

Tadeusz Lewaszkiewicz   
Uniwersytet im. Adama Mickiewicza, Poznań  
[lewtaad@interia.pl](mailto:lewtaad@interia.pl)

## PROBLEM UTOŻSAMIANIA LICZBY PORZĄDKOWEJ Z RANGĄ W POLSKICH SŁOWNIKACH FREKWENCYJNYCH

**Słowa kluczowe:** językoznawstwo statystyczne, słowniki frekwencyjne, utożsamianie liczb porządkowych z rangą wyrazów

**Keywords:** statistical linguistics, frequency dictionaries, equation of ordinal numbers and word rank

Językoznawstwo od dawna korzysta w badaniach z metod statystycznych. Pierre Guiraud (1966), wybitny znawca statystyki językoznawczej, napisał na początku pierwszego rozdziału swojej książki:

Statystyka językoznawcza jest spadkobierczynią tradycji bardzo starych: już gramatycy aleksandryjscy opracowali wykaz hapax legomena Homera, a masoreci policzyli wszystkie wyrazy Biblii. Od XIX wieku, kiedy to powstała gramatyka historyczna, badania języków oparto na ogromnych inwentarzach liczbowych, które od 20 lat zastąpiła wyrozumowana analiza statystyczna (Guiraud 1966: 9).

Ocena częstości występowania elementów językowych dotyczy wszystkich podsystemów języka: fonetyczno-fonologicznego, morfologicznego, składniowego i leksykalnego. Badania statystyczne obejmują również zagadnienie stylu, metrykę, wersyfikację itd. W językoznawstwie polonistycznym dane statystyczne wykorzystywało w swoich badaniach wielu lingwistów, np. Rolf Hammerl i Jadwiga Sambor (1990), Władysław Kuraszkiewicz (1932, 1958, 1966), J. Sambor (1969, 1971, 1972, 1974a, 1974b, 1975), Maria Zarębina (1977, 1978). Informacje statystyczne zawierają różne słowniki, np. słowniki pisarzy (jak SJAM), słowniki okresów historycznojęzycznych (np. SPXVI) itd.

Specjalny typ słowników z informacjami statystycznymi stanowią słowniki frekwencyjne. Dysponuje nimi wiele języków świata. Różnią się one bardzo pod względem koncepcji. Ich charakterystyka nie jest celem artykułu<sup>1</sup>. Ograniczę się do omówienia i oceny list rangowych w poszczególnych tomach: *Słownictwa współczesnego języka polskiego. List frekwencyjnych* (t. I–V, Warszawa 1972/1974–1977, dalej: SWJP LF)<sup>2</sup> oraz *Słownika frekwencyjnego polszczyzny współczesnej* (t. I–II, Kraków 1990, dalej: SFPW), ponieważ interesuje mnie struktura list rankingowych w tych słownikach, zwłaszcza listy rangowe sporządzone według wskaźnika F, tj. ze względu na częstość absolutną haseł wyrazowych. **Każdy tom oparty jest na materiale 100 tysięcy słów, które wylosowano z obszernych kanonów źródeł. Łącznie takie próbki stanowią 500 tysięcy słów.**

Podstawę materiałową SFPW stanowi pięć tomów SWJP LF (tj. 500 tysięcy słów), SWPP LF oraz słownik frekwencyjny Haliny Zgółkowej (1983). Poszczególne tomy SWJP LF obejmują: t. I – teksty popularnonaukowe; t. II – drobne wiadomości prasowe; t. III – publicystykę; t. IV – prozę artystyczną; t. V – dramat artystyczny. Publikacja H. Zgółkowej (1983) dotyczy polszczyzny potocznej<sup>3</sup>.

Ida Kurcz zajmowała się psychologią i psycholingwistyką. Andrzej Lewicki reprezentował językoznawstwo polonistyczne i ogólne. J. Sambor była główną przedstawicielką językoznawstwa statystycznego w Polsce; ukończyła studia ekonomiczne, w ich ramach zaliczyła kurs wyższej matematyki i zajęcia z metod statystycznych. Jerzy Woronczak był literaturoznawcą i leksykografem; propagował metody matematyczne w badaniach literaturoznawczych. Władysław Masłowski pracował jako dziennikarz, stenografista i prasoznawca. H. Zgółkowa jest autorką publikacji z różnych dziedzin językoznawstwa polonistycznego. SFPW zredagowali: I. Kurcz, A. Lewicki, J. Sambor, Krzysztof Szafran i J. Woronczak. K. Szafran zajmuje się lingwistyką komputerową i przetwarzaniem języka naturalnego. W okresie prac nad SFPW był młodym pracownikiem naukowym, nie miał jeszcze doktoratu, który uzyskał w 1994 r. na podstawie rozprawy doktorskiej przygotowanej pod kierunkiem prof. Zygmunta Saloniego. Redaktorem naukowym SFPW był właśnie

1 W niniejszym artykule nie zamierzam oceniać całości słowników frekwencyjnych (w tym list frekwencyjnych), lecz tylko listy rangowe.

2 W 1972 r. ukazał się tom sygnałowy, który w 1975 r. stał się tomem trzecim.

3 Słownik frekwencyjny H. Zgółkowej (1983) oparty jest również na materiale 100 tysięcy słów, które nie pochodzą jednak z wylosowanych próbek, lecz z kanonu tekstów o objętości 100 tysięcy wyrazów. Z punktu widzenia językoznawstwa statystycznego materiał ten jest mniej wartościowy niż materiał zawarty w pięciu tomach SWJP LF. Brak jednak wzmianki na ten temat w *Przedmowie* i *Wstępie* do SFPW. Mimo to wzmiankowany słownik polszczyzny potocznej zasługuje na pozytywną ocenę, ponieważ wykonanie go z zastosowaniem metody autorów SWJP LF (którzy dodatkowo korzystali z pomocy współpracowników) niewątpliwie przerosłoby możliwości czasowe jednej osoby. H. Zgółkowa jest także autorką i współautorką innych słowników frekwencyjnych polszczyzny (Zgółkowa, Bułczyńska 1987; Zgółkowa, Zgółka, Szymoniak 1991; Zgółkowa 2016). W tych słownikach materiał 100 tysięcy wyrazów również nie pochodzi z próbek losowych.

Z. Saloni. Uczony ten studiował zarówno polonistykę i językoznawstwo, jak i matematykę. W poszczególnych tomach SWJP LF nie podano nazwiska redaktora naukowego; prawdopodobnie nikt nie pełnił takiej funkcji.

W drugim tomie SFPW znajdują się trzy listy rangowe. Ich strukturę autorzy słownika objaśniają w ten sposób:

Budowa wszystkich list rangowych jest identyczna; różnią się one tym, ze względu na jaki wskaźnik uporządkowane są zamieszczone w nich hasła: F (częstość absolutna), U (częstość względna) i D (wskaźnik równomierności i rozkładu w stylach). Na wszystkich listach hasła uporządkowano według danego wskaźnika w porządku malejącym. Na pozycję listy rangowej składają się trzy elementy: numer kolejny (ranga), nazwa hasła oraz wartość wskaźnika. W wypadku, gdy konkretna wartość wskaźnika przyporządkowana jest kilku hasłom, o ich kolejności (i związanych z nią numerach na liście) decyduje porządek alfabetyczny (SFPW I: lii–liii).

Wskaźnik D stosuje się w tych słownikach frekwencyjnych, które oparte są na jednakowych statystycznie próbach słownictwa reprezentujących różne style funkcjonalne.

Poważne zastrzeżenia budzi przyjęte założenie, że przy tej samej wartości wskaźnika F, U lub D o randze hasła (tj. numerach na liście) „decyduje porządek alfabetyczny”. Swoje wątpliwości zilustruję wybranymi danymi z *Listy rangowej według wskaźnika F (Częstość absolutna)* (SFPW II: 802–859).

Do pozycji 68 na liście rangowej (zajmuje ją wyraz *drugi* – 673 poświadczenia) znajduje się 68 haseł, ponieważ każde hasło cechuje odrębna frekwencja. Pozycję 69 zajmuje słowo *jednak*, zaś pozycję 70 – *pod*. Oba wyrazy mają po 668 poświadczeń, ale o różnych miejscach na liście rangowej zadecydował porządek alfabetyczny. Tak samo jest przy kolejnych hasłach o tej samej frekwencji, np. 88 *człowiek*, 89 *jakiś* – częstość absolutna obu haseł: 500. Przy niższych frekwencjach haseł o tej samej częstości przybywa – ze względu na porządek alfabetyczny mają one na liście rangowej rangę niższą od kilku do bardzo wielu miejsc, np. hasło o frekwencji 56 jest 17 (ze względu na kolejność alfabetyczną wyraz *atmosfera* ma rangę 1148, zaś *zdarzać się* 1164 – o 17 pozycji niższą). Hasła o frekwencji 5 zajmują na liście rangowej pozycje od 7623 (*adiutant*) do 8738 (*żużel*) – różnica 1115 miejsc na liście rangowej; hasła o frekwencji 4 znajdują się na pozycjach od 8739 (*abstrahować*) do 10 355 (*żywot*) – różnica 1616 miejsc na liście rangowej. Różnice na liście rangowej haseł o frekwencji 3, 2 i 1 byłyby znacznie większe. Taką samą strukturę mają listy rangowe według wskaźnika F w pięciu tomach SWJP LF oraz w słownikach: H. Zgólkowej (1983), H. Zgólkowej i Katarzyny Bułczyńskiej (1987), H. Zgólkowej, Tadeusza Zgółki i Krzysztofa Szymoniaka (1991), H. i T. Zgólków (1992). Duże różnice dotyczą również haseł na listach rangowych według wskaźników U i D w SFPW.

Poszczególne tomy SWJP LF przejrzałem powierzchownie już w latach 70. ubiegłego wieku. Wówczas nie zainteresował mnie problem struktury list rangowych

i roli porządku alfabetycznego haseł w określaniu ich pozycji rangowej. Kiedy po 1990 r. (prawdopodobnie w 1991 lub 1992 r.) zapoznałem się z dwoma tomami SFPW, byłem zaskoczony, że hasła o tej samej frekwencji mają różną rangę ze względu na porządek alfabetyczny. Nie uznałem tego za błędną decyzję, ponieważ wydawało mi się, że autorytet zespołu autorskiego słownika jest gwarancją poprawności przyjętej koncepcji list rangowych. Wziąłem też pod uwagę fakt, że redaktorem naukowym SFPW jest Z. Saloni (wybitny językoznawca, który studiował także matematykę), jak również w finalizacji przygotowań do druku tego słownika uczestniczyli pracownicy Instytutu Informatyki Uniwersytetu Warszawskiego – Janusz Bień i K. Szafran.

Moje wątpliwości w sprawie poprawności struktury list rangowych stopniowo narastały, ale wciąż nie byłem pewien, czy mam rację. Przed rokiem 2010 zaplanowałem w Instytucie Filologii Słowiańskiej UAM serię prac magisterskich o anglicyzmach we współczesnej prasie czeskiej, bułgarskiej, serbskiej i chorwackiej. Kilka-naście osób miało wyekscerpować z dzienników lub periodyków słowiańskich ok. 5 tysięcy anglicyzmów (tj. również wyrazów powtarzających się), ułożyć słownik frekwencyjny anglicyzmów i dokonać odpowiedniego opisu materiału leksykalnego. Uznałem wówczas, że przyporządkowanie hasłom o tej samej frekwencji różnych rang ze względu na porządek alfabetyczny to decyzja całkowicie nieracjonalna. Wobec tego zaleciłem magistrantom, aby przygotowali dwie listy rangowe: listę rangową A (wzorowaną na SWJP LF i SFPW) oraz listę rangową B (tj. przypisującą hasłom o tej samej frekwencji identyczną rangę). Przypuszczałem, że moje stanowisko jest słuszne, ale nie wiedziałem, jak dowieść, że autorzy i współpracownicy SWJP LF oraz SFPW popełnili błąd. Ostatecznie doszedłem do wniosku, że istotne w ocenie struktury list rangowych SWJP LF i SFPW może być porównanie koncepcji zastosowanej w tych słownikach z koncepcjami w słownikach frekwencyjnych języków słowiańskich i niesłowiańskich. Jednak w ciągu najbliższych kilkunastu lat nie przeprowadziłem tych badań.

Pod koniec listopada 2022 r. odbyła się obrona pracy doktorskiej Marka Chlebowskiego *Angielskie zapożyczenia leksykalne i derywaty od anglicyzmów leksykalnych w języku górnołużyckim (na tle angielsko-czeskich i angielsko-polskich kontaktów językowych)*. Jako promotor również i w tej pracy zaproponowałem zamieszczenie list rangowych A i B według wskaźnika F. Lista rangowa typu A znalazła się w pracy doktorskiej (podobnie jak w kilkunastu pracach moich magistrantów) tylko dlatego, że lista rangowa o takiej koncepcji funkcjonowała w polskim językoznawstwie statystycznym jako spełniająca warunki naukowości.

W recenzji wzmiankowanej pracy doktorskiej prof. Henryk Duda (KUL) zwrócił uwagę, że lista rangowa A jest właściwie zbędna, ponieważ porządek alfabetyczny nie może decydować o randze konkretnego hasła. Krytyczna uwaga była słuszna, ale jednak nie zostało przekonująco podważone stanowisko nauki polskiej w tej kwestii. Wobec tego postanowiłem zrealizować mój dawny plan porównania koncepcji polskich list rangowych (w szczególności *Listy rangowej według wskaźnika F* w SFPW)

z koncepcjami list rangowych słowiańskich i niesłowiańskich słowników frekwencyjnych. Gdyby się okazało, że słowniki frekwencyjne innych języków przyznają hasłom o tej samej frekwencji identyczną rangę, to wówczas stanowisko polskie byłoby odosobnione i należałoby poddać je krytyce. Ograniczę się do ukazania struktury list rangowych sporządzonych ze względu na wskaźnik F (częstość absolutną).

Opublikowano zaledwie kilka recenzji SWJP LF oraz SFPW. Pięć dotyczyło tomu poświęconego tekstom publicystycznym, który wydano w 1972 r. (taka data widnieje na okładce, na stronie tytułowej jest rok 1971). Był to tom dyskusyjny wydrukowany w małym nakładzie; po raz drugi słownik ukazał się (zapewne ze zmianami redakcyjnymi) jako trzeci tom SWJP LF w 1975 r. Oto nazwiska recenzentów: M. Zarębina (1973), Bronisław Rocławski (1973), Alicja Zagrodnikowa (1973), Krystyna Kowalikówna (1974), Witold Mańczak (1974), Igor W. Kaspin (1975). Walery Pisarek (1976) poświęcił swoją recenzję czterem tomom SWJP LF wydanym w latach 1974–1976. Wydaje się, że jedynie Gerd Hentschel (1992) zrecenzował dwutomowy SFPW **i była to recenzja w pełni pozytywna**<sup>4</sup>.

Spośród wymienionych recenzentów tylko K. Kowalikówna skrytykowała utożsamianie w SWJP LF (na podstawie tomu wydanego w 1972 r.) liczby porządkowej (tj. kolejnego numeru) z rangą. Stanowisko autorki jest jednoznaczne:

Zastrzeżenia może budzić fakt, że każdemu wyrazowi przypisano osobną rangę, tzn. że numer kolejny na liście jest jednocześnie rangą wyrazu. Wydaje się to niesłuszne, bowiem ranga odpowiada numerowi kolejnemu tylko w wypadku wyrazów o wysokich frekwencjach. W wypadku wyrazów o częstości niskiej ujęcie autorów Słownictwa [...] jest wręcz mylące, bo np. wyrazy o rangach 1506–1635 (jest ich 131) posiadają tę samą frekwencję (Kowalikówna 1974: 58).

Do zajęcia takiego stanowiska przyczyniła się niewątpliwie znajomość książki W. Pisarka (1972), której omówienie i ocena zajmuje pierwszą część artykułu recenzyjnego. K. Kowalikówna pisze, że słownik frekwencyjny W. Pisarka zawiera następujące informacje: liczba porządkowa, ranga, wyraz, frekwencja względna i frekwencja bezwzględna. W słowniku W. Pisarka „[w]yrazy o tej samej częstości mają tę samą rangę. W ramach danej rangi wyrazy podane są w układzie alfabetycznym” (Kowalikówna 1974: 56). Autorka cytuje z książki W. Pisarka rozumienie frekwencji względnej: „Stanowi ona iloczyn frekwencji bezwzględnej i wskaźnika dyspersji [...]” (Pisarek 1972: 88). Dyspersja to „[p]arametr liczbowy określający rozkład częstości wyrazu w ramach poszczególnych słów” (Kowalikówna 1974: 56). Na frekwencję względną powołują się także inne słowniki innosłowiańskie i niesłowiańskie.

Zadziwiające jest to, że W. Pisarek w swojej recenzji nawet nie wspomniał, że rozumie inaczej niż autorzy SWJP LF rangę wyrazów w słowniku frekwencyjnym. **Prawdopodobnie obawiał się, że jego krytyczne wypowiedzi skonfliktują go**

4 Nie mam informacji o innych recenzjach SFPW.

**z autorami SWJP LF.** Trudno zrozumieć też stanowisko pozostałych recenzentów SWJP LF (zwłaszcza W. Mańczaka), którzy również nie skrytykowali stanowiska SWJP LF, chociaż przecież zajmowali się językoznawstwem statystycznym.

Starałem się skorzystać z jak największej liczby słowiańskich i niesłowiańskich słowników frekwencyjnych. Udało mi się dotrzeć do niewielu słowników w wersji papierowej i również do niewielu słowników zamieszczonych w Internecie. Uważam jednak, że ich liczba wystarczy do wyciągnięcia odpowiednich wniosków.

W trzech słownikach języków słowiańskich przyjęto inną zasadę ustalania numerów rangowych na listach frekwencyjnych hasel leksykalnych niż w polskich słownikach frekwencyjnych. Wyrazy o tej samej frekwencji nie mają odmiennych numerów rangowych, tj. o randze hasel nie decyduje porządek alfabetyczny.

Najstarszy jest słowacki słownik frekwencyjny Jozefa Mistríka (1969). Jego autor we *Wstępie* (Mistrík 1969: 131) pisze, że pierwsza liczba informuje o liczbie porządkowej słowa, druga zaś o randze słowa. Następnie dodaje, że liczba oznaczająca kolejność słowa na liście pokrywa się z rangą słowa przy najczęstszych słowach, przy coraz niższych częstościach słów znacznie więcej wyrazów ma tę samą rangę. I tak np. 27 wyrazów ma frekwencję 50, a 304 wyrazy – frekwencję 10. Wyrazy o tej samej frekwencji mają identyczną rangę. Słownik poświadcza 9568 liczb porządkowych, ale tylko 429 rang (ibid.). Mniej istotne w interesującej nas sprawie jest to, że J. Mistrík wyróżnia oprócz frekwencji absolutnej słów także frekwencję relatywną, tj. frekwencję względną. **Warto wspomnieć, że autorzy SFPW (t. I: xii) znali ten słownik.**

Bułgarski słownik języka potocznego (Nikolova 1987) wymienia na liście rangowej (*Рангов списък*) hasła o frekwencji do 10 poświadczeń. Hasła o tej samej frekwencji mają identyczną rangę.

Chorwacki słownik frekwencyjny (Moguš et al. 1999) stosuje konsekwentnie zasadę tej samej rangi przy hasłach o identycznej frekwencji. Lista rangowa według wskaźnika F obejmuje 569 pozycji. Ranga 1 dotyczy czasownika *biti* o frekwencji 56 194, rangę 569 mają hasła poświadczone raz w wykorzystanym korpusie. Są one wymienione na s. 197–298; ich lista zawiera prawie 14 tys. hasel wyrazowych. Rangi od 1 do 127 mają tylko pojedyncze hasła; po raz pierwszy rangę wspólną 128 mają trzy wyrazy: *kakav*, *lice*, *mali*. Później wspólną rangę hasła mają coraz częściej, np. rangę 154 – wyrazy o frekwencji 558 (*organizacija*, *razvoj*, *vrlo*). Przy niższych frekwencjach wspólna ranga dotyczy kilkunastu, kilkuset, kilku tysięcy lub nawet kilkunastu tysięcy hasel (hasła o frekwencji 1). Zastosowanie zasady porządku alfabetycznego oznaczałoby, że rang jest ponad 20 tysięcy, co budziłoby podejrzenie, że taka koncepcja listy rangowej według wskaźnika F nie jest racjonalna.

Można by wobec tego stwierdzić, że struktura polskich list rangowych (nie tylko dotycząca częstości absolutnej) ma charakter nienaukowy, że ustalenie rangi na podstawie porządku alfabetycznego hasel jest pozbawione sensu.

Problem polega jednak na tym, że stanowisko polskiej lingwistyki statystycznej nie jest odosobnione. Na tej samej zasadzie – jak SWJP LF i SFPW – listę rangową



sporządził *Новый частотный словарь русской лексики* (Lâševskaâ, Šarov 2009). Jest to słownik młodszy o prawie 30 lat i oparty na korpusie 100 milionów słów (korpus SFPW to 600 tysięcy słów), korzystający z najnowszego dorobku językoznawstwa statystycznego. O utożsamianiu numeru porządkowego hasel z ich rangą świadczą następujące cytaty ze *Wstępu* (brak oznaczenia stron w edycji elektronicznej):

частотный ранг слова (то есть порядковый номер в общем частотном списке), [...]. Здесь для каждой леммы указана ее общая частота и ранг (порядковый номер) в общем списке (dict.ruslang.ru).

Słownik frekwencyjny języka czeskiego (Čermák et al. 2004) wyróżnia dwa rodzaje rang: rangę FRQ i rangę ARF. Ranga FRQ dotyczy miejsca na liście w porządku alfabetycznym według frekwencji absolutnej. Ranga ARF odnosi się również do miejsca na liście, ale ze względu na frekwencję względną. FRQ informuje o frekwencji absolutnej, ARF zaś o frekwencji względnej. Słownik zawiera dwie listy frekwencyjne wyrazów pospolitych.

Pierwsza lista została ułożona według frekwencji absolutnej: *Frekvenční slovník apelativ (pořadí podle absolutní frekvence)* (s. 373–467). Na pierwszym miejscu znajduje się ranga FRQ (pokrywająca się z liczbą porządkową słów), na drugim ranga ARF (dotycząca rangi względnej słów), na trzecim FRQ (informująca o frekwencji absolutnej). Ranga FRQ obejmuje 20 tysięcy pozycji hasłowych, tj. pokrywa się z liczbą hasel od frekwencji 2 792 011 do frekwencji 200, przy czym uwzględnia tylko 10 hasel z tą frekwencją. Na przykład: *stoupat* o frekwencji absolutnej 4617 ma rangę FRQ 2419 i rangę ARF 2038; ostatni wyraz na liście – *trošičku* ma frekwencję absolutną 200, rangę FRQ 20 000 i rangę ARF 17 335. Druga lista to *Frekvenční slovník apelativ (pořadí podle průměrné redukované frekvence)* (s. 470–563). Na pierwszym miejscu znajduje się ranga ARF (ranga frekwencji względnej i jednocześnie liczba porządkowa hasła), na drugim ranga FRQ (ranga frekwencji absolutnej), na trzecim ARF (frekwencja względna). Przykład: wyraz *pohon* ma frekwencję względną 438, rangę ARF 7177 i rangę FRQ 4571.

Wydaje się, że na tych dwóch listach (określanych jako słowniki) częstość absolutna nie jest wartością najważniejszą i dlatego ten typ rangi jest utożsamiany z liczbami porządkowymi. Ważniejsza jest obecność słów w poszczególnych stylach funkcjonalnych (zwłaszcza udział ilościowy), co jest traktowane jako frekwencja względna. Ranga według frekwencji względnej jest także podporządkowana kolejności alfabetycznej.

Starszy czeski słownik frekwencyjny (Jelínek et al. 1961) nie wprowadza pojęcia rangi.

Rangę utożsamia z liczbą porządkową niemiecki słownik frekwencyjny (Meier 1964). To dzieło leksykograficzne jest oparte na słowniku Friedricha Wilhelma Kaedinga (1897), właściwie pierwszym na świecie słowniku frekwencyjnym. Według Helmuta Meiera „[d]ieses Rangbuch der geläufigsten deutschen Wortformen, geordnet

nach dem Grade ihrer Häufigkeit, ist das Ergebnis einer jahrelangen Umarbeitung des »Häufigkeitswörterbuches« von KAEDING” (Meier 1964: 111).

Zaskakują numery rang we francuskim słowniku frekwencyjnym (Vander Beke 1932). Wyrazy *abandonner* (s') i *abasourdir* o frekwencji 21 i 12 mają rangę 9; *abstenir* (s') i *abstraction* o frekwencji 13 mają rangę 12 i 7; wyrazy *détester* i *détour* o frekwencji 33 i 25 mają identyczną rangę 20. Z pewnością nie utożsamia się w tym słowniku liczby porządkowej hasła z jego rangą.

W rosyjskim słowniku frekwencyjnym (Steinfeldt/Štejnfeldt 1963) ranga hasła na liście frekwencyjnej jest ustalana w specyficzny sposób. Na przykład wyrazy o frekwencji 333 (*место, после*) mają rangi 162 i 187, wyrazy o frekwencji 320 i 318 (*дворъ, дызъ*) mają tę samą rangę 128, 11 wyrazów o frekwencji 67 ma różne rangi: 43, 55, 39, 33, 52, 45, 57, 44, 23, 33, 28 (w tym wyrazy *общество* i *участник* mają identyczną rangę 33). Struktura list frekwencyjnych i rangowych w tym słowniku zasługuje na szczegółowe przeanalizowanie.

Łaciński słownik frekwencyjny (Diederich 1939) stosuje przy hasłach o tej samej frekwencji kolejną numerację na liście (uwzględniając porządek alfabetyczny), np. *fero* 367 i *tamen* 367 zajmują na liście pozycje 59 i 60, *amo* 202 i *teneo* 202 – pozycje 124 i 125. Zapewne jednak numery nie oznaczają rangi, lecz jedynie liczby porządkowe.

Hiszpański słownik frekwencyjny (Juilland, Chang-Rodríguez 1964) na liście frekwencyjnej wyróżnia trzy rubryki: Usage<sup>5</sup>, Frequency, Dispersion. Przy lewym marginesie znajdują się liczby – od 1 do 5024. Nie wiadomo, czy oznaczają one rangę, czy tylko są liczbami porządkowymi. Zawsze przy tej samej frekwencji hasła są wymieniane w porządku alfabetycznym i występują przy nich kolejne liczby. Na takiej samej koncepcji oparty jest rumuński słownik frekwencyjny (Juilland et al. 1965). **W SFPW (t. 1: xi) znajduje się wzmianka, że autorzy wzorują się na tych dwóch i innych słownikach frekwencyjnych języków romańskich.**

Istnieje konieczność przeanalizowania i porównania struktury list rangowych w większej liczbie słowników frekwencyjnych, zwłaszcza z przełomu XX i XXI w. Nie miałem możliwości dotarcia do innych słowników niż powyżej wymienione. Szczególnie istotne byłoby uwzględnienie wydanych w Lipsku słowników Uwe Quasthoffa. Chodzi o następujące dzieła leksykograficzne: *Frequency dictionary Afrikaans* (2019), *Frequency dictionary Czech* (2017), *Frequency dictionary Danish* (2021), *Frequency dictionary Esperanto* (2014), *Frequency dictionary Georgian* (2018), *Frequency dictionary Hungarian* (2013), *Frequency dictionary Icelandic* (2012), *Frequency dictionary Indonesian* (2015), *Frequency dictionary Russian* (2017), *Frequency*

5 Posłużę się cytatem z J. Mistrika, aby objaśnić termin *usage*. Według leksykografa (1969: 17): „[...] relativná frekvencia, tzw. usage (je to absolútna frekvencia násobená percentom udávajúcim disperziu slova)”.



*dictionary Ukrainian* (2017), *Frequency dictionary Vietnamese* (2018), *Frequency dictionary Zulu* (2020).

Z. Saloni – jako redaktor naukowy SFPW – zaakceptował utożsamianie liczby porządkowej haseł z rangą. Jednak w *Encyklopedii językoznawstwa ogólnego* opowiedział się – jak się wydaje – za takim rozumieniem rangi, które odróżnia ją od liczby porządkowej:

**Ranga.** Kolejny numer elementu jęz. na liście porządkującej różne jednorodne elementy tekstu (np. litery, fonemy, leksemy) wg odpowiadających im części. Lista taka nosi nazwę listy rangowej. Element najczęstszy w tekście ma r. 1, drugi z kolei co do częstości – r. 2 itd. Jeśli kilka elementów ma jednakową częstość, na liście rangowej należy dla nich przeznaczyć taką liczbę miejsc, ile jest tych elementów, tj. przypisać wspólną r. alternatywną (EJO: 442).

\*\*\*

Można by przyjąć, że utożsamianie liczb porządkowych słów z ich rangami według wskaźnika F ma charakter nienaukowy. Ale przecież tak postępowali nawet wybitni językoznawcy. Czy wobec tego należy im zarzucić niekompetencję? Osobiście uważam, że utożsamianie liczb porządkowych z numeracją rang ma charakter nienaukowy<sup>6</sup>. **Sądzę, że podobne jest stanowisko K. Kowalikówny (1974) oraz autorów list frekwencyjnych: Mieczysława Balowskiego (2001), Leonardy Mariak (2023) oraz Katarzyny Ostrowskiej (2023), którzy nie popełniają tego błędu. Jednak H. Zgółkowa w swoim najnowszym słowniku frekwencyjnym (Zgółkowa 2016) akceptuje takie rozumienie rangi wyrazów, które ja uważam za nienaukowe.**

Nie wiadomo, jaka „filozofia” metodologiczna legła u podstaw utożsamiania na listach frekwencyjnych liczb porządkowych haseł leksykalnych z ich rangami. W tej sprawie warto zaproponować dyskusję z udziałem logików i lingwistów interesujących się językoznawstwem teoretycznym i statystycznym.

### Wykaz wykorzystanych opracowań i słowników frekwencyjnych

BALOWSKI M., 2001, *Lista frekwencyjna poezji Jerzego Ortana*, Prochowice.

ČERMÁK F., KŘEN M., BLATNÁ R., 2004, *Frekvenční slovník češtiny*, Praha.

6 Jeśli utożsamianie liczb porządkowych z numerami rang jest rzeczywiście nienaukowe, to badania językoznawców korzystających z takich słowników i akceptujących ich podstawy teoretyczne są obciążone pewnym marginesem błędów. Z tego stwierdzenia nie wynika jednak, że niepozytywnie oceniam takie słowniki jako całość. Są one mimo to wartościowymi pomocami w badaniach językoznawczych, choćby ze względu na listy frekwencyjne i różnego typu informacje gramatyczne. Innymi słowy, zasługują na pozytywną ocenę.

- DIEDERICH P.B., 1939, *The Frequency of Latin Words and their Endings*, Chicago.
- JELÍNEK J., BEČKA J.V., TĚŠITELOVÁ M., 1961, *Frekvence slov, slovních druhů a tvarů v českém jazyce*, Praha.
- JUILLAND A., CHANG-RODRÍGUEZ E., 1964, *Frequency Dictionary of Spanish Words*, „The Romance Languages and their Structures. First Series. S 1”, London.
- JUILLAND A., EDWARDS P.M.H., JUILLAND I., 1965, *Frequency Dictionary of Rumanian Words*, „The Romance Languages and their Structures. First Series. R 1”, London.
- KAEDING F.W., 1897, *Häufigkeitwörterbuch der deutschen Sprache. Festgestellt durch einen Arbeitsausschuss der deutschen Stenographiesysteme*, Steglitz bei Berlin.
- LÁŠEVSKÁ O.N., ŠAROV S.A., 2009, *Novyj častotnyj slovar' russkoj leksiki*, elektronnaâ ver-siâ izdaniâ: *Častotnyj slovar' sovremennogo russkogo âzyka (na materialah Nacional'nogo korpusa russkogo âzyka)*, Moskva, [on-line:] <http://dict.ruslang.ru>.
- MARIAK L., 2023, *Rodzina*, „Słownictwo Pism Stefana Żeromskiego”, t. 17, Szczecin.
- MEIER H., 1964, *Deutsche Sprachstatistik*, t. I–II, Hildesheim.
- MISTRÍK J., 1969, *Frekvencia slov v slovenčine*, Bratislava.
- MOGUŠ M., BRATANÍĆ M., TADIĆ M., 1999, *Hrvatski čestotni rječnik*, Zagreb.
- NIKOLOVA C., 1987, *Čestoten rečnik na b"lgarskata razgovorna reč*, Sofiâ.
- OSTROWSKA K., 2023, *Słownictwo wybranych polskich reportaży książkowych (2004–2018)*, t. 1: *Listy frekwencyjne*, Kielce.
- SFPW: I. Kurcz, A. Lewicki, J. Sambor, K. Szafran, J. Woronczak, *Słownik frekwencyjny pol-szczyzny współczesnej*, red. Z. Saloni, t. I–II, Kraków 1990.
- STEINFELDT E. / ŠTEJNFEL'DT È.A., 1963, *Russian Word Count / Častotnyj slovar' sovremen-nogo russkogo literaturnogo âzyka. 2500 naibolee upotrebitel'nyh slov. Posobie dlâ prepo-davatelej russkogo âzyka*, Tallin.
- SWJP LF: I. Kurcz, A. Lewicki, J. Sambor, J. Woronczak, *Słownictwo współczesnego języka polskiego. Listy frekwencyjne*, t. I, cz. 1–2: *Teksty popularnonaukowe*, Warszawa 1974; t. II, cz. 1–2: *Drobne wiadomości prasowe*, Warszawa 1974; t. III, cz. 1–2: *Publicystyka*, Warsza-wa 1975<sup>7</sup>; t. IV, cz. 1–3: *Proza artystyczna*, Warszawa 1976; t. V, cz. 1–2: *Dramat artystyczny*, Warszawa 1977.
- SWPP LF: A. Lewicki, W. Masłowski, J. Sambor, J. Woronczak, *Słownictwo współczesnej publicystyki polskiej. Listy frekwencyjne*, Warszawa 1972.
- VANDER BEKE G.E., 1932, *French Word Book*, „Publications of the American and Canadian Committees on Modern Languages”, t. 15, New York.
- ZGÓŁKOWA H., 1983, *Słownictwo współczesnej polszczyzny mówionej. Lista frekwencyjna i rangowa*, „Filologia Polska. Uniwersytet im. Adama Mickiewicza w Poznaniu”, nr 28, Poznań.
- ZGÓŁKOWA H., 2016, *Słownictwo dzieci w wieku przedszkolnym w latach 2010–2015. Listy frekwencyjne*, Poznań.
- ZGÓŁKOWA H., BUŁCZYŃSKA K., 1987, *Słownictwo dzieci w wieku przedszkolnym. Listy fre-kwencyjne*, „Filologia Polska. Uniwersytet im. Adama Mickiewicza w Poznaniu”, nr 33, Poznań.
- ZGÓŁKOWA H., ZGÓŁKA T., 1992, *Słownictwo współczesnej poezji polskiej*, t. 1–2, „Filologia Polska. Uniwersytet im. Adama Mickiewicza w Poznaniu”, nr 50, Poznań.

---

7 Sygnałne wydanie tego tomu ukazało się w 1972 r.

ZGÓŁKOWA H., ZGÓŁKA T., SZYMONIAK K., 1991, *Słownictwo polskich tekstów rockowych. Listy frekwencyjne*, „Filologia Polska. Uniwersytet im. Adama Mickiewicza w Poznaniu” 49, Poznań.

## Literatura

- EJO: K. Polański (red.), *Encyklopedia językoznawstwa ogólnego*, Wrocław 1993.
- GUIRAUD P., 1966, *Zagadnienia i metody statystyki językoznawczej*, przeł. M. Kniaginowa, Warszawa.
- HAMMERL R., SAMBOR J., 1990, *Statystyka dla językoznawców*, Warszawa.
- HENTSCHEL G., 1992, [recenzja] Kurcz, I. (et al.): *Słownik frekwencyjny polszczyzny współczesnej*. Kraków 1990, „Die Welt der Slaven. Internationale Halbjahresschrift für Slavistik” 37, s. 394–398.
- KASPIN I.W., 1975, [recenzja] A. Lewicki, W. Masłowski, J. Sambor, J. Woronczak, *Słownictwo współczesnej publicystyki polskiej. Listy frekwencyjne*, Warszawa 1972, s. 538, „Język Polski” LV, z. 1, s. 49–51.
- KOWALIKÓWNA K., 1974, *Na marginesie badań nad frekwencją wyrazów we współczesnej polszczyźnie*, [recenzja] W. Pisarek, *Frekwencja wyrazów w prasie: wiadomości, komentarze, reportaże*, „Biblioteka Wiedzy o Prasie” seria B, t. 16, Kraków 1972, „Język Polski” LIV, z. 1, s. 53–60.
- KURASZKIEWICZ W., 1932, *Studja nad polskimi samogłoskami nosowymi*. Rezonans nosowy, „Rozprawy Wydziału Filologicznego – Polska Akademia Umiejętności”, t. 63, nr 3, Kraków.
- KURASZKIEWICZ W., 1958, *Statystyczne badania słownictwa polskich tekstów XVI w.*, [w:] P. Zwoliński (red.), *Z polskich studiów slawistycznych. Prace językoznawcze i etnogenetyczne na IV Międzynarodowy Kongres Slawistów w Moskwie 1958*, Warszawa, s. 240–257.
- KURASZKIEWICZ W., 1966, *Uwagi o statystyce w Słowniku*, [w:] M.R. Mayenowa, F. Peplowski (red.), *Słownik polszczyzny XVI wieku*, t. I, Wrocław, s. XIV–XXV.
- MAŃCZAK W., 1974, [recenzja] A. Lewicki, W. Masłowski, J. Sambor i J. Woronczak, *Słownictwo współczesnej publicystyki polskiej. Listy frekwencyjne*, Warszawa 1972, „Rocznik Slawistyczny” XXXV, cz. 1, s. 68–71.
- PISAREK W., 1972, *Frekwencja wyrazów w prasie. Wiadomości – komentarze – reportaże*, „Biblioteka Wiedzy o Prasie. Seria B”, t. XVI, Kraków.
- PISAREK W., 1976, [recenzja] I. Kurcz et al., *Słownictwo współczesnego języka polskiego. Listy frekwencyjne. T. I–IV*, Warszawa 1972–1976, „Polish Psychological Bulletin” 7, z. 1, s. 64–65.
- ROCLAWSKI B., 1973, [recenzja] A. M. Lewicki et al., *Słownictwo współczesnej publicystyki polskiej. Listy frekwencyjne*, Warszawa 1972; W. Pisarek, *Frekwencja wyrazów w prasie. Wiadomości – komentarze – reportaże*, Kraków 1972, „Literary. Magazyn społeczno-kulturalny Wybrzeża” 8, s. 34.
- SAMBOR J., 1969, *Badania statystyczne nad słownictwem (na materiale „Pana Tadeusza”)*, „Z Dziejów Form Artystycznych w Literaturze Polskiej”, t. 12, Wrocław.
- SAMBOR J., 1971, *Z zagadnień gramatyki w słowniku frekwencyjnym współczesnego języka polskiego*, „Biuletyn Polskiego Towarzystwa Językoznawczego” XXIX, s. 117–129.

- SAMBOR J., 1972, *Słowa i liczby. Zagadnienia językoznawstwa statystycznego*, Wrocław.
- SAMBOR J., 1974a, *Słownictwo bardzo częste w pięciu stylach współczesnej polszczyzny pisanej*, cz. I, „Poradnik Językowy” nr 9, s. 466–475.
- SAMBOR J., 1974b, *Słownictwo bardzo częste w pięciu stylach współczesnej polszczyzny pisanej*, cz. II, „Poradnik Językowy” nr 10, s. 533–537.
- SAMBOR J., 1975, *O słownictwie statystycznie rzadkim. Na materiale derywatów we współczesnej publicystyce polskiej*, „Rozprawy Uniwersytetu Warszawskiego” 85, Warszawa.
- SJAM: K. Górski, S. Hrabec (red.), *Słownik języka Adama Mickiewicza*, t. I–X, Wrocław 1962–1980.
- SPXVI: *Słownik polszczyzny XVI wieku*, red. komitet redakcyjny (t. I–IV), M.R. Mayenowa (t. V–XXXIV), K. Mrowcewicz (t. XXXV–XXXIX), Wrocław 1966–.
- ZAGRODNIKOWA A., 1973, [recenzja] A. Lewicki, W. Masłowski, J. Sambor, J. Woronczak, *Słownictwo współczesnej publicystyki polskiej. Listy frekwencyjne*, Warszawa 1972, s. 583, „Zeszyty Prasoznawcze” XIV, nr 3 (57), s. 121–122.
- ZARĘBINA M., 1973, [recenzja] A. M. Lewicki, W. Masłowski, J. Sambor, J. Woronczak, *Słownictwo współczesnej publicystyki polskiej. Listy frekwencyjne*, Warszawa 1972, „Język Polski” LIII, z. 4, s. 300–302.
- ZARĘBINA M., 1977, *Wyrazy obce w „Panu Tadeuszu”*, „Prace Językoznawcze – Polska Akademia Nauk. Instytut Językoznawstwa”, t. 83, Wrocław.
- ZARĘBINA M., 1978, *Zastosowanie metod statystycznych do badania zapożyczeń w gwarach pogranicznych*, [w:] *Z polskich studiów slawistycznych. Seria 5. Prace na VIII Międzynarodowy Kongres Slawistów w Zagrzebiu 1978*, t. 1: *Językoznawstwo*, Warszawa, s. 325–332.

## **The Problem of Identifying an Ordinal Number with a Ranking Polish Frequency Dictionaries**

### **Abstract**

The frequency dictionaries listed in the title contain three ranking lists based on the F index (absolute frequency), U index (relative frequency) and D index (measure of uniformity and distribution in styles). There are serious objections to the method used, according to which, when the F, U or D index values are equal, the entry rank (i.e. the number in the list) is decided by the alphabetical order. In case of words with high frequencies, each word is usually assigned a separate rank, because they differ in frequency. With increasingly lower frequencies, the same frequency is shared by a few, a dozen, a few hundred or even thousands of words (for example frequency 2 or 1). Words with the same frequency should have the same rank. In Polish frequency dictionaries, however, their rank differs depending on the alphabetical order. Hence, the ordinal number of a word is made equal to its rank. I believe that such method falls short of scholarly standards, because words with the same frequency differ considerably in rank. For example, in the Polish frequency dictionary (*Słownik frekwencyjny polszczyzny współczesnej*, Kraków 1990), words with frequency 4 are assigned ranks from 8739 (*abstrahować*) to 10355 (*żywoł*) – a difference of 1616 places on the ranking list. For words with frequency 3, 2 and 1, the differences in the ranking list would be much greater.

The ordinal number of words is not equated with their rank in, among others, frequency dictionaries of Slovak and Croatian as well as the Bulgarian dictionary of colloquial speech. Nevertheless, many frequency dictionaries of various languages do equate the number and the rank. Therefore, a discussion involving linguists interested in statistical linguistics is necessary.