

# Wybrane elementy opisu języka specjalistycznego (na przykładzie języka retoryki)

Łukasz Karpiński  
Uniwersytet Warszawski

## **Abstract**

The article attempts to demonstrate the procedure of parametric analysis based on the collected textual corpora in rhetoric. The theoretical part outlines the systems which inform the practical analysis of the research, namely lexical-communicative and terminological system, praxeology, as well as terminographic engineering. The selected indices of the rhetorical corpus are analysed on the basis of comparative corpora which represent other thematic stylistic levels.

**Keywords:** microstructure, macrostructure, metastructure of dictionary, terminological system, lexical-communicative system, terminographic praxeology, terminographic engineering

## **Abstrakt**

Artykuł zawiera omówienie analizy parametrycznej, przeprowadzonej na zebranych korpusie tekstów z zakresu retoryki. W części teoretycznej przedstawiono zarys koncepcji systemu leksykalno-komunikacyjnego, systemu terminologicznego, prakseologii oraz inżynierii terminograficznej, w zakresie której pozostają przedstawione praktyczne działania. Wybrane wskaźniki opisujące korpus retoryczny zostały przy tym omówione na tle korpusów porównawczych, reprezentujących inne poziomy tematyczne i stylistyczne.

**Słowa kluczowe:** mikrostruktura, makrostruktura, metastruktura słownika, system terminologiczny, system leksykalno-komunikacyjny, prakseologia terminograficzna, inżynieria terminograficzna

Powszechnie wiadomo, że język jest niezbędnym nośnikiem informacji i jednocześnie narzędziem opisu świata. Problem pojawia się, gdy dana społeczność wprowadza do języka zbędne jednostki pod wpływem chwilowej mody, manieri lub dla wyróżnienia się z ogółu. Szczególnie ten ostatni czynnik odgrywa dość istotną rolę w komunikacji, gdyż elokucyjna celowość i stosowność użycia odpowiednich określeń i podporządkowanie stylistyki celom funkcjonalnym ma zazwyczaj na celu uzyskania określonego efektu komunikacyjnego. Zjawisko to jest widoczne od czasów starożytnych, jednakże jego natężenie jest zmienne i uwarunkowane wieloma czynnikami kulturowymi, historycznymi, politycznymi czy społecznymi. Arystoteles (2001, 1358) wyróżnił trzy rodzaje mów, ukierunkowanych na różne sytuacje komunikacyjne<sup>1</sup>. Dwa pierwsze rodzaje odnoszą się do sytuacji specjalistycznych, natomiast trzeci rodzaj – retoryka popisowa – może być identyfikowany w szerokiej gamie tekstów tworzonych i wygłaszanych zarówno przez władców, polityków, przedsiębiorców, specjalistów od reklamy jak i innych osób, chcących za pomocą komunikacji werbalnej osiągnąć określony cel<sup>2</sup>.

Odwołując się dalej do klasycznej organizacji wypowiedzi (Arystoteles 2001, 1358–1359), wg Arystotelesa każda mowa, (a traktując ten porządek współcześnie – każdy tekst) jest tworzony w oparciu o pięć podstawowych elementów: inwencję, kompozycję, elokucję, mnemonikę i actio. Z punktu widzenia leksykografii to właśnie elokucja stanowi najbardziej interesujący element, gdyż zasady odpowiedniego doboru słownictwa i stosowania poprawnej stylistyki stanowią elementy rejestrowane w różnorodnych kompendiach słownikowych. Przytaczany przy okazji pojęcia elokucji klasyczny cytat z „Ogólników” C. K. Norwida jest tutaj trafny nie tylko ze względu na podstawowe znaczenie, odnoszące się do literackiego aksjomatu, iż słowa służą do nazywania tego, co

<sup>1</sup> Retoryka doradcza (łac. *genus deliberatorium*) polega na doradzaniu lub odradzaniu działania w zakresie jakiegoś problemu prywatnego lub publicznego. Retoryka sądowa (łac. *genus indiciale*) obejmuje wypowiedzi perswazyjne w zakresie oskarżania lub obrony, w związku z czym mogą dotyczyć przedstawienia dowodów winy lub ich zbijania. Retoryka popisowa (*genus demonstrativum*) obejmuje wypowiedzi perswazyjne zawierające pochwałę albo naganę osób, rzeczy czy wartości.

<sup>2</sup> por. Lichański J., *Miejsca wspólne – perspektywa polska. Przegląd problemów i prolegomena dalszych badań*. W: Forum Artis Rhetoricae, 16–17, 2009, s. 6–23.

człowiek myśli i czuje, a więc wyrażania wszelkich uczuć, doznań, stanów i przekonań. „Ponad wszystkie wasze uroki (...) Odpowiednie dać rzeczy słowo!”

C.K. Norwid pisze u schyłku romantyzmu, przychyłając się jakby do postulatów następczej epoki – pozytywizmu, głoszącego, iż jedynie prawdziwą wiedzą jest ta, która może być zdobyta tylko dzięki pozytywnej weryfikacji teorii za pomocą empirycznej metody naukowej. W polskiej literaturze początek pozytywizmu jest jasno określony – to upadek powstania styczniowego, 1863 rok. Być może niedługo na gruncie leksykografii za przełomowy będzie uważany rok 2010, kiedy po raz ostatni ukazała się drukowana wersja encyklopedii *Britannica*. Przeniesienie tak fundamentalnego dzieła na grunt mediów elektronicznych jest symbolicznym zamknięciem okresu „papierowej” leksykografii i potwierdzeniem współczesnych kierunków rozwoju tej dziedziny, tj. definitywnym przejściem od papierowych fiszek do elektronicznych baz terminologicznych o nieograniczonej pojemności. Oznacza to, iż własności przypisywane elokucji mogą być badane i określane przy pomocy czysto empirycznych metod, a przy zastosowaniu odpowiedniej metodologii dokładnie opisywane cyfrowo<sup>3</sup>. Niniejszy artykuł ma za zadanie przedstawić zarys problematyki rejestracji zmian zachodzących w języku w leksykograficznych cyfrowych bazach danych na językowym materiale z zakresu retoryki oraz dla celów porównawczych na innych, kontrastujących korpusach tekstowych.

W kontekście kultury języka poprawne i jasne wysłowienie się stanowi podstawę komunikacji, jednakże wszelkie ozdobniki i modne wyrażenia są zazwyczaj używane w celu osiągnięcia określonego efektu i to zazwyczaj nie tylko komunikacyjnego. Ważne są jednak rozmiary tego zjawiska. Jakie wnioski wyprowadzi współczesny świadomy użytkownik języka polskiego, skoro już na przełomie lat 60-tych i 70-tych XX wieku językoznawcy pisali: „Treścią zainteresowań językowych szerokich kręgów inteligencji jest niepokój, niepokój o to, że język polski znajduje się w stanie zagrożenia, że ze wszystkich stron czyhają na niego niebezpieczeństwa w postaci różnego rodzaju zanieczyszczeń, naleciałości, dialektyzmów, wyrażen żargonowych, barbaryzmów, że toczy go choroba sztampy dziennikarskiej, że żre go mól cynizmu na modłę młodzieżową, że zabija go moda na wyspecjalizowany żargon naukowy. (...) Sytuację (językową) po drugiej wojnie światowej charakteryzuje przede wszystkim przejście do warstw wykształconych setek tysięcy ludzi obciążonych pewnymi relikami gwarowymi czy żargonowymi. Ludzie trawieni troską

<sup>3</sup> Więcej o parametryzacji znaczeniowej tekstu, wybranych wskaźnikach tekstowych patrz np. Karpiński Ł., *Wybrane założenia komputerowej analizy tekstów i gromadzenia danych*. W: *Języki specjalistyczne*, t. 9, s. 9–23 lub Karpiński Ł., *Analiza parametryczna tekstu a translacja maszynowa – wybrane zagadnienia*. W: „The Linguistic Academy Journal of Interdisciplinary Language Studies”, 2 (2012), s. 103–114.

o „czystość języka” zauważają to i niepokoją się” (Wierzbicka, Wierzbicki 1969, 34–35). Cóż powiedzieć o współczesnych, zglobalizowanych czasach, w których wyrażenie *euro-sceptycznymerchandiserwdresie zoxy-biodegradowalnymgabarytemwSUV-ie pod dyskontem* może być przez niektórych użytkowników języka polskiego uważane wręcz za naturalne.

W tym miejscu należy postawić pytanie, w jaki sposób i na podstawie jakich wytycznych rejestrować zmiany w systemie językowym, w tym nowopowstałe neologizmy i zapożyczenia. Przypomnijmy, iż w polskiej tradycji językoznawczej działalność związana z tworzeniem różnego rodzaju leksykonów była rozdzielana pomiędzy kilka jednostek nauki. Według koncepcji Witolda Doroszewskiego leksykologia (a więc działalność badająca znaczenie i funkcjonowanie wyrazów) oraz leksykografia (leksykologia stosowana) stanowią działy językoznawstwa, rozpatrywane łącznie ze względu na ich ściśle powiązane pola semantyczne. Historycznie i generycznie rzecz ujmując, leksykografia wyodrębniła się znacznie wcześniej od leksykologii, była też jedną z kilku dyscyplin, których przedmiot stanowiła refleksja nad słowem, przy czym refleksji tej towarzyszyło gromadzenie i porządkowanie materiału leksykalnego (Miodunka 1989, 9). W zakres kompetencji leksykografii wchodzi więc działania związane z metodyką gromadzenia i analizy materiału językowego, typologia słowników, model pracy nad określonym typem leksykonu oraz porządek, sposób i narzędzia potrzebne do przeprowadzenia odpowiedniej syntezy zebranych danych w postaci opisu znaczenia w haśle. Do narzędzi można współcześnie zaliczyć dostęp do wszelkich możliwych korpusów językowych oraz informatyczne narzędzia analizy tekstowej, odpowiednie oprogramowanie, zapewniające pracę z bazami danych, edytowanie poszczególnych haseł słownika oraz przygotowanie gotowego materiału do wydania w formie tradycyjnej – papierowej lub elektronicznej.

Koncepcja słownika zawiera się w opracowaniu jego mikrostruktury, makrostruktury i metastruktury. Mikrostruktura to ogół zadań związanych ze sformułowaniem ogólnej idei słownika, ukierunkowaniu go na określonego odbiorcę, określeniu bazy konceptualnej i szeroko pojętej harmonizacji danych. Uwzględnienie odpowiedniego ujęcia problemu – a więc czy będzie to słownik językowy, encyklopedyczny, dydaktyczny czy terminologiczny<sup>4</sup>, jakiej objętości i typu (przekładowy, etymologiczny, minimum,

---

<sup>4</sup> *Słowniki językowe* (SJ) koncentrują się na opisie kontekstów znaczeniowych danego wyrażenia oraz podaniu możliwie wyczerpującego zakresu jego użycia. Definicje i objaśnienia w SJ powinny mieć zwięzły charakter. W artykule hasłowym występują kolokacje lub zdania standardowe, ilustrujące użycie, a także kwalifikatory stylistyczne i znaczeniowe oraz komentarz gramatyczny. Nadrzedną funkcją SJ jest prezentowanie i korygowanie wiedzy o języku. *Słowniki encyklopedyczne* (SE) przedstawiają otaczającą rzeczywistość w możliwie najszerszym ujęciu, z uwzględnieniem wszelkich treści kulturowych i historycznych. SE mają wszechstronnie służyć różnym typom użytkowni-

frekwencyjny itd.) wraz z ustaleniem granic opracowywanej dziedziny wiedzy daje w efekcie precyzyjne określenie idei słownika oraz przeniesienie tych założeń do (elektronicznej) bazy danych. To dostosowanie zebranych danych do budowy artykułów hasłowych w słowniku odnosi się do mikrostruktury słownika. Z kolei metastruktura słownika, rozumiana jako ogół zjawisk diachronicznych i synchronicznych, uwzględnianych w procesie tworzenia działu leksykograficznego odnosi się do dwóch pojęć<sup>5</sup>:

- *systemu leksykalno-komunikacyjnego*, tj. wielostopniowego, interdyscyplinarnego, spójnego, przekształcalnego i ewoluującego układu zdefiniowanych relacji i sprzężeń odniesiony do materialnej lub abstrakcyjnej postaci informacji językowej<sup>6</sup>;
- *systemu terminologicznego*, rozumianego jako zbiór elementów i działań zachodzących między nimi, mających na celu stworzenie i aktualizację uniwersalnej bazy danych terminologicznych, składającej się z uszeregowanych tematycznie parametrów w celu uzyskania optymalnej możliwości selekcji informacji w zależności od potrzeb użytkowych, jak najdokładniej odzwierciedlających strukturę leksykalną, gramatyczną i systemową danego języka (specjalistycznego).

ków. Zawartość tematyczna jest zróżnicowana, zależnie od zakładanej objętości i tematyki. Artykuł hasłowy ma w odpowiedni sposób opisywać aktualny stan całościowej wiedzy na temat tytułowego pojęcia. *Słowniki dydaktyczne* (SD) prezentują podejście pośrednie pomiędzy SE i ST, ze względu na podyktowane procesami nauczania obszernie objaśnienia odpowiednich pojęć i zjawisk. SD ma reprezentować dostosowany do wymogów dydaktycznych typ myślenia i działania zawodowego w określonej dziedzinie. Hasła powinny występować w odpowiednio szerokim paradygmacie terminologicznym i być przedstawione jako element systemu języka specjalistycznego. Słownik tego typu powinien być ściśle skorelowany z procesem nauczania. *Słownik terminologiczny* (ST) skupia się na przedstawieniu wybranej jednostki języka jako elementu większego systemu językowego. Cechą charakterystyczną jest brak nacechowania emocjonalnego – ewentualne kwalifikatory stylistyczne mają obowiązywać użytkownika do użycia wyrażenia w dokładnie określonym kontekście. ST muszą być tak konstruowane, aby odzwierciedlały podstawowe cechy języka specjalistycznego, tj.: elastyczność, produktywność, zmienność i otwartość systemu. ST mogą być budowane dla każdego zakresu tematycznego, można nimi objąć praktycznie cały zakres wiedzy ludzkiej. Por. Karpiński 2008, 59–62.

<sup>5</sup> Koncepcja „systemu leksykalno-komunikacyjnego”, „systemu terminologicznego” oraz zdefiniowanie „prakseologii terminograficznej” i „inżynierii terminograficznej” stanowi jeden z elementów przygotowywanej obecnie większej rozprawy naukowej. Wybrane elementy zostały przy tym przedstawione w pracach: Karpiński 2012, 2014a, 2014b.

<sup>6</sup> System jest wielostopniowy, gdyż można w nim wyróżnić kolejne etapy rozwoju i wydzielić mikrocykle metodologiczne: rozpoznanie problemu, modelowanie i heurystyczne poszukiwanie rozwiązań oraz wyprowadzanie określonych wniosków na podstawie poczynionych obserwacji; interdyscyplinarny – gdyż wkracza poza badania ściśle lingwistyczne; spójny – gdyż zbudowany jest z czytelnych metodologicznie elementów, ułożonych w kolejności „od szczegółu do ogółu; relacje i sprzężenia dotyczą przetwarzania informacji językowej.

Pierwszy z systemów stymuluje zadania związane z *prakseologią terminograficzną*, która stanowi w pewien sposób rozwinięcie funkcjonującego pojęcia *zasady pracy terminograficznej* (por. STP 2005, 158) i oznacza humanistyczną (językoznawczą) część działalności specjalistycznej, formułującą merytorycznie optymalną realizację paradygmatu naukowego w zakresie analizy tezaurusów i tworzenia systemów terminograficznych, stanowiących podstawę różnorodnych opracowań słownikowych. Jest to działalność nadrzędna i pierwotna w stosunku do inżynierii terminograficznej (powiązanej z drugim z systemów) ze względu na kontrolowanie całego cyklu badawczego, formułowanie zasad wdrażanych w praktyce terminograficznej oraz prymarny językoznawczo charakter dyscypliny. Postulowany prymarny charakter wynika z faktu, iż to językoznawcy kontrolują, czy informatycy prawidłowo kodują cechy języka i zarządzają zasobami zebranymi w bazach danych. Informatyka jest narzędziem, terminologia jest odpowiedzialna za opracowanie koncepcyjne.

Drugi system odnosi się do czynności z zakresu inżynierii terminograficznej, także stanowiących rozwinięcie dotychczasowych zasad pracy terminograficznej, idących w kierunku matematycznej (informatycznej) części działalności specjalistycznej, stosującej w praktyce tworzenia systemów terminograficznych wytyczne i zasady opracowane w prakseologicznym cyklu badawczym. Inżynierię terminograficzną można traktować jako działalność wtórną w stosunku do prakseologii terminograficznej, gdyż stanowi zbiór działań praktycznych (technologicznych) polegających na projektowaniu, konstrukcji, modyfikacji i utrzymaniu efektywnych kosztowo rozwiązań dla praktycznych problemów z wykorzystaniem wiedzy naukowej oraz technicznej. Doświadczenia zdobyte w praktyce tworzenia i modyfikowania systemów terminologicznych stanowią integralny element, sygnał zwrotny dla doskonalenia prakseologii terminograficznej.

Powyższe rozdzielenie działań nie jest przyczynkiem do dzielenia prac leksykograficznych na kolejne subdyscypliny, lecz uporządkowanie działań i powierzanie ich odpowiednim specjalistom, gdyż wyraźnie część prac powinna być kierowana przez językoznawców, a część – przez przedstawicieli informatyki we współpracy z leksykografami.

Odnosząc się do tytułowych poszukiwań danych w zakresie cyfrowego opisu języka specjalistycznego należy wspomnieć, iż do głównych procesów weryfikacji materiału terminologicznego można zaliczyć następujące działania:

- analiza i weryfikacja korpusów tekstowych;
- określenie rejestru stylistycznego;
- badanie frekwencji wyrażen;
- określenie wstępnych parametrów modalności tekstu.

W przypadku nasuwających się wątpliwości językowych warto także wspomagać proces weryfikacyjny bezpośrednim przeglądem zawartości Internetu, jednakże nie powinno to stanowić podstawy prac terminograficznych.

## Korpus tekstowy

Wydawnictwa wprowadzające na rynek nowe (lub aktualizowane) słowniki coraz częściej z dumą podkreślają, iż baza językowa w tych kompendiach została zaktualizowana i zweryfikowana przy pomocy korpusu językowego. Inna metoda kontroli słownictwa to zaawansowane korzystanie z usług wszelkiego rodzaju wyszukiwarek zbierających dane z ogólnodostępnej sieci Internet. Korpus powinien zawierać znaczną liczbę słów tak, aby spełniał takie podstawowe warunki jak reprezentatywność, aktualność, jakość i autentyczność. Jeśli przyjąć, że dla współczesnego języka polskiego można wyróżnić około 120000 jednostek znaczeniowych, to uwzględniając średnią długość wyrazu polskiego, współczynnik ilości spacji w tekście, reprezentatywną ilość wystąpień danego wyrażenia w korpusie otrzymuje się w uproszczeniu regułę, iż statystycznie na jedno hasło słownikowe przypada minimum 1000 znaków tekstu<sup>7</sup>.

Zebrany materiał językowy powinien odwzorowywać interesujący obszar językowy w odpowiednich proporcjach, przy czym zebrane dane powinny stanowić przykłady jak najbardziej naturalnej komunikacji językowej między ludźmi w różnych formach, z zachowaniem właściwych reguł kompozycji korpusu. Poniżej zamieszczono zestawienie wyników analizy statystycznej korpusu tekstów z zakresu szeroko pojętej retoryki i czterech korpusów porównawczych<sup>8</sup>.

<sup>7</sup> Por. dokładne opracowanie współczynników w: Karpiński 2014a, s. 59.

<sup>8</sup> Korpus „Retoryka” został zebrany na bazie dostępnych wybranych tematycznie artykułów z czasopisma „Forum Artis Rhetoricae” i materiałów tematycznych dostępnych w formie elektronicznej na stronach internetowych wybitnych postaci tej dyscypliny i poświęconych retoryce; korpus liczy 332356 znaków. Korpus „Filozofia” zawiera podręcznik W. Tatarkiewicza „Filozofia” i liczy 3387761 znaków. Korpus „Ogólny” zawiera teksty przygotowywane przez studentów do analizy frekwencyjnej na zajęciach z „Terminologii i języków specjalistycznych” i można go określić jako próbę złożenia reprezentatywnego korpusu dla celów dydaktycznych; zawiera teksty literackie, potoczne, zapis dialogów filmowych, publicystykę ogólną i naukową; objętość korpusu – 717114 znaków. Korpus „Bankowość” zawiera „Ustawę z dnia 29 sierpnia 1997 r. Prawo bankowe. (tekst jednolity)” i zawiera 432828 znaków. Korpus oznaczony jako „Biblia” zawiera tekst całej Księgi Rodzaju i liczy 185728 znaków. Przedstawione w artykule dane oparto o „szczegółowy raport wyników”, generowany przez autorski program do analizy tekstów – PANTEXT wer. 2.0.

**Tabela 1. Podstawowe wskaźniki i parametry statystyczne oceny tekstu:**

wskaźnik:	Retoryka	Filozofia	Ogólny	Bankowość	Biblia
znaków/wyraz	6,653	6,202	6,116	6,624	5,19
średnia częstość wyrazu	7,371	20,23	6,063	28,34	8,022
BLT Guirauda	19,830	17,599	27,45	5,859	15,075
BLT Mistrika	2,713	0,988	3,299	0,705	2,493
IOS Guirauda	0,486	–	0,491	0,281	0,459
IOS Mistrika	1,318	–	1,619	0,198	1,144
SS Mistrika	13,389	–	10,946	39,04	13,983
IF10%	42,81%	51,74%	72,86%	51,75%	40,22%

BLT – bogactwo leksykalne tekstu, IOS – indeks oryginalności słownictwa,

SS – stereotypowość słownictwa<sup>9</sup>, IF10% – pierwsze 10% wyrazów na liście frekwencyjnej.

Powyższa tabela wskazuje, iż korpus retoryki operuje stosunkowo najdłuższymi formami wyrazowymi (występują długie formy wyrazowe, np. *epistemologiczny*, *transcendentalnie*, *antyarystokratyczny*, *agonistyczny*, *nikomachejski*, *pragmatyngwistyczny*) i razem z tekstami korpusu ogólnego i Księgą Rodzaju jest tworzony przez wypowiedzi stosunkowo bogate językowo, zbyt często nie powtarzające się (na co wskazuje indeks stereotypowości słownictwa). Korpus „ogólny” uzyskał wysoki wskaźnik w zakresie bogactwa leksykalnego ze względu na różnorodność tematyczną, motywowaną chęcią stworzenia korpusu reprezentatywnego jakościowo. Wysoki wskaźnik częstości wyrazu dla korpusu „Filozofia” i „Bankowość” wynika z monotematyczności tych tekstów, dodatkowo wskaźnik IF10% podkreśla ich specjalistyczność. Z kolei najwyższy indeks stereotypowości słownictwa i najniższy – oryginalności słownictwa w korpusie bankowym świadczy o pewnej schematyczności danego typu tekstu prawnego i dążeniem do precyzyjnego i jednoznacznego opisanie danego zagadnienia. W formie uzupełnienia można dodać, iż IOS Mistrika zbliżony do 1,2 predestynuje tekst do kategorii publicystycznej, a wartości zbliżone do 1,8 – do kategorii artykułów naukowych. Wartości wyższe niż 1,8 świadczą o coraz większej specjalizacji i naukowej stylistyce tekstu.

Pierwszy wskaźnik, dotyczący średniej długości wyrazu należy odnosić do (jak się wydaje na obecnym etapie badań) wartości granicznej wynoszącej w przybliżeniu 6,2 wyrazu. Wartości niższe odpowiadają zazwyczaj tekstom ogólnym, natomiast wskaźnik ten rośnie wraz ze specjalizacją tekstu. Odpowiada to zasadzie, że im termin jest szczegółowszy i dokładniejszy, tym dłuższą postać może przyjmować. Nie bez zna-

<sup>9</sup> Opis indeksów BLT, IOS, SS por. np. Kamińska-Szmaj, 1989.



czenia jest tu także wpływ modeli słowotwórczych, uwzględniających prefiksy, sufiksy i wielokomponentowość.

Jeśli chodzi o rekordowy w tym zestawieniu wskaźnik IF10% dla korpusu ogólnego należy zestawić go z listą frekwencyjną w celu porównania jakościowego składu zbioru najczęściej występujących form. Zestawienie ilościowe gramatycznych form samodzielnych na listach pokazuje poniższa tabela:

**Tabela 2. Formy samodzielne w korpusach tekstowych.**

wskaźnik:	Retoryka	Filozofia	Ogólny	Bankowość	Biblia
ilość form wyrazów samodzielnych w pierwszej „50”	20	22	11	26	21
ilość form wyrazów samodzielnych w pierwszej „100”	64	58	53	75	42
IF10%	42,81%	51,74%	72,86%	51,75%	40,22%

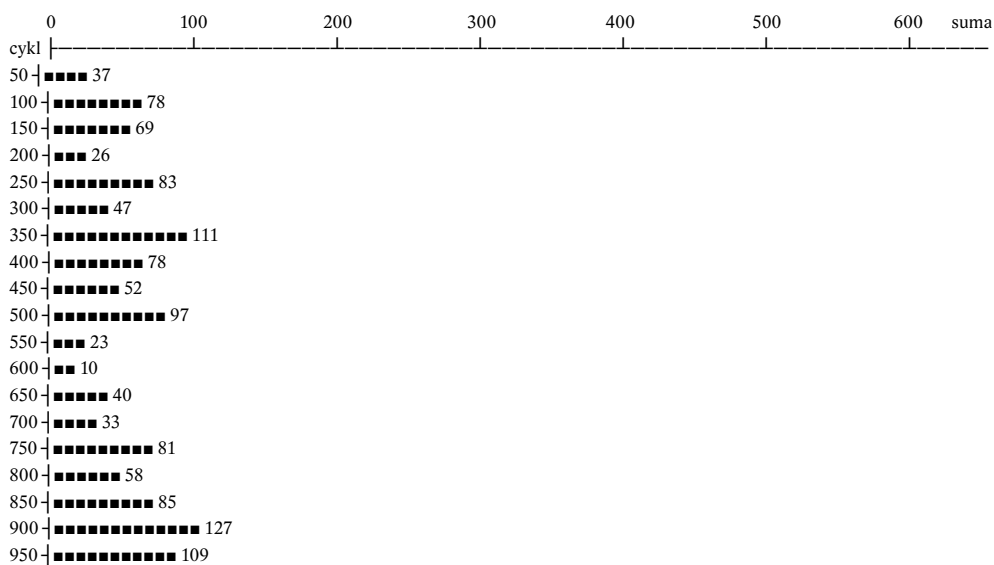
W przypadku korpusu ogólnego widać, iż bogactwo i różnorodność językowa jest osiągnięta głównie poprzez użycie form niesamodzielnych, głównie spójników i przyimków, okupujących pierwsze kilkadziesiąt miejsc na liście, natomiast w drugiej pięćdziesiątce wyrazów widać sukcesywne pojawianie się samodzielnych form znaczeniowych. Jest to charakterystyczne zjawisko dla tekstów języka ogólnego. Formę pośrednią pomiędzy korpusem ogólnym a korpusami specjalistycznymi wykazuje fragment „Księgi Rodzaju”. Jak wykazał przegląd listy frekwencyjnej, tekst ten równomiernie przeplata formy samodzielne i niesamodzielne. Z kolei korpus tekstów retorycznych wpisuje się w grupę reprezentującą teksty specjalistyczne. Ilość form samodzielnych zarówno w pierwszej setce listy frekwencyjnej jest porównywalna, przy czym najbardziej nasycony treścią znaczeniową okazuje się być korpus bankowy – w istocie jest to treść ustawy i pokazane w niniejszym artykule wskaźniki świadczą o właściwej budowie tego tekstu.

## Różne rejestry językowe / modalność tekstu

Badając zarówno korpusy językowe jak i strony internetowe można napotkać różne odmiany stylistyczne języka. Korpus umożliwia jedynie odtworzenie kontekstów, natomiast zasoby internetowe, a w szczególności archiwa grup dyskusyjnych stanowią obszerny

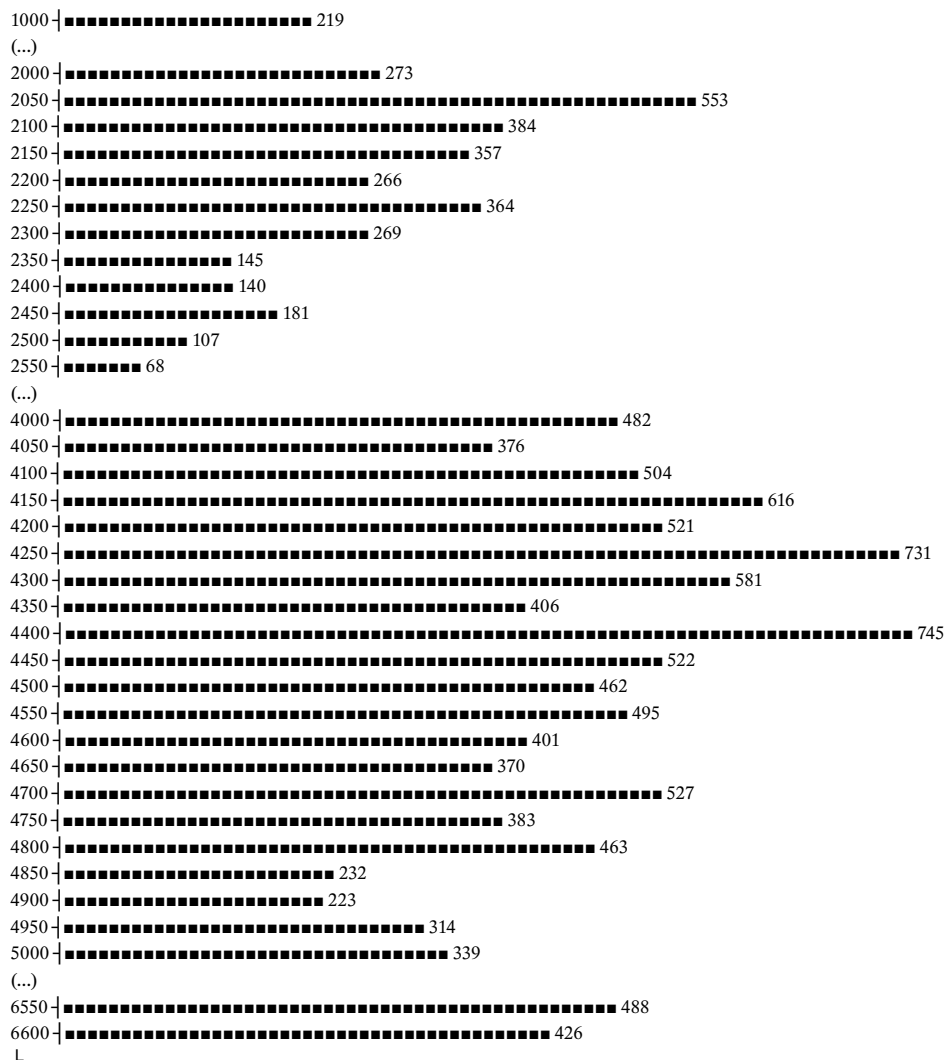
zbiór różnorodnych i specyficznych informacji (opinie, wnioski, spostrzeżenia, doświadczenia różnych osób – użytkowników Internetu). Przeżywające obecnie rozkwit blogi internetowe również zawierają interesujące zbiory tekstów, pisanych zarówno przez anonimowe osoby jak i ludzi z pierwszych stron gazet. Jest to cenne źródło języka „mówionego”. W takim przypadku znane są parametry stylistyczne i kulturowe autorów, co może pomóc przy identyfikacji i analizie kontekstów znaczeniowych badanych słów (wyrażeń).

W przypadku maszynowej analizy korpusowej autorska ocena modalności tekstu oparta jest o zdefiniowane pojęcia kwalifikacyjne, pogrupowane wg arystotelesowskiej zasady percepcji świata na klasy (akcydensy), a parametry podzielone są na te, które tradycyjnie odnoszą się do czasu, miejsca, ilości, jakości, położenia, posiadania, relacji, działania i doznania. Aby uniknąć zarzutów stawianych klasycznemu podejściu, tj. możliwości dużych odstępstw od reguł czy sztucznej modyfikacji danych empirycznych w celu dostosowania ich do kategorii (Brożek 2007: 6) wprowadzone zostało wewnętrzne zróżnicowanie parametrów wg ich stopnia natężenia<sup>10</sup>. Poniżej przedstawiono wybrane fragmenty wykresu obrazującego sumaryczne występowanie ww. cech w korpusie retorycznym<sup>11</sup>.



<sup>10</sup> Więcej na temat opracowanego systemu maszynowej parametryzacji tekstu patrz w Karpiński Ł., 2012.

<sup>11</sup> Wykres pokazuje sumaryczne występowanie tekstu w kolejno, cyklicznie pobieranych przez program fragmentach tekstu. 50 cykli nie oznacza automatycznie 50 kolejnych zdań, gdyż uwzględniane są także inne znaczniki edytorskie oraz znaki interpunkcyjne.



Łącząc powyższy wykres z wykazem, jaki fragment tekstu występuje w danym cyklu można wyodrębnić teksty o różnej funkcji i stylistyce. W przedziale 50–1000 cykli przeważają teksty dotyczące genezy i funkcji retoryki, w przedziale 2000–2200 występuje tekst szkoleniowy, adresowany do konkretnych słuchaczy i dotyczący metod prowadzenia sporów. Przedział cykli 4000–4500 odnosi się do artykułu naukowego pisanego dodatkowo w formie obszernych zdań złożonych, zawierającego interpretację i ocenę działalności retoryków starożytnych. Obydwa te czynniki znacząco podniosły wskaźnik nacechowania tekstu. Korpus retoryczny odnotowuje 33341 parametrów w 6600 cyklach pobrania

tekstu, co daje średnio 5,05 jednostki na cykl. Dla porównania korpus ogólny uzyskał średnio 7,56 jednostki, a będący na przeciwnym biegunie tekst ustawy o bankowości – 3,72 jednostki. Na podstawie tego wskaźnika można wnioskować o charakterze i funkcji danego tekstu oraz szacować stylistyczną ascetyczność lub bogactwo stylistyczno-funkcjonalne.

## Frekwencja wyrażień

Obliczenie częstotliwości wyrazów występujących w określonym zbiorze tekstów stanowi jedno z najczęściej stosowanych narzędzi lingwistyki korpusowej. Listy frekwencyjne, czyli spisy wyrazów występujących w zbiorze tekstu wraz z przypisaną do każdego z nich częstotliwością występowania, stają się nierzadko podstawą tworzenia słowników. Jest to obecnie temat coraz szerzej znany, w związku z czym poniżej przedstawiono bardziej szczegółowe zestawienie, zawierające frekwencję występowania wyrazów w tekstach z podziałem na czasowniki, przymiotniki i rzeczowniki. W tabeli zawarto do dziesięciu wyrazów w każdej kategorii z podaniem miejsca na liście (do 100-nego):

**Tabela 3. Podstawowe części mowy w korpusach tekstowych.**

	<b>Czasowniki</b>	<b>przymiotniki</b>	<b>rzeczowniki</b>
<i>Retoryka</i>	myśleć [5]* być [8] należeć [27] mówić [37] poznać [54]* pytać [81]* stosować [95]*	retoryczny [4] estetyczny [11] filozoficzny [23] określony [31] stylistyczny [40] polityczny [44] pierwszy [45] sprawny [48] piękny [52] ludzki [57]	retoryka [2] prawda [6] teoria [9] argumentacja [10] sztuka [12] mowa [14] przeciwnik [16] słowo [17] wypowiedź [18] nauka [21]
<i>Filozofia</i>	być [1] mieć [15] wiedzieć [28] myśleć [34] móc [44]* istnieć [56] należeć [80]	pierwszy [18] bogaty [21] systematyczny [26] starożytny [30] nowy [46] chrześcijański [59] średniowieczny [62] istotny [65] główny [67] ogólny [71]	filozofia [2] świat [6] nauka [7] rzecz [10] poznanie [11] teoria [12] rozumowanie [13] stanowisko [14] wiek [16] dusza [17]

<i>Korpus ogólny tekstów prasowych</i>	być [10] mieć [32] móc [19]* działać [46] mówić [67] chcieć [84]	wszystkie [26] pierwszy [53] nietypowy [66] czarny [73] polski [82] duży [93] słoneczny [96]	stan [43] praca [51] rok [54] piwo [60] ludzie [61] znaczenie [62] życie [63] miejsce [69] lat [70]* całość [75]
<i>Bankowość</i>	móc [28]* określać [66] dotyczyć [84]	informatyczny [30] terminowy [36] obowiązujący [38] nadzorczy [47] polski [52] uchylony [61] krajowy [62] zagraniczny [74] bankowy [82] członkowski [98]	art. [1] ustawa [2] bank [3] ust. [4] pkt [5] dzień [6]* Dz.U. [9] nadzór [10] komisja [16] podmiot [21]
<i>Biblia (Księga Rodzaju)</i>	być [2] rzec [3] dać [23] powiedzieć [26] panować [37]* odpowiedzieć [39] mówić [46] nazwać [50] przekonać [60] mieszkać [73]*	wszystkie [9] większy [11] cały [41] domowy [57]* wszelkie [85] polny [95]	Jakub [4] Bóg [5] ziemia [6] Józef [10] Abraham [12] ojciec [13] żona [14] pan [15] syn [17] kraj [19]

\* – oznacza, iż na wynik mogło wpłynąć nałożenie się pochodnych form gramatycznych

W tabeli z korpusu „Retoryka” zwraca uwagę stosunkowo duża liczba rzeczowników na szczycie tabeli przy stosunkowo małej liczbie czasowników – zaledwie 7 na 100 miejsc. Zwraca uwagę duża liczba form rzeczownikowych, również w dalszej części listy, co wskazuje na bogactwo językowe, duży stopień terminologizacji dziedziny i jej ukierunkowanie na funkcje językowe charakterystyczne zarówno dla terminów teoretycznych jak i empirycznych (funkcja denotacyjna, dydaktyczna, dyferencjalna, informacyjna, instrumentalna, komunikatywna, kumulacyjna i nominatywna, por. Lukszyn, Zmarzer 2006, 27–28) oraz odniesienie do wybranych funkcji przynależnym terminom teoretycznym (funkcja diagnostyczna, ideologiczna, kognitywna i prognostyczna, por. *ibid.*, 28–29), co jest zgodne z ogólną definicją tej dziedziny.

Na liście korpusu „Filozofia” zwraca uwagę podobna struktura semantyczna jak w korpusie retorycznym, w tym ukierunkowanie dydaktyczne wyrażone występującymi czasownikami. Fakt, iż źródłem korpusu jest przekrojowy podręcznik do filozofii, znalazł swoje odzwierciedlenie w występujących na liście przymiotnikach.

„Korpus ogólny” odróżnia się od powyższych o wiele szerszym zakresem tematycznym podanych form zarówno samodzielnych jak i niesamodzielnych, o czym była mowa wcześniej. Korpus „Bankowość” stanowi przykład korpusu specjalistycznego, charakteryzującym się występowaniem dużej liczby usystematyzowanych terminów, których znaczenie zawężane jest zazwyczaj poprzez dodanie przymiotnika do głównego rzeczownika. Zwraca również uwagę występowanie akronimów, charakterystycznych dla języka prawnego. Charakter tekstu, jakim jest ustawa, powoduje, iż na pierwszy plan wysuwa się precyzyjne definiowanie pojęć i zjawisk, natomiast na plan dalszy schodzi określanie czynności, jakie wiążą się z daną dziedziną – w badanej części listy frekwencyjnej występują tylko trzy czasowniki. Korpus „Biblia – Księga Rodzaju” ma charakter księgi historycznej i opisuje wydarzenia w sposób chronologiczny, od powstania świata do dziejów głównych patriarchów. Występujące na liście czasowniki odzwierciedlają strukturę i przekaz tekstu, opisującego relacje Boga (Stwórca) z ludźmi. Stosunkowo nieduża liczba przymiotników wskazuje na cechy jakościowe, na które położony jest nacisk. Z kolei lista rzeczowników odzwierciedla „patriarchalny” charakter tekstu.

W uzupełnieniu poniżej podano występowanie w tekstach czasowników modalnych, wskazujące na postawę nadawcy tekstu wobec przekazywanej treści:

**Tabela 4. Czasowniki modalne w korpusach**

<i>cz. mod.:</i>	<b>Retoryka</b>	<b>Filozofia</b>	<b>Ogólny</b>	<b>Bankowość</b>	<b>Biblia</b>
<i>móc*</i>	95*	509*	182*	232*	52*
<i>Można</i>	62	207	273	6	3
<i>Wolno</i>	2	–	2	5	9
<i>wypada</i>	–	3	1	–	–
<i>musieć</i>	18	134	45	4	4
<i>trzeba</i>	18	–	15	–	1
<i>należy</i>	98	276	81	23	8
<i>woleć</i>	–	174	–	2	1
<i>pragnąć</i>	11	49	2	–	3
<i>chcieć</i>	17	154	6	–	9
$\Sigma$ cz. mod.	321	1598	602	272	90

\* program nie rozróżniał homonimów „może” – partykuły i formy czasownika.

Powyższe zestawienie wskazuje, iż najbardziej nacechowanym emocjonalnie jest „korpus ogólny”, za nim kolejno korpus retoryczny oraz filozoficzny, co jest zgodne z ogólnym postrzeganiem ich tematyki. Zwraca uwagę bardzo niska liczba form modalnych (szczególnie przy pominięciu nieobiektywnej formy *móc/może*) w dziedzinie bankowości, co świadczy o specjalizacji i neutralnym nacechowaniu tekstu ustawy. Korpus „Biblia – Księga Rodzaju” podanymi wynikami broni tezy o historycznym charakterze tekstu, wzmocnionej dodatkowo teologicznym założeniem, iż czynów Boga się nie ocenia.

Analiza samego korpusu retorycznego wskazuje również na głęboką spójność systemu językowego i jasno określa genealogię tej dziedziny. Do badania tego zjawiska użyto rzeczowniki imienne oraz formy pochodne odnoszące się do kluczowych postaci retoryki:

Pojęcie główne i ich pochodne	Miejsce pojęcia głównego na liście frek.	Ilość wystąpień	% wyrazów w korpusie
filozofia	23	94	0,308%
Arystoteles	55	67	0,220%
Platon	74	54	0,177%
scholastyk	2374	2	0,007%
sofiści	64	141	0,463%
Sokrates	423	15	0,049%
Kant	625	10	0,033%
mieszane ww.		2	0,007%
Σ ww. pojęć	–		1,263%

W ramach podsumowania charakterystyki korpusu retorycznego warto odnotować obecność terminów, cytatów i maksym łacińskich. Po zsumowaniu wartości okazało się, iż procentowy udział leksyki łacińskiej w zebranych korpusie stanowi 2,55%. W zestawieniu z pozostałymi wskaźnikami zebranego korpusu, można wnioskować, iż retoryka operuje stosunkowo bogatym słownictwem, a teksty mogą być dość dobre w odbiorze przez czytelników, gdyż bliżej im do stylistyki publicystycznej niż (popularno-)naukowej. Pewnym mankamentem wydaje się być wysoki wskaźnik średniej długości wyrazów, wynikający z bogactwa leksykalnego i, jak wynika z analizy listy frekwencyjnej, skłonności do tworzenia form przymiotnikowych z przedrostkami, które to formy mogą negatywnie wpływać na odbiór tekstu.

## Weryfikacja poprawności wyrażenia hasłowego.

Częstym problemem przy opracowywaniu hasła słownikowego jest ustalenie, które z posiadanych w bazie danych wyrażen jest prawidłowe w danym kontekście. Rola korpusów językowych czy zasobów sieci jest tutaj oczywista. Przydatnym jest również znalezienie odpowiedniego kontekstu, w jakim dane wyrażenie (fraz) występuje. W przypadku wyrażen wyekscerpowanych z retorycznego korpusu tekstowego wyniki weryfikacji przedstawiają się następująco:

wyrażenie:	il. wystąpień w korpusie [miejsce na liście frek.]	il. wystąpień w Internecie	il. wystąpień w korpusie PWN (ogr.)	il. wystąpień w korpusie PAN (pełny)
Agonistyczny	2 [2911]	38 100	–	1
Amplifikacja	19 [325]	120 000	maksymalna	1
antydoktrynalizm	1 [>2980]	(razem) 236 (osobno) 6310	–	–
antyrelatywistyczny	1 [>2980]	2 790	–	–
cyrograficzność	1 [>2980]	39	–	–
epistemologiczny	9 [703]	159 000	–	–
gnozeologiczny	2 [2937]	51 100	–	–
Hiperbole	5 [1162]	1 090 000	maksymalna	–
Hyperbole	1 [>2980]	strony angielskie	–	–
metapolityczny	1 [>2980]	4 630	–	–
najfatalniejszy	1 [>2980]	18 800	maksymalna	1
niedeprecjonowanie	1 [>2980]	(razem) 1670 (osobno) 220 000	–	1
odfilozoficznianie	1 [>2980]	2 (1977–2012)	–	–
pierwotniejszy	1 [>2980]	7820	–	1
pragmalingwistyczny	1 [>2980]	5 420	–	–
quintilianowski	3 [1950]	brak	–	–
rekryminacja	1 [>2980]	46 800	–	–
zretoryzowany	4 [1512]	892	–	–
Retoryka	451 [2]	705 000	maksymalna	120

Konfrontując wybrane wyrażenia z zasobami Internetu i zawartością polskich korpusów tekstowych nasuwa się pytanie o wzajemną zgodność korpusu tekstów retorycznych z uchodzącymi za normatywne korpusami PWN i PAN. W pewnych przypadkach



również wyniki wyszukiwania w zasobach internetowych każe zastanowić się nad płynnością i względnością granic poprawności stylistycznej wypowiedzi.

Przedstawiona w zarysie analiza leksykograficzna tekstów z dziedziny retoryki, na tle porównawczym, ma na celu przekonanie, iż lingwistyka parametryczna i maszynowy pomiar różnorodnych współczynników może pozwolić na skuteczny i coraz bardziej trafny opis tekstu<sup>12</sup>. Oprócz zastosowania ww. metod w translacji maszynowej, podstawowe ich zastosowanie powinno odnosić się do prakseologii i inżynierii terminograficznej. Osoba (zespół) tworzący dany słownik stoi zwykle przed nie lada wyzwaniem, aby końcowy efekt prac był aktualny, rzetelny i kompletny oraz uwzględniał wszystkie aktualne tendencje w języku. W haśle słownikowym występuje kilka parametrów mikrostruktury, które stanowią niejednokrotnie o „być albo nie być” danego słowa w języku. Jest to po pierwsze kwalifikator stylistyczny i specjalistyczny, definicja i przykład użycia (kolokacja). Umiejętne stosowanie tych podstawowych parametrów wraz z zachowaniem możliwie największej obiektywności w procesie weryfikacji zebranego materiału jest wyrazem wysokiego kunsztu warsztatowego, który musi następnie zmierzyć się z oczekiwaniami i wymaganiami wydawcy, a następnie odbiorcy docelowego, który w słowniku szukać będzie odpowiedzi na pojawiające się pytania związane z dowolną dziedziną działalności człowieka, w tym także tytułową retoryką, a narzędzia informatyczne mają wspomagać poszukiwanie właściwych odpowiedzi na te pytania.

## Bibliografia

- Arystoteles. 2009. *Metafizyka*. (tłum. K. Leśniak). Warszawa: PWN.
- Brożek, Anna. 2007. „O kategoriach i kategoryzacjach”. *Roczniki filozoficzne*, tom LV, nr 1–2007, Warszawa. 5–22.
- Chaciński, Bartek. 2005. *Wyczesany słownik najmłodszej polszczyzny*, Kraków: Wydawnictwo Znak.
- Głowiński, Michał. 2009. *Nowomowa i ciągi dalsze*, Kraków: Universitas.

<sup>12</sup> Dla celów porównawczych podano zestaw parametrów, uzyskanych w wyniku analizy niniejszego artykułu: średnio znaków/wyraz – 7,43; średnia częstość wyrazu – 3,31; BLT Guirauda/Mistrika 14,24/6,04; IOS Guirauda/Mistrika 0,54/0,25; SS Mistrika 6,07; IF10% – 49,45%; wyrazów samodzielnych w 50/100 listy: 35/72; średnie nacechowanie modalne 1,69; pierwszych 10 rzeczowników jest na miejscach w przedziale [5]-[30], pierwszych 10 przymiotników wyst. w przedziale [21]-[48], występuje także 7 czasowników [22]-[90].

- Kamińska-Szmaj, Irena. 1989. „Słownictwo tekstów popularnonaukowych w ujęciu statystycznym”. *Rozprawy Komisji Językowej*, t. XVI, Wrocław: Wrocławskie Towarzystwo Naukowe, 69–87.
- Karpiński, Łukasz. 2008. *Zarys leksykografii terminologicznej*, Warszawa: KJS UW.
- Karpiński, Łukasz. 2009. „Wybrane założenia komputerowej analizy tekstów i gromadzenia danych”. *Języki Specjalistyczne* 9. Wyraz – tekst – interpretacja, pod red. M.Kornackiej i M.Górnica, Warszawa: KJS UW, 9–23.
- Karpiński, Łukasz. 2012. „Analiza parametryczna tekstu a translacja maszynowa – wybrane zagadnienia”. *The Linguistic Academy Journal of Interdisciplinary Language Studies*, Warszawa: LSW, 103–114.
- Karpiński, Łukasz. 2014 a. „The Evolution of Databases from the Perspective of Linguistic Constructivism”. *Komunikacja Specjalistyczna*, t.7–2014, Warszawa: IKLA WLS UW. 51–70.
- Karpiński, Łukasz. 2014 b. „Baza danych w konstruowaniu lingwistycznym”. *Narzędzia pracy lingwistycznej*, pod red. W.Zmarzer, T.Michy, Warszawa: IR UW. 199–226.
- Lichański, Jakub. 2009. „Miejsca wspólne – perspektywa polska. Przegląd problemów i prolegomena dalszych badań”. *Forum Artis Rhetoricae*, 16–17, 6–23.
- Lukszyn, Jerzy, Zmarzer, Wanda. 2001. *Teoretyczne podstawy terminologii*, Warszawa WLSiFW.
- Mielczarski, Cyprian. 2009. „Dziedzictwo klasycznej sofistyki i kultura współczesna”. *Forum Artis Rhetoricae*, Numer 16–17, 7–26.
- Miodunka, Tadeusz. 2002. *Podstawy leksykologii i leksykografii*, Warszawa: PWN, 1989.
- Schopenhauer, Artur. *Erystyka czyli sztuka prowadzenia sporów*, Warszawa: Oficyna Wydawnicza Alma-Press.
- Wierzbicka, Anna, Wierzbicki Paweł. 1968. *Praktyczna stylistyka*, Warszawa: Wiedza Powszechna.