

Co to są słowa książkowe? O potrzebie uściślenia kwalifikatorów słownikowych na przykładzie wyrazów opatrzonych kwalifikatorem *książk.* w „Słowniku języka polskiego” pod redakcją Mieczysława Szymczaka

Was sind ‘słowa książkowe’? Zur Definition von Wörterbuchqualifikatoren am Beispiel der Wörter mit dem Qualifikationsmerkmal ‚książk.’ in dem von Mieczysław Szymczak bearbeiteten Wörterbuch

Der Beitrag soll die Relevanz des Konzeptes von Qualifikationsmerkmalen der Wörter in Wörterbüchern betonen. Diese Qualifikatoren pragmatischer Art sollten objektiviert werden. Eine Möglichkeit diesbezüglich bieten die PELCRA-Suchmaschine und die PROFIL-Funktion in NKJP (dt. Nationalem Korpus der polnischen Sprache), die eine Analyse der stilistischen Distribution der Treffer ermöglicht. Die Überlegungen betreffen die Recherchemöglichkeiten der Verwendung von statistischen und Computerwerkzeugen zur Überprüfung von Wörterbuchqualifikatoren. Die Methode wird anhand einer Stichprobe von 100 eindeutigen Wörtern getestet, die im Polnischsprachigen Wörterbuch von Mieczysław Szymczak als ‚książk.’ gekennzeichnet sind, weil es eines der mysteriösesten Qualifikationsmerkmale ist.

Schlüsselwörter: Computerlinguistik, Qualifikatoren, Lexikographie, Wörterbücher

What are ‘Bookish’ Words? About the Need to Specify Dictionary Labels on the Example of *książk.*-Words in the “Polish Language Dictionary” by Mieczysław Szymczak

The article aims to explain the importance of qualitative methods in assigning style labels to words in dictionaries. Style labels, as elements of pragmatic information, should be as objective as possible, but in practice they often lack objectivity because they are assigned to words intuitively. A chance to give them more rigor is offered by the PELCRA search engine in the National Corpus of Polish which computes stylistic distribution of search words. The considerations concern the research possibilities that the use of statistical and computer tools in verifying dictionary qualifiers. The method is tested on a sample of 100 words marked as ‘bookish’ in the “Polish Language Dictionary” by Mieczysław Szymczak, because it is one of the most mysterious qualifiers.

Keywords: computational linguistics, stylistic distribution, dictionary labels, lexicography

Author: Karolina Kompa, University of Warsaw, ul. Krakowskie Przedmieście 26/28, 00-927 Warszawa, Poland, e-mail: karolina.kompa@gmail.com

Received: 15.2.2019

Accepted: 23.5.2019

Do przyjrzenia się tematyce kwalifikatorów słownikowych skłoniła mnie okolicencka lektura „Uniwersalnego słownika języka polskiego” (dalej USJP) pod redakcją Stanisława Dubisza z 2003 roku. Trafiłam na definicję leksemu *romansować* ‘mieć z kimś

romans – przygodę erotyczną, flirtować’ i zwróciłam uwagę na kwalifikator, którym go opatrzone – *książk*. Podeszłam do tej oceny bardzo sceptycznie – słowo wydawało się powszechne w użyciu, zwłaszcza że znalazłam je w jednym z tomów popularnonaukowej serii dla dzieci i młodzieży. Podział wydał mi się ugruntowany historycznie, a nie kontekstowo. Niejasne wydały mi się kryteria brane pod uwagę przy opracowywaniu haseł słownikowych i na jakiej podstawie zapadają decyzje o przyporządkowywaniu kwalifikatorów słownikowych danym jednostkom leksykalnym. Postanowiłam zweryfikować wybrany przez siebie kwalifikator za pomocą narzędzia, jakim jest wyszukiwarka PELCRA przy Narodowym Korpusie Języka Polskiego.

Kwalifikatory to napisy, zwykle skróty, stosowane w słownikach, zawierające informacje o zakresie używania danego wyrazu lub jego różnorodnemu nacechowaniu (np. stylistycznym, chronologicznym, ekspresywnym). W słownikach ogólnych, nazywanych zwykle słownikami języka polskiego, zazwyczaj przed definicją można znaleźć kwalifikator informujący o zakresie użycia danego wyrazu – to oznacza, że zakres jest ograniczony do pewnej odmiany (lub odmian) języka, czyli że wyraz jest nacechowany, nie jest neutralny i nie można go użyć w każdej sytuacji komunikacyjnej (Engelking/Markowski/Weiss 1989:300–309).

Słowniki ogólne mają charakter mniej lub bardziej normatywny, pomagają więc odbiorcy rozstrzygnąć, w jakim kontekście można użyć danego wyrazu, a w którym – nie. Kwalifikatory, jako elementy charakterystyki pragmatycznej wyrazów hasłowych, powinny być precyzyjnie używane i starannie objaśnione. Zwykle we wstępach słowników poświęca się im zaskakująco niewiele miejsca, co nie pozwala w pełni zrozumieć ich natury. Listy kwalifikatorów bywają zaskakująco długie, ale kryteria wyboru są zwykle nieobecne i nieprecyzyjne. Kwalifikatory zazwyczaj są zbyt ogólne lub zbyt szczegółowe, nie mają wyjaśnionych zakresów lub te zakresy się pokrywają (por. „Słownik języka polskiego” pod red. Witolda Doroszewskiego – dalej SJPDo – z prawdopodobnie synonimicznymi kwalifikatorami *dentystyka* i *stomatologia*; pojawiają się też tam kwalifikatory spoza spisu we wstępie, np. *provincjonalizm*), a definicje nie zawsze są wystarczająco precyzyjne (Majdak 2003:7–19). Jak podkreślają autorzy (Engelking/Markowski/Weiss 1989:300–309) w polskiej literaturze leksykograficznej brakowało opracowań porządkujących kwalifikatory, takich, które precyzują zakres ich użycia; nie było także określonej ich teorii. Proponują podział na typy kwalifikatorów – chronologiczne, geograficzne, stylistyczne, ekspresywne, zawodowe, środowiskowe i frekwencyjne.

Zgodnie z propozycją (Engelking/Markowski/Weiss 1989:300–309) kwalifikator *książkowy* należałoby dołączyć do grupy kwalifikatorów stylistycznych, co powoduje przypisanie mu pewnego nacechowania. Autorzy odróżniają nacechowanie stylistyczne od ekspresywnego na podstawie oddziaływania wyrazu na sfery umysłu (oceny intelektualne bądź emocjonalne). W tych założeniach kwalifikator *książkowy* znalazłby się obok trzech innych kwalifikatorów stylistycznych: *podniosłego*, *poetyckiego* oraz *oficjalnego*.

W zaproponowanym podziale można znaleźć informacje dotyczące kanału komunikacyjnego, czyli sposobu, w jaki tekst zostaje przekazany odbiorcy (Górski/Łaziński 2012:13–22), por. „*podniosły* – [...] Ograniczony do tekstów pisanych i nielicznych mówionych”, wraz z wymienionymi gatunkami, jednak kwalifikator *książkowy* wydaje się najbardziej zagadkowy:

książkowy — wyraz o zabarwieniu erudycyjnym, a więc taki, którego zrozumienie wymaga pewnego przygotowania, wiedzy na poziomie przynajmniej szkoły średniej. W większości są to wyrazy obce (np. *komercjalizacja*, *weryfikacja*, *konsumować*, *pluralistyczny*¹).

Mimo podjętej próby systematyzacji kwalifikatorów brakuje w tej definicji informacji o typach tekstów charakteryzujących się występowaniem wyrazów *książkowych*. Dowiadujemy się, jak słowo książkowe powinno wyglądać i jaki etap edukacyjny może być niezbędny do jego zrozumienia. Oczywiście sama nazwa kwalifikatora może być pewną wskazówką, jednak nie informuje nas ona, w jakich typach tekstów wyrazy książkowe mogą występować, czy mogą one przenikać do tekstów innych niż te, które zostały wydane w formie książki. W USJP znajdujemy kolejną odpowiedź nakierowującą na szeroko rozumianą książkowość. Według podanej tam definicji kwalifikator *książk.* dotyczy wyrazów rzadko występujących w polszczyźnie, raczej nieprzenikających do warstwy potocznej lub wyrazów występujących w mowie inteligencji humanistycznej. Sygnalizuje ona złożony charakter opisywanego kwalifikatora, jednak nie przedstawia precyzyjnych kryteriów, na podstawie których leksykograf podejmuje decyzję.

Szczegółowe przyjrzenie się wyrazom z określonym kwalifikatorem pod kątem kanału, za pomocą którego mogą się one rozprzestrzeniać, a nawet – popularyzować, może przynieść wiele korzyści. W uproszczeniu, według wcześniejszych założeń, kategoryzacje bazują na podziałach fizycznych (związanych z geografiami) lub charakterystyce użytkownika. Kanał nie jest najistotniejszy w ocenie stylistycznej wyrazu, jednak pozwala wnioskować na temat prawdopodobieństwa spotkania określonego typu wyrazu w źródle. Związek między kanałem a typem tekstu jest luźny, choć interesujący ze względu na sposoby transmisji określonych typów. Dostępność do tekstów zróżnicowanych leksykalnie (albo niezróżnicowanych) wpływa na to, jak dane jednostki mogą być postrzegane przez odbiorców.

Jak zatem zarysować zbiór, wewnątrz którego znajdują się wyrazy określane przez leksykografów jako *książkowe*? Odpowiedzią na to pytanie jest językoznawstwo korpusowe. Już Linde w przełomowym „Słowniku języka polskiego” (1807–1814) wykorzystał 850 tekstów źródłowych, reprezentujących różne odmiany piśmiennictwa polskiego od XVI

¹ W „Słowniku języka polskiego” pod red. M. Szymczaka żaden z tych wyrazów nie jest opisany kwalifikatorem *książk.*, choć rok wydania słownika zbliża się do daty publikacji artykułu, więc można by sądzić, że odczucia badaczy będą częściowo spójne.

po koniec XVIII w. Obecnie nie trzeba prowadzić papierowej kartoteki, aby zgromadzić odpowiednią próbkę materiału.

Mieczysław Szymczak w przedmowie do „Słownika języka polskiego” z 1978 r. (dalej SJPSz) podkreśla ważkość opracowania słownika współczesnego języka polskiego, który obejmowałby wszystkie odmiany polszczyzny ogólnej. Zdaje sobie sprawę z mnogości wyzwań, które takiemu przedsięwzięciu towarzyszą, i sugeruje wykorzystanie maszyn matematycznych, zwłaszcza w pierwszym etapie pracy, czyli zbieraniu materiału. Można powiedzieć, że marzenie Szymczaka się ziściło – w XXI w. dysponujemy narzędziem, które pozwala nie tylko magazynować materiał, lecz także błyskawicznie go przeanalizować. Praca nad hasłem słownikowym nie polega już na piętrzeniu karteczek z licznymi wyimkami z tekstów, które mają stanowić korpus. Istnieją takie użycia wyrazów, których – bez wglądu w kontekst – nie jesteśmy w stanie dobrze wyjaśnić i opisać. Dlatego obszerna baza tekstów i możliwość weryfikacji poglądów mogą okazać się niezbędne w obiektywizacji badań nad kwalifikatorami słownikowymi. Nasuwa się wątpliwość, czy badacz, który przypisuje konkretny kwalifikator konkretnej jednostce językowej, ma rację. Czy leksykografia nie opiera się wyłącznie na intuicji osób, których polszczyzna może nie odzwierciedlać uzusu i nie są w stanie przyjąć perspektywy przeciętnego użytkownika języka polskiego.

Nieścisłości związanych z kwalifikatorami słownikowymi jest oczywiście więcej – te same wyrazy mogą mieć różne kwalifikatory, w zależności od autora słownika. Tym razem chciałabym jednak skupić się jednym kwalifikatorze – i zaproponować jego weryfikację na wyrazach jednoznacznych. W tym celu wyekscerpowałam 100 jednostek leksykalnych ze „Słownika języka polskiego” pod red. Mieczysława Szymczaka. Trzytomowe wydanie z 1988 r. podzieliłam na sto równych części i z każdej wybrałam pierwszy spotkany wyraz monosemiczny opatrzony interesującym mnie kwalifikatorem. Pomocnym narzędziem okazało się również wydanie elektroniczne tego słownika z 1998 r., które usprawniło cały proces, dzięki funkcji wyszukiwania wyrazów za pomocą danego kwalifikatora. W badaniu wzięłam pod uwagę wyrazy, które w Narodowym Korpusie Języka Polskiego (dalej NKJP) miały frekwencję powyżej zera. Podczas całego badania ani razu nie zdarzyło się, aby w danej części słownika nie znajdował się wyraz kwalifikowany jako *książkowy*. Ekscerpcję utrudniały natomiast kwalifikatory nakładające się na siebie – stosunkowo często kwalifikatorowi *książk.* towarzyszył *przen.*, przez co należało tę jednostkę pominąć (podwójne kwalifikatory sugerują niejednoznaczność) i w badaniach uwzględnić pierwszy dostępny i jednoznaczny pod względem dystrybucji stylistycznej leksem.

„Słownik języka polskiego” pod red. Mieczysława Szymczaka przyjąłam za podstawę materiałową z dwóch powodów. Po pierwsze, chciałam przeanalizować wyłącznie wyrazy monosemiczne – czyli takie, które mają tylko jedno znaczenie. Po drugie, szukałam wyrazów opatrzonych wyłącznie jednym kwalifikatorem. Nowsze wydania słowników, np. USJP, zawierają więcej kwalifikatorów lub są całkiem ich pozbawione

(por. „Inny słownik języka polskiego” pod red. Mirosława Bańko)². Zależało mi na wybraniu typowego słownika ogólnego, wystarczająco dużego, aby wybrać z niego potrzebny materiał. Ponadto przydatna okazała się wersja elektroniczna słownika („Komputerowy słownik języka polskiego” z 1998), dzięki której wyszukiwanie wyrazów okazało się znacznie szybsze niż w wersji tradycyjnej.

Szymczak ani we wstępie, ani w informacjach szczegółowych nie poświęca zbyt wiele miejsca objaśnianiu natury kwalifikatorów. Informuje jedynie, że służą one do charakterystyki wyrazów pod względem chronologicznym, stylistycznym oraz społeczno-środowiskowym. Ogólność tych wskazówek oraz rok wydania słownika tylko pozornie nie sprzyjają współczesnym badaniom. Dzięki oszczędności środków (czyli oszczędnemu przypisywaniu kwalifikatorów poszczególnym leksemom) w tym słowniku mogłam go potraktować jako źródło wyrazów *książkowych*, które poddałam dalszej analizie. Posłużyłam się wyszukiwarką PELCRA w Narodowym Korpusie Języka Polskiego. Zrównoważona wersja korpusu liczy ponad 250 milionów słów tekstowych. PELCRA dysponuje funkcją Profil, która pozwala na analizę dystrybucji stylistycznej wyrazów (Pęzik 2012). Wybór jednoznacznych jednostek leksykalnych umożliwił zbliżenie się do obiektywizacji w tym badaniu – wyszukiwane wyrazy nie generowały kolejnych wątpliwości, mogłam więc mieć pewność, że znaczenia będą książkowe. Czy słusznie?

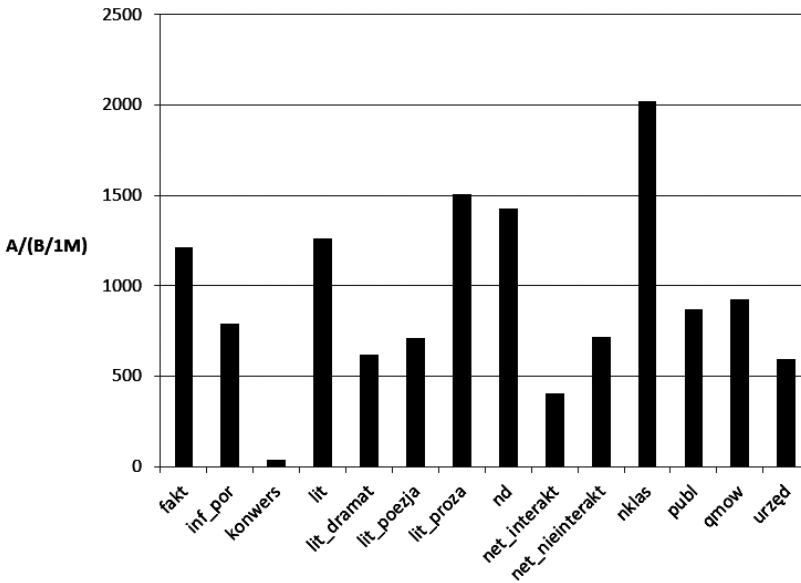
Wśród leksemów opatrzonych kwalifikatorem *książk.* w SJPSz znalazłam takie, których intuicyjnie nie przydzieliłabym do tej kategorii. Były to np. *korektywa*, *rekreacja*, *wojenka* czy *urokliwy*. Z różnych przyczyn te wyrazy mogły zmienić nacechowanie stylistyczne, a przeszukanie korpusu pod tym kątem umożliwiło weryfikację moich podejrzeń. Okazało się, że wyniki są niejednorodne. Blisko połowa ze stu analizowanych słów dociera do odbiorców za pośrednictwem kanału innego niż książka. Są to różne rodzaje prasy – pogrupowane przez autorów korpusu na „prasę_miesięcznik”, „prasę_tygodnik” i „prasę_inne” – oraz przestrzeń internetowa czy nawet kanał „mówiony” (w nim na uwagę zasługuje przysłówek *totalnie*, w SJPSz uznany przecież za wyraz *książkowy*). Podobnie z wynikami uzyskanymi po analizie typów funkcjonalnych używanych w NKJP – spodziewałam się przede wszystkim znaczącej grupy reprezentującej literaturę piękną, jednak te wystąpienia oscylują na wysokości ok. 30%. Dominowały za to słowa gatunkowo przynależące do literatury faktu, publicystyki, typu naukowo-dydaktycznego, quasi-mówionego, a także znalezione w zbiorze źródeł internetowych nieinteraktywnych (jak strony domowe, blogi bez komentarzy).

Taki wybór jednostek leksykalnych poddawanych analizie miał na celu zwiększenie obiektywności badania oraz sprawdzenie przyjętej metody badawczej. Okazało się, że

² Na uwagę zasługuje także „Słownik współczesnego języka polskiego” pod red. Bogusława Dunaja (1996), jednak nie wyróżnia się w nim kwalifikatora *książk.*, który wydaje mi się najbardziej interesujący.

to rozwiązanie nie jest wolne od wad, co szczególnie jest widoczne na wykresie nr 1. Współczynnik $A/(B/1M)$ oznacza sumę wystąpień stu leksemów z kwalifikatorem *książk.* (A) podzieloną przez iloraz wszystkich jednostek leksykalnych danego typu zgromadzonych w NKJP (B) i podzielonych przez milion ($1M$). Tak przedstawiony wynik pozwala oszacować frekwencję wyrazów *książkowych* na milion wyrazów w konkretnym typie (np. nklas – niejasny typ funkcjonalny tekstu objaśniany przez NKJP jako „inne”). O ile większość słupków oscyluje wokół podobnych wyników (dominują typy okołoliterackie, czyli: lit – literatura, lit_proza – literatura proza, fakt – literatura faktu), to nie da się przeoczyć typu nklas. Na podstawie wykresu i bez znajomości szczegółowych wyników badań można stwierdzić, że słowa książkowe są charakterystyczne dla tego typu. Byłoby to znaczące uogólnienie i rodzaj pułapki, w którą łatwo wpaść, gdy przed oczami ma się jedynie wyimek.

W typie nklas występuje 2 245 020 jednostek (por. lit – 30 230 684), po zsumowaniu wystąpień z losowo wybranych stu leksemów tych oznaczonych w SJPSz jako *książkowe* będzie aż 4533. Tak przedstawione wyniki nie oddają w pełni problemu, którym jest częściowo losowo dokonywany wybór. Otóż o tak wysokim wskaźniku wyrazów książkowych zadecydowały trzy jednostki: *lecz* – 2053, *bowiem* – 1662 oraz *rzec* – 283. Po ich odjęciu średnia frekwencja pozostałych jednostek książkowych nie byłaby tak imponująca (z 4533 wystąpień – zaledwie 535), ponieważ wynosiłaby niecałe 24 wystąpienia na milion.



Wykres 1. Dystrybucja stylistyczna wyrazów książkowych w NKJP

Wykorzystanie korpusu w badaniach językoznawczych – i tworzeniu słowników – pozwala na większą funkcjonalność, dzięki ograniczeniu przeszukiwania tekstów

do wybranych typów. Oczywiście za klasyfikację stylistyczną w NKJP odpowiadają badacze, jednak nie oceniają oni poszczególnych jednostek leksykalnych pod kątem stylistycznym, tylko teksty – o charakterze wyrazu decyduje jego frekwencja w określonych typach tekstów, i dopiero taka weryfikacja może okazać się istotną przesłanką do stworzenia podziału na kwalifikatory słownikowe. Definicje słownikowe miałyby wtedy cel praktyczny i zwracałyby uwagę na szczegóły drugorzędne, tak istotne w interpretacji językowego obrazu świata przeciętnego użytkownika języka polskiego. Publikacje nieoparte na korpusie prawdopodobnie bazują na wcześniejszych, niezaktualizowanych wydaniach słowników. Praca oparta na przedawnionym materiale traci na użyteczności. Interesujące byłoby porównanie wyekscerpowanych przeze mnie wyrazów z innymi słownikami ogólnymi, tymi nowszymi, i zweryfikowanie kwalifikatorów przypisanych przez leksykografów.

Metoda weryfikacji kwalifikatorów za pomocą analizy dystrybucji stylistycznej wyrazów mogłaby doprowadzić do powstania całkiem nowego systemu kwalifikatorów słownikowych, np. z ich podtypami. Oczywiście należy mieć na uwadze, że sformułowanie nowych kryteriów i ocena za pomocą korpusu nie pozwoliłyby uniknąć niekonsekwencji w opisie leksemów, ponieważ różnice niekiedy są subtelne, ale z pewnością umożliwiłoby znacznie ograniczenie niespójności w obrębie publikacji (nawet gdyby to rozwiązanie zastosować w tylko jednej). Na podstawie poczynionych obserwacji można stwierdzić, że status kwalifikatora *książk.* jest dość niejasny. Nie tylko brakuje ogólnej metodologii opisu książkowości, lecz także uwidacznia się brak wypracowanych i konsekwentnie stosowanych reguł. Losowy wybór wyrazów, którego się podjęłam, nie pozwala na przedstawienie pełnej problematyki – na liście nie mogły pojawić się wyrazy wieloznaczne czy takie, które należało pominąć ze względu na kolejność alfabetyczną. Przyjęte przeze mnie rozwiązanie umożliwiło za to przetestowanie opisywanej wcześniej metody. Otwiera to pole do kolejnych badań, nad którymi z pewnością warto się pochylić.

Wykaz literatury

- ENGELKING, Anna, MARKOWSKI, Andrzej i Elżbieta WEISS. „Kwalifikatory w słownikach – próba systematyzacji”, *Poradnik Językowy* z. 5 (1989): 300–309. Print.
- GÓRSKI, Rafał L. i Marek ŁAZIŃSKI. „Typologia tekstów w NKJP”. *Narodowy Korpus Języka Polskiego*. Red. Adam Przepiórkowski, Mirosław Bańko, Rafał L. Górski, Barbara Lewandowska-Tomaszczyk. Warszawa: PWN, 2012, 13–22. Print.
- MAJDAK, Magdalena. „Inne kwalifikatory w *Innym słowniku języka polskiego*. Wybrane zagadnienia”. *Zeszyt 2 Studenckich Prac Naukowych Koła Młodych Językoznawców*. Red. Ewa Kołodziejek. Szczecin: Uniwersytet Szczeciński, 2003, 7–19. Print.
- PĘZIK, Piotr. „Wyszukiwarka PELCRA dla danych NKJP”. *Narodowy Korpus Języka Polskiego*. Red. Adam Przepiórkowski, Mirosław Bańko, Rafał L. Górski i Barbara Lewandowska-Tomaszczyk. Warszawa: PWN, 2012, 253–274. Print.
- Komputerowy słownik języka polskiego*. Warszawa: PWN, 1998.

Słownik języka polskiego. Red. Witold Doroszewski. Warszawa, 1958–1969. Print.

Słownik języka polskiego. Red. Mieczysław Szymczak. Warszawa, 1978–1981. Print.

Uniwersalny słownik języka polskiego. Red. Stanisław Dubisz. Warszawa, 2003. Print.

ZITIERNACHWEIS:

KOMPA, Karolina. „Co to są słowa książkowe? O potrzebie uściślenia kwalifikatorów słownikowych na przykładzie wyrazów opatrzonych kwalifikatorem książk. w Słowniku języka polskiego pod red. Mieczysława Szymczaka.” *Linguistische Treffen in Wrocław* 15, 2019 (I): 295–302. DOI: 10.23817/lingtreff.15-24.