

Kamil Wabnic

Uniwersytet Wrocławski
kamil.wabnic@uwr.edu.pl
ORCID: 0000-0001-9762-3420

Jan Krzywdziński

Uniwersytet Wrocławski
jan.krzywdzinski@uwr.edu.pl
ORCID: 0000-0002-4723-4619

SŁOWNIK FREKWENCYJNY W STULECIE URODZIN JEGO TWÓRCY

Prace nad polskim słownikiem frekwencyjnym rozpoczęły się w latach sześćdziesiątych dwudziestego wieku i zostały zwieńczone publikacją najpierw list frekwencyjnych w pięciu tomach (Kurcz i in. 1974–1977), a następnie dwutomowego *Słownika frekwencyjnego polszczyzny współczesnej* (Kurcz i in. 1990). Redaktorami list frekwencyjnych byli: Ida Kurcz, Andrzej Lewicki, Jadwiga Sambor i Jerzy Woronczak (w wypadku trzeciego tomu – Władysław Maśłowski). Słownik redagowali: Ida Kurcz, Andrzej Lewicki, Krzysztof Szafran, Jadwiga Sambor i Jerzy Woronczak.

Jednym z twórców słownika był Jerzy Woronczak, którego setną rocznicę urodzin obchodziliśmy w listopadzie 2023 roku. Z tej okazji postanowiliśmy przyrzeć się recepcji jego pracy, a także krótko opisać metody wykorzystywane w pracach nad słownikiem.

1. METODOLOGIA KORPUSU

Wielokrotnie poruszano już temat opracowania korpusu słownika frekwencyjnego oraz opisywano jego metodologię. Szczegółowo robią to m.in. Irena Kamińska-Szmaj (1990), Elżbieta Awramiuk (2002) czy Monika Czerepowicka i Zygmunt Saloni (2004).

Słownik został opracowany na podstawie zbioru tekstów pisanych pochodzących z lat 1963–1967. Teksty – według intencji twórców – obejmowały pięć stylów współczesnego (wówczas) pisanego języka polskiego: popularnonaukowy, publicystyczny,

drobnych wiadomości prasowych, prozy artystycznej i dramatu artystycznego. Teksty miały być reprezentatywne dla języka pisanego. Następnie na podstawie tak dobranego zbioru przygotowano korpus, który obejmował po 100 tys. wyrazów z każdego stylu. Następnie każdy subkorpus zrównoważono także wewnątrznie¹. Korpus został zatem na kilku poziomach poddany procesom, które miały na celu jego zrównoważenie (zob. Kurcz i in. 1974: 4–5). Ogółem korpus obejmował więc pół miliona jednostek; dla porównania: subkorpus zrównoważony NKJP (Przepiórkowski i in. 2012) liczy około 300 milionów segmentów. Przy takim porównaniu można stwierdzić, że na słownik frekwencyjny składało się mało segmentów, jednak należy mieć na uwadze ówczesne stadium rozwoju technologii informatycznej oraz pionierski charakter przeprowadzonych prac. Dla wskazania frekwencji i dystrybucji jednostek najczęstszych taki zbiór jest w zupełności wystarczający.

Porównanie propozycji definicyjnych i uwzględnienie aktualnego stanu wiedzy warunkuje zdefiniowanie korpusu językowego jako cyfrowo utrwalonego w celu przetwarzania zbioru danych językowych, który jest opatrzony metadanymi i anotacją. Teksty – jako składnik definicyjny – można postrzegać jako całe teksty lub ich fragmenty (por. Johns, Waller 2015: 5–6). Konieczność definicyjnego uwzględnienia anotacji łączy się ze zautomatyzowaną procedurą anotowania morfosyntaktycznego, co jest standardowym procesem we wczesnej obróbce danych. W definicjach korpusu pojawia się również zwykle warunek objętości zbioru – wskazuje on, że korpusy powinny liczyć tysiące lub miliony tokenów² (Hunston 2022: 5–6; Johns, Waller 2015: 5–6; Baker i in. 2006: 48–49). Istotniejsza od objętości jest jednak metodologiczna użyteczność korpusu³, która sprawia, że można na jego podstawie weryfikować i falsyfikować hipotezy badawcze oraz odpowiadać na pytania badawcze.

Korpus słownika frekwencyjnego został utrwalony cyfrowo. Teksty wchodzące w jego skład zostały wpisane do maszyny liczącej ODRA 1204 znajdującej się w Instytucie Matematycznym Uniwersytetu Wrocławskiego, były one również później znakowane przy pomocy maszyny ELLIOTT 803 (Kurcz i in. 1974: 35) znajdującej się w Instytucie Filologii Polskiej Uniwersytetu Wrocławskiego.

¹ Np. w wypadku subkorpusu stylu popularnonaukowego teksty dobrano tak, aby reprezentowały cztery typy nauk: społeczne, biologiczne, techniczne i matematyczno-fizyczne. Każdy typ stanowi 25% próbek tekstów (Kurcz 1974: 9–11).

² Liczba tokenów – lub też: segmentów, jednostek – to podstawowa miara objętości korpusu. Token jest rozumiany jako jakiś typ wyrazu – w zależności od stosowanej metodologii (zob. Hunston 2022: 5–6).

³ Zwykle użyteczność korpusu jest związana z warunkami jego zrównoważenia i reprezentatywności, jednakże nie zawsze są one odpowiednie lub wystarczające (zob. Hunston 2022: 30–37) i jako takie nie mogą stanowić elementów definiujących korpus. Zrównoważenie korpusu może podlegać zawieszeniu, np. gdy celem twórcy korpusu jest uzyskanie poświadczenia bardzo rzadkich jednostek leksykalnych (por. Zabawa 2019).

Próbki tekstów znajdujące się w korpusie są opisane metadanymi, które ukazały się w kolejnych tomach zawierających listy frekwencyjne (Kurcz i in. 1974–1977). Metadane te obejmują: autora i tytuł tekstu, rok wydania oraz liczbę próbek danej publikacji, które znajdują się w korpusie.

Korpus został również poddany ręcznej anotacji gramatycznej, która obejmowała znakowanie: dziewięciu klas gramatycznych⁴ oraz informacji morfologicznych (a w wypadku przyimków – także syntaktycznych) dostosowanych do poszczególnych klas (Kurcz i in. 1974: 40–60). Umożliwiło to oddzielną analizę frekwencji lematów i form.

2. DALSZE PRACE NAD KORPUSEM SŁOWNIKA FREKWENCYJNEGO

W późniejszym okresie – po przeniesieniu prac nad słownikiem do Warszawy w 1985 roku (Saloni 1991) – zmodyfikowano korpus oraz niektóre podstawowe założenia słownika, np. rozdzielono klasę partykuło-przysłówek na partykuły i przysłówki oraz włączono do słownika informacje ze słownika frekwencyjnego Haliny Zgółkowej (1983). Korpus posłużył do wydania na jego podstawie *Słownika frekwencyjnego polszczyzny współczesnej* (zob. Kurcz i in. 1990).

Następnie prace nad korpusem kontynuowano w celu udostępnienia go szerszemu gronu odbiorców. Wydano go na płycie CD oraz zamieszczono w internecie na licencji GNU (zob. Bień, Woliński 2001, 2003; Ogrodniczuk 2003a).

3. RECEPCJA SŁOWNIKA

Wyniki badań nad frekwencją słownictwa były wykorzystywane w licznych pracach badaczy z różnych subdyscyplin językoznawstwa. Badania te wywarły istotny wpływ na rozwój glottodydaktyki i stanowią materiał, na podstawie którego praktycy i teoretycy nauki polskiego jako obcego określają zasób słownictwa dla różnych poziomów znajomości polszczyzny (zob. Martyniuk 1991; Skrzypek 2009; Seretny 2011, 2018; Skrzypek, Singleton 2013; Zarzeczny 2014; Futera 2015; Mirocha 2019; Błasiak-Tytuła 2023; Shevchuk-Kliuzheva, Błasiak-Tytuła 2023). Ponadto informacje o frekwencji leksemów są wykorzystywane w pracach dotyczących edukacji i akwizycji języka (zob. Bakuła 1997; Dąbrowska 2008; Milewski 2009; Łobos 2013; Zgółkowa 2016).

Istotne pole badawcze stanowią również prace dotyczące komparatystyki języków (zob. Dukiewicz 1991; Kupść 2009; Mańczak 2010; Kwiatek 2014; Zydorowicz i in.

⁴ Rzeczowniki, przymiotniki, liczebniki, zaimki, czasowniki, przyimki, wykrzykniki, partykuło-przysłówki oraz spójniki.

2016), ich rozwoju (zob. Jankowiak 1997; Majewska 2006; Brehmer i in. 2017), interferencji międzyjęzykowych (zob. Witaszek-Samborska 1993; Wierzbička 1997; Mańczak-Wohlfeld 2005; Fałowski 2023) i przekładoznawstwa (zob. Sworowska 2023).

Słownik oraz listy frekwencyjne są także wykorzystywane jako źródło leksyki do badań i teorii z zakresu logopedii i pedagogiki, gdzie służą wyodrębnianiu wyrazów do terapii, diagnozy i stymulacji rozwoju mowy (zob. Grabias 1994, 1997; Perlak, Jarema 2003; Jodzio, Leszniewska-Jodzio 2006; Michalik 2006a, 2006b; Matuszczak i in. 2009; Jednoróg i in. 2010; Kaproń-Charzyńska, Kamper-Warejko 2016; Bieńkowska 2017). Podstawa takiego działania pozostała niezmienna – założono, że najczęstsze słownictwo jest słownictwem najbardziej podstawowym i najlepiej znanym przeciętnemu użytkownikowi języka. Również psychologia i kognitywistyka korzystają ze słownictwa uzyskiwanego na podstawie list frekwencyjnych – np. w badaniu pamięci, umysłowego przetwarzania języka (zob. Nowicka i in. 2006, 2009; Prochowicz 2007; Parzuchowski, Szymkow-Sudziarska 2008; Perlak i in. 2008; Śmiński 2010; Hanczakowski i in. 2013; Nieznański 2013, 2014; Kotlewska, Nowicka 2016) oraz osądu (Lu, Nieznański 2020; Lu i in. 2023).

Ponadto nie sposób wskazać wszystkich prac, w których dane dotyczące frekwencji leksemów lub ich poszczególnych form są kluczowe albo co najmniej niezwykle istotne dla dowodzenia naukowego. Słownik frekwencyjny oraz jego korpus stanowiły dla wielu prac podstawowe źródło zarówno danych frekwencyjnych, jak i przykładów występowania badanych zjawisk w rzeczywistych tekstach języka polskiego. Korpus słownika był namiastką korpusu referencyjnego przed stworzeniem Narodowego Korpusu Języka Polskiego (Przepiórkowski i in. 2012). Jako przykład takich prac można wskazać: dyskusję toczącą się wokół polskich liczebników głównych 1–4 oraz ich wymagań syntaktycznych (zob. Linde-Usiekniewicz, Rutkowski 2003, 2007; Rutkowski 2003; Hammarström 2004), opracowania z zakresu badania literatury (zob. Zembaty-Michalakowa 1982; Święcicka 2002; Sokólska 2005) czy też liczne prace Zygmunta Frajzyngiera, w których fragmenty z korpusu stanowią przykład różnych zjawisk językowych opisywanych z perspektywy językoznawstwa ogólnego na tle innych języków naturalnych (zob. Frajzyngier, Curl 2000a, 2000b; Frajzyngier, Shay 2003/2013; Frajzyngier 2005, 2008, 2010).

W funkcji korpusu referencyjnego korpus słownika był wykorzystywany w deskryptywnych badaniach korpusowych języka. Przykład mogą stanowić badania składniowe Ewy Jędrzejko (1993) i Magdaleny Derwojedowej (2000) oraz morfologiczne, np. dotyczące morfologii imiennej Wiesława Stefańczyka (2007, 2015), czy też supletyvizmu Bogdana Szymanka (2008). Ponadto, mimo że korpus gromadzi wyłącznie fragmenty tekstów pisanych, badacze upatrują w jego części opracowanej na podstawie tekstów dramatów możliwość badania języka mówionego (zob. Śledź 2001; Awramiuk 2002).

Korpus słownika wielokrotnie służył jako punkt odniesienia dla naukowców zajmujących się rozwojem badań korpusowych w Polsce. Często jest przywoływany

w pracach metodologicznych z zakresu językoznawstwa korpusowego (zob. Woliński, Przepiórkowski 2001; Kida 2002; Linde-Usiekniewicz, Łaziński 2003; Ogrodniczuk 2003b; Hajnicz i in. 2007; Lewandowska-Tomaszczyk 2011; Zaborowski, Przepiórkowski 2012; Woźniak 2017; Świdziński, Rutkowski 2022) oraz opisujących nowo powstałe narzędzia (zob. Miłkowski 2010; Mandera i in. 2015) i zasoby (zob. Pawłowski 2006; Szymczyk i in. 2012).

Bibliografia

- Przepiórkowski, A., Bańko, M., Górski, R.L., Lewandowska-Tomaszczyk, B. red. 2005. *Narodowy Korpus Języka Polskiego*. Warszawa: Wydawnictwo Naukowe PWN.
- Awramiuk, E. 2002. *O możliwościach wykorzystania Słownika frekwencyjnego polszczyzny współczesnej w badaniach nad słownictwem polszczyzny mówionej*. W: *Polszczyzna Mazowska i Podlasia, część VI, Słownictwo i frazeologia polszczyzny mówionej*, red. H. Sędziak, s. 43–50. Łomża: Łomżyńskie Towarzystwo Naukowe im. Wagów.
- Baker, P., Hardie, A., McEney, T. 2006. *A Glossary of Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Bakuła, K. 1997. *Kształcenie językowe w szkole podstawowej w świetle współczesnych teorii psychologicznych i lingwistycznych: projekt nowej metodyki*. Wrocław: Wydawnictwo Uniwersytetu Wrocławskiego.
- Bień, J., Woliński, M. 2003. *Wzbogacony korpus Słownika frekwencyjnego polszczyzny współczesnej*. W: *Prace językoznawcze dedykowane Profesor Jadwidze Sambor*, red. J. Linde-Usiekniewicz, R. Huszcza, s. 6–10. Warszawa: Wydział Polonistyki Uniwersytetu Warszawskiego.
- Bieńkowska, K. 2017. Wykorzystanie siatki pól semantyczno-leksykalnych w programie terapeutycznym służącym kształtowaniu języka. Propozycja. *Poradnik Językowy* 6, s. 84–95.
- Błasiak-Tytuła, M. 2023. Wspomaganie procesu adaptacji językowej dzieci z doświadczeniem migracji w nowej wspólnocie językowej. *Annales Universitatis Paedagogicae Cracoviensis. Studia ad Didacticam Litterarum Polonarum et Linguae Polonae Pertinentia* 14, s. 241–260. DOI: 10.24917/20820909.14.1.
- Brehmer, B., Kurbangulova, T., Winski, M. 2016. *Measuring lexical proficiency in Slavic heritage languages: A comparison of different experimental approaches*. W: *Slavic languages in psycholinguistics. Chances and challenges for empirical and experimental research*, red. T. Anstatt, A. Gattnar, C. Clasmeier, s. 225–256. Tübingen: Gunter Narr Verlag.
- Christian, J., Waller, D. 2015. *Corpus Linguistics for Grammar: A guide for research*. London – New York: Routledge.
- Dąbrowska, E. 2008. The later development of an early-emerging system: The curious case of the Polish genitive. *Linguistics* 46 (3), s. 629–650. DOI: 10.1515/LING.2008.021.
- Derwojedowa, M. 2000. *Porządek linearny składników zdania elementarnego w języku polskim*. Warszawa: Dom Wydawniczy Elipsa.
- Dukiewicz, L. 1991. *O niektórych wyznacznikach głosowych podobieństw i różnic między językami*. W: *Problemy opisu gramatycznego języków słowiańskich*, red. M. Grochowski, s. 35–38. Warszawa: Instytut Języka Polskiego PAN.
- Fałowski, A. 2023. Ile jest słów pochodzenia ukraińskiego we współczesnej polszczyźnie ogólnej? *LingVaria* 18 (1(35)), s. 63–72. DOI: 10.12797/LV.18.2023.35.04.

- Frajzyngier, Z. 2005. Grammaticalization of switch reference. *Studies in African Linguistic Typology* 64, s. 113–130. DOI: 10.1075/tsl.64.07fra.
- Frajzyngier, Z. 2008. Grammaticalization of propositional and phrasal relators in Chadic. *Afrika und Übersee* 88 (1–2), s. 79–102.
- Frajzyngier, Z. 2010. Towards a non-aprioristic syntactic theory. *Faits des langues. Les Cahiers: Revue de Linguistique* 2, s. 9–40. DOI: 10.1163/19589514-035-036-02-900000003.
- Frajzyngier, Z., Curl, T.S. 2000a. *Reflexives: Forms and functions. Volume 1*. Amsterdam – Philadelphia: John Benjamins Pub Co.
- Frajzyngier, Z., Curl, T.S. 2000b. *Reciprocals: Forms and functions. Volume 2*. Amsterdam – Philadelphia: John Benjamins Pub Co.
- Frajzyngier, Z., Shay, E. 2003. *Explaining language structure through systems interaction*. Amsterdam: John Benjamins Publishing Company. DOI: 10.1075/tsl.55 (wyd. polskie: 2013, *Współdziałanie podsystemów w strukturze języka*, Warszawa: Wydawnictwa Uniwersytetu Warszawskiego).
- Futera, A. 2015. *Pollyanna czy polskiego – dobry, dobrze, zły i źle w niektórych podręcznikach do języka polskiego jako obcego*. W: *Bogactwo językowe i kulturowe Europy w oczach Polaków i cudzoziemców*. Łódź: Wydawnictwo Uniwersytetu Łódzkiego. DOI: 10.18778/8088-034-4.05.
- Grabias, S. 1994. *Język w nauczaniu niesłyszących. Zasady programowania systemu komunikacyjnego*. W: *Głuchota a język*, red. S. Grabias, s. 185–221. Lublin: Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej.
- Grabias, S. 1997. *Język w zachowaniach społecznych: podstawy socjolingwistyki i logopedii*, wyd. 2. Lublin: Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej.
- Hajnicz, E., Dębowski, Ł., Wiech, M. 2007. Przykładowe zastosowanie gradacyjnej analizy danych w badaniach lingwistycznych. *Prace Instytutu Podstaw Informatyki Polskiej Akademii Nauk* 1005, s. 1–40.
- Hammarström, H. 2004. Properties of lower numerals and their explanation: A reply to Paweł Rutkowski. *Journal of Universal Language* 5 (2), s. 1–20.
- Hanczakowski, M., Pasek, T., Zawadzka, K., Mazzoni, G. 2013. Cue familiarity and ‘don’t know’ responding in episodic memory tasks. *Journal of Memory and Language* 69 (3), s. 368–383.
- Hunston, S. 2022. *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9781139524773.
- Jankowiak, L.A. 1997. *Prasłowiańskie dziedzictwo leksykalne we współczesnej polszczyźnie ogólnej (na podstawie „Słownika Prasłowiańskiego” t. 1–7)*. Warszawa: Slawistyczny Ośrodek Wydawniczy.
- Jednoróg, K., Marchewka, A., Tacikowski, P., Grabowska, A. 2010. Implicit phonological and semantic processing in children with developmental dyslexia: Evidence from event-related potentials. *Neuropsychologia* 48 (9), s. 2447–2457.
- Jędrzejko, E. 1993. *Nominalizacje w systemie i w tekstach współczesnej polszczyzny*. Katowice: Wydawnictwo Uniwersytetu Śląskiego.
- Jodzio, K., Leszniewska-Jodzio, B. 2006. Wybiórczość zaburzeń leksykalno-semantycznych w afazji na przykładzie rozumienia pojedynczych słów. *Czasopismo Psychologiczne* 12 (2), s. 155–164.
- Kaleta, Z. 1995. *Gramatyka języka polskiego dla cudzoziemców*. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.
- Kaproni-Charzyńska, I., Kamper-Warejko, J. 2016. *Kształtowanie się systemu językowego w zakresie kategorii przypadku rzeczowników u dzieci 5- i 6-letnich*. Toruń: Wydawnictwo Naukowe Uniwersytetu Mikołaja Kopernika.

- Kida, J. 2002. Lingwistyka matematyczna a badania językowo-stylistyczne. W: *Składnia, stylistyka, struktura tekstu: księga jubileuszowa dedykowana Profesor Teresie Ampel*, red. M. Krauz, K. Ożóg, s. 112–124. Rzeszów: Wydawnictwo Uniwersytetu Rzeszowskiego.
- Kotłowska, I., Nowicka, A. 2016. Present-self, past-self and the close-other: neural correlates of assigning trait adjectives to oneself and others. *European Journal of Neuroscience* 44 (4), s. 2064–2071.
- Kupść, A. 2009. Adjectives: Constructions vs. Valence. *Aspects of Natural Language Processing: Essays Dedicated to Leonard Bolc on the Occasion of His 75th Birthday*, s. 241–269.
- Kurcz, I., Lewicki, A., Sambor, J., Woronczak, J. 1974–1977. *Słownictwo współczesnego języka polskiego: listy frekwencyjne*, t. 1, 2, 4, 5. Warszawa: Instytut Języka Polskiego PAN.
- Kurcz, I., Lewicki, A., Sambor, J., Szafran, K., Woronczak, J. 1990. *Słownik frekwencyjny polszczyzny współczesnej*, t. 1–2. Kraków: Instytut Języka Polskiego PAN.
- Kwiatkiewicz, E. 2014. *Contrastive analysis of English and Polish surveying terminology*. Cambridge: Cambridge Scholars Publishing.
- Lewandowska-Tomaszczyk, B. 2011. Nowe wyzwania w jakościowej i ilościowej metodologii analizy języka. *Biuletyn Polskiego Towarzystwa Językoznawczego* 67, s. 139–164.
- Lewicki, A., Masłowski, W., Sambor, J., Woronczak, J. 1975. *Słownictwo współczesnego języka polskiego: listy frekwencyjne*, t. 3. Warszawa: Instytut Języka Polskiego PAN.
- Linde-Usiekiewicz, J., Łaziński, M. 2003. „Słownik frekwencyjny polszczyzny współczesnej” w badaniach nad słownictwem polszczyzny dzisiejszej. W: *Prace językoznawcze dedykowane Profesor Jadwidze Sambor*, red. J. Linde-Usiekiewicz, R. Huszcza, s. 112–124. Warszawa: Wydział Polonistyki UW.
- Linde-Usiekiewicz, J., Rutkowski, P. 2003. O zależności między właściwościami składniowymi i frekwencyjnymi liczebników głównych w polszczyźnie. W: *Prace językoznawcze dedykowane Profesor Jadwidze Sambor*, red. J. Linde-Usiekiewicz, R. Huszcza, s. 125–140. Warszawa: Wydział Polonistyki UW.
- Linde-Usiekiewicz, J., Rutkowski, P. 2007. *Frequency of Use of Polish Numerals Does Not Influence Their Syntax*. W: *Proceedings of the 22nd Annual Northwest Linguistics Conference*, red. N. Carter, L. Hadic-Zabala, A. Rimrott, D.R. Storoshenko, s. 175–86. Burnaby: Simon Fraser University.
- Lu, Y., Nieznański, M. 2020. The base rate neglect in episodic memory. *Memory* 28 (2), s. 270–277.
- Lu, Y., Nieznański, M., Obidziński, M. 2023. The Effects of Expected Value and Episodic Memory on Preference Reversals. *Review of Behavioral Economics* 10 (2), s. 105–161.
- Łobos, A. 2013. *Czytelność tekstu w wybranych podręcznikach do nauczania muzyki w szkole podstawowej*. W: *Odmiany polszczyzny w szkole: teoria i praktyka*, red. H. Synowiec, M. Kubarek, s. 202–219. Katowice: Wydawnictwo Uniwersytetu Śląskiego.
- Majewska, M.B. 2006. *Rzeczownikowe homonimy heterogeniczne: analiza synchroniczna i diachroniczna*. Kraków: Wydawnictwo LEXIS.
- Mandera, P., Keuleers, E., Wodniecka, Z., Brysbaert, M. 2015. Subtlex-pl: subtitle-based word frequency estimates for Polish. *Behavior Research Methods* 47, s. 471–483.
- Mańczak, W. 2010. Rzekoma różnica między zaimkami celtyckimi a polskimi. *Linguistica Copernicana* 4 (2), s. 35–42.
- Mańczak-Wohlfeld, E. 2005. English borrowings in two contemporary Polish lexicons of anglicisms. *Studia Turcologica Cracoviensia* 10, s. 269–277.

- Martyniuk, W. 1991. *Grundbaustein Polnisch. Propozycja programu nauczania języka polskiego jako obcego*. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.
- Matuszczak, M., Krejtz, I., Orylska, A., Bielecki, M. 2009. Trening pamięci operacyjnej u dzieci z zespołem hiperkinetycznym (ADHD) (Working memory training in children with ADHD). *Czasopismo Psychologiczne* 15, s. 87–103.
- Michalik, M. 2006a. *Jak diagnozować język ucznia? – propozycje dla logopedów i nauczycieli*. *Studia Logopaedica* 1, s. 490–521.
- Michalik, M. 2006b. Konteksty pragmatyczne definicji dziecięcych w dyskursie zaburzo- nym. *Studia Logopaedica* 1, s. 169–182.
- Milewski, S. 2009. Wyrazy n-fonemowe i n-sylabowe w mowie dorosłych adresowanej do nie- mowląt. *Psychologia Rozwojowa* 14 (2), s. 43–56.
- Miłkowski, M. 2010. Developing an open-source, rule-based proofreading tool. *Software: Prac- tice and Experience* 40 (7), s. 543–566.
- Mirocha, P. 2019. Listy frekwencyjne jako narzędzie ewaluacji leksyki w podręcznikach do języka polskiego jako obcego. Analiza i ocena podręcznika *Krok po kroku 1*. *Kwartalnik Polonicum* 31/32, s. 38–46.
- Nieznański, M. 2013. Effects of resource demanding processing on context memory for context- -related versus context-unrelated items. *Journal of Cognitive Psychology* 25 (6), s. 745–758.
- Nieznański, M. 2014. The role of reinstating generation operations in recognition memory and reality monitoring. *Polish Psychological Bulletin* 45 (3), s. 363–371.
- Nowicka, A., Jednoróg, K., Wypych, M., Marchewka, A. 2009. Reversed old/new effect for inten- tionally forgotten words: An ERP study of directed forgetting. *International Journal of Psy- chophysiology* 71 (2), s. 97–102.
- Nowicka, A., Marchewka, A., Szatkowska, I. 2006. Lateralization of repetition effects in event- related potentials to words in left- and right-handed women. *Neuroscience Letters* 393 (2–3), s. 150–154.
- Ogrodniczuk, M. 2003a. *Nowa edycja wzbogaconego korpusu słownika frekwencyjnego*. W: *Języko- znawstwo w Polsce. Stan i perspektywy*, red. S. Gajda, s. 181–190. Opole: Uniwersytet Opolski.
- Ogrodniczuk, M. 2003b. *Rozszerzenie opisów morfologicznych w tekstach korpusu Słownika frekwencyjnego polszczyzny współczesnej*. W: *Prace językoznawcze dedykowane Profesor Jadwidze Sambor*, red. J. Linde-Usiekiewicz, R. Huszcza, s. 125–140. Warszawa: Wydział Polonistyki UW.
- Parzuchowski, M., Szymkow-Sudziarska, A. 2008. Well, slap my thigh: Expression of surprise fa- cilitates memory of surprising material. *Emotion* 8 (3), s. 430–434.
- Pawłowski, A. 2006. Chronological analysis of textual data from the Wrocław Corpus of Polish. *Poznań Studies in Contemporary Linguistics* 41, s. 9–29.
- Perlak, D., Jarema, G. 2003. The recognition of gender-marked nouns and verbs in Polish-spe- aking aphasic patients. *Cortex* 39 (3), s. 383–403.
- Perlak, D., Feldman, L.B., Jarema, G. 2008. Defining regularity: Does degree of phonological and orthographic similarity among Polish relatives influence morphological processing? *The Mental Lexicon* 3 (2), s. 239–258.
- Prochwicz, K. 2007. Dostępność znaczeń w warunkach powtarzalnej prezentacji. Badania nad zjawiskiem wysycania semantycznego. *Studia Psychologica* 2, s. 142–155.
- Rutkowski, P. 2003. On the Universal Neuropsychological Basis of the Syntax of Numerals. *Journal of Universal Language* 4 (2), s. 147–182.

- Saloni, Z. 1991. Słownik frekwencyjny polszczyzny współczesnej. *ComputerWorld* 1 listopada 1991. Online: <https://www.computerworld.pl/news/Sloownik-frekwencyjny-polszczyzny-wspolczesnej,315352.html> [dostęp: 28.05.2024].
- Seretny, A. 2011. *Kompetencja leksykalna uczących się języka polskiego jako obcego w świetle badań ilościowych*. Kraków: Wydawnictwo UJ.
- Seretny, A. 2018. Badania leksykalne w (glotto)dydaktyce polonistycznej. *Język Polski* 98 (2), s. 42–60. DOI: 10.31286/JP.98.2.3.
- Shevchuk-Kliuzheva, O., Błasiak-Tytuła, M. 2023. The vocabulary skills of Ukrainian pre-schoolers in Poland. *Cognitive Studies / Études cognitives* 23, s. 1–12.
- Sibiga, Z. 1999. *Sprawność językowa uczniów szkół polonijnych Stanów Zjednoczonych*. Rzeszów: Wydawnictwo Wyższej Szkoły Pedagogicznej.
- Skrzypek, A. 2009. Phonological short-term memory and L2 collocational development in adult learners. *EUROSLA Yearbook* 9 (1), s. 160–184.
- Skrzypek, A., Singleton, D. 2013. *Phonological Short-Term Memory and the Operation of Cross-Linguistic Factors in Elementary and Pre-Intermediate Adult L2 Learners' Collocational Usage*. W: *Language Learning, Discourse and Communication: Studies in Honour of Jan Majer*, red. Ł. Szubko-Sitarek, P. Salski, P. Stalmaszczyk, s. 193–214. Cham – Heidelberg – New York – Dordrecht – London: Springer.
- Sokółska, U. 2005. *Leksykalno-stylistyczne cechy prozy Melchiora Wańkowicza (na materiale reportaży z lat 1961–1974)*. Białystok: Wydawnictwo Uniwersytetu w Białymstoku.
- Stefańczyk, W.T. 2007. *Kategoria rodzaju i przypadku polskiego rzeczownika: próba synchronicznej analizy morfologicznej*. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.
- Stefańczyk, W.T. 2015. Zmiany we współczesnym polskim systemie fleksyjnym (na materiale rzeczownikowym i przymiotnikowym). *LingVaria* 20, s. 97–106.
- Sworowska, A. 2023. *Językowe wykładniki emocji we współczesnych szwedzkich tekstach literackich i ich polskich przekładach*.
- Szymanek, B. 2008. A New Look at Suppletion in Polish. *Roczniki Humanistyczne* 56 (5), s. 75–92.
- Szymczyk, B., Zakowicz, W., Stemplewska-Zakowicz, K. 2012. Automatyczna analiza tekstu: polska adaptacja programu LIWC Jamesa Pennebaker. *Przegląd Psychologiczny* 55 (2), s. 195–209.
- Śledź, A. 2001. Przerwywnikowe użycia pewnych wariantów słowa *wiedzieć* w mowie potocznej. *Prace Językoznawcze Uniwersytetu Warmińskiego-Mazurskiego* 3, s. 135–148.
- Śmigórski, K. 2010. Różnice w przebiegu procesów poznania utajonego u osób wysoce i nisko uzdolnionych twórczo. Badania w paradygmacie maskowanego poprzedzania semantycznego. *Przegląd Psychologiczny* 53 (4), s. 497–510.
- Świdziński, M., Rutkowski, P. 2022. Korpus ogólny jako model danego języka naturalnego: korpusy języków fonicznych a korpus polskiego języka migowego. *Poradnik Językowy* 792 (3), s. 7–22.
- Święcicka, M. 2002. Kreatywne budowanie wypowiedzi potocznej w utworach literackich dla młodzieży. W: *Czynności tworzenia i rozumienia wypowiedzi*, red. J. Porayski-Pomsta, s. 307–323. Warszawa: Dom Wydawniczy Elipsa.
- Wierzbicka, A. 1997. *Understanding cultures through their key words: English, Russian, Polish, German, and Japanese*. New York – Oxford: Oxford University Press.
- Witaszek-Samborska, M. 1993. *Zapóżyczenia z różnych języków we współczesnej polszczyźnie: na podstawie słowników frekwencyjnych*. Poznań: Wydawnictwo Poznańskiego Towarzystwa Przyjaciół Nauk.

- Woliński, M., Przepiórkowski, A. 2001. *Projekt anotacji morfosyntaktycznej korpusu języka polskiego*. Warszawa: Instytut Podstaw Informatyki PAN.
- Woźniak, M. 2017. *Jak znaleźć igłę w stogu siana. Automatyczna ekstrakcja wielosegmentowych jednostek leksykalnych z tekstu polskiego*. Kraków: Instytut Języka Polskiego PAN.
- Zabawa, M. 2019. O roli samodzielnie przygotowanych korpusów w badaniach językoznawczych (na przykładzie korpusu wykorzystującego zasoby internetowe). *Półrocznik Językoznawczy Tertium* 4 (1), s. 212–232. DOI: 10.7592/10.7592/Tertium2019.4.1.Zabawa.
- Zaborowski, B., Przepiórkowski, A. 2012. Tagset conversion with decision trees. *Advances in Natural Language Processing: 8th International Conference on NLP, JapTAL 2012, Kanazawa, Japan, October 22-24, 2012. Proceedings*, s. 144–155.
- Zarzewny, G. 2014. *Leksyka w ujęciu ilościowym (na materiale wybranych wrocławskich podręczników)*. W: *40 lat wrocławskiej glottodydaktyki polonistycznej. Teoria i praktyka*, red. A. Dąbrowska, U. Dobesz, s. 373–384. Wrocław: Oficyna Wydawnicza ATUT.
- Zembały-Michalakowa, M. 1982. *Poezja Juliana Przybosa w świetle badań statystyczno-językowych na tle porównawczym*. Wrocław: Wydawnictwo Uniwersytetu Wrocławskiego.
- Zgółkowska, H. 1983. *Słownictwo współczesnej polszczyzny mówionej. Lista frekwencyjna i rangowa*. Poznań: Wydawnictwo Naukowe Uniwersytetu Adama Mickiewicza.
- Zgółkowska, H. 2016. Słownictwo dzieci w wieku przedszkolnym. Dwa pokolenia – dwa światy. *Poradnik Językowy* 10, s. 9–19.
- Zygorowicz, P., Orzechowska, P., Jankowski, M., Dziubalska-Kończyk, K., Wierzchoń, P., Pietrala, D. 2016. *Phonotactics and morphonotactics of Polish and English. Theory, Description, Tools and Applications*. Poznań: Wydawnictwo Naukowe Uniwersytetu Adama Mickiewicza.

The Polish frequency dictionary on the centenary of its creator's birth

Summary

The article aims at discussing the reception of *Słownik frekwencyjny polszczyzny współczesnej* [‘The frequency dictionary of contemporary Polish’] and of the frequency lists published in *Słownictwo współczesnego języka polskiego*. The article pays particular attention to the current impact of the work, which was originally inspired, among others, by Jerzy Woronczak. The influence of the frequency dictionary is observable in various fields and subdisciplines of linguistics, including glottodidactics, comparative studies, speech therapy, pedagogy, psychology, and cognitive science. Additionally, the article outlines the general methodological assumptions of the corpus of the frequency dictionary, which is presented as an early form of corpus linguistics.

Keywords: corpus linguistics – frequency dictionary – Jerzy Woronczak – language corpus – frequency list.

Adj. Marta Falkowska