

WITOLD KIERAŚ\*

NATALIA ZAWADZKA-PALUEKTAU\*\*

INSTYTUT PODSTAW INFORMATYKI POLSKIEJ AKADEMII NAUK, WARSZAWA

# Słowa kluczowe w dyskursie polskiej polityki na przestrzeni stu lat

Słowa kluczowe: lingwistyka korpusowa, słowa kluczowe, dyskurs polityczny, analiza diachroniczna.

doi: <https://doi.org/10.31286/JP.00304>

## Wstęp

Słowa kluczowe i ich poszukiwanie w tekstach mają długą tradycję w badaniach filologicznych, kulturoznawczych czy społecznych. Już w latach 30. XX wieku John Rupert Firth (1935), w odniesieniu do prac Bronisława Malinowskiego, postulował badanie dystrybucji słów charakteryzujących daną społeczność, występujących w określonych kontekstach, niosących pewne skojarzenia i wartości. Wśród prekursorów analizy słów kluczowych wymienia się też Raymonda Williama (1976) i Annę Wierzbicką (np. 1997), którzy widzieli w nich odbicie danej kultury. Jednocześnie badacze ci opierali się na introspektywnym, a więc całkowicie subiektywnym ich wyborze.

Wraz z rozwojem lingwistyki korpusowej słowa kluczowe były wszakże coraz silniej kojarzone z podejściem kwantytatywnym, a metody ich wydobywania z tekstów opierały się na względnej częstotliwości słów. Zastosowanie metodologii ilościowych pozwoliło na otrzymywanie bardziej wiarygodnych wyników na podstawie obiektywnej, systematycznej i kompletnej analizy wskazującej na statystycznie najważniejsze jednostki leksykalne, które nie zawsze pokrywają się z intuicją badaczy. Definicja *kluczowości* przestała też ograniczać się do aspektu kulturowego, a zaczęła obejmować wszystkie jednostki wskazujące na najistotniejsze różnice pod względem treści i stylu między porównywanymi korpusami (Baker 2004: 347).

W istocie badacze rozgraniczają często te dwa zastosowania analizy słów kluczowych i skupiają się na słowach gramatycznych, traktowanych raczej jako wyznaczniki stylu, lub na słowach wskazujących na treść. Analiza pierwszego typu bywa wykorzystywana do identyfikowania cech gatunkowych zbioru tekstów. Na przykład Christopher Tribble (2000) w wyniku porównania korpusu powieści romansów z ogólnym korpusem języka angielskiego zaobserwował w tym pierwszym cechy charakterystyczne dla języka mówionego: słowami kluczowymi

\* wkieras@ipipan.waw.pl; ORCID: 0000-0002-8062-5881

\*\* n.zawadzka@ipipan.waw.pl; ORCID: 0000-0003-4969-2039

były bowiem m.in. zaimki osobowe i nazwy własne, podczas gdy tzw. negatywne słowa kluczowe wskazywały na mniejszą proporcję konstrukcji złożonych, niż można się spodziewać w języku pisanym.

Z drugiej strony, w analizie dyskursu nacisk kładzie się często na słowa należące do klas otwartych – rzeczowniki, przymiotniki, czasowniki – lub wręcz odrzuca się zupełnie jednostki gramatyczne. Przykładowo Denise Milizia (2010: 128) skupia się na wybranych rzeczownikach, argumentując, że wyznaczniki stylu „nie wnoszą nic zaskakującego”<sup>1</sup>, a Alison Duguid (2010) ogranicza się do analizy przymiotników wartościujących w swoim badaniu informalizacji prasy brytyjskiej. Inni jednakowoż zwracają uwagę, że słowa należące do zamkniętych klas mogą również wskazywać na cechy dyskursu (m.in. Rayson 2008: 531). Jonathan Culpeper (2009) na przykład sugeruje, że częste użycie słów takich jak: *if, be, or, yet* oraz *would* pomaga ukazać Julię z dramatu Williama Szekspira jako osobę nerwową, niespokojną.

Plennym obszarem badań z wykorzystaniem słów kluczowych jest analiza dyskursu politycznego, czyli programów wyborczych (Rayson 2008), przemów (Milizia 2010), debat parlamentarnych (Sarfo-Kantankah 2018) czy konferencji prasowych (Partington i in. 2013). Badania takie pokazują, co politycy w danym okresie uważają za ważne (ale także jakie tematy omijają), na jakie wartości się powołują, jak ukazują siebie i swoją partię, a jak – swoich politycznych oponentów i wyborców. Przykładowo Paul Rayson (2008) wnioskuje, że Liberalni Demokraci spodziewali się, że przegrają wybory w Wielkiej Brytanii w 2001 roku, ponieważ słowem kluczowym w ich programie wyborczym było *would*, wskazujące na odniesienia do sytuacji hipotetycznych, podczas gdy Partia Pracy częściej opisywała swoje plany jako definitywne za pomocą *will*. D. Milizia (2010) natomiast podkreśla, że priorytetem (przynajmniej deklarowanym) w polityce Tony’ego Blaira w latach 2005–2007 stała się ekologia, o czym miało świadczyć częste używanie słów *climate* (klimat) i *change* (zmiana).

Według naszej najlepszej wiedzy badacze polskiego dyskursu politycznego nie korzystali dotychczas z ilościowych metod identyfikacji słów kluczowych – wciąż dominuje w tym obszarze wcześniejsze podejście, oparte na intuicyjnym, ręcznym procesie selekcji. Anna Siewierska-Chmaj (2005: 100, 105) w jedynej dostępnej pracy poświęconej analizie wyłącznie exposé polskich premierów (do 2004 roku) dowodzi, że w pierwszych latach po uzyskaniu niepodległości „słowa sztandarowe o największej frekwencji to [...] *pokój i wojna*”, a także: *dobro ogółu, oszczędność, sanacja, wolność, przyjaźń, patriotyzm* oraz *kryzys, bezrobocie, komunizm, bolszewizm, anarchia i samowola*<sup>2</sup>. W PRL-u odznaczają się, jej zdaniem, *pokój i socjalizm*, ale także: *współpraca, bezpieczeństwo, przyjaźń, jedność i demokracja*, którym przeciwstawia słowa/pojęcia: *imperializm, kapitalizm, kolonializm, zimna wojna, rewizjonizm, kryzys i spekulacja* (Siewierska-Chmaj 2005: 106–112). Natomiast za charakterystyczne dla okresu po transformacji autorka uznaje przede wszystkim słowa *bezpieczeństwo, współpraca i Europa*, ale też: *wolność, dobro wspólne, demokracja, dialog, prawda, odpowiedzialność* oraz

1 W oryginale: „offer few surprises”.

2 Należy zaznaczyć, że w swojej pracy A. Siewierska-Chmaj (2005) bada słowa sztandarowe w znaczeniu, jakie przypisuje im Walery Pisarek (2002), a nie słowa kluczowe. W obszarze analizy dyskursu politycznego są to jednak pojęcia bardzo zbliżone.

*bezrobocie, inflacja, kryzys, komunizm i korupcja*. Michał Banczerz (2003: 180) stwierdza, że w okresie III RP „szczególnie celebrowanymi słowami-kluczami są: *normalność, polskość, europejskość, katolicyzm i przedsiębiorczość*”, podczas gdy Lilianna Lakomy (2008), ograniczając się do lat 2004–2005, wskazuje na słowa: *solidarność, liberalizm, układ, afery, godność, bezrobocie, demokracja i kłamstwo*. Na jeszcze inne zestawy słów powołują się badacze języka poszczególnych polityków, np. Donalda Tuska (Grzesiński 2011) czy Jarosława Kaczyńskiego (Jakubowska 2011).

Niniejszy artykuł wykorzystuje metodę automatycznej ekstrakcji słów kluczowych w celu opisanego zmian, jakim dyskurs polskiej polityki podlegał na przestrzeni ostatniego stulecia. Za materiał badawczy posłuży korpus przemówień programowych (*exposé*) polskich premierów od 1919 do 2019 roku. Są to „wypowiedzi programowe premierów, prezentujące ogólną linię polityczną rządu, zagadnienia bieżącej polityki społeczno-gospodarczej, plany i zamierzenia rządu na okres kadencji” (Siewierska-Chmaj 2005: 11–12). Ponieważ ich celem jest ogólne zarysowanie obszarów i strategii działania nowego rządu, stanowią one miarodajny wyznacznik tematów i wartości uznawanych w danym okresie za najistotniejsze dla klasy politycznej.

## 1. Korpus

Zdecydowana większość tekstów składających się na niniejszy materiał badawczy pochodzi z Korpusu Dyskursu Parlamentarnego (KDP; Ogrodniczuk 2018). Ponieważ jednak zasoby KDP wciąż nie są kompletne, zwłaszcza jeśli chodzi o dane sprzed 1989 roku, brakujące przemówienia (głównie z okresu II Rzeczypospolitej) zostały zaczerpnięte z książki A. Siewierskiej-Chmaj (2005). Korpus został podzielony na trzy części odpowiadające głównym okresom historycznym w Polsce po odzyskaniu niepodległości: II RP, PRL oraz III RP.

Do pierwszego zbioru należy 18 wystąpień wygłoszonych między rokiem 1919 i 1936. Jest to najmniejszy z trzech podkorpusów, zawiera 65,4 tys. segmentów. Niemal wszystkie wystąpienia mają różnych autorów, wyjątkami są dwukrotni premierzy Wincenty Witos i Władysław Grabski (obaj występowali w tej roli przed parlamentem w 1920 i 1923 roku). Drugi, najobszerniejszy (156,7 tys. segmentów) zbiór obejmuje 24 teksty: od przemówienia Edwarda Osóbki-Morawskiego ze stycznia 1945 (pierwszego z czterech w tym roku) aż do wstępnego *exposé* Czesława Kiszczaka z 1989 roku zamykającego okres PRL. W okresie PRL znacznie więcej jest przypadków wielokrotnego premierostwa, a więc i w zbiorze są autorzy występujący kilkakrotnie: E. Osóbka-Morawski (pięć przemówień w latach 1945–1946), Józef Cyrankiewicz (osiem przemówień, 1947–1969), Piotr Jaroszewicz (trzy wystąpienia, 1970–1976) oraz Wojciech Jaruzelski (dwa wystąpienia – w 1981 i 1982 roku). Trzeci zbiór obejmuje 20 wystąpień z lat 1989–2019, czyli trzech dekad od Tadeusza Mazowieckiego do Mateusza Morawieckiego. Łącznie podkorpus ten zawiera 130,7 tys. segmentów. W zestawieniu są wystąpienia trzech dwukrotnych premierów: Waldemara Pawlaka (1992 i 1993), D. Tuska (2007 i 2011) oraz M. Morawieckiego (2017 i 2019), a także dwa przemówienia Marka Belki z maja i czerwca 2004 roku. Cały korpus składa się zatem z 62 przemówień premierów z lat 1919–2019 o łącznej wielkości 352,9 tys.

segmentów, przy czym okres II RP obejmuje około 18,75 procent materiału, PRL – 44,7 procent, III RP zaś – 36,55 procent.

Wszystkie trzy zbiory tekstów połączono i zindeksowano jako jeden korpus w aplikacji Korpusomat (Kieraś i in. 2018)<sup>3</sup>. Następnie korpus został automatycznie oznakowany i wyeksportowany z aplikacji w postaci plików XML w celu przeprowadzenia analizy statystycznej.

## 2. Metodologia

Dzięki automatycznemu oznakowaniu tekstów za pomocą Korpusomatu wszystkie słowa tekstowe mogły zostać sprowadzone do ich form hasłowych. Dodatkowo ich zapis został znormalizowany do zapisu wyłącznie małymi literami. W takiej postaci zostały poddane następnie analizie ilościowej i tak też będą przytaczane w dalszej części artykułu oraz na wykresach.

Do wydobycia słów kluczowych użyto metody TF-IDF (*term frequency – inverse document frequency*), która dla każdego słowa w zbiorze tekstów wylicza wartość stanowiącą jedno z możliwych ujęć idei kluczowości. Metoda bierze pod uwagę dwa parametry: liczbę wystąpień słowa w całym korpusie oraz liczbę różnych dokumentów z korpusu, w których występuje ono co najmniej raz. W efekcie TF-IDF przypisuje wysokie wartości słowom występującym często w niewielkiej liczbie tekstów, niskie zaś – słowom ogólnie rzadkim i występującym we wszystkich lub prawie wszystkich tekstach.

Do analizy wybraliśmy sto najwyższej ocenionych tą metodą słów z całego zbioru, ograniczając się do słów pełnoznaczących, a zatem reprezentujących tylko wybrane klasy gramatyczne: rzeczowników, przymiotników, przysłówków i czasowników (w tym imiesłowów i odsłowników), a także (wyjątkowo) zaimków osobowych. Następnie przedstawiliśmy każde przemówienie w postaci wektora liczb będących wartościami TF-IDF dla każdego z wytypowanych słów. W ten sposób reprezentacje wszystkich tekstów stanowią wektory rozpięte w przestrzeni stuwymiarowej. Im bliżej w tej przestrzeni znajdują się dwa wektory reprezentujące odpowiednie dwa teksty, tym więcej mają wspólnych słów o wysokiej wartości TF-IDF; i odwrotnie – im odleglejsze są od siebie dwa wektory, tym mniej współdzielą słów kluczowych wytypowanych przez algorytm.

Taki sposób wielowymiarowej reprezentacji danych może się wydać mało intuicyjny, ale w rzeczywistości jest powszechnie wykorzystywany w wielu dziedzinach badawczych, w tym również w językoznawstwie kwantytatywnym i korpusowym (Gries 2022). Do interpretacji danych w takiej postaci opracowano szereg metod analizy wielowymiarowej, w tym stosowaną tutaj analizę głównych składowych (*principal component analysis – PCA*). Elementarnym zadaniem PCA jest taki obrót układu współrzędnych, aby kolejne współrzędne uszeregować ze względu na wariancję od największej do najmniejszej<sup>4</sup>. Tak uszeregowane współrzędne nazywa się składowymi głównymi. Dzięki takiemu przetworzeniu danych liczbowych przy ich interpretacji można położyć nacisk na te składowe, które obrazują największe zróżnicowanie

3 Korpus jest dostępny pod adresem: <https://korpusomat.pl/corpus/6353>.

4 Dokładniejsze i jednocześnie przystępne omówienie analizy głównych składowych znaleźć można m.in. w popularnym podręczniku metod statystycznych w językoznawstwie R. Haralda Baayena (2008).

w całym zbiorze. Skupienie się na pierwszych dwóch lub trzech składowych pozwala też łatwo zwizualizować wyniki na wykresach dwu- i trójwymiarowych. Podobnie postąpili np. Maciej Eder i Rafał Górski (2019), badając tą samą metodą typologię tekstów w NKJP.

Jednym z podstawowych sposobów na sprawdzenie, czy z wielowymiarowego zbioru danych takiego jak macierz TF-IDF dla zbioru tekstów *exposé* wyłania się jakaś struktura, jest hierarchiczna analiza skupień, czyli metoda polegająca na łączeniu wektorów reprezentujących poszczególne teksty w zbiory oparte na ich relatywnie bliskim położeniu w przestrzeni wielowymiarowej. Analiza typuje w ten sposób najbardziej charakterystyczne dla korpusu skupienia i pozwala ocenić, czy da się im przypisać jakąś interpretację. Jest to pierwszy, najodleglejszy i najbardziej ogólny, widok tego zbioru danych.

W drugim etapie zidentyfikowane jednostki kluczowe zostały podzielone na kategorie tematyczne. Przy grupowaniu słów kierowaliśmy się funkcjami, jakie pełnią w analizowanym dyskursie (nie zaś definicjami słownikowymi), które weryfikowaliśmy za pomocą narzędzi konkordancji i kolokacyjnych profili słów w Korpusomacie. Taka agregacja słów minimalizuje wpływ pojedynczych jednostek i jest bardziej miarodajna. Jak stwierdza Alan Partington (2010: 87):

obecność na liście pojedynczego słowa może być kwestią przypadku lub charakterystycznego stylu danego autora. Dopiero zbiory podobnych lub związanych ze sobą słów mogą stanowić prawdziwy dowód na zmiany w praktyce językowej czy dyskursywnej<sup>5</sup>.

Następnie porównaliśmy częstości względne (CW) – w przeliczeniu na 10 tys. słów – jednostek kluczowych i całych grup tematycznych w odpowiednich podkorpusach (normalizacja umożliwia porównanie zbiorów o różnych rozmiarach) oraz przeanalizowaliśmy ich funkcje dyskursywne.

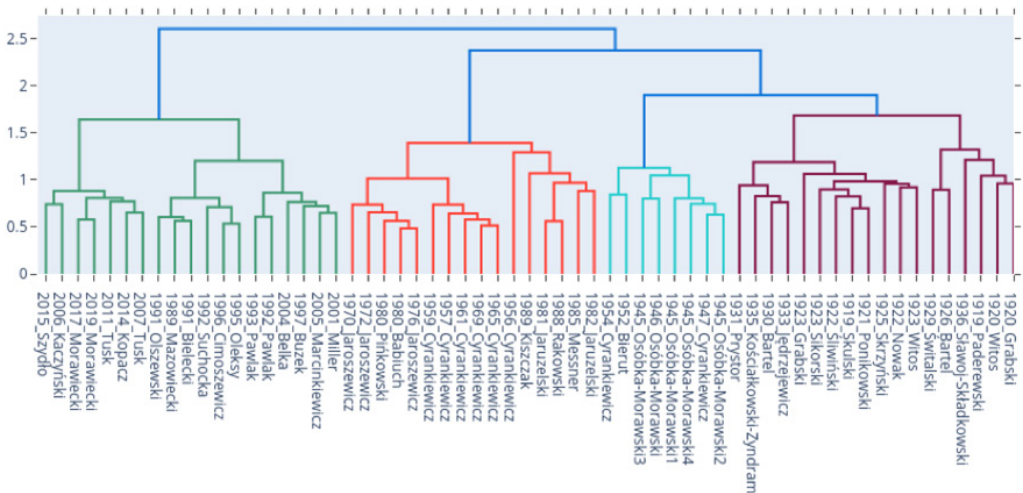
### 3. Wyniki analizy

Na wykresie 1 prezentujemy wyniki analizy skupień w postaci dendrogramu, czyli wykresu przypominającego drzewo, którego liśćmi są teksty, a wierzchołki pośrednie reprezentują stopień podobieństwa (bliskości w przestrzeni wielowymiarowej) między nimi<sup>6</sup>. Widać na nim cztery charakterystyczne skupienia. Najbardziej odrębne z nich stanowią przemówienia z okresu III RP. Wyraźnie odróżniają się też od siebie *exposé* z II RP i PRL, a wśród tych ostatnich dodatkowo wydzielają się: okres od 1945 do 1954 roku (który można powiązać z okresem

5 W oryginale: „The presence in the list of an individual item may simply be a question of happenstance or the quirk of an individual writer's style. It is when sets of similar or related items are uncovered that real evidence of changes in linguistic or discourse practice may be found”.

6 Dendrogram hierarchicznej analizy skupień jest też dostępny wśród materiałów dodatkowych. Znajdują się tam również interaktywny wykres stu słów kluczowych i tekstów *exposé* rozpiętych w przestrzeni trzech najważniejszych składowych PCA (jako materiał pomocniczy) oraz tabela częstości (absolutnych i względnych) wszystkich zidentyfikowanych słów kluczowych w poszczególnych podkorpusach.

stalinizmu; dominują w nim przemówienia jednego autora – E. Osóbki-Morawskiego) i pozostały okres od 1956 do 1989. Warto przy tym zauważyć, że okres III RP dzieli się wyraźnie na dwie części: od 1989 do 2005 oraz od 2006 do 2019 roku – podział ten można wiązać zarówno ze wstąpieniem Polski do Unii Europejskiej, jak i z dojściem do władzy ugrupowania pod przywództwem J. Kaczyńskiego i faktycznym początkiem rządów dwóch silnie spolaryzowanych formacji postsolidarnościowych: Prawa i Sprawiedliwości oraz Platformy Obywatelskiej. Analiza pokazuje zatem, że polski dyskurs polityczny podlegał na przestrzeni czasu przemianom, które można powiązać z czynnikami historycznymi: transformacjami systemowymi i innymi wydarzeniami decydującymi o kształcie życia politycznego, gospodarczego i społecznego w Polsce. Potwierdza ona również zasadność zastosowanego podziału korpusu według okresów historycznych, wskazując na ich odrębne cechy leksykalne, i stanowi przyczynek do dalszej analizy różnic między nimi.



Wykres 1. Dendrogram hierarchicznej analizy skupień macierzy TF-IDF na tekstach expose

W wyniku analizy zidentyfikowanych słów kluczowych wyodrębniliśmy czternaście kategorii tematycznych. Ze względu na ograniczoną objętość artykułu poniżej omówione zostaną tylko kategorie zawierające co najmniej pięć jednostek, wskazujące na najważniejsze trendy w analizowanym dyskursie (tab.)<sup>7</sup>. Jednocześnie zweryfikowaliśmy też, że wszystkie słowa zajmujące pierwszych dwadzieścia pozycji pod względem frekwencji w każdym z podkorpusów należą do jednej z najliczniejszych kategorii (co oznacza, że nie „odrzućliśmy” słów o wysokiej frekwencji, które nie tworzyły grup powyżej pięciu jednostek).

7 W poniższej analizie nie zostały uwzględnione m.in. dwa słowa odnoszące się do kolejności (*pierwszy, ostatni*), dwa przymiotniki opisujące odległości (*bliski, daleki*) oraz słowa odnoszące się do czasu (*czas, okres*).



Tabela. Częstości względne grup słów kluczowych

KATEGORIE SŁÓW KLUCZOWYCH		SUMA CZĘSTOŚCI WZGLĘDNYCH		
		II RP	PRL	III RP
Strategia rozwoju		527	650	683
	odniesienia do działań: <i>chcieć, tworzyć, podjąć, prowadzić, zostać, decyzja, działanie, działalność, plan, program, reforma, projekt, realizacja, droga, walka, wysiłek</i>	195	254	294
	cel działań: <i>zwiększyć, zapewnić, poprawa, rozwój, interes, wzrost, zmiana, cel, kierunek</i>	99	186	161
	ogólne rzeczowniki: <i>dziedzina, rzecz, stan, strona, sytuacja, sposób, warunek, warunki, zakres, zasada, stosunek, środek</i>	227	191	166
	przeszkody w działaniu: <i>problem</i>	6	19	17
Organizacja państwowo-polityczna		205	298	277
	odniesienia do konkretnych elementów składowych: <i>izba, organizacja, administracja, stanowisko, minister, ustawa, związek, rada</i>	105	169	132
	ogólne: <i>państwowy, polityka, polityczny, publiczny, system, socjalistyczny</i>	100	129	145
Gospodarka		119	218	137
	<i>gospodarczy, gospodarka, produkcja, przemysł, rynek, miejsce, rolny, rolnictwo</i>			
Naród/społeczeństwo		114	201	188
	<i>człowiek, naród, narodowy, obywatel, polak, społeczeństwo, społeczny</i>			
Świat zewnętrzny		55	79	112
	<i>europa, europejski, międzynarodowy, stosunki, zagraniczny, świat</i>			
Wartości		65	98	68
	<i>bezpieczeństwo, pokój, pomoc, siła, współpraca</i>			

Uwaga: podane wartości frekwencji względnych zostały zaokrąglone do jedności.

Najliczniejszą wyróżnioną grupę (pod względem typów i okazów) stanowią słowa opisujące, zgodnie z definicją exposé A. Siewierskiej-Chmaj (2005: 11–12), „plany i zamierzenia rządu na okres kadencji”, ujęte z reguły w terminach pozytywnych. Premierzy deklarują, że *chcą*<sup>8</sup>: *tworzyć, podejmować i prowadzić różne decyzje, działania, działalności, plany, programy, projekty, reformy i wysiłki. Celem czy też kierunkiem owych działań jest poprawa, rozwój, interes (Polski), wzrost i zmiana*. Procesy te określane są metaforycznie jako *droga*, co sugeruje stopniowe dążenie do obranego celu (tę samą metaforę konceptualną realizują też słowa *cel* i *kierunek*), a także *walka*, co z kolei zwraca uwagę na napotykaną *problemy*, a jednocześnie heroizuje grupę rządzącą, spełniając ważną funkcję w procesie pozytywnej autoprezentacji. Potencjalne trudności natomiast są wyraźnie marginalizowane: wskazuje na nie tylko jedno słowo kluczowe (*problem*) o stosunkowo niskiej częstotliwości (II RP: CW = 6; PRL: CW = 19; III RP: CW = 17). Wreszcie do grupy tej włączone zostały rzeczowniki, które są raczej ogólne i nacechowane neutralnie, ale które również w omawianych tekstach wskazują na strategię działania nowo powstałego rządu (często zresztą, jak dowodzą poniższe przykłady, występują w bliskim sąsiedztwie innych słów z tej grupy), takie jak: *dziedzina, rzecz, stan, sytuacja, zakres czy warunek/warunki*, np.:

W zakresie polityki finansowej pierwszym celem jest uporządkowanie stosunków walutowych [...] (Leopold Skulski, 1919)<sup>9</sup>.

Nie jestem fachowcem od spraw ekonomicznych, mam jednakże wolę stworzenia warunków działania dla najlepszych w kraju specjalistów w dziedzinie gospodarki (C. Kiszcak, 1989).

Najczęściej słowa z tej kategorii używane były przez premierów peerelowskich (637 razy na 10 tys. słów), ale ich użycie po 1989 roku spadło tylko nieznacznie do 572. Ten sam trend dystrybucyjny można również zaobserwować dla prawie każdej ze szczegółowych podkategorii, oprócz ogólnych rzeczowników, które występują częściej w przemówieniach z II RP i które w tym okresie zdecydowanie dominują nad słowami z pozostałych podkategorii – w istocie *dziedzina* jest w tym podkorpusie najczęściej używanym słowem odnoszącym się do planów i strategii nowego rządu (CW = 27), np.:

Oświadczam Panom, że i obecnie nie uchylę się od najtrudniejszego zadania walki z kryzysem nie tylko gospodarczym, lecz i psychicznym, i nie wątpię, że w tej dziedzinie nie odmówią mi Panowie harmonijnej współpracy (Marian Zyndram-Kościałkowski, 1935).

W pozostałych dwu podzbiorach najczęstszym słowem z tej kategorii jest *rozwój* (o częstościach względnych równych 47 dla podkorpusu PRL i 40 dla III RP), używany raczej w ogólnych

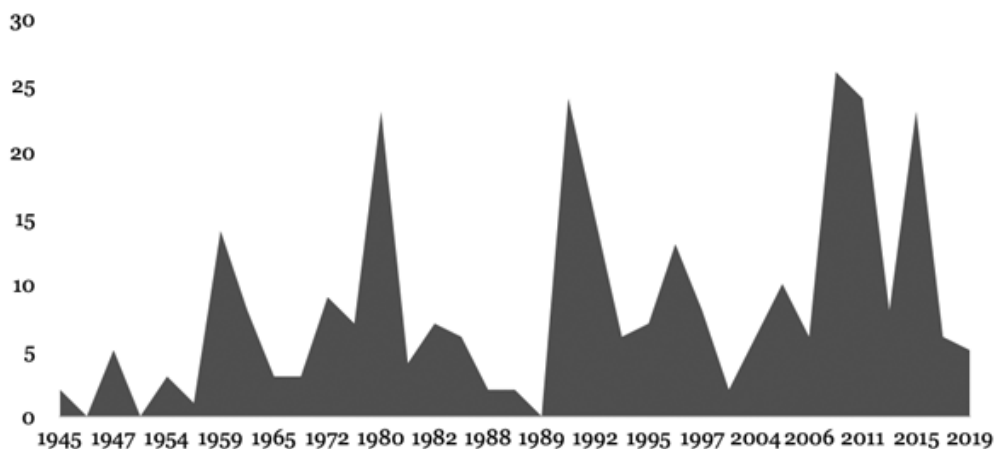
8 Czasownik *chcieć* szczególnie zyskuje na znaczeniu w III RP (aż 524 wystąpienia; CW = 66), zwłaszcza w przemówieniach premierów D. Tuska i M. Morawieckiego. Dla porównania: jego częstości względne w podkorpusach II RP i PRL wynoszą, odpowiednio, 20 i 13.

9 Wyróżnienia w cytatach pochodzą od autorów.



kontekstach (*rozwój kraju/Polski/gospodarki*). Drugim ulubionym słowem premierów peerelowskich z tej kategorii jest *plan* (*x-letni, gospodarczy, tegoroczny*). To, jak bardzo jest ono charakterystyczne dla tego okresu, widać szczególnie wtedy, gdy porówna się częstość jego występowania w PRL-u z frekwencją w dwóch pozostałych okresach: w podkorpusie PRL pojawia się ono pięć razy częściej (CW = 32). W III RP *plan* zostaje wyparty przez *program* – jest to drugie najczęstsze słowo w tym podzbiorze (CW = 31). A zaraz po nim mamy *cel* – jest to również drugie pod względem frekwencji względnej słowo kluczowe w podkorpusie II RP (CW = 26).

Uwagę przykuwa też słowo *zmiana*: potrzeba zmian, najwyraźniej marginalna w czasie II RP, zyskuje na znaczeniu po wojnie (CW = 13), ale dopiero w III RP otrzymuje rangę priorytetu (prawie dwa razy więcej wzmianek w stosunku do poprzedniego okresu). Warto jednak zaznaczyć, że nie jest to stały trend, a wzrost częstości użycia słowa *zmiana* wyraźnie koreluje z wydarzeniami historycznymi (rok 1980 i lata 1991–1992) oraz, szczególnie w czasach najnowszych, z faktycznymi zmianami obozu rządzącego. W tym drugim okresie za najbardziej wyraziste piki odpowiada nie tylko hasło wyborcze Prawa i Sprawiedliwości z roku 2015, w którym partia ta wróciła do władzy (*dobra zmiana*)<sup>10</sup>, ale też mają w tym swój udział przemówienia z okresu sprawowania władzy przez Platformę Obywatelską (zastąpiła PiS u władzy w 2007 roku). Co ciekawe, *zmiana* pojawia się z podobną częstością w obu przemówieniach ówczesnego premiera D. Tuska – z 2007 roku, zaraz po odzyskaniu władzy (26 wystąpień), i cztery lata później, w 2011 roku (24 wystąpienia). Podobnej tendencji nie można zaobserwować w przemówieniach premiera M. Morawieckiego (PiS) po 2015 roku – potrzeba zmian wyraźnie wówczas maleje (lub może uznano, że cele zmian zostały już osiągnięte).



Wykres 2. Zmiana w PRL i III RP

<sup>10</sup> Określenie *dobra zmiana* pojawia się w przemówieniu D. Tuska z 2007 roku aż cztery razy, z czego trzy razy we frazie *szansa na dobrą zmianę*. Dla porównania *dobra zmiana* w przemówieniu M. Morawieckiego (2017) pojawia się trzy razy, a w przemówieniu Beaty Szydło (2015) – raz.

Poza analizą konkordancji słowa należące do omówionej wyżej grupy raczej nie wskazują jednoznacznie na pola, które miałyby podlegać owym działaniom usprawniającym. Natomiast analiza pozostałych słów kluczowych pokazuje, że takim obszarem jest przede wszystkim gospodarka kraju: odnosi się do niej osiem słów kluczowych tworzących oddzielną, trzecią pod względem relatywnej częstości segmentów, kategorię. Najwięcej odniesień do gospodarki pojawia się w przemówieniach premierów peerelowskich (CW = 218), a słowa *gospodarczy* (CW = 48) i *gospodarka* (CW = 43) stanowią pierwsze i trzecie pod względem frekwencji słowa kluczowe w tym podkorpusie w ogóle (bez podziału na kategorie semantyczne). *Gospodarczy* jest też drugim pod względem częstości słowem kluczowym w zbiorze przemówień z II RP (CW = 46) i trzecim w III RP (CW = 41). Różne jest jednak ich zachowanie kolokacyjne – w podkorpusie II RP *gospodarczy* modyfikuje głównie słowa: *życie*, *polityka*, ale także *trudność* i *kryzys*; w okresie PRL-u nacisk położony jest na *współpracę* oraz *stosunki gospodarcze* (inne silne kolokacje to: *sytuacja*, *życie* i *plan*); natomiast premierzy III RP wolą mówić o: *wzroście*, *rozwoju* i *sukcesie gospodarczym*, wspominają też o *współpracy* (inne, bardziej ogólne, kolokacje to m.in. *działalność* i *polityka*). *Gospodarka*, czyli drugie najczęstsze słowo kluczowe w tej kategorii dla okresów powojennych (w podkorpusie III RP liczy 38 wystąpień na 10 tys. słów), ma zaledwie pięć wystąpień na 10 tys. słów w podkorpusie II RP. Co za tym idzie, nie wchodzi w silne związki kolokacyjne. W PRL-u *gospodarka* jest modyfikowana przez słowa: *narodowy*, *cały*, *nasz*, *uspołeczniony*, *komunalny* i *planowy*; natomiast w III RP przez: *rynkowy*, *polski*, *żywnościowy*, *morski*, *nowoczesny* i *narodowy* – wskazuje to na istotne różnice ideologiczne w sposobie ujmowania kwestii ekonomicznych oraz na odmienne priorytety.

Drugą najliczniejszą kategorię słów kluczowych stanowią ogólne odwołania do organizacji państwowo-politycznej (*polityka*, *polityczny*, *publiczny*, *system*, *socjalistyczny*) oraz odniesienia do konkretnych organów, organizacji i stanowisk w państwie (np.: *izba*, *minister*, *związek*, *rada*). Tu również można zauważyć istotne rozbieżności między trzema analizowanymi okresami. Najbardziej wyraźną z nich stanowi różnica w częstości użycia słowa *socjalistyczny*, które jest charakterystyczne jedynie dla PRL-u, podczas gdy jego frekwencje względne w pozostałych dwu podkorpusach są bliskie zeru. Inne słowa, które pojawiają się częściej w tym podkorpusie w porównaniu z pozostałymi dwoma (choć różnice te nie są już aż tak znaczące), to *rada* – przede wszystkim *narodowa* (82 współwystąpienia) i *robotnicza* (17 współwystąpień) – oraz *związek* (głównie *zawodowy* i *radziecki*).

W podkorpusie przemówień premierów III RP najczęstszym słowem z tej grupy (ale też ogólnie, nie uwzględniając podziału na kategorie semantyczne) jest *polityka*. Wyraźna jest tendencja wzrostowa odwołań do polityki jako takiej w analizowanym okresie: od 20 wzmianek (w przeliczeniu na 10 tys. słów) w podkorpusie II RP, przez 33 wzmianki w PRL, do 46 w zbiorze zawierającym przemówienia najnowsze. W każdym z nich słowo *polityka* modyfikuje w pierwszym rzędzie przymiotnik *zagraniczna*. Jednocześnie pozostałe modyfikatory różnią się znacznie między podkorpusami. W II RP *polityka* jest też: *pokojowa*, *wewnętrzna*, *gospodarcza*, *nasza*, *finansowa* i *polska*; w PRL-u: *nasza*, *rolna*, *gospodarcza*, *polska*, *dotychczasowa*, *podatkowa*, *społeczna* i *inwestycyjna*; natomiast w III RP: *społeczna*, *gospodarcza*, *socjalna*, *aktywna*,

*rolna, polska, historyczna, nasza i prorodzinna*. Innymi słowy, świeżo po odzyskaniu niepodległości priorytetem dla polityki jest utrzymanie pokoju, ale kwestie ekonomiczno-finansowe także wydają się istotne; w PRL-u nacisk kładzie się na identyfikację obozu rządzącego jako autorów określonej strategii politycznej; natomiast w III RP kluczowe znaczenie otrzymują działania w zakresie polityki społecznej/socjalnej/prorodzinnej.

W tym ostatnim okresie ważną kwestią staje się też tak zwana *polityka historyczna* (brak jakichkolwiek wzmianek o niej w pozostałych dwu podkorpusach), aczkolwiek, jak pokazują wyniki konkordancji tej frazy, głównie dla jednego polityka – D. Tuska (7 wzmianek w przemówieniu z 2007 roku). Ostatnia, ósma wzmianka, pochodzi z wypowiedzi J. Kaczyńskiego z 2006 roku, lecz można zauważyć, że politycy ci mają bardzo zbieżną wizję *polityki historycznej* Polski: obaj przypisują jej duże znaczenie, dostrzegając w niej zarazem źródło konfliktu:

Jest rzeczą niezwykle ważną, niezwykle istotną dla polskiej przyszłości, dla wykonania przesłania, które sformułowałem na początku, to wszystko, co zostało podjęte przez ministerstwo kultury, a co zwykle w pewnym skrócie nazywa się *polityką historyczną*. Znam głosy krytyków. Nie damy się zepchnąć z tej drogi (J. Kaczyński, 2006).

Będziemy realizować *politykę historyczną* nie po to, żeby drażnić kogokolwiek; będziemy chcieli przy pomocy polityki historycznej wzmocnić wizerunek Polski jako kraju, który w swojej historii zawsze miłował wolność i który wiedział, kiedy miał tę szansę, jak mądrze ją wykorzystać. W ostatecznym rachunku to właśnie taka *polityka historyczna* może być i będzie zasadniczym elementem oddziaływania Polski wobec naszych sąsiadów, także spoza Unii Europejskiej. [...] Chciałbym tutaj na marginesie powiedzieć jeszcze jedno zdanie właśnie o *polityce historycznej*, o tzw. *polityce historycznej* i o znaczeniu prawdy historycznej w naszych relacjach międzynarodowych, ale także tutaj, krajowych. Żaden rząd w Polsce nie powinien *polityki historycznej*, prawdy historycznej szukać, a później używać przeciwko komuś. *Polityka historyczna* ma służyć dziedzictwu i pamięci, a nie walce politycznej (D. Tusk, 2007).

Przy czym w wypowiedzi D. Tuska widoczna jest chęć zachowania pewnego dystansu wobec tego zagadnienia (tzw. *polityka historyczna, na marginesie*). Żaden z polityków nie precyzuje, kto właściwie miałby być *polityce historycznej* przeciwny. W dyskursie tego rodzaju (wyobrażonym) polemistą jest zazwyczaj oponent polityczny, a jednak J. Kaczyński i D. Tusk, którzy pozycjonują się jako liderzy dwu przeciwnych obozów, przypisują jej tak samo duże znaczenie. Być może zatem sposób prowadzenia polityki historycznej czy też interpretacja wydarzeń historycznych stanowią zarzewie konfliktu, do którego nawiązują mówcy, ale ich szczegóły nie zostają ujawnione albo dlatego, że zostały uznane za wiedzę powszechną, albo z powodów strategicznych (np. chęć pozostawienia swojej pozycji w nieodmówieniu).

Najczęstsze słowo kluczowe z tej kategorii dla podkorpusu II RP (i trzecie najbardziej kluczowe słowo w tym zbiorze w ogóle) – *państwowy* – wykazuje natomiast odwrotną tendencję

diachroniczną do słowa *polityka*. Jego częstość spada z 45 wzmianek, przez 22 wzmianki (na 10 tys. słów) w exposé peerelowskich, do zaledwie 13 w III RP. Rzeczowniki, które modyfikuje w tym pierwszym podzbiorze, to przede wszystkim: *aparatuskonieczność i życie* (powyżej dziesięciu współwystąpień); w PRL-u są to: *administracja, aparat i władza*; natomiast w III RP: *przedsiębiorstwo, aparat, administracja i służba*. Po wojnie przymiotnik ten używany jest więc ściśle w odniesieniu do mechanizmów funkcjonowania państwa (i w mniejszym stopniu do jego aktywności gospodarczej), z kolei w czasie II RP odsyła również do procesów szerszych, bardziej ogólnych (*życie, konieczność*), być może obejmujących ogół społeczeństwa, np.:

Da to trwalszą, niż dotychczasowa, podstawę naszego życia państwowego i stanie się jednym z decydujących kroków w historii umacniania się naszych sił i wartości (Janusz Jędrzejewicz, 1933).

Kolejną grupę słów kluczowych stanowią odniesienia do narodu/społeczeństwa, przy czym ponownie najczęściej takich słów występuje w przemówieniach z okresu PRL-u (CW = 201). Słowa z tej kategorii podlegają znacznemu zróżnicowaniu, jeżeli chodzi o ich frekwencje w poszczególnych podkorpusach. *Naród* jest najczęstszym słowem w tej kategorii w podkorpusach II RP (CW = 33) i PRL (CW = 54; wyraźnie to widać zwłaszcza w okresie stalinizmu), najrzadszym zaś w podkorpusie III RP (CW = 13)<sup>11</sup>. Premierzy z tego ostatniego okresu w zestawieniu ze swoimi poprzednikami zdecydowanie preferują za to słowo *Polak* (CW = 39) i, bardziej ogólnie, *człowiek* (CW = 33); szczególnie ten pierwszy rzeczownik jest w poprzednich okresach mocno niedoreprezentowany (4 i 5 użyć na 10 tys. słów w podkorpusach II RP i PRL, podczas gdy *człowiek* pojawia się w tych podzbiórach, odpowiednio, 16 i 15 razy), por.:

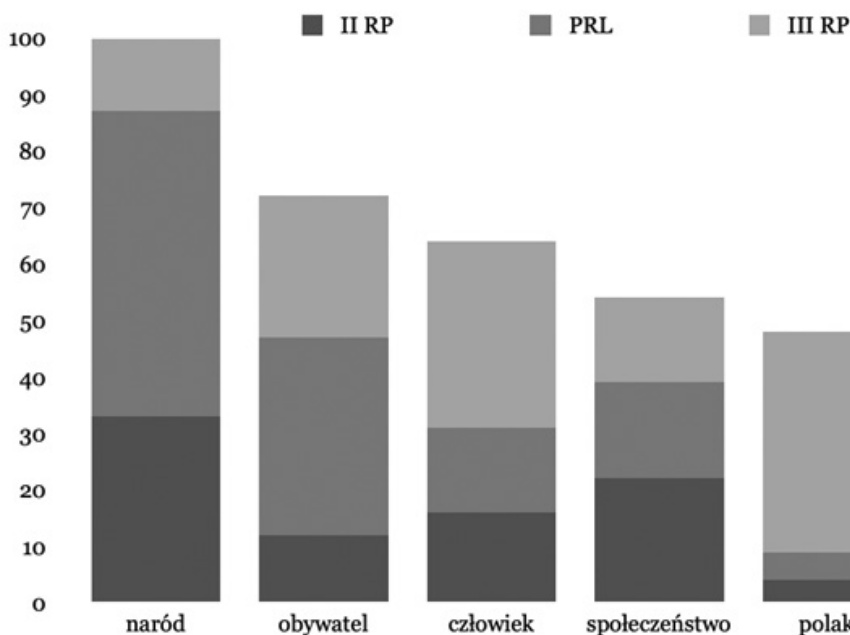
Jeśliśmy istotnie wielkim narodem, to powinniśmy również skończyć z tymi rozlicznymi przewiskami dla wszystkich sił naszych zbrojnych (Ignacy Paderewski, 1919).

Polsce potrzebne są głębokie reformy, ale muszą być to reformy robione z ludźmi i dla ludzi. Polacy są realistami i nie oczekują cudów (W. Pawlak, 1993).

Inne alternatywne terminy to *obywatel* i *społeczeństwo*: ten pierwszy jest preferowany raczej przez premierów peerelowskich (CW = 35)<sup>12</sup> w porównaniu z pozostałymi dwoma podkorpusami (II RP: CW = 12; III RP: CW = 25), ten drugi natomiast notuje niewielką tendencję spadkową w analizowanym okresie, jak można zaobserwować na wykresie 3. Przymiotniki *społeczny* i *narodowy* występują najczęściej w podzbiorze PRL (CW = 40 i CW = 35, odpowiednio) i w podzbiorze III RP (CW = 42 i CW = 21, odpowiednio), stosunkowo rzadko zaś w exposé z II RP (CW = 15 i CW = 12, odpowiednio).

11 Aczkolwiek słowo to nie zawsze opisuje Polaków.

12 Należy zaznaczyć, że rzeczownik *obywatel* jest w tym zbiorze tekstów używany również jako forma grzecznościowa w odniesieniu do konkretnych osób.



Wykres 3. Najczęściej używane terminy określające osoby/grupy osób

Kolejną kategorię stanowią słowa, które pozycjonują Polskę wobec świata zewnętrznego, takie jak: *Europa*,  *europejski*,  *międzynarodowy*,  *stosunki*,  *zagraniczny* oraz  *świat*. Występują one w sumie najczęściej w podkorpuse III RP, a najrzadziej w podkorpuse II RP. Najwyraźniej ta tendencja widoczna jest w stosunku do słów  *Europa* i  *europejski*: frekwencja przymiotnika rośnie od 2, przez 4, aż do 35 wzmianek na 10 tys. słów w trzech poszczególnych okresach, a frekwencja rzeczownika – od 7, przez 14, do 23 wzmianek. Jak się okazuje, po transformacji  *Europa* i  *europejskość* stają się ważnymi punktami odniesienia dla polskiej polityki. Pozostałe słowa z tej kategorii mają z kolei podobne frekwencje względne we wszystkich trzech podkorpusech.

Ostatnia kategoria słów kluczowych skupia słowa, które można uznać za odnośniki do wartości, są to:  *bezpieczeństwo*,  *pokój*,  *pomoc*,  *siła* i  *współpraca* (nie zaobserwowano natomiast wśród słów kluczowych terminów odnoszących się do antywartości, na które powołuje się m.in. A. Siewierska-Chmaj (2005)). Odniesienia do wartości dominują nieznacznie w podkorpuse PRL, przy czym najistotniejsza dla tego okresu wydaje się  *siła* (CW = 25), np.:

W tym powszechnym wzroście aktywności społecznej, w rosnącej zdolności mobilizacyjnej mas, w pogłębiającej się jedności narodu – tkwi niezwykła siła naszego ustroju, siła naszego Państwa Ludowego (Bolesław Bierut, 1952).

Niewiele mniej wzmianek mają: *współpraca* (CW = 22), *bezpieczeństwo* (CW = 19) oraz *pokój* (CW = 18). Potwierdza to (częściowo) wyniki analizy A. Siewierskiej-Chmaj (2005: 106), która również identyfikuje te trzy ostatnie słowa jako kluczowe dla exposé premierów peerelowskich, aczkolwiek przypisuje dużo większe znaczenie słowu *pokój*<sup>13</sup>. Większe dysproporcje w użyciu słów denotujących wartości można natomiast zaobserwować w pozostałych dwu podkorpusach. W II RP *siła* ma 27 wzmianek na 10 tys. słów, *pokój* – tylko 15, natomiast pozostałe pojawiają się mniej niż 10 razy. Tymczasem w III RP najważniejszą wartością zdaje się *bezpieczeństwo* (CW = 27; potwierdza to również analiza A. Siewierskiej-Chmaj (2005)), po nim zaś następują: *współpraca* (CW = 20), *siła* (CW = 12) i *pomoc* (CW = 11). *Pokój* odnotowano tylko dwa razy na 10 tys. słów w tym podzbiorze. Także analiza kolokacji i konkordancji słów opisujących wartości wskazuje na ich odmienne pojmowanie/konteksty użycia w poszczególnych podkorpusach. Na przykład *bezpieczeństwo* w PRL-u jest przeważnie: *zbiorowe, europejskie, trwałe, międzynarodowe, publiczne i nasze*, a w III RP jest to *bezpieczeństwo: energetyczne, osobiste, zdrowotne, zewnętrzne, wewnętrzne, narodowe i finansowe*, a ponadto mamy *bezpieczeństwo: obywatela, kontyentu, państwa, dziecka, kraju, rodziny, Polski i człowieka*.

W odniesieniu do poprzednich badań warto też wspomnieć o niewielkiej liczbie przymiotników wartościujących, które zidentyfikowano w wyniku niniejszej analizy. Podczas gdy np. M. Banczerz (2003: 177) pisze o okresie PRL-u, że:

główną cechą nowomowy stanowiła dążność do wartościowania. Znaczy to, że każdemu słowu odnoszącemu się do rzeczywistości nadawano wyrazisty znak wartości. [...] Nie można było przedstawić elementów rzeczywistości, nie oceniając ich jednocześnie<sup>14</sup>,

analiza słów kluczowych wskazała na zaledwie cztery przymiotniki, które można by uznać za wartościujące, i to tylko w określonych kontekstach: *duży, dobry, nowy i podstawowy*. Są one (nieznacznie) częstsze w podkorpusie III RP niż w podkorpusie PRL. Nie zidentyfikowano zaś ani jednego słowa kluczowego odnoszącego się do sfery religijnej, które uważane były za ważne w okresach zarówno II (Macała 2004), jak i III RP (Banczerz 2003).

## Podsumowanie

Analiza słów kluczowych pozwoliła na prześledzenie zmian językowych w polskim dyskursie politycznym na przestrzeni stu lat oraz zidentyfikowanie jego stałych cech. Ponadto wskazała na pozajęzykowe – społeczne, kulturowe, polityczne – przemiany, które znalazły odzwierciedlenie w analizowanym dyskursie. Zastosowanie metod statystycznych na obszernej, reprezentatywnej próbie badawczej przyniosło bardziej miarodajne, wiarygodne wyniki w porównaniu z poprzednimi jakościowymi analizami, opierającymi się na subiektywnym, intuicyjnym wyborze jednostek kluczowych. W istocie niniejsza analiza niejednokrotnie

13 Aczkolwiek autorka podkreśla też, że opracowana przez nią hierarchia słów kluczowych jest „bardzo ogólna” (Siewierska-Chmaj 2005: 100).

14 Inne badania z kolei wskazują na obecność cech nowomowy, w tym skłonności do wartościowania, np. w dyskursie PiS-u (Gwara 2016).



podaje w wątpliwość wyniki poprzednich badań: zdecydowana większość słów zidentyfikowanych jako kluczowe w wyniku jedynej dotychczas analizy exposé polskich premierów (Siewierska-Chmaj 2005), jak i w innych jakościowych analizach polskiego dyskursu politycznego (np.: Bancercz 2003; Lakomy 2008), nie znalazła potwierdzenia w tym badaniu. Wreszcie praca ta wskazuje na wartość analiz porównawczych i diachronicznych, które zapewniają spojrzanie na dane zjawisko z szerszej perspektywy. Można wręcz stwierdzić, że dopiero w zestawieniu z innym okresem (innym zjawiskiem, inną społecznością językową itp.) jest możliwa właściwa ocena znaczenia poczynionych obserwacji.

Niniejsze badanie ma też ograniczenia. W odniesieniu do materiału badawczego reprezentatywność korpusu zwiększyłaby się po dodaniu innych tekstów należących do dyskursu politycznego. Jeżeli zaś chodzi o decyzje metodologiczne, podział korpusu na mniejsze podzbiory pozwoliłby na wyciągnięcie precyzyjniejszych wniosków. Nie ma też w środowisku naukowym konsensusu co do najbardziej optymalnej statystycznej miary kluczowości (por. np. propozycje Stephena Jeaco (2020) czy Stefana Th. Griesa (2021)) – zastosowanie innej miary z pewnością wpłynęłoby na otrzymane wyniki.

## Bibliografia

- Baayen R.H. 2008: *Analyzing linguistic data. A practical introduction to statistics using R*, Cambridge University Press, Cambridge.
- Baker P. 2004: *Querying keywords. Questions of difference, frequency and sense in keywords analysis*, „Journal of English Linguistics”, vol. 32(4), s. 346–359.
- Bancercz M. 2003: *Język polityki i polityków. Wczoraj i dziś*, „Pisma Humanistyczne”, nr 5, s. 172–181.
- Culpeper J. 2009: *Words, parts-of-speech and semantic categories in the character-talk of Shakespeare's Romeo and Juliet*, „International Journal of Corpus Linguistics”, vol. 14(1), s. 29–59.
- Duguid A. 2010: *Newspaper discourse informalisation. A diachronic comparison from keywords*, „Corpora”, vol. 5(2), s. 109–138.
- Eder M., Górski R.L. 2019: *Typologia tekstów oparta na miarach kwantytatywnych. Studium korpusowe o zróżnicowaniu polszczyzny*, „Język Polski” XCIX, z. 3, s. 5–17.
- Firth J.R. 1935: *Technique of semantics*, „Transactions of the Philological Society”, vol. 34(1), s. 36–72.
- Gries S.Th. 2021: *A new approach to (key) keywords analysis. Using frequency, and now also dispersion*, „Research in Corpus Linguistics”, vol. 9(2), s. 1–33.
- Gries S.Th. 2022: *How to use statistics in quantitative corpus analysis*, [w:] A. O'Keeffe, M.J. McCarthy (red.), *The Routledge handbook of corpus linguistics*, Routledge, London.
- Grześniński D. 2011: *Język wypowiedzi publicznych Donalda Tuska w latach 2005–2007*, „Refleksje. Pismo naukowe studentów i doktorantów Wydziału Nauk Politycznych i Dziennikarstwa Uniwersytetu im. Adama Mickiewicza”, nr 3, s. 113–125.
- Gwara E.E. 2016: *Nowomowa – stare narzędzie do nowych celów. Jak skutecznie używać nowomowy do deprecjacji „wroga”?*, [w:] K.T. Konecki (red.), *Emocje i polityka. Sceny z życia polskiego parlamentu*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź, s. 175–209.
- Jakubowska A. 2011: *Język wypowiedzi publicznych Jarosława Kaczyńskiego w latach 2007–2010*, „Refleksje. Pismo naukowe studentów i doktorantów Wydziału Nauk Politycznych i Dziennikarstwa Uniwersytetu im. Adama Mickiewicza”, nr 3, s. 127–140.
- Jeaco S. 2020: *Key words when text forms the unit of study. Sizing up the effects of different measures*, „International Journal of Corpus Linguistics”, vol. 25(2), s. 125–154.
- Kieraś W., Kobyliński Ł., Ogrodniczuk M. 2018: *Korpusomat – a tool for creating searchable morphosyntactically tagged corpora*, „Computational Methods in Science and Technology”, vol. 24(1), s. 21–27.

- Lakomy L. 2008: *Komunikacja perswazyjna w języku polityki na przykładzie polskich kampanii prezydenckich*, praca doktorska, Uniwersytet Śląski, Katowice.
- Macala J. 2004: *Polska katolicka w myśli politycznej II RP*, Wydawnictwo Adam Marszałek, Toruń.
- Milizia D. 2010: *Keywords and phrases in political speeches*, [w:] M. Bondi, M. Scott (red.), *Keyness in texts*, John Benjamins Publishing Company, Amsterdam–Philadelphia, s. 127–146.
- Ogrodniczuk M. 2018: *Polish Parliamentary Corpus*, [w:] D. Fišer, M. Eskevich, F. de Jong (red.), *Proceedings of the LREC 2018 Workshop ParlaCLARIN. Creating and using Parliamentary Corpora*, European Language Resources Association, Paris, s. 15–19.
- Partington A. 2010: *Modern Diachronic Corpus-Assisted Discourse Studies (MD-CADS) on UK newspapers. An overview of the project*, „Corpora”, vol. 5(2), s. 83–108.
- Partington A., Duguid A., Taylor C. 2013: *Patterns and meanings in discourse. Theory and practice in Corpus-Assisted Discourse Studies (CADS)*, John Benjamins Publishing, Amsterdam–Philadelphia.
- Pisarek W. 2002: *Polskie słowa sztandarowe i ich publiczność*, Towarzystwo Autorów i Wydawców Prac Naukowych „Universitas”, Kraków.
- Rayson P. 2008: *From key words to key semantic domains*, „International Journal of Corpus Linguistics”, vol. 13(4), s. 519–549.
- Sarfo-Kantankah K.S. 2018: *It's about people. Identifying the focus of parliamentary debates through a corpus-driven approach*, „Corpora”, vol. 13(3), s. 393–430.
- Siewierska-Chmaj A. 2005: *Język polskiej polityki. Politologiczno-semantyczna analiza exposé premierów Polski w latach 1919–2004*, Wydawnictwo Wyższej Szkoły Informatyki i Zarządzania w Rzeszowie, Rzeszów.
- Tribble C. 2000: *Genres, keywords, teaching. Towards a pedagogic account of the language of project proposals*, [w:] L. Burnard, T. McEnery (red.), *Rethinking language pedagogy from a corpus perspective. Papers from the third international conference on teaching and language corpora*, Peter Lang, Frankfurt am Main, s. 75–90.
- Wierzbicka A. 1997: *Understanding cultures through their key words*, Oxford University Press, Oxford.
- Williams R. 1976: *Keywords. A vocabulary of culture and society*, Fontana Press, London.

---

## Summary

### Keywords in Polish political discourse over the last 100 years

Keywords: corpus linguistics, key words, political discourse, diachronic analysis.

The paper presents the results of a quantitative analysis of a corpus comprising Polish prime ministers' speeches given between 1919 and 2019. Its aim is to trace potential changes Polish political discourse has undergone over the last 100 years, which might also reflect wider extra-linguistic changes. To this end, it employs tools associated with corpus linguistics, specifically keywords analysis. The study identifies three main linguistically distinct periods in the recent history of Polish political discourse, but it also points to trends characterized by stability rather than change, reflecting the requirements of the genre. The use of quantitative methods puts into perspective some of the findings of previous qualitative studies by providing a broader and more nuanced diachronic view of Polish political discourse.