych odbiorców. W dalszej kolejności referat na temat *Marka w pracy muzeum w Polsce. Dlaczego jest tak jak jest?* wygłosił Leszek Jodliński, który opisał strategię kreowania marki na przykładzie wybranych muzeów. W swoim wystąpieniu nie tylko zwrócił uwagę na wizerunek osób pracujących w instytucjach kultury, lecz także omówił zagadnienie, jakie wymogi prawne instytucje te stawiają przed osobami chcącymi zajmować się działalnością kulturotwórczą. Następny prelegent, Mikołaj Winkiel z „Brand24”, w swoim wystąpieniu zatytułowanym *Dlaczego marki rozmawiają w sieci z klientami zamiast ich olać?* zaprezentował aplikacje do monitorowania działalności marki w Internecie. Z referatem pt. *Nie tylko Obama – o celebrytyzacji polskiej polityki* wystąpiła Agata Kempa z Uniwersytetu Jagiellońskiego. Z kolei prelekcję zatytułowaną *Media a wojna informacyjna* przedstawił Robert Bryniak z Uniwersytetu Śląskiego w Katowicach. Konferencję zakończyło wystąpienie Konrada Kitaszewskiego z firmy „Jak sprzedawać na LinkedIn”, który referatem zatytułowanym *Jak uwolnić potencjał LinkedIn i wykorzystać go w social sellingu* wprowadził słuchaczy w tematykę nowego i nieco wymagającego medium społecznościowego specjalizującego się w kontaktach zawodowo-biznesowych.

Całodniowe spotkanie było okazją do inspirującej wymiany myśli i doświadczeń dotyczących szeroko rozumianej komunikacji nie tylko między prelegentami, lecz także licznie przybyłymi słuchaczami. Tematyka konferencji spotkała się z dużym zainteresowaniem odbiorców, którzy mieli okazję zetknąć się zarówno z pracami doświadczonych praktyków, jak i badaniami teoretyków. Nie zabrakło również świeżego spojrzenia na poruszane kwestie dzięki głosom studentów, którzy chętnie dzielili się swoimi spostrzeżeniami na temat budowania relacji marki z konsumentem we współczesnym świecie. Przygotowane wydarzenie nie wyczerpało jednak potencjału tematycznego poruszanych zagadnień, dlatego też na następny rok organizatorzy zapowiedzieli kolejne spotkanie zogniskowane wokół szeroko pojętej komunikacji.







Sprawozdanie z konferencji naukowej „Dialog bez barier. Skuteczna komunikacja w kontekście EBP – praktyki opartej na dowodach” (Katowice, 8 czerwca 2019)

A Report from the Academic Conference „Dialog bez barier.
Skuteczna komunikacja w kontekście EBP – praktyki opartej na dowodach”
(Katowice, 8 June 2019)

Konferencja naukowa „Dialog bez barier. Skuteczna komunikacja w kontekście EBP [Evidence-Based Practice] – praktyki opartej na dowodach” odbyła się 8 czerwca 2019 roku w Centrum Informacji Naukowej i Bibliotece Akademickiej (CINiBA). Była ona współorganizowana przez Instytut Języka Polskiego im. Ireny Bajerowej, CINiB-ę, Fundację Wiedzy i Dialogu Społecznego „Agere Aude” oraz Koło Naukowe Młodych Logopedów Uniwersytetu Śląskiego w Katowicach. Konferencja została poświęcona zagadnieniom związanym z szeroko pojętą komunikacją, jej usprawnianiem oraz zaburzeniami poprawności jej przebiegu. Prelegenci, w tym między innymi przedstawiciele Uniwersytetu Śląskiego w Katowicach, Państwowej Wyższej Szkoły Filmowej, Telewizyjnej i Teatralnej w Łodzi oraz goście z zagranicznych uczelni (Uniwersytetu Leibniza w Hanowerze w Niemczech oraz brytyjskiego University of Sheffield), zaprezentowali zgromadzonemu audytorium referaty, stanowiące różnorodne ujęcia zagadnień, sygnalizowanych w tytule konferencji.

Wydarzenie rozpoczęło się uroczystym powitaniem przybyłych prelegentów oraz słuchaczy przez przedstawicielki organizacji zaangażowanych w przygotowanie konferencji – reprezentującą Instytut Języka Polskiego Ewę Ficek oraz przedstawicielkę CINiB-y, Katarzynę Wyszyńską. Dalej spotkanie zostało podzielone na cztery sesje. Podczas kolejnych trzech prelegentki i prelegenci wygłaszali półgodzinne referaty, dające możliwość spojrzenia na tematykę komunikacji z różnorodnych perspektyw – między innymi językoznawczej, logopedycznej oraz związanej z dydaktyką języków obcych czy bilingwizmem. Charakter ostatniej sesji był natomiast odmienny – podczas jej trwania uczestnicy mogli wziąć udział w premierowej projekcji filmu *When I Stutter – Kiedy się jąkam*, poświęconego tematyce tego zaburzenia płynności mowy. Mogli też wysłuchać wywiadu z twórcami obrazu.

Podczas pierwszej z sesji, moderowanej przez Katarzynę Węsierską, referaty wygłosiły kolejno trzy prelegentki reprezentujące Uniwersytet Śląski w Katowicach. Pierwsza z nich, Kamila Kuros-Kowalska, w swoim wystąpieniu zatytułowanym *Opóźniony rozwój mowy a nowe technologie* poruszyła tematykę związaną z korzyściami, a przede wszystkim zagrożeniami, jakie niosą z sobą zdobywcze techniki w odniesieniu do rozwoju mowy dziecka, przedstawiając argumenty zarówno za stosowaniem nowoczesnych technologii przez najmłodszych, jak i przeciw takiej możliwości. Następnie prelegentka przedstawiła wyniki

badania, przeprowadzonego z udziałem grup dzieci: z opóźnionym oraz z prawidłowym rozwojem mowy. Miało ono na celu odniesienie przebiegu rozwoju ich mowy do częstotliwości użytkowania urządzeń multimedialnych. Uzyskane wyniki wyraźnie wskazywały na niekorzystny wpływ nadużywania elektroniki na różne aspekty rozwoju dziecka, w tym także rozwój interakcji społecznych. Kolejny referat, wygłoszony przez Ewę Ficek, nosił tytuł *Drugie oblicze terapii? Czyli dialog terapeutyczny okiem językoznawcy/dyskursologa (i nie tylko)*. W swoim wystąpieniu prelegentka analizowała między innymi kwestie barier komunikacyjnych, które mogą zaistnieć podczas terapii, przedstawiła słuchaczom różne ujęcia i przykłady takiego procesu, z którymi mogli zetknąć się w wielu publikacjach oraz, co dla wielu niespodziewane, w popularnych programach telewizyjnych, takich, jak *Big Brother* czy *Kuchenne Rewolucje*. Jako ostatnia podczas pierwszego z paneli wystąpiła Ewa Biłtas-Pleszak, wygłaszając referat zatytułowany *W stronę komunikacji otwartej – o coachingu, mentoringu, tutoring i innych niebezpiecznych rzeczach*, w ramach którego przybliżyła audytorium różnorodne techniki i metody, które, stosowane zarówno w życiu prywatnym, jak i zawodowym, mogą znacząco usprawnić komunikację międzyludzką. Jego część poświęciła także zagadnieniu wprowadzania zmian, będących istotnym impulsem do rozwoju człowieka i do inicjowania różnego rodzaju przełomów w jego życiu, często jawiących się jako tytułowe *niebezpieczne rzeczy*, wdrażanie ich bowiem zwykle wiąże się z ryzykiem i narażaniem się na krytykę, jednak, jak zaznaczyła prelegentka, niepodejmowanie takich działań w związku z różnymi obawami znacznie ogranicza nasze możliwości i zawęża horyzonty.

Znaczną część drugiej z sesji, moderowanej przez Ewę Ficek, stanowiły referaty zagranicznych gości. Jako pierwsza wystąpiła reprezentująca niemiecki Uniwersytet Leibniza w Hanowerze Aleksandra Syczewska, absolwentka logopedycznych studiów licencjackich oraz tego kierunku studiów magisterskich na uniwersytecie w Niemczech, a obecnie doktorantka hanowerskiego uniwersytetu. Jej wystąpienie, noszące tytuł *Terapia bilingwalna dzieci polsko-niemieckich z SLI na przykładzie zdolności narracyjnych*, zostało poświęcone prezentacji i analizie wyników badań, przeprowadzanych w ramach pracy doktorskiej, których celem było wskazanie, czy możliwa jest praca terapeutyczna z bilingwalnym dzieckiem zmagającym się ze specyficznymi zaburzeniami językowymi równocześnie w obu znanych mu językach, kładąc przy tym nacisk na wspomniane umiejętności narracyjne, które zwykle u dzieci z SLI wymagają odpowiedniej stymulacji. Jak wykazało owo badanie, takie działanie jest dobrze przyjmowane przez uczestników terapii i przynosi zadowalające rezultaty. Następnie Marta Węsierska z Uniwersytetu w Sheffield zaprezentowała wyniki swoich badań, których celem było porównanie postaw polskich i brytyjskich studentów logopedii wobec gielkotu – tzw. mowy beładnej, zaburzenia płynności mowy, charakteryzującego się jej zbyt szybkim i nieregularnym tempem, na temat którego wiedza, nawet wśród osób związanych z tematyką logopedyczną, wciąż bywa niewystarczająca. Wynik badania, uzyskany dzięki uzupełnieniu przez polskich i brytyjskich uczestników kwestionariusza badania opinii publicznej POSHA-CI (Public Opinion Survey of Human Attributes – Cluttering) potwierdził ów stan rzeczy, wykazując, że poziom wiedzy i kompetencji polskich studentów jest nieco niższy w porównaniu do brytyjskiej grupy badawczej. Podczas panelu wspólnie wystąpiły także Aleksandra Boroń, logopeda-praktyk, prowadząca prywatne gabinety w Sycowie i Zgorzelcu, oraz Karolina Wytrych, studentka Uniwersytetu Śląskiego w Katowicach – specjalności logopedycznej na kierunku komunikacja promocyjna

i kryzysowa. Podczas wystąpienia, noszącego tytuł *Sukces w terapii jąkania – opinie „podwójnych” ekspertów*, prelegentki rozważały, co oznacza ów sukces i dla osoby jąkającej się, i dla prowadzącego terapię, jakie czynniki są niezbędne do jego odniesienia i utrzymania zadowalającego efektu. Następnie przedstawiły one wyniki badania ankietowego, którego respondentami byli tzw. *podwójni eksperci* – osoby będące specjalistami w zakresie zaburzeń płynności mowy zarówno ze względu na swoje wykształcenie oraz praktykę zawodową, jak i osobiste doświadczenia związane z jąkaniem się, a także czynny udział w działalności grup samopomocowych, również w roli liderów. Respondenci w ramach badania zostali poproszeni o spojrzenie na powyższe zagadnienie sukcesu terapeutycznego zarówno z perspektywy biorącego udział w terapii klienta, jak i prowadzącego ją logopedy-terapeuty. Dzięki temu mogli oni tym skuteczniej zdefiniować oczekiwania, cele oraz wyznaczniki skuteczności terapii jąkania. Po tej serii referatów ponownie wyznaczony został czas na zadawanie pytań i dyskusję.

Następnie rozpoczęła się sesja trzecia, moderowana przez Kamilę Kuros-Kowalską. Jako pierwsi podczas przedostatniego panelu wystąpili Mikołaj Witkowski oraz Dariusz Jakubowski, reprezentujący Uniwersytet Śląski w Katowicach. Temat ich wystąpienia odnosił się do zagadnienia komunikacji w sposób nieco odmienny od pozostałych, głównym bowiem przedmiotem zainteresowania prelegentów było porozumiewanie się z przedstawicielami odmiennej kultury w języku obcym oraz niekonwencjonalne sposoby jego przyswajania, co zostało wskazane w tytule – *Wykorzystanie piosenek w dydaktyce języka arabskiego – ewaluacja studencka*. Jak wykazywali prelegenci, takie niestandardowe metody nauczania oraz wykorzystanie kreatywności w tym procesie może wywierać nań bardzo pozytywny wpływ. Dodatkowo przybliżyli oni słuchaczom nieco szerszą kwestię nauki i użycia języka arabskiego – jak się okazuje, znacznie skomplikowaną ze względu na występowanie dużej liczby dialektów, których przyswojenie wymaga dobrej znajomości kontekstu społecznego i kulturowego. Jako druga podczas sesji wystąpiła Katarzyna Węsierska, związana z Uniwersytetem Śląskim oraz Centrum Logopedycznym w Katowicach. Swoje wystąpienie poświęciła zagadnieniu udzielania wsparcia dziecku jąkającemu się, ze zwróceniem szczególnej uwagi na niezwykle istotnych uczestników tego procesu – jego rodziców. Po przybliżeniu audytorium opinii osób jąkających się, dotyczących tego, jakiego wsparcia potrzebują i oczekują od najbliższego otoczenia, prelegentka przedstawiła wyniki badania przeprowadzonego wśród amerykańskich i polskich rodziców dzieci z zaburzeniami płynności mowy. W badaniu sondowano opinie ekspertów na temat najmniej oraz najbardziej wspierających zachowań i osób lub grup. Jego wyniki wykazały zarówno wiele punktów wspólnych, jak i różnic w opiniach rodziców z Polski i Stanów Zjednoczonych. Jako ostatnia wystąpiła Agnieszka Płusajska-Otto, związana z Łódzką Państwową Wyższą Szkołą Filmową, Telewizyjną i Teatralną. Jej prezentacja, nosząca tytuł *O komunikacji interpersonalnej, czyli jak być słyszany, rozumiany i chętnie słuchany. Refleksje studentów Szkoły Filmowej w Łodzi*, tym razem dotyczyła komunikacji nie tylko jako procesu, międzyludzkiej wymiany myśli i odczuć, lecz głównie jej widocznego i odczuwalnego aspektu – doskonalenia poprawnej artykulacji i emisji, ze zwróceniem szczególnej uwagi na funkcje z nimi związane, takie jak właściwa postawa ciała. Dodatkowo wystąpienie to od pozostałych wyraźnie odróżniał jego warsztatowy charakter. Tym razem audytorium zostało wprowadzone w tajniki właściwej emisji głosu oraz pracy z nim i nad nim, a także technik oddechowych i relaksacyjnych.

Po tym wystąpieniu zakończyła się główna część konferencyjna i rozpoczęła się kolejna, równie istotna część wydarzenia. Podczas jej trwania miało miejsce szczególne wydarzenie – polska premiera filmu Johna Gomeza *When I Stutter – Kiedy się jąkam*. Jest to nieco ponad-godzinny materiał o dokumentalnym charakterze, w ramach którego jąkający się: kobiety i mężczyźni dzielą się swoimi doświadczeniami, opowiadają o wyzwaniach i ograniczeniach, związanych z towarzyszącą im niepełnością mowy. Co niezwykle wartościowe, film ten daje widzowi możliwość faktycznego identyfikowania się z bohaterami i wczucia w ich sytuację oraz wrażenia dzięki wizualizacjom towarzyszącym ich historiom. Dodatkowo po zakończeniu projekcji uczestnicy wydarzenia mieli możliwość wzięcia udziału w wywiadzie z twórcami filmu. Wywiad ten prowadzony był w czasie rzeczywistym, z wykorzystaniem komunikatora Skype, dzięki czemu nie tylko prowadzący, wymienieni wyżej moderatory panelu, lecz także obecni na sali mogli uzyskać odpowiedzi na nurtujące ich pytania.

Był to ostatni z zaplanowanych elementów konferencji „Dialog bez barier”, po którego zakończeniu organizatorki wydarzenia – Katarzyna Węsierska, Ewa Ficek oraz Kamila Kuros-Kowalska – dokonały oficjalnego podsumowania, dziękując aktywnym uczestnikom oraz zebranych słuchaczom i tym samym zamykając to wyjątkowe wydarzenie, które spotkało się z pozytywnym odbiorem osób biorących w nim udział. Za szczególne walory i wartości konferencji należy uznać przede wszystkim zróżnicowanie wystąpień, zarówno ze względu na ich formę, bo choć znaczna część z nich miała charakter badawczy, znalazły się wśród nich także prelekcje zawierające elementy warsztatowe, jak i ze względu na treść. Choć tematy oraz główne przedmioty zainteresowania poszczególnych prezentacji znacznie różniły się od siebie, wciąż najistotniejszą kwestią pozostawała komunikacja i dialog, a punktem wspólnym różnych ujęć tych zagadnień był wyraźny cel – dążenie do lepszej, bardziej otwartej, skuteczniejszej komunikacji z drugim człowiekiem.



Sprawozdanie z konferencji naukowej „Perspektywy rozwoju badań lingwistycznych” (Katowice, 27–28 czerwca 2019)

A Report from the Academic Conference „Perspektywy rozwoju badań lingwistycznych”
(Katowice, 27–28 June 2019)

W dniach 27–28 czerwca 2019 roku Instytut Języka Polskiego im. Ireny Bajerowej Uniwersytetu Śląskiego w Katowicach zorganizował konferencję naukową „Perspektywy rozwoju badań lingwistycznych”, w której udział wzięli pracownicy naukowci IJP. Celem spotkania było zaprezentowanie planów badawczych śląskich językoznawców oraz dyskusja nad możliwościami i obszarami współczesnych badań lingwistycznych. Zgłoszono wystąpienia zarówno indywidualne, jak i zespołów badawczych.

Pierwsze prezentacje dotyczyły językoznawstwa historycznego. Zespół badawczy, w którego skład wchodzi Barbara Mitrenga, Kinga Wąsińska, Katarzyna Sujkowska-Sobisz, Joanna Przyklenk, Karolina Lisczyk i Beata Duda, reprezentowała kierująca nim Magdalena Pastuch. W referacie pt. *Od nieoficjalności do potoczności. Pragmalingwistyka historyczna w badaniach korpusowych* lingwistka podkreśliła znaczenie tego typu badań oraz wykorzystania narzędzi cyfrowych we współczesnym językoznawstwie, zwróciła też uwagę na korzyści z łączenia badań synchronicznych i diachronicznych.

Magdalena Pastuch, Barbara Mitrenga i Kinga Wąsińska tworzą również osobny zespół zajmujący się pragmalingwistyką w badaniach diachronicznych i stawiający sobie za cel opisanie między innymi wybranych markerów potoczności oraz przeobrażeń ekspresywnizmów, analizując tworzony korpus tekstów polskich dramatów z lat 1772–1939. Autorki projektu przedstawiły inspiracje materiałowe oraz metodologiczne badania tekstów dawnych w ujęciu pragmalingwistycznym, wymieniły też przykładowe zadania analityczne, jak badanie aktów mowy czy elementów struktury tekstu.

Kolejny projekt badawczy, tym razem indywidualny, zreferował Artur Rejter. Uczony kontynuuje i rozwija analizę polskiego dyskursu naukowego wieków dawnych, skupiając się na czasie jego kształtowania się, czyli wiekach od XVI do XVIII. W badaniach uwzględnia dość dobrze już opracowane kwestie leksyki i terminologii, zajmuje go również problematyka wartościowania w tekstach naukowych XVI wieku. W wystąpieniu podkreślił wymiar genologiczny dyskursu naukowego oraz rolę czynników strukturalnotekstowych w jego kształtowaniu, co stanowi pole kolejnych badań.

Aleksandra Janowska podejmuje badania naukowe nad składnią historyczną, zajmując się między innymi translokacją elementów pomiędzy kategoriami gramatycznymi, wpływem zjawisk słowotwórczych na leksykalne i odwrotnie. Drugim polem zainteresowań

naukowych językoznawczyni są zmiany semantyczne i składniowe czasowników mentalnych w polszczyźnie.

Kolejna prezentacja dotyczyła badań Aleksandry Niewiary z zakresu lingwistyki kulturowej i aksjologii lingwistycznej oraz memetycznej ewolucji kulturowej. Badaczka kontynuuje prace nad zbiorowymi wyobrażeniami o narodach będących konceptualizacjami kulturowymi, kategoriami pamięci zbiorowej, istotnymi w procesie wyłaniania się aktualnych dla danego czasu tożsamości wspólnotowych.

Katarzyna Wyrwas zaprezentowała główne założenia rozpoczętych badań nad słownictwem budowanym w polszczyźnie oraz pierwsze wnioski z analiz materiału staropolskiego i średniopolskiego, w którym uwidaczniają się zarówno leksyka rodzima, odziedziczona z prasłowiańskiego, a następnie rozwijana dzięki derywacji słowotwórczej, jak i zapożyczenia z kolejnych języków, z którymi polszczyzna wchodziła w kontakty.

Kolejny blok tematyczny skupiał się wokół mediów. Iwona Loewe wskazała obszary wspólne lingwistyki mediów, prasoznawstwa i genologii mediolingwistyki. Przedstawiła ideę planowanej publikacji – słownika terminów i pojęć mediolingwistycznych. W projekcie udział zadeklarowało kilkoro specjalistów z Polski (Uniwersytet Śląski w Katowicach, Uniwersytet Pedagogiczny w Krakowie, Uniwersytet Rzeszowski, Uniwersytet Warmińsko-Mazurski w Olsztynie), a także badacze z Niemiec, Rosji i Słowacji.

Następnie odtworzono film z wypowiedzią nieobecnej na konferencji Eweliny Tyc. Prelegentka streściła główne wątki swoich badań nad telewizją regionalną w kontekście globalizacji. Lingwistka przekonywała, że zauważalny jest zwrot ku lokalności, a telewizja regionalna w odróżnieniu od ogólnopolskiej ma większe możliwości ocalania od zapomnienia dawnych i współczesnych regionalnych wytworów kultury duchowej i materialnej. W dalszych badaniach językoznawczyni zamierza uwzględnić między innymi podejmowane tematy, prezentowane miejsca, wydarzenia, bohaterów oraz medialny obraz świata.

Zagadnieniami logopedycznymi zainteresowała zebranych Danuta Pluta-Wojciechowska obecnie zajmująca się naukowo trudnościami w uczeniu się polskich głosek przez cudzoziemców poznających polski jako język drugi. W wystąpieniu zapoznała obecnych z problematyką swych badań prowadzonych w Lipsku na 32-osobowej grupie studentów, w tym szczególnie badań nad *interspeech*, czyli pozycją spoczynkową języka. Lingwistka przedstawiła także postulaty zmian w kształceniu logopedów i standardy współpracy z ortodontami.

Kolejną referentką była Kamila Kuros-Kowalska, która omówiła prowadzone przez siebie badania nad wpływem emigracji i dwujęzyczności na kompetencje językowe młodzieży – uczniów szkół polonijnych w Wielkiej Brytanii w wieku od 11 do 18 lat. Lingwistkę interesowała odpowiedź na pytanie: Czy fakt przebywania w środowisku wielokulturowym zrywa więź z żywym językiem polskim? Narzędzie badawcze stanowił test rozumienia najpopularniejszych słów młodzieżowych z ostatnich lat.

Karolina Lisczyk, prezentując referat pt. *Akceptacja, szacunek, tolerancja w ujęciu semantycznym*, zarysowała plan badań o charakterze interdyscyplinarnym, obejmujących przegląd interesujących ją pojęć w ujęciu socjologicznym, psychologicznym, filozoficznym, religijnym. Ważnymi elementami procedury badawczej są także rekonesans słownikowy oraz analiza współczesnych wypowiedzi potocznych (polskich i obcych).

Wystąpienie Mirosławy Siuciak było skupione na dwóch wątkach wynikających z tradycji badań prowadzonych w Instytucie. Pierwszy temat stanowiły stosunki językowe na

Górnym Śląsku w aspekcie historycznym i współczesnym, w tym między innymi badania nad tożsamością i świadomością językową Górnoślązaków w XIX i na początku XX wieku. Drugi nurt badań, który zaproponowała lingwistka, dotyczył opracowania polszczyzny XX wieku w perspektywie diachronicznej, jako zamkniętego już okresu historycznego, lecz jednocześnie takiego, co do którego dysponujemy nadal kompetencją językową oraz bogatym materiałem przykładowym w postaci tekstów drukowanych, nagrań, korpusów cyfrowych itd.

Kolejną prelegentką była Jolanta Tambor, która opowiedziała o prowadzonych przez siebie badaniach systemowych odmian regionalnych i języków mniejszościowych. Językoznawczyni zwróciła uwagę na duże zainteresowanie regionalnością i naszym regionem w nauce światowej. Przypomniała przy tym między innymi projekt badawczy na temat niemiecko-polskich ogólnojęzykowych podobieństw leksykalnych we współczesnym śląskim etnolekcie mówionym. Badanie, które ma na celu ustalenie stopnia udziału germanizmów w obecnie używanej śląszczyźnie, jest prowadzone w ramach grantu i na mocy umowy międzyuniwersyteckiej przez zespół naukowców z Uniwersytetu im. Carla von Ossietzkiego w Oldenburgu i Uniwersytetu Śląskiego w Katowicach.

Agnieszka Piela zaprezentowała kontynuowane przez siebie badania historycznojęzykowe związane z dawną warstwą frazeologiczną, otwierające szerokie perspektywy badawcze i pozwalające sięgać do różnych dyscyplin językoznawczych, na przykład etymologii, semantyki czy słowotwórstwa. Badaczka publikuje szkice na temat wybranych pól leksykalno-frazeologicznych, poszukuje źródeł polskich frazeologizmów oraz objaśnia ich etymologię, przygląda się także wygasłym w polszczyźnie połączeniom frazeologicznym.

Następny referat wygłosili Danuta Pluta-Wojciechowska i Barbara Sambor. Poruszyły w nim problem badawczy, jakim jest patofonetyka, której przedmiotem są indywidualne, nienormatywne realizacje fonemów wynikające z zaburzeń anatomicznych i czynnościowych. Właściwa diagnoza logopedyczna powinna być w takich wypadkach przeprowadzana na podstawie precyzyjnego opisu w odniesieniu do systemu fonologicznego. Logopedki poinformowały zebranych o powstającej publikacji *Vademecum wywoływania głosek* – praktycznej pomocy dla logopedów.

Nieobecna na konferencji Katarzyna Węsierska przygotowany film rozpoczęła od przedstawienia ustaleń badaczy na temat genezy jąkania się, a następnie omówiła planowane badania nad porównaniem sprawności językowej i zdolności narracyjnych dwóch grup dzieci w wieku przedszkolnym (4–7 lat) – mówiących płynnie i jąkających się. Badania, pierwsze tego typu na gruncie polskim, śląska logopedka prowadzi wraz z Martą Węsierską z Uniwersytetu w Sheffield (Wielka Brytania) z użyciem wystandaryzowanych narzędzi diagnostycznych, jak *Test rozwoju językowego*, *Test do badania narracji prostej* i in.

Pierwszy dzień konferencji zwieńczyła uroczystość, podczas której pożegnano odchodzące na emeryturę wybitne językoznawczynie – Ewę Jędrzejko i Joannę Sobczykową. Laudacje wygłosili Krystyna Kleszczowa i Ewa Biłas-Pleszak, przypominając sylwetki obu badaczek polszczyzny oraz ich bogaty dorobek naukowy.

Drugi dzień obrad otworzyło wystąpienie Aldony Skudrzyk pt. *Ramy interpretacyjne polskiego publicznego dyskursu medialnego*. Lingwistka poruszyła w nim kwestie problematyki lokalnej w zestawieniu z globalnością, zagadnienia świadomości językowo-kulturowej społeczności lokalnych czy wpływu nowych mediów na zmiany sposobu myślenia młode-

go pokolenia. Omówione zostały także założenia grantu na temat ram interpretacyjnych w dyskursie elektronicznym, który służy między innymi badaniu zjawiska globalnego zacierania się różnic w konceptualizacji świata, pokazując przy tym różnice lokalne.

Bernadetta Ciesek, Beata Duda, Ewa Ficek, Joanna Przyklenk i Katarzyna Sujkowska-Sobisz zaprezentowały się jako grupa badawcza zajmująca się fenomenem nowej narracji historycznej z perspektywy lingwistycznych badań nad dyskursem. Referująca plany badawcze grupy Beata Duda przybliżyła zebranych historię archiwistyki społecznej w Polsce, podkreślając, że rolą *oral history* jest badanie ideologicznej matrycy świadomości kulturowej społeczeństwa poprzez rozwijanie idei historii, ponieważ zadaniem archiwistyki społecznej jest budowanie pamięci zbiorowej.

W imieniu kolejnego zespołu z założeniami projektu badawczego pn. „Język / styl / dyskurs artystyczny – historia i współczesność” zapoznał odbiorców Artur Rejter. Badania są kontynuacją tradycji prowadzonych w IJP od kilkadziesiątu lat prac nad językiem artystycznym. Plany naukowe na najbliższe lata obejmują dramat współczesny wraz z jego problematyką społeczną i kulturową, polimodalnością oraz zagadnieniami adaptacji, a także kwestie onomastycznoliterackie. W skład zespołu wchodzi Artur Rejter, Beata Kiszka-Pytel, Joanna Przyklenk i Wioletta Wilczek.

Następnie odtworzono nagranie Aleksandry Kalisz, prezentującej projekt pn. „Dyskurs prywatny i jego gatunki mowy w obliczu social mediów. Analiza przestrzeni prywatnej w ujęciu genologicznym”. W planowanych badaniach językoznawczyni zamierza ustalić parametry dyskursu prywatnego oraz wyodrębnić gatunki mowy typowe dla tego zdarzenia komunikacyjnego w tzw. pokoleniu Z, co pozwoli na ustalenie świadomości genologicznej wspólnoty komunikacyjnej, dla której w powszechnej opinii najważniejsze jest budowanie relacji społecznych w Sieci.

Materiał wideo dostarczony przez Adama Warzechę pt. *O czym mówi miasto? Przyczynek do badań lingwistyki miasta* wprowadził obecnych w planowane badania, które pozwolą ustalić, co decyduje o jakości życia w miastach oraz jak przebiega komunikowanie w przestrzeni miejskiej. Pomoc ma w tym analiza komunikatów werbalnych o charakterze wizualnym, jak słowa, nazwy, polecenia, hasła na rozmaitych nośnikach (neony, murale, szyldy, znaki, plakaty itp.). Korpus tekstów powstanie dzięki badaniu terenowemu, będą też przeprowadzone badania społeczne na temat postrzegania miasta. Projekt poza walorami poznawczymi może znaleźć zastosowanie praktyczne, umożliwiając opracowanie instrumentów do wykorzystania w zarządzaniu wizerunkiem obszarów miejskich.

Następnie Tomasz Nowak przedstawił zgromadzonym zarys przedsięwzięcia związanego z lingwistyką doświadczalną. Na wstępie badacz zastanawiał się nad pytaniem o to, która lingwistyka jest prawdziwa, na co odpowiedzią może być sąd, iż prawdziwą lingwistykę stanowi ta adekwatna deskrypcyjnie. Wyjaśnił różnice w podejściu filologicznym i kognitywistycznym, zaprezentował ponadto różne modele metodologiczne aplikowane w analizie materiału językowego. Językoznawcę interesują obecnie zwłaszcza eksperymenty neuro-lingwistyczne związane z badaniem aktywności neuronów w różnych obszarach mózgu.

Kolejne wystąpienie – Jacka Warchali i Adama Warzechy – dotyczyło publikacji na temat dyskursów perswazyjnych. Jedną z nich jest opracowanie na temat perswazyjności miasta na przykładzie Bielska-Białej. Projekt ten ma na celu ukazanie, w jaki sposób miasto organizuje życie mieszkańców, jak dba o swoją historię. Drugi projekt skupia się wokół

historii śląskiej socjolingwistyki. Naukowcy zamierzają również opublikować monografię wieloautorską dla studentów komunikacji promocyjnej i kryzysowej na temat komunikacji perswazyjnej, PR, promocji, wizerunku i społecznych uwarunkowań języka.

Karolina Tomala zaprezentowała główne wątki prowadzonych przez siebie badań nad nazwami żeńskimi w polszczyźnie. Badaczka proponuje historyczne spojrzenie na nazewnictwo kobiet, ujmując *feminativa* jako problem natury nie tylko językowej, lecz także społecznej, i próbując ustalić między innymi przyczynę nacechowania większości współczesnych leksemów tego typu.

Emilia Bańczyk w zaprezentowanym nagraniu zreferowała pomysł na badania, jakimi planuje objąć matki w internetowych grupach dyskusyjnych, opisując sposoby komunikacji tej grupy. Językoznawczyni wysunęła propozycję, by nazywać język matek w Internecie *matkolektem*, choć zastrzegła, że o zasadności stosowania takiej nazwy powinny zdecydować przeprowadzone obserwacje i analizy.

W ostatnim wystąpieniu Ewa Ficek przypomniała obszary współczesnego dyskursu terapeutycznego już przez siebie zbadane, w tym między innymi prozatorskie obrazy terapii, narracje autobiograficzne czy kwestie dominacji kultury terapii w widowiskach medialnych. Lingwistka uznała za konieczne w tak zaprojektowanych badaniach otwarcie się na ustalenia innych dyscyplin naukowych, jak socjologia, psychologia, antropologia kulturowa, filozofia czy poradoznawstwo, lecz zadeklarowała prowadzenie oglądu naukowego wyłącznie z perspektywy językoznawczej, a ściślej: dyskursologicznej.

Dwudniowe spotkanie naukowe zamknęła dyskusja nad stanem i problematyką badań w Instytucie Języka Polskiego im. Ireny Bajerowej Uniwersytetu Śląskiego w Katowicach. W podsumowaniu konferencji dyrektor IJP Mirosława Siuciak wyraziła uznanie dla wielotorowości zainteresowań pracowników Instytutu oraz bogactwa podejść metodologicznych, podkreślając, że podejmowane badania indywidualne i grupowe oraz publikowane prace naukowe przekładają się na rozpoznawalność i doskonałą opinię śląskiego środowiska polonistycznego zarówno w kraju, jak i za granicą.







Sprawozdanie z konferencji naukowej „Nie/porozumienie, nie/tolerancja, w(y)kluczenie w języku i kulturze” (Wisła, 18–20 września 2019)

A Report from the Academic Conference „Nie/porozumienie, nie/tolerancja, w(y)kluczenie w języku i kulturze”
(Wisła, 18–20 September 2019)

W dniach 18–20 września 2019 roku w Wiśle odbyła się konferencja naukowa zatytułowana „Nie/porozumienie, nie/tolerancja, w(y)kluczenie w języku i kulturze”. Wydarzenie zostało zorganizowane przez Zakład Historii Języka Polskiego oraz Zakład Lingwistyki Tekstu i Dyskursu Instytutu Języka Polskiego im. Ireny Bajerowej Uniwersytetu Śląskiego w Katowicach. Wzięli w nim udział badacze reprezentujący ośrodki naukowe z całej Polski. W konferencji uczestniczyła także dyrektor Instytutu Języka Polskiego im. Ireny Bajerowej – Mirosława Siuciak, która wraz z organizatorami Ewą Biłas-Pleszak, Arturem Rejterem, Katarzyną Sujkowską-Sobisz i Wiolettą Wilczek dokonała otwarcia spotkania naukowego.

Uczestnicy mogli wziąć udział w obradach plenarnych i w sekcjach panelowych, podczas których wygłoszono prawie trzydzieści referatów. Wystąpienia pierwszego dnia zostały podzielone na dwie części. Rozważania naukowe otworzył referat Urszuli Sokólskiej (Uniwersytet Łódzki), która przedstawiła wnioski płynące z analizy semantycznej słów *tolerancja* i *tolerancjny*. Punkt wyjścia stanowiły opinie reprezentantów różnych środowisk (politolog, filozof, duchowny) na temat tolerancji. Referentka poddała oglądowi badawczemu definicje słownikowe dwóch wymienionych leksemów i zwróciła uwagę na różnorodność elementów występujących w tychże definicjach, między innymi takich, jak sposób myślenia, religia, sposób postępowania czy sfera seksualna. Kolejny wykład wygłosił Mariusz Rutkowski (Uniwersytet Warmińsko-Mazurski), który poruszył temat strategii wykluczenia za pomocą nazw własnych. Prelegent przedstawił przykłady zmiany, eufemizowania oraz celowego pomijania onimów na różnych poziomach dyskursów politycznego, historycznego i medialnego. Dowiodło to, że wykluczenie za pomocą nazw własnych jest skutecznym działaniem społecznym znanym od czasów starożytnych. Magdalena Steciąg (Uniwersytet Zielonogórski), której wystąpienie zamykało pierwszą część obrad, poddała analizie język demonstracji publicznych kobiet w Polsce w latach 2016–2017. Autorka odniosła strategię antydyskryminacyjne i dyskredytujące obecne w przekazie werbalnym protestujących kobiet do poprawności politycznej, biorąc pod uwagę aspekty: semiotyczny, aksjologiczny i etykietalny.

Po krótkiej przerwie kontynuowano obrady plenarne. Rafał Zarębski (Uniwersytet Łódzki) w referacie *Murzyn, Czarny, Kolorowy – o nazwach ludności czarnoskórej w dawnej i współczesnej polszczyźnie* dokonał analizy funkcjonalnej leksemów należących do pola

semantycznego 'nazwy człowieka o ciemnej/czarnej skórze'. Przeprowadzona obserwacja zobrazowała, jak słowa, będące pierwotnie neutralnymi, pod wpływem uwarunkowań kulturowych zyskały negatywne konotacje i zaczęły funkcjonować jako narzędzia wykluczenia. Tomasz Lisowski (Uniwersytet im. A. Mickiewicza) uczynił przedmiotem swojej analizy własny przekład koreańskiej piosenki *Achim iseul*, która stała się protest songiem i symbolem niezgody na niesprawiedliwość rządów dyktatury wojskowej. Referent ukazał trudne zadanie, jakie staje przed tłumaczem, który za pomocą rodzimych środków językowych stara się oddać bogactwo sensów wpisanych w tekst należący do obcej mu kultury – w tym wypadku kultury koreańskiej. Jako ostatnia pierwszego dnia zabrała głos Irmina Kotlarska (Uniwersytet Zielonogórski), która poddała oglądowi badawczemu materiały do nauki języka angielskiego dla polskich emigrantów wojennych w Wielkiej Brytanii w latach 40. XX wieku. Mówczyni zaprezentowała różnorodne zabiegi i środki językowe służące przedstawieniu różnic kulturowych między kulturą polską a anglosaską. Wystąpienie dowiodło, że sposób doboru i prezentacji elementów kulturowych sprawiał, iż omawiane teksty stawały się źródłem postaw i wartości godnych naśladowania wśród emigrantów.

Referaty drugiego dnia zostały podzielone na dwie sekcje, z których pierwsza składała się z trzech wystąpień. Wykład Anny Lenartowicz-Zagrodnej (Uniwersytet Łódzki) dotyczył relacji między polszczyzną a językami obcymi (łacina, greką, niemieckim) prezentowanych przez leksykografów. Referentka skupiła się na elementach, które zbliżały polszczyznę do języków obcych i które ją wyróżniały wśród nich, a także na zadaniach i miejscu języka polskiego w życiu. Lucyna Warda-Radys (Uniwersytet Gdański) poruszyła temat osvajania hejtu w języku polskim i kulturze. Badaczka przeanalizowała rozwój semantyczny samego leksemu *hejt*, a także powstanie rodziny wyrazowej i polskich odpowiedników omawianego wyrazu. Ponadto uwaga została zwrócona na rosnące przyzwolenie społeczne na publiczne dawanie upustu negatywnym emocjom. Obrady w tej sekcji zamknęła Karolina Liszyk (Uniwersytet Śląski w Katowicach), która przedmiotem swojego wystąpienia uczyniła zakresy znaczeniowe, konteksty użycia, a także konstrukcje składniowe i połączenia leksykalne, w których używane są wyrazy *tolerancja* i *tolerancyjny*. Prelegentkę interesowała szczególnie potoczna odmiana polszczyzny.

Obrady w drugiej (równoległej) sekcji otworzył referat Pawła Grafa (Uniwersytet im. A. Mickiewicza) *Interpersonalne trombosis. O współczesnych aporiach semantyczno-komunikacyjnych*. Przedmiotem analizy stał się kryzys komunikacyjny naznaczający współczesność, którego powodem są liczne aporie: cielesne, etyczne, estetyczne, językowe i tożsamościowe determinujące rozumienie siebie w świecie oraz rozumienie rzeczywistości. Beata Grochala (Uniwersytet Łódzki) w wystąpieniu pokazała, w jaki sposób (zwłaszcza w perspektywie mediolingwistycznej) przejawia się dyskurs wykluczenia w sporcie. Referentka zaprezentowała przykłady manifestacji, które odnaleźć można na płaszczyźnie struktur językowych, semantyzacji pewnych pojęć i tematyzacji (lub jej braku) określonych wydarzeń. Jako ostatnia w tej sekcji głos zabrała Kamila Kuros-Kowalska (Uniwersytet Śląski w Katowicach), która poruszyła zagadnienie wielokulturowości i jej wpływu na język. Lingwistka na podstawie rozumienia najpopularniejszych słów młodzieżowych z ostatnich lat wśród uczniów szkół polonijnych w Wielkiej Brytanii starała się odpowiedzieć na pytanie o to, czy po powrocie do kraju emigranci będą w stanie porozumieć się z rówieśnikami w rodzimym języku.

Po krótkiej przerwie obrady w dwóch równoległych sekcjach były kontynuowane. Sekcja pierwsza obejmowała cztery referaty. Pierwszy z nich wygłosiły Danuta Kowalska i Magdalena Pietrzak (Uniwersytet Łódzki), które, wykorzystując genologiczne i dyskursywne instrumentarium badawcze, prześledziły karierę Heleny Modrzejewskiej w Warszawie i miejscu w(y)kluczenia w tym kontekście. Wykład Magdaleny Graf (Uniwersytet im. A. Mickiewicza) poświęcony został różnym aspektom komunikacji na forach i blogach o tematyce dotyczącej zaburzeń odżywiania. Rozważania te zostały osadzone we współczesnym dyskursie cielesnym. Prelegentkę interesowała sfera apelatywna oraz warstwa onimiczna. Jako trzeci referat wygłosiła Renata Bura (Uniwersytet Jagielloński), która ogładowi badawczemu poddała objaśnienia nawiasowe w tekstach górnołużyckich. Referentka rozważała, czy objaśnienia te można traktować jako przejaw tolerancji, czy konieczności. Omawiany problem został przeanalizowany w dwóch perspektywach: synchronicznej i diachronicznej. Ostatni (w tej sekcji) wykład – dotyczący problemu powodów i form wykluczenia ujawnionych na podstawie analizy tekstów pochodzących z kronik wiejskich szkół Pomorza – wygłosiła Aneta Lewińska (Uniwersytet Gdański), która analizy językowe odniosła do problemów społecznych – narodowościowych i tożsamościowych – mieszkańców Pomorza w XX wieku.

W równoległej sekcji obrady otworzył referat Doroty Szogun (Uniwersytet Zielonogórski) *Wykluczeni wśród wykluczonych. Chorzy psychicznie w dyskursie publicznym*. Referentka poddała analizie dwa sposoby łączenia choroby z podmiotem: poprzez użycie czasownika *mieć* i poprzez użycie czasownika *być*. Wśród chorób psychicznych dominuje drugi sposób łączenia, a co za tym idzie, w dyskursie publicznym nie choroba jest zagrożeniem, tylko chory, a przekonanie o potrzebie jego wykluczenia społecznego przez izolację – powszechne. Kolejny referat – także związany z leksyką należącą do pola semantycznego choroby psychicznej – wygłosiła Dorota Hamerlok (Uniwersytet Śląski w Katowicach), która poddała analizie znaczenia leksemów *kretyn* i *maniak*. Prelegentka przyjrzała się temu, w jaki sposób kontekst sytuacyjny wpływa na znaczenie i nacechowanie emocjonalne analizowanych jednostek leksykalnych. Jako następna głos zabrała Agnieszka Jedziniak (Uniwersytet Śląski w Katowicach). Referentka przyjrzała się językowym granicom w(y)kluczenia. Analizie zostały poddane etymologia i semantyka leksemów *inny* i *obcy*, co pozwoliło na dostrzeżenie relacji zachodzących pomiędzy tymi wyrazami. Jako ostatni wykład wygłosił Paweł Baryś (Uniwersytet Śląski w Katowicach), którego interesował negatywny wizerunek islamu i jego wyznawców w polszczyźnie. Uwaga zwrócona została na dominację dyskursu wykluczenia nad dyskursem tolerancji.

Ostatniego dnia konferencji pierwsza część wystąpień została podzielona na dwie równoległe sekcje. Pierwszą z nich – wpisującą się w nurt rozważań nad fenomenem archiwistyki społecznej – otwierał wykład Artura Rejtera (Uniwersytet Śląski w Katowicach) poświęcony archiwum społecznemu wykluczonych oraz możliwości jego badania z uwzględnieniem różnych instrumentów współczesnej lingwistyki. Analizy zostały przeprowadzone na podstawie materiału zaczerpniętego z biblioteki i archiwum Queerstory działającego przy Stowarzyszeniu Lambda Warszawa. Ze względu na złożony charakter dyskursu wykluczonych w wystąpieniu uwzględnione zostały nie tylko komunikaty werbalne, ale także ikoniczne. Referat Beaty Dudy, Ewy Ficek i Katarzyny Sujkowskiej-Sobisz (Uniwersytet Śląski w Katowicach) dotyczył archiwum społecznego jako korpusu. Autorki ukazały specyfikę archiwum

językowego i różne wymiary porozumienia wewnątrz i na zewnątrz przestrzeni archiwalnej. Zaprezentowane zostało także narzędzie umożliwiające analizy zasobów archiwalnych, tj. korpus spotkań ze świadkami historii stworzony przy współpracy z konsorcjum CLARIN-PL. Beata Duda i Ewa Ficek w swoim wystąpieniu poruszały z kolei temat cyfrowych archiwów biograficznych stanowiących nową przestrzeń kształtowania się pamięci zbiorowej. Refleksji poddane zostały wybrane cyfrowe „miejsca pamięci”, poświęcone wybitnym postaciom kultury, literatury i sztuki. Referat stanowił także próbę odpowiedzi na pytania dotyczące między innymi tego, kim są twórcy i korzystający z archiwów, kogo się w nich upamiętnia i w jaki sposób sylwetki tych osób są budowane.

Druga sekcja obejmowała dwa referaty. Pierwszy z nich wygłosiła Katarzyna Wyrwas (Uniwersytet Śląski w Katowicach), która zdefiniowała mowę nienawiści, przedstawiła jej obszary tematyczne oraz językowe wykładniki. Objęta refleksją została zawartość postów i komentarzy na facebookowym fanpage'u Ośrodka Monitorowania Zachowań Rasistowskich i Ksenofobicznych, wśród których liczną grupę stanowią przejawy agresji językowej. Waldemar Czachur (Uniwersytet Warszawski) wygłosił referat dotyczący pojednania polsko-niemieckiego w latach 90. XX wieku. Ważną rolę w kontekście tych rozważań odegrało pojęcie kontradycyjności jako zjawiska społecznego i lingwistycznego. Na przykładzie dyskursu polsko-niemieckiego referent ukazał mechanizmy jego działania i jego językowe wykładniki.

Na obrady plenarne złożyły się dwa ostatnie wystąpienia konferencyjne. Pierwsze z nich wygłosiła Iwona Burkacka (Uniwersytet Warszawski), która poruszyła zagadnienie słowotwórczych sposobów nazywania i opisywania bohaterów powieści Joanny Bator, których wygląd lub zachowanie różni się od dominujących w społeczeństwie, a także słowotwórczych sposobów nominalizacji nietypowych zwierząt i istot fantastycznych. Referentka wykazała, że w analizowanych derywatach utrwalone są opisy włączania bohaterów we wspólnotę lub wypełnianie ról społecznych, z kolei w sposobach ich stosowania można dostrzec mentalność kreowanych postaci. Ostatni wykład w czasie konferencji wygłosiła Marta Wójcicka (Uniwersytet Marii Curie-Skłodowskiej). Został on poświęcony (nie)porozumieniom mediologicznym na podstawie dyskusji, jaką wywołała seria plakatów Tęczowa Matka Boska autorstwa Elżbiety Podleśnej. Referentka zaproponowała traktowanie wspomnianego plakatu jako nośnika pamięci, a punktem wyjścia rozważań stała się typologia pomników – nośników pamięci Regisa Debreya. Umożliwiło to spojrzenie na omawiany tekst kultury z trzech różnych punktów widzenia, a także pozwoliło dostrzec, wynikające z takiego sposobu oglądu, różne (często utrudniające porozumienie) interpretacje i dyskursy.

Wszystkie zaprezentowane referaty zainspirowały słuchaczy do ożywionej dyskusji dotyczącej zagadnień poruszanych w czasie spotkania naukowego. Uczestnicy czerpali z różnych paradygmatów, metodologii i teorii, a referaty niejednokrotnie miały charakter interdyscyplinary. Zagadnienia zostały zaprezentowane w zróżnicowanych ujęciach: językoznawczym, literaturoznawczym i kulturoznawczym. Ta różnorodność dowodzi, że obszar badań zarysowany przez organizatorów stale cieszy się zainteresowaniem. Niniejszy tekst przybliży jedynie w ogólnym wymiarze prezentowane wystąpienia teoretyczne i analityczno-interpretacyjne, które zostaną zebrane w tomie pokonferencyjnym przygotowywanym przez organizatorów.



Noty o Autorach

Iliana Genew-Puhalewa, dr hab., pracuje w Instytucie Sławiastyki Zachodniej i Południowej Uniwersytetu Warszawskiego. Zainteresowania badawcze: językoznawstwo porównawcze i stosowane, terminoznawstwo, etnolingwistyka, kultura Bałkanów. Autorka monografii *Терминологията на Европейския съюз. Съпоставка на българската, гръцката, полската и английската терминология на правото на околната среда* (Katowice 2015) oraz wielu artykułów, m.in.: *Koncept Europa (EYPΩΠH) w nowogreckim językowym obrazie świata* (w: *Leksykon aksjologiczny Słowian i ich sąsiadów*. T. 2: Europa. Red. J. Bartmiński, W. Chlebda. Lublin 2018; współautor: Alexandra Ioannidou), *Аксиологизация на съвременни идеи и термини – примерът на КЛИМАТ* („Български език“ 2019, nr 1).

Dorota Hamerlok, mgr, doktorantka w zakresie językoznawstwa na Wydziale Humanistycznym Uniwersytetu Śląskiego w Katowicach. Zainteresowania badawcze: historia języka, semantyka, fonologia współczesnej polszczyzny. Autorka m.in. następujących opracowań naukowych: *Status tzw. samogłosek q w wybranych modelach fonologicznych* („Linguarum Silva”. T. 6: *Problemy języka i tekstu w perspektywie historycznej i współczesnej*. Red. B. Mitrenga. Katowice 2017), *MasterChef XVII w. – studium porównawcze pierwszej polskiej książki kulinarnej „Compendium ferculorum” oraz współczesnych publikacji o tematyce kulinarnej* (w: *Kuchnia w języku i kulturze dawniej i dziś*. Red. T. Gęsina, W. Wilczek. Katowice 2018; współautorka: Alicja Bronder), *Etymologia i rozwój semantyczny czasownika „wariować” w polszczyźnie* (w: *Slavica iuvenum XX. Mezinárodní setkání mladých slavistů*. Ostrava 2019).

Larysa Kysliuk, dr hab., pracuje w Instytucie Języka Ukraińskiego Narodowej Akademii Nauk Ukrainy. Zainteresowania badawcze: słowotwórstwo, leksykologia, leksykografia języka ukraińskiego. Autorka monografii *Сучасна українська словотвірна номінація: ресурси та тенденції розвитку* (Київ 2017) oraz wielu artykułów naukowych, m.in.: *Динаміка словотвірної норми сучасної української мови* (w: *Є.А. Карпіловська, Л.П. Кислюк, Н.Ф. Клименко, В.І. Критська, Т.К. Пуздирева, Ю.В. Романюк: Вплив суспільних змін на розвиток української мови*. Відп. ред. Є.А. Карпіловська. Київ 2017), *Нові тенденції адаптування інішомовної лексики. Словотворча активність неозапозичень* (w: *Динамічні процеси в сучасному українському лексиконі*. Red. Н.Ф. Клименко, Є.А. Карпіловська, Л.П. Кислюк. Київ 2008), *Особливості актуалізації словотворчих ресурсів в українській інтернет-комунікації* (w: *Wortbildung und Internet*. Hrsg. B. Tošović, A. Wonisch. Graz 2016).

Mariusz Leńczuk, dr, pracuje w Instytucie Języka Polskiego PAN. Zainteresowania badawcze: edycja tekstów staropolskich, leksykografia, paleografia. Autor monografii *Staropolskie przekazy kanonu Mszy Świętej. Wariantowość leksykalna* (Warszawa 2013) oraz wielu artykułów naukowych, m.in.: *Marginalia rękopiśmienne z XV wieku. Wezwania i westchnienia* (w: *Marginalia w książce dawnej i współczesnej*. Red. B. Mazurkova. Katowice 2019), *O wydawaniu polskich glos z łacińskich rękopisów*.

pisów średniowiecznych (w: *Staropolskie spotkania językoznawcze. Jak wydawać teksty dawne?* Red. K. Borowiec, D. Masłej, T. Mika, D. Rojszczak-Robińska. Poznań 2017), *Dwa staropolskie zaklęcia z XV w.* („Studia Źródłoznawcze” 2017, T. LV).

Marek Łaziński, dr hab., prof. UW, pracuje w Instytucie Języka Polskiego Uniwersytetu Warszawskiego. Zainteresowania badawcze: aspektologia, lingwistyka korpusowa i genderowa, grzeczność językowa. Autor m.in. monografii *O paniach i panach. Polskie rzeczowniki tytułarne i ich asymetria rodzajowo-płciowa* (Warszawa 2006), *Wykłady o aspekcie polskiego czasownika* (Warszawa 2020) oraz redaktor *Słownika zapożyczeń niemieckich w polszczyźnie* (Warszawa 2008).

Anna Majewska-Tworek, dr hab., pracuje w Instytucie Filologii Polskiej Uniwersytetu Wrocławskiego. Zainteresowania badawcze: lingwistyka stosowana, glottodydaktyka, współczesna polszczyzna mówiona, wymowa polska jako obca. Autorka wielu opracowań, m.in.: *Wymowa polska jako obca. Rozwiązywanie trudności artykulacyjnych w grupie międzynarodowej na rocznym kursie przygotowawczym* („Acta Universitatis Lodzianis. Kształcenie Polonistyczne Cudzoziemców” 2019, z. 26), *Ofensywny futbol, ofensywna polityka historyczna i ofensywny palec... Przemiany w obrębie znaczenia i wartościowania neosemantyzmu z Obserwatorium Językowego UW* („Orbis Linguarum” 2018, z. 51), *Niepłynność wypowiedzi w oficjalnej odmianie polszczyzny. Propozycja typologii* (Wrocław 2014), *Szura, szumi i szeleści. Ćwiczenia fonetyczne nie tylko dla cudzoziemców* (Wrocław 2010).

Magdalena Makowska, dr, pracuje w Uniwersytecie Warmińsko-Mazurskim w Olsztynie. Zainteresowania badawcze: mediolingwistyka, multimodalność, komunikacja wizualna. Autorka wielu artykułów naukowych, m.in.: *Einige Bemerkungen zum Infografischen in der visuellen Publizistik* („Studia Germanica Posnaniensia” 2018, nr 39), *Visuelle Narration. Zum journalistischen Storytelling aus der Sicht der Medienlinguistik* („Linguistik Online 93” 2018, nr 6), *Tekst multimodalny w glottodydaktyce* („Neofilolog” 2018, nr 50(2)), *„Coś pękło, coś się skończyło”. O sposobach werbalizacji i wizualizacji żałoby w tekstach medialnych powstałych po śmierci prezydenta Gdańska Pawła Adamowicza* („Prace Językoznawcze” 2019, XXI(3)).

Agnieszka A. Niekrewicz, dr hab., pracuje w Akademii im. Jakuba z Paradyża w Gorzowie Wielkopolskim. Zainteresowania badawcze: leksykologia, pragmatyka, mem internetowy, hashtag. Autorka monografii *Rzeczowniki staropolskie we współczesnej polszczyźnie ogólnej* (Gorzów Wielkopolski 2011) i *Od schematyzmu do kreatywności. Język memów internetowych* (Gorzów Wielkopolski 2015) oraz artykułów naukowych, m.in. *Środki językowo-obrazowe służące etykietowaniu polityków w memach internetowych* („Poznańskie Studia Polonistyczne” 2016, XXIII(43)), *Leksyka i frazeologia memetyczna we współczesnej polszczyźnie* („Prace Filologiczne” 2017, LXXI).

Jarosław Pacuła, dr hab., prof. UŚ, pracuje w Instytucie Językoznawstwa Uniwersytetu Śląskiego w Katowicach. Zainteresowania badawcze: językoznawstwo diachroniczne, historia socjolektów, semantyka leksykalna, polszczyzna w sytuacjach ekstremalnych. Autor m.in. monografii *Polszczyzna w GUŁagu. Leksyka* (Bielsko-Biała 2018) i *Auschwitz – słowa i milczenie. Studium leksykalno-semantyczne* (Bielsko-Biała 2020) oraz wielu artykułów naukowych, m.in. *Znaki czasu. O „koniu Andersa” i nazwach innych „metod śledczych” z okresu Polski Ludowej* („Studia Filologiczne UJK” 2019, nr 32), *O kilku judaikach w polskim socjolekcie przestępczym* („Linguistica Copernicana” 2019, nr 16).

Piotr Pęzik, dr hab., pracuje w Instytucie Anglistyki Uniwersytetu Łódzkiego. Zainteresowania badawcze: językoznawstwo korpusowe, językoznawstwo komputerowe, frazeologia, przetwarzanie

języka naturalnego, systemy wyszukiwawcze. Autor m.in. monografii *Facets of Prefabrication. Perspectives on Modelling and Detecting Phraseological Units* (Łódź 2018).

Joanna Przyklenk, dr, pracuje w Instytucie Językoznawstwa Uniwersytetu Śląskiego w Katowicach. Zainteresowania badawcze: historia języka polskiego, badania tekstu, stylu i dyskursu w ujęciu lingwistycznym, glottodydaktyka, lingwistyka cyfrowa, język artystyczny. Autorka monografii *Staropolska kronika jako gatunek mowy* (Katowice 2009) oraz wielu artykułów naukowych, m.in.: *Digital Humanities in Poland from the Perspective of the Historical Linguist of the Polish Language: Achievements, Needs, Demands* („Digital Scholarship in the Humanities” 2018, Vol. 33, Issue 4. Oxford University Press; współautorki: Magdalena Pastuch, Beata Duda, Karolina Lisczyk, Barbara Mitrenga, Katarzyna Sujkowska-Sobisz), *Distance toward Foreigners in the Polish Public Discourse of the Polish-Lithuanian Commonwealth* („East European & Balkan Studies” 2018, Vol. 42, No. 4), *Pytając o zmianę językową. Językoznawstwo historyczne a socjolingwistyka* (w: *Historia języka w XXI wieku. Stan i perspektywy*. Red. M. Pastuch, M. Siuciak, przy współpr. K. Wąsińskiej, W. Wilczek. Katowice 2018).

Katarzyna Romaniuk, lic., absolwentka komunikacji promocyjnej i kryzysowej (specjalności poradnictwo logopedyczne), obecnie studentka filologii polskiej oraz Studiów Podyplomowych Kwalifikacyjnych Logopedii i Medialnej Emisji Głosu na Uniwersytecie Śląskim w Katowicach oraz przewodnicząca i założycielka Koła Naukowego Młodych Logopedów Uniwersytetu Śląskiego. Zainteresowania naukowe: logopedia, zaburzenia płynności mowy.

Kinga Wąsińska, dr, pracuje w Instytucie Językoznawstwa Uniwersytetu Śląskiego w Katowicach. Zainteresowania badawcze: językoznawstwo diachroniczne, zmiany semantyczne, badania korpusowe, pragmalingwistyka historyczna. Autorka monografii *Słownictwo mentalne w historii polszczyzny. Studium słowotwórczo-leksykalne* (Katowice 2015), współredaktorka 5. numeru czasopisma „Forum Lingwistyczne” (2018) oraz autorka artykułów naukowych, m.in. *Przeobrażenia semantyczne wyrazów polskich związanych ze sferą mentalną* (Katowice 2016), *Wizerunek idealnego mężczyzny w świetle leksyki dawnej* (Kraków 2017). Brała udział w redakcji tomu *Historia języka w XXI wieku. Stan i perspektywy* (Katowice 2018).

Björn Wiemer, dr hab., pracuje w Uniwersytecie Johanna Gutenberga w Moguncji. Zainteresowania badawcze: aspektologia, morfoskładnia, modalność, kontakty językowe, zmiany językowe. Autor m.in. monografii: *Diskursreferenz im Polnischen und Deutschen – aufgezeigt an der narrativen Rede ein- und zweisprachiger Schüler* (München 1997), *Resultativa in den nordslawischen und baltischen Sprachen (Bestandsaufnahme unter arealen und grammatikalisierungstheoretischen Gesichtspunkten)* (München–Newcastle 2005; współautor: Markus Giger), *Catching the Elusive: Lexical Evidentiality Markers in Slavic Languages (a Questionnaire Study and Its Background)* (Berlin 2018).

Szymon Witczak, lic., absolwent komunikacji promocyjnej i kryzysowej (specjalności poradnictwo logopedyczne), obecnie student kultury mediów. Zainteresowania badawcze: marketing, komunikacja, media społecznościowe.

Joanna Wrześień-Kwiatkowska, mgr, pracuje w Uniwersytecie Johanna Gutenberga w Moguncji. Zainteresowania badawcze: lingwistyka korpusowa, statystyka, zarządzanie językiem.

Katarzyna Wyrwas, dr hab., prof. UŚ, pracuje w Instytucie Językoznawstwa Uniwersytetu Śląskiego w Katowicach. Zainteresowania badawcze: genologia lingwistyczna, leksyka, kultura języka. Autorka

Noty o Autorach

m.in. monografii: *Skarga jako gatunek mowy* (Katowice 2002), *Mały słownik terminów teorii tekstu* (Warszawa–Kraków 2005; współautorka: Katarzyna Sujkowska-Sobisz), www.poradniajęzykowa.pl (Katowice 2007, współautorstwo i redakcja), *Opowiadania potoczne w świetle genologii lingwistycznej* (Katowice 2014).

Monika Zaśko-Zielińska, dr hab., prof. UW, pracuje w Instytucie Filologii Polskiej Uniwersytetu Wrocławskiego. Zainteresowania badawcze: lingwistyka korpusowa, stylistyka, lingwistyka kryminalistyczna, współczesna polszczyzna mówiona. Autorka monografii *Przez okno świadomości: gatunki mowy w świadomości użytkowników języka* (Wrocław 2002) i *Listy pożegnalne: w poszukiwaniu lingwistycznych wyznaczników autentyczności tekstu* (Wrocław 2013) oraz wielu artykułów, m.in.: *Zastosowanie językoznawstwa korpusowego do badań nad dyskursem epistolarnym* (w: *Dyskurs i jego odmiany*. Red. B. Witosz, K. Sujkowska-Sobisz, E. Ficek. Katowice 2016), *Ocena stabilności ekspresywnej jednostek leksykalnych – propozycja opisu* („Orbis Linguarum” 2018, nr 51).

Redaktor tekstów w języku polskim

Barbara Konopka

Redaktor tekstów w języku rosyjskim

Swietłana Biczak

Redaktor techniczny

Małgorzata Pleśniar

Opracowanie graficzne, projekt okładki

AiR Anna Osadnik i Ryszard Stotko

Przygotowanie okładki do druku

Paulina Dubiel

Korektor

Lidia Szumigała

Łamanie

Edward Wilk

Publikacja objęta programem MNiSW „Wsparcie dla czasopism naukowych”

ISSN 2450-2758

(wersja elektroniczna)

Publikacja na licencji Creative Commons

Uznanie autorstwa-Na tych samych warunkach

4.0 Międzynarodowe (CC BY-SA 4.0)

<https://creativecommons.org/licenses/by-sa/4.0/deed.pl>



Wydawca

Wydawnictwo Uniwersytetu Śląskiego

ul. Bankowa 12B, 40-007 Katowice

www.wydawnictwo.us.edu.pl

e-mail: wydawnictwo@us.edu.pl

Wydanie I. Ark. druk. 12,25. Ark. wyd. 16,0.

Egzemplarz bezpłatny

Więcej o książce

ISSN 2450-2758

