

TADEUSZ PIOTROWSKI* | UNIwersYTET WROCLAWSKI

Ilościowa struktura haseł Słownika warszawskiego w aspekcie porównawczym

Słowa kluczowe: leksykografia, słownik, historia, leksykologia.

doi: <http://dx.doi.org/10.31286/JP.01022>

Słownik warszawski (dalej SW), czyli *Słownik języka polskiego* pod redakcją Jana Karłowicza, Adama Antoniego Kryńskiego i Władysława Niedźwiedzkiego, wydany w latach 1900–1927, zawiera wszechstronne informacje o największej liczbie jednostek leksykalnych języka polskiego. Tak charakteryzowali ów słownik historycy leksykografii polskiej i języka polskiego, z powyższym sądem zgodzi się także zapewne każdy, komu potrzebny opis rzadkich wyrazów języka polskiego czy tych, które nie są obecnie używane. Chciałem podkreślić, że SW to słownik, a nie lista wyrazów, obecnie zaś istnieje wiele spisów wyrazów, które są charakteryzowane przez autorów jako największe zbiory leksykalne¹ etc., lecz brak w nich informacji o włączanych jednostkach. W literaturze o SW nie znajdziemy jednak zwykle dokładnych danych o zasobach słownika, liczbie haseł i ich kwalifikacji do różnych grup słownictwa (por. syntezę w Majdak 2008). W niniejszym artykule opiszę liczbę haseł SW, ich strukturę, a także porównam ją z zasobem największego słownika następującego po nim – *Słownika języka polskiego* pod redakcją Witolda Doroszewskiego (1958–1967, dalej SJPdor).

W większości współczesnych opisów SW można znaleźć dane pochodzące od W. Doroszewskiego, który w bardzo wpływowym zarysie historii leksykografii polskiej swojego pióra twierdził, że „Ogólna liczba zarejestrowanych w słowniku wyrazów wynosi ok. 280 000” (1954: 42). Przytaczałem i ja tę liczbę kilka razy w swoich publikacjach, czasem bez odpowiedniego cytatu, a obecnie niekiedy niesłusznie podaje się moje nazwisko jako autora tego stwierdzenia, np. na stronie internetowej *Leksykografia polska* (Majdak 2011). Doroszewski był redaktorem planowanego suplementu do SW, do którego materiały zostały zniszczone w czasie II wojny światowej – pisze o tym we wstępie do SJPdor – można więc sądzić, że SW znał bardzo dobrze. W przytoczonym zdaniu, jak również we wspomnianej publikacji, brakuje wszakże danych, które pozwoliłyby na jego jednoznaczną interpretację. Nie podał Doroszewski metody liczenia haseł, nie określił też dokładnie, co rozumie przez frazę „wyraz zarejestrowany w słowniku”, jeżeli tak będziemy określać wyrazy będące główkami hasła (inaczej: nazwami

* tadeusz.piotrowski@uwr.edu.pl; ORCID: 0000-0002-9315-6543

1 Por. np. fotokorpus.pl, <http://nfjp.pl/>. „Niniejsze repozytorium biblioteczne przedstawia największy zbiór leksykalny polszczyzny XX w. wraz z poświadczeniem cytatowym”. Repozytorium ma informację o formach, a dokładniej ciągach znaków, występujących w tekstach języka polskiego, nie o jednostkach leksykalnych.

hasła), to wypada przypomnieć, że w SW występują hasła różnego typu, przede wszystkim główne i odsyłaczowe, a także wyrazy wewnątrz haseł: jako ich warianty, synonimy etc. Czy Doroszewski pisał o wszystkich takich wyrazach? Nie wiemy. W swej książce (Piotrowski 2001) starałem się oszacować liczbę haseł (ich nazw) w największych słownikach języka polskiego na podstawie niewielkiej próbki, stąd też wyniki obarczone są bardzo dużym błędem. Tą drogą doszedłem do wniosku, że SW zawiera ponad 230 000 haseł, wśród nich głównych jest około 150 000.

Obecnie możemy podać znacznie dokładniejszą charakterystykę zasobu SW. Listę haseł SW przygotował zespół projektu naukowego „Automatyczna analiza fleksyjna polszczyzny XIX wieku”, który był realizowany w latach 2013–2017 ze środków Narodowego Centrum Nauki². Zespół tworzyli: Magdalena Derwojedowa (kierownik projektu), Joanna Bilińska, Witold Kieraś, Monika Kwiecień, Danuta Skowrońska z Uniwersytetu Warszawskiego oraz Robert Wołosz z Uniwersytetu w Pécsu na Węgrzech (Derwojedowa 2017). Lista haseł SW posłużyła do dostosowania współczesnych narzędzi leksykograficznych i korpusowych do automatycznej analizy XIX-wiecznych tekstów. Dane i narzędzia projektu, zgodnie z najlepszymi tradycjami nauki³, zostały udostępnione bezpłatnie na stronach internetowych⁴. Wśród nich, na podstronie <http://www.fi9.uw.edu.pl/download-category/swar/>⁵, znajdują się pliki z listą haseł *a fronte* oraz *a tergo*. Pliki zostały sporządzone przez R. Wołosza. W korespondencji prywatnej ich autor wraz z M. Derwojedową podkreślali, że listy są w domenie publicznej, co oznacza, że każdy może z nich korzystać bez przeszkód; mimo wszystko poprosiłem o możliwość ich użycia we własnej publikacji, zgodę tę otrzymałem.

Zanim opiszę, jaka była metoda przygotowania listy i ile pozycji liczy, na chwilę wrócę do przeszłości. W lutym 2008 roku w ramach e-biblioteki Uniwersytetu Warszawskiego udostępniono skan stron SW⁶; zeskanowane strony, dzięki inicjatywie Janusza Bienia, poddano interpretacji programem OCR, dokładnie programem FineReader w wersji 10, otrzymując w wyniku dokument tekstowy, który można analizować jak korpus w przeglądarce dawnych słowników⁷. Na swoją prośbę w październiku 2008 roku otrzymałem od J. Bienia pliki z wynikami OCR – chciałem spróbować wydobyć z nich m.in. listę haseł, opierając się na swoim doświadczeniu przy przetwarzaniu tekstów własnych słowników. Dostępne były dwa pliki: plik tekstowy (ok. 50 MB) oraz plik w formacie xml (ok. 500 MB). Osoby pracujące z programami OCR wiedzą doskonale, że nawet w przypadku tekstów współczesnych o bardzo dobrej jakości druku

2 Projekt numer UMO DEC-2012/07/B/HS2/00570.

3 Podkreślam ten fakt, ponieważ z dużym zaniepokojeniem obserwuję, jak wielu kolegów, choć korzysta z pieniędzy publicznych w ramach etatów na uczelniach państwowych lub projektów, traktuje nie tylko wyniki, ale i metody ich osiągnięcia jak swoją prywatną własność. To całkowite zaprzeczenie podstaw działalności naukowej.

4 Adres internetowy: <http://www.fi9.uw.edu.pl>.

5 Stan z początku 2022 r.

6 SW, wersja 1; <http://ebuw.uw.edu.pl/dlibra/publication?id=254>; obecnie <https://crispaa.uw.edu.pl/object/files/416370/display/Default>. Skan jest dostępny w postaci plików PDF, JPEG, DJVU, można przeglądać go w przeglądarce internetowej lub pobrać pliki do wykorzystywania na własnym komputerze.

7 <https://szukajwsownikach.uw.edu.pl/sownik-warszawski/query/>, wcześniej: <http://poliqarp.wbl.klf.uw.edu.pl/sownik-warszawski/>. W 2021 r. J. Bień sygnalizował w korespondencji prywatnej, że ten zasób może przestać działać.

w wyniku automatycznej interpretacji danych obrazowych pojawia się wiele błędów. W przypadku publikacji tak starych błędów musi być bardzo dużo⁸. Dla przykładu cytuję fragment słownika (artykuły hasłowe *Adnotować* i *Adnotowanie*) w formacie tekstowym, nie wskazując błędów jako oczywistych. Nie jest oczywisty znak „x” na początku hasła *Adnotować* – oznacza on krytyk, używany jako kwalifikator w słowniku. O kwalifikatorach piszę niżej.

xAdnotować, uje, owal, xAnnotować zrobić
adnotacji,, zanotować, zaznaczyć, zapisać: Kościuszeko
kazał robić mapę. adnotując, gdzie i ile
żołnierza stało. <Łć. adnolare a. aunotaro>
xAdnotowanie, a, blin., ezyнноśł cz. Adno-
towaó.

W formacie XML fragment słownika, jedynie dla hasła *adnotowanie*, wygląda następująco:

```
<LINE>
<WORD coords="425,1390,727,1346">xAdnotowanie,</WORD>
<WORD coords="766,1391,799,1354">a,</WORD>
<WORD coords="840,1387,937,1344">blin.,</WORD>
<WORD coords="975,1391,1151,1344">ezyнноśł</WORD>
<WORD coords="1188,1383,1238,1357">cz.</WORD>
<WORD coords="1280,1384,1395,1347">Adno-</WORD>
</LINE>
<LINE>
<WORD coords="370,1429,497,1392">towaó.</WORD>
</LINE>
```

Ze względu na stopień skomplikowania zadania i występujące błędy nie umiałem sobie poradzić z odpowiednim przetworzeniem plików. Te dwie cechy tekstu powstałego ze skanów SW sprawiają, że z poprawnym wyszukiwaniem haseł nie radzi sobie też wyszukiwarka J. Bienia, przygotowana przez profesjonalnych informatyków: nie można w niej znaleźć ani hasła *Adnotować*, ani *Adnotowanie*⁹. Słownik można znaleźć także w zasobach biblioteki cyfrowej Polona, dla której najwyraźniej powtórnie wykonano OCR skanu lub sporządzono nowy skan, ponieważ jest lepszej jakości. Wyszukiwarka tekstowa tej biblioteki odnajduje hasło *Adnotować*¹⁰.

Wracając do opisywanej listy haseł – R. Wołosz ma bardzo duże doświadczenie w przetwarzaniu tekstów języka polskiego, dlatego ciekawiło mnie, w jaki sposób uzyskał listę haseł.

8 Wypada zauważyć, że pierwsze wydanie SW wydrukowano na papierze znacznie lepszej jakości niż reprint z roku 1954. Coraz bardziej żółknące kartki przedruku niedługo zaczną się rozpadać.

9 Nie można też znaleźć *xadnotować* ani *xAdnotować*. Oba hasła można wyszukać z użyciem wyrażeń regularnych, konkretnie np. wyrażenia „*notowa.*”, które jednak daje bardzo dużo wyników.

10 Strona: <https://polona.pl/item/sownik-jezyka-polskiego-t-1-a-g.NDEWOTEwNzU/6/#info:search:Adnotowa%C4%87> (dostęp: 31 listopada 2021).

Poniższy opis powstał na podstawie korespondencji prywatnej z Wołoszem, który uprzejmie przeczytał niniejszy tekst. Wyabstrahowanie listy haseł za pomocą narzędzi informatycznych nie powiodło się także jemu. Użyto więc metody najbardziej pracochłonnej, ale i najbardziej niezawodnej: nazwy haseł wpisali ręcznie (z klawiatury) jego żona i on. Stąd też mogą się pojawiać błędy innego typu niż przy przetwarzaniu maszynowym, mogły np. zostać pominięte jednostkowe hasła (do dnia 2 listopada 2017 r. brakowało np. hasła [*Cieruty*]). Potencjalnie mogły również wystąpić błędy literowe, choć nie znalazłem ich ani przy wrywkowej lekturze listy, ani przy jej przetwarzaniu. Jakość powstałej listy haseł była weryfikowana, jedną z metod weryfikacji było sortowanie mechaniczne nazw haseł według kolejności alfabetycznej, dlatego np. można było stwierdzić, że niektóre hasła powtarzają się w tekście oryginalnym, czy to ze względu na błąd zecera, czy autorów (jak hasło *ostrożenie*).

Metod ręcznego wprowadzania tekstu dzieł leksykograficznych do pamięci komputera używa się w krajach znacznie bogatszych od Polski, np. tak digitalizowano *Oxford English Dictionary* (Simpson, Weiner 1989; por. Piotrowski 1994), natomiast wersja cyfrowa słownika języka niemieckiego braci Grimm (1854–1961) powstała na podstawie dwóch wersji: tekst wprowadzany był niezależnie ręcznie przez dwa zespoły pracowników chińskiej firmy¹¹, a dzięki maszynowemu porównaniu obu tekstów można było efektywniej poprawić nieuniknione błędy. Bezбłędność kopii cyfrowej tego słownika określa się *alla maniera tedesca* na 99,997 procent. Słownik udostępniany jest bezpłatnie na stronie *Deutsches Wörterbuch von Jacob Grimm und Wilhelm Grimm*.

Poniżej przedstawię analizę liczbową listy nazw haseł wykonanej przez R. Wołosza, listę będą określał skrótem RW. Zapisano ją w pliku tekstowym, w którym każda nazwa hasła wraz z dodatkowymi elementami umieszczona jest w odrębnym wierszu, a dokładniej każda nazwa hasła oddzielona jest od innej znakami – w rozumieniu znaków komputerowych – określanymi jako „powrót karetki” (to znak ASCII nr 13, konwencjonalnie oznaczany jako [CRLF]). Większość analiz przeprowadziłem z użyciem edytora tekstowego dla programistów TextPad 7.5, który ma rozbudowane możliwości wyszukiwania ciągów znaków, co obejmuje także wyrażenia regularne, oraz wyboru linii zawierających takie wyrażenia.

Lista RW haseł SW zawiera nie tylko nazwy wszystkich haseł, ale też poszczególne elementy słownikowe, które je charakteryzują. Niektóre używane były przez redaktorów SW w drukowanym słowniku, inne zaś wprowadził R. Wołosz. W SW wiele haseł jest opisywanych za pomocą znaków literowych lub graficznych, w skrócie – określeń hasła (dalej tak będą nazywane). Zapewne można je nazywać kwalifikatorami, w książce Magdaleny Majdak (2008: 143) o SW znajdujemy dyskusję o tym, czy faktycznie są to kwalifikatory. Wszystkie określenia, które znajdują się w liście haseł SW, zestawione zostały w tabeli 1; w innym miejscu podam dokładne dane liczbowe. Uwagi wprowadzone przez R. Wołosza są wydzielone jako jego komentarze. Każde określenie zapisane jest w postaci dwuznaków, składającego się z ukośnika i oznaczenia

11 Około 300 milionów znaków tekstu słownika zostało dwukrotnie wpisane przez firmę TQY Double Key z Nankinu w Chinach; „Daher wurden die rund 300.000.000 Zeichen des ¹DWB im Double-Keying-Verfahren von der Firma TQY Double Key in Nanjing/China erfasst” (<http://dwb.uni-trier.de/de/die-digitale-version/volltextdigitalisierung/>, dostęp: 31 listopada 2021).

(np. „*”), jedno hasło może mieć kilka określeń. Opisy zaczerpnąłem z książki M. Majdak (2008: 154), brak w niej jednak określenia „?” (znak zapytania), używanego w SW, być może dlatego, że nie występuje on w spisach skrótów w SW w odpowiednich tomach. Określenie „/!” opisane jest cytatem z SW.

Tabela 1. Określenia w liście haseł SW

ZNAK W RW	ZNAK W SW	OPIS	PRZYKŁAD Z RW
OKREŚLENIA PRZED NAZWĄ HASŁA			
/!	!	„Wyraz, którego unikać należy”	/! afranszować
/#		Komentarz RW: hasło odsyłaczowe	/# stamęt
/&		Komentarz RW: odsyłacz do erraty w SW	/& aksynit <zob. Sprostowania, tom I>
/*	*	Hipotetyczna forma wyrazu (nie występuje w etymologii)	/* niebaczek
/?		Sygnał wątpliwości redaktorów RW	/? Podumajka
/@		Komentarz RW, uszczegółowiony przez głosę umieszczoną w nawiasach trójkątnych (patrz niżej)	/@ pobudzenie <powtarza się>
/+	+	Wyraz staropolski	/+ pedard
/1	1	Homograf	/1 i
/2	2	Homograf	/2 i
/3	3	Homograf	/3 i
/4	4	Homograf	/4 i
/5	5	Homograf	/5 i
/6	6	Homograf	/6 /+ i
/g	[]	Wyraz gwarowy	/g abceng
/x	x	Wyraz mało używany	/x abe
OKREŚLENIA PO NAZWIE HASŁA			
?	?	Sygnał wątpliwości (?) autorów SW	/g dlacynu?
<>		Uwagi RW	pobudzenie <powtarza się>
{abc}		Nie umieszczone poprawnie w ciągu alfabetycznym	abiturjent {abc}

Ponieważ redaktorzy SW nie wyjaśnili konwencji opisu, nie wiemy dokładnie, co oznaczają używane określenia. Nie wiemy na przykład, jak interpretować symbol wskazujący na użycie danej jednostki w okresie staropolskim, w których stuleciach mogła być używana. Takim samym komentarzem możemy opatrzyć każde określenie. Należy też zauważyć, że określenia uwzględnione przez zespół R. Wołosza odnoszą się jedynie do hasła jako całości, czyli danej formy wyrazowej wraz z poszczególnymi znaczeniami i innymi elementami hasła. W hasłach, np. przy poszczególnych znaczeniach, występują także określenia – kwalifikatory – które wskazują np. na dziedzinę, w której używa się danego znaczenia. Te kwalifikatory nie zostały uwzględnione przez R. Wołosza, nie mogłem więc ich uwzględnić i ja.

Należy jednak zwrócić uwagę, że lista nazw haseł, przygotowana prymarnie do dalszego użycia w przetwarzaniu maszynowym, nie jest wiernym odwzorowaniem listy haseł SW. W SW wszystkie nazwy haseł pisane są dużą literą (*Niebaczek*), w RW zaś konsekwentnie małymi literami (*niebaczek*). W RW nazwy haseł są poprawione, przy czym zmiany są sygnalizowane odesłaniem do podstawy zmiany przez komentarz w nawiasach trójkątnych, po których następuje skrót „zob.”; takich odesłań jest 116. Odsyła się czytelnika do erraty w samym SW (np. „/& /x bachmistrzostwo <zob. Sprostowania, tom I>”) lub do zapisów w innych słownikach (np.: „/# /x szałszanowe ziele <zob. Linde>”, „/# /g warnacht <zob. Karłowicz>”) czy też do form użytych w innym miejscu w samym SW (np. „/# /+ świętoświęty <zob. w cytacie świętoświętny>”). Wreszcie nazwy haseł umieszczone są w poprawnym porządku alfabetycznym, a w SW występuje 903 jednostek w niewłaściwym miejscu – opatrzone są określeniem „{abc}”.

Liczba wszystkich nazw haseł w liście RW wynosi 287 655; wynika z tego, że W. Doroszewski podał w zasadzie prawidłowe dane – pedantycznie można stwierdzić, że poprawnym zaokrągleniem liczby haseł byłoby 290 000, nie 280 000. W tabeli 2 wykazuję, jaka jest liczba haseł oznaczonych odpowiednimi określeniami. Wszystkich haseł, w których występuje jakieś określenie, a dokładniej haseł, w których opisie pojawia się znak „/”, jest 111 640, a takich, gdzie ów znak się nie pojawia – 176 015. Oprócz jednostek określonych dwuznakiem ze znakiem „/” jako pierwszym podałem niżej także liczbę uwag R. Wołosza umieszczonych w spisie. Ze względu na liczbę jednostek nie można było dokładnie zweryfikować uzyskanych podzbiorów.

Tabela 2. Hasła w SW

ZNAK W RW	OPIS	LICZBA JEDNOSTEK	LICZBA PROCENTOWO (100% = 287 655)
/!	„Wyraz, którego unikać należy”	1599	0,56
/#	Komentarz RW: hasło odsyłaczowe	252	0,09
/&	Komentarz RW: odsyłacz do erraty	52	0,02
/*	Hipotetyczna forma wyrazu (użyta przed nazwą hasła, nie występuje w etymologii)	162	0,06

ZNAK W RW	OPIS	LICZBA JEDNOSTEK	LICZBA PROCENTOWO (100% = 287 655)
/?	Sygnal wątpliwości redaktorów RW	315	0,11
?	Sygnal wątpliwości redaktorów SW	8	0,01
/@	Komentarz RW: powtarza się	68	0,02
/+	Wyraz staropolski	25 292	8,79
/1	Homograf	1918	0,43
/2	Homograf	1918	0,43
/3	Homograf	117	0,03
/4	Homograf	18	0,01
/5	Homograf	1	0,00
/6	Homograf	1	0,00
/g	Wyraz gwarowy	51 473	17,89
/x	Wyraz mało używany	30 283	10,53
<>	Uwagi RW	1627	0,57
	Jednostki nieoznaczone	176 015	59,13

Kusi oczywiście, aby spróbować zinterpretować te liczby. Jako punkt odniesienia będę traktował dane z SJPDor. Na wstępie natykamy się na trudność, ponieważ wszystkich haseł w tym słowniku mamy 127 232 lub 125 632. Pierwsza liczba podana jest na podstawie danych ze słownika *a tergo*, opracowanego cyfrowo przez R. Wołosza¹², drugą podają autorzy hasła o SJPDor na znanej stronie internetowej o słownikach polskich (Saloni, Bańko 2011). Zastanawiająca jest ta różnica, równo 1 600 jednostek, jednak nie możemy tu dociekać, skąd się wzięła. Na pewno nie wynika z liczby haseł w suplemencie do SJPDor, te hasła są uwzględniane w spisach. Dalej będę traktował tę większą liczbę jako właściwą, nie przesądając, czy tak jest faktycznie, ponieważ lista z większą liczbą jednostek ma postać cyfrową i zawiera oznaczenia kwalifikatorów, dlatego mogę ją analizować w taki sam sposób jak listę haseł z SW. Tabela 3, oparta na liście haseł R. Wołosza, zawiera zestawienie liczbowe haseł w SJPDor w rozbiciu na grupy haseł opatrzone kwalifikatorami. Do celów analizy nie musiałem przekształcać listy *a tergo* w listę *a fronte*.

12 Plik udostępniony jest na stronie autora http://verob.hu/a_tergo/a_tergo.zip.

Tabela 3. Hasła w SJPdor

ZNAK W SJPdor	OPIS	LICZBA JEDNOSTEK	LICZBA PROCENTOWO (100% = 127 232)
d	Dawne	11 587	9,1
s	Suplement	9850	7,7
g	Gwarowe	877	0,7
p	Przestarzałe	5250	4,1
i	Indywidualne	982	0,8
	Ogółem haseł z kwalifikato- rami z suplementem	28 546	22,4
	Ogółem haseł z kwalifikato- rami bez suplementu	18 696	14,7
	Bez kwalifikatora	108 536	85,3

Jeżeli spojrzymy na dane z obu słowników dotyczące jednostek, które nie są opatrzone żadnym określeniem, to w SW mamy prawie 180 000 takich jednostek, a w SJPdor około 110 000. SW rejestruje więc około 70 000 jednostek nieopatrzonych określeniami. Z jednej strony można sądzić, że w obu słownikach jednostki nieopatrzone kwalifikatorami są porównywalne, leksykografowie traktowali je jako neutralne, z drugiej strony oczywiście można uznać, że użycie kwalifikatorów znacznie się różniło między zespołami redaktorów i np. w SJPdor stosowano je częściej. Na podstawie swojego doświadczenia leksykograficznego skłonny jestem przyjąć pierwszą interpretację, choć nie mam żadnych dowodów empirycznych. A 70 000 jednostek to ogromna liczba. Uważam, że nie wynika ona z tego, że w ciągu około 60 lat¹³ użytkownicy języka polskiego przestali posługiwać się 70 000 jednostek – wszak wiele nowych jednostek przybyło – ale z zasadniczej różnicy polityki redakcyjnej. Autorzy SW mieli ambicję rejestrować możliwie wszystkie jednostki języka polskiego, niezależnie od ich nacechowania, a także wprowadzali jednostki potencjalne, a redaktor SJPdor programowo ograniczał włączane słownictwo do tego, które aprobował i które w jego opinii mogło być przydatne użytkownikowi¹⁴. Dokładniej tę kwestię omawiam w publikacjach z 1994 i 2001 roku. To kryterium bardzo subiektywne. Bardzo ciekawa byłaby analiza jednostek, których uwzględnienie bądź pominięcie różnicuje oba słowniki. Można co prawda porównać maszynowo obie listy, ale aby taka procedura dała sensowny wynik, najpierw należałoby znormalizować obie listy, przede wszystkim

13 Licząc od początku wydawania SW do umownego zakończenia prac nad SJPdor.

14 Wiele decyzji musiało pozostawać w zgodzie z aktualną polityką ideologiczną.

zaś ujednoczyć pisownię nazw haseł; Bańko zaś już w 1988 roku pisał o trudnościach związanych z tego rodzaju porównaniami.

Pewne spojrzenie na różnice między siatkami haseł w obu słownikach może dać zestawienie właśnie domniemanych jednostek potencjalnych¹⁵ – najczęściej można je określić jako wyrazy, których nie zna autor danego opisu (por. dyskusję w: Majdak 2008: 73–79) lub których nie aprobuje, i to są właśnie kryteria subiektywne. W SW występuje wiele takich haseł, co oburzało redaktora SJPDor (por. dyskusję w: Doroszewski 1954). Poniżej przyjrzyć się małej próbie takich wyrazów. W SW występuje 49 704 haseł zakończonych na *-nie*, wyłączając leksem *nie*, wśród nich wiele jest rzeczowników odczasownikowych (*przerąbanie*) i wiele przysłówków odprzymiotnikowych (*mitrężnie*), można sądzić, że duża ich liczba to formy potencjalne, utworzone na podstawie intuicji redaktorów SW. Wśród nich 33 032 hasła nie są opatrzone określeniami w znaczeniu zdefiniowanym wyżej, czyli, w przybliżeniu, kwalifikatorami. W SJPDor z kolei analogicznych form można znaleźć 13 479, w tym form bez określeń 12 794 (wraz z tymi notowanymi w suplemencie). Na przykład forma *mitrężnie* znajduje się w tym słowniku, bez żadnego poświadczenia, z powołaniem się na L; formy *przerąbanie* brak. Obie występują w tekstach literackich, jak i nieliterackich współczesnych¹⁶, np. *mitrężnie* u Bolesława Leśmiana (1956/1978/2021) („o ile całość nagromadza się stopniowo i *mitrężnie*”), *przerąbanie* u Marcina Mortki (2021) („*Przerąbanie* się przez gąszcz prastarych drzew”); *przerąbanie* występuje współcześnie nie tylko jako odsłownik, ale też (rzadko) jako przysłówek, zapewne w funkcji eufemizmu („*przerąbanie* udany” – Krawczyk 2018; „*przerąbanie* głośny” – Shadow Laguna II 2019). SW ma więc prawie 2,5 razy więcej analogicznych form w stosunku do SJPDor, można mniemać, że wiele wprowadzono dzięki inwencji autorów, ten sąd zweryfikuję za chwilę. Tabela 4 zawiera osiem wyrazów zakończonych na *-nie*, które w obu słownikach występują na listach nazw haseł zakończonych na *-nie*, począwszy od hasła *śniadanie*. Jednostki ułożone są w kolejności *a tergo*.

Tabela 4. Osiem haseł zakończonych na *-nie* w SW i SJPDor

SW	SJPDor
śniadanie	śniadanie
pośniadanie	siadanie
siadanie	zasiadanie
nasiadanie	obsiadanie

15 Co ciekawe, politykę włączania wyrazów potencjalnych, czyli takich, które nie mają poświadczenia w analizowanych tekstach lub słownikach, przyjęli także autorzy *Słownika gramatycznego języka polskiego* (Woliński i in. 2020). Wszystkich leksemów jest 456 294, w tym 316 231 leksemów poświadczonych w tekstach, co obejmuje także 112 713 nazw własnych, czyli 203 518 leksemów ma poświadczenia tekstowe.

16 Oczywiście wchodzimy tu na grunt rozważań, czy jednostki tworzone systemowo, jak odsłowniki, powinny być rejestrowane w słownikach. Systemowość procesów słotwórczych jest jednak często podważana, a podnoszona jest kwestia nieprzewidywalności ich wyników.

SW	SJPDor
zasiadanie	przesiadanie
pozasiadanie	osiadanie
obsiadanie	dosiadanie
poobsiadanie	posiadanie

Po sprawdzeniu okazuje się, że wszystkie wyrazy notowane w SW używane są w tekstach, łącznie z wyrazem *pozasiadanie*, dziwacznym według mojej intuicji. Ten wyraz występuje tylko w funkcji odślownika w tekstach (np. „Czy te tępe głowy nie mają [sic!] chrapki na *pozasiadanie* w radzie nadzorczej szpitala” – Majka 2021). Forma *pośniadanie*, także osobliwa, występuje nie tylko jako odślownik od *pośniadać*, ale i jako rzeczownik ‘pośiłek po śniadaniu’ („Śniadanie, *pośniadanie*, obiad, no i teraz podwieczorek” – Wywiad 2021). Słownik starszy obejmuje więc bardzo rzadkie słownictwo, które faktycznie występuje w tekstach współczesnych, niezależnie od jego subiektywnej osobliwości.

W niniejszym artykule uściśliłem szacunki, podawane w literaturze przedmiotu, odnoszące się do liczby haseł w Słowniku warszawskim, a także ustaliłem liczbę haseł opatrzonych różnymi określeniami, porównując te dane do danych z SJPDor. Niestety żaden z omawianych słowników nie został w pełni zdigitalizowany, można się zapoznać z ich treścią, jedynie oglądając wizerunki poszczególnych stron. Teksty powstałe z OCR są użyteczne, ale mają bardzo niewielką funkcjonalność, pełna wersja cyfrowa pozwoliłaby badaczom na zapoznanie się z jednostkami, które prawdopodobnie występują w treści haseł, lecz nie zostały uwzględnione jako osobne hasła (co do SW por. Wawrzyńczyk 2009). Ciekawe byłoby zwłaszcza zapoznanie się ze wszystkimi ciągami wyrazów bliskoznacznych, które występują w SW, np. w hasłach *kołtun*, *kręgosłup*¹⁷ prawdopodobnie są to wyłącznie rzeczowniki. Najwyraźniej są dwie cyfrowe wersje tekstowe SW, wspomniane wyżej, i być może znajdzie się odważny badacz, który będzie umiał sporządzić z nich wersję użytkową.

Bibliografia

- Bańko M. 1988: *Projekt porównawczej siatki haseł słowników języka polskiego*, [w:] Z. Saloni (red.), *Studia z polskiej leksykografii współczesnej*, Zakład Narodowy im. Ossolińskich, Wrocław, s. 75–95.
- Derwojedowa M. 2017: *Automatyczna analiza fleksyjna polszczyzny XIX wieku* (online: <http://www.fi9.uw.edu.pl>, dostęp: 21 października 2021).
- Deutsches Wörterbuch von Jacob Grimm und Wilhelm Grimm, digitalisierte Fassung im Wörterbuchnetz des Trier Center for Digital Humanities*, Version 01/21 (online: <https://www.woerterbuchnetz.de/DWB>, dostęp: 31 listopada 2021).
- Doroszewski W. 1954: *Z zagadnień leksykografii polskiej*, Państwowy Instytut Wydawniczy, Warszawa.
- Fotokorpus.pl (online: <http://nfpj.pl>, dostęp: 21 października 2021).

17 Intrygujące jest to, że są zestawy synonimów przy wyrazach oznaczających tabuizmy, np. *lechtaczka*, *pica*.

- Krawczyk M. 2018: wpis z 27.07.2018 r. (online: <https://www.instagram.com/p/BlufuBguSR/?hl=pl>, dostęp: 11 lutego 2022).
- Leśmian B. 1956/1978/2021: *Klechdy polskie*, Instytut Wydawniczy Pax, Warszawa (online: <https://wolnelektury.pl/katalog/>, dostęp: 3 lutego 2022; pierwodruk: B. Leśmian, *Klechdy polskie*, słowo wstępne B. Przyłuski, Katolicki Ośrodek Wydawniczy „Veritas”, Londyn 1956).
- Majdak M. 2008: *Słownik warszawski. Koncepcja – realizacja – recepcja*, nakładem Wydziału Polonistyki Uniwersytetu Warszawskiego, Warszawa.
- Majdak M. 2011: Słownik języka polskiego, *Warszawa 1900–1927*, red. Jan Karłowicz, Adam Antoni Kryński, Władysław Niedźwiedzki (online <http://www.leksykografia.uw.edu.pl/slowniki/35/sownik-jezyka-polskiego-warszawa-1900-1927>, dostęp: 31 listopada 2021).
- Majka 2021: komentarze (online: <https://www.grojec24.net/news-zbigniew-konczak-rezyguje-z-funkcji-prezesa-pcm-w-groju,2146.html>, dostęp: 5 maja 2021).
- Mortka M. 2021: *Głodna puszcza*, Wydawnictwo SQN, Kraków (online: <https://books.google.pl>, dostęp: 3 lutego 2022).
- Piotrowski T. 1994: *Z zagadnień leksykografii*, Wydawnictwo Naukowe PWN, Warszawa.
- Piotrowski T. 2001: *Zrozumieć leksykografię*, Wydawnictwo Naukowe PWN, Warszawa.
- Saloni Z., Bańko M. 2011: Słownik języka polskiego, *Warszawa 1958–1969*, red. Witold Doroszewski (online: <http://www.leksykografia.uw.edu.pl/slowniki/31/sownik-zyka-polskiego-warszawa-1958-1969>, dostęp: 7 lutego 2022).
- Shadow Laguna II 2019: wpis i wideo z 3.09.2019 r. (online: <https://www.facebook.com/ShadowLaguna2/>, dostęp: 11 lutego 2022).
- Simpson J., Weiner E.S. 1989: *The Oxford English dictionary*, wyd. 2, Clarendon, Oxford.
- SJPdor: *Słownik języka polskiego*, red. W. Doroszewski, t. 1–4, Wiedza Powszechna, Warszawa 1958–1962, t. 5–11, Państwowe Wydawnictwo Naukowe, Warszawa 1963–1969.
- SW: *Słownik języka polskiego*, red. J. Karłowicz, A.A. Kryński, W. Niedźwiedzki, t. 1–8, nakładem prenumeratorów i Kasy im. Mianowskiego, Warszawa 1900–1927 (przedruk: 1952–1953, Państwowy Instytut Wydawniczy, Warszawa). Wersja cyfrowa 1 (online: <https://crispa.uw.edu.pl/object/files/416370/display/Default>), wersja cyfrowa 2 (online: <https://polona.pl/item/sownik-jezyka-polskiego-t-1-a-g,NDEwOTewNzU/6/#info:search:Adnotowa%C4%87>) (dostęp: 31 listopada 2021).
- Wawrzyńczyk J. 2009: *Autosuplement do Słownika warszawskiego*, Sorus, Poznań (online: https://www.academia.edu/6255721/Jan_Wawrzy%C5%84czyk_Autosuplement_do_S%C5%82ownika_warszawskiego, dostęp: 3 lutego 2022).
- Woliński M., Saloni Z., Wołosz R., Gruszczyński W., Skowrońska D., Bronk Z. 2020: *Słownik gramatyczny języka polskiego*, wyd. 4, Warszawa (online: <http://sgjp.pl>, dostęp: 31 listopada 2021).
- Wywiad 2021 (online: <http://www.domseniorawielun.pl/wywiad.html>, dostęp: 5 maja 2021).

Summary

Quantitative structure of entries in *Słownik warszawski* in a comparative approach

Keywords: lexicography, dictionary, history, lexicology.

In the paper the exact number of entries in a large dictionary of Polish (called *Słownik warszawski*, 1900–1927) is given, following a computer analysis of a list of its entries. Furthermore, the structure of entries is analyzed, using selected qualifiers, and the data compared to those in a more recent large dictionary of Polish, edited by Witold Doroszewski (published 1958–1969).