

Rafał Mazur 
Uniwersytet Jagielloński, Kraków
r.mazur@uj.edu.pl

WOJNA ROSYJSKO-UKRAIŃSKA W POLSKIM DYSKURSIE POLITYCZNYM NA TWITTERZE – METODY ILOŚCIOWEJ ANALIZY TREŚCI

Słowa kluczowe: analiza ilościowa, analiza treści, dyskurs polityczny, modelowanie tematyczne (LDA), duże modele językowe (LLM), przetwarzanie języka naturalnego, Twitter
Keywords: quantitative analysis, content analysis, political discourse, Topic Modeling (LDA), Large Language Models (LLM), Natural Language Processing (NLP), Twitter

Wprowadzenie

Choć pełnoskalowa zbrojna agresja Rosji na Ukrainę 24 lutego 2022 r. była dla wielu wydarzeniem niespodziewanym, to wywołała natychmiastową reakcję środowiska międzynarodowego i szybko stała się jednym z najważniejszych i najpilniej śledzonych tematów w ramach agendy politycznej. Rezolucje i deklaracje potępiające rosyjską inwazję niedługo po wybuchu wojny przyjęły m.in.: Parlament Europejski (Dz.U.UE.C.2023.229.30), Zgromadzenie Ogólne ONZ (A/ES-11/L.1), a także państwa członkowskie NATO. Wiele krajów i organizacji międzynarodowych wprowadziło ponadto wobec Rosji zakrojone na niespotykaną dotąd skalę sankcje gospodarcze oraz zadeklarowało militarne i humanitarne wsparcie dla broniącej się Ukrainy. Agresję Federacji Rosyjskiej na Ukrainę potępiły też uchwały polskiego sejmu (M.P. 2022 poz. 284) i senatu (M.P. 2022 poz. 1043). Do rosyjskiej inwazji odniósł się także prezydent Andrzej Duda, który w orędziu do narodu mówił:

Nasi sąsiedzi – Ukraińcy – zdają dzisiaj najtrudniejszy możliwy egzamin, ryzykując własnym życiem. Egzamin, którego stawką jest niepodległość ich państwa. Bronią nie tylko swojej wolności, ale wolności nas wszystkich – Europejczyków. Mogą w tym liczyć na pełne wsparcie ze strony Polski (*Orędzie*).

Wydarzenia na Ukrainie były komentowane również przez inne najważniejsze osoby w państwie, w tym premiera Mateusza Morawieckiego oraz marszałków sejm i senatu – Elżbietę Witek i Tomasza Grodzkiego. Co więcej, wszystkie największe stacje telewizyjne i portale medialne zaczęły prowadzić relacje z przebiegu działań wojennych. Można powiedzieć, że temat wojny toczącej się tuż za granicami naszego kraju niemal natychmiast zdominował polską przestrzeń medialną i stał się podstawowym punktem odniesienia zarówno dla komentatorów politycznych, jak i dla samych polityków. Konflikt ten nie tylko zrewidował dotychczasowe podejście do kwestii bezpieczeństwa i polityki zagranicznej, lecz także wywołał intensywną dyskusję na temat solidarności międzynarodowej. W obliczu tak złożonego kontekstu analiza ilościowa wypowiedzi polityków staje się istotnym narzędziem pozwalającym zrozumieć sposób, w jaki kształtowane są polityczne narracje dotyczące wojny.

Celem niniejszego artykułu jest identyfikacja głównych tematów obecnych w treści przekazów publikowanych przez największe polskie partie polityczne na portalu społecznościowym Twitter¹ przez pierwsze pół roku od rozpoczęcia pełnoskalowej rosyjskiej inwazji na Ukrainę. W tym celu wykorzystałem cztery podejścia metodologiczne: analizę frekwencyjną (pozwalającą określić, które pojęcia i terminy występowały w badanych tekstach najczęściej), analizę jednostek nazewniczych (wskaźującą na obecność w badanym dyskursie konkretnych osób, instytucji i miejsc), modelowanie tematyczne przy użyciu algorytmu LDA (umożliwiające identyfikację głównych tematów dyskursu) oraz analizę semantyczną z wykorzystaniem dużego modelu językowego (pozwalającą na wychwycenie głębszych zależności i relacji między wypowiedziami). Zakładam, że pozwoli to na wskazanie dominujących narracji wykorzystywanych przez polityków do kreowania obrazu wojny, a także na uchwycenie złożoności badanego dyskursu. W artykule postaram się zatem odpowiedzieć na pytanie, na jakie aspekty wojny politycy zwracali szczególną uwagę w swoich wypowiedziach.

Materiał badawczy

Materiał badawczy zgromadziłem przy użyciu stworzonego w języku Python autorskiego programu komputerowego wykorzystującego bibliotekę Twint², która

¹ Od lipca 2023 r. Twitter działa pod nazwą X.

² Użycie tej biblioteki pozwala na uzyskanie lepszych efektów przy pobieraniu archiwalnych tweetów w porównaniu z wykorzystaniem API Twittera. Ze względu na strukturę portalu i sposób

pozwala na pobieranie publicznie dostępnych danych z witryny Twittera. W ten sposób zgromadziłem łącznie 13 289 tweetów opublikowanych w okresie od 24 lutego do 24 sierpnia 2022 r. na oficjalnych kontach pięciu największych polskich ugrupowań politycznych: Prawa i Sprawiedliwości (pisorgpl), Platformy Obywatelskiej (Platforma_org), Lewicy (__Lewica), Polskiego Stronnictwa Ludowego (nowePSL) oraz Konfederacji (KONFEDERACJA_). Z pobranych tekstów wyselekcjonowałem 3619 tweetów, które zawierały co najmniej jeden z następujących ciągów znaków³: [Uu]krai, [Rr]osj, [Rr]osyj, [Ww]ojn, [Ww]ojen, [Uu]chodź, [Ss]ank, [Zz]elensk, [Zz]elensk⁴, [Pp]utin⁵. Na potrzeby badania założyłem, że treść tych tweetów jest związana z wydarzeniami w Ukrainie. Żeby ułatwić przeszukiwanie korpusu, skompilowałem go przy użyciu aplikacji webowej Korpusomat.pl służącej do ekstrakcji informacji o frekwencji słów, fraz i kolokacji oraz najczęstszych kontekstów wystąpienia wybranych leksemów (Kieraś, Kobyliński 2021).

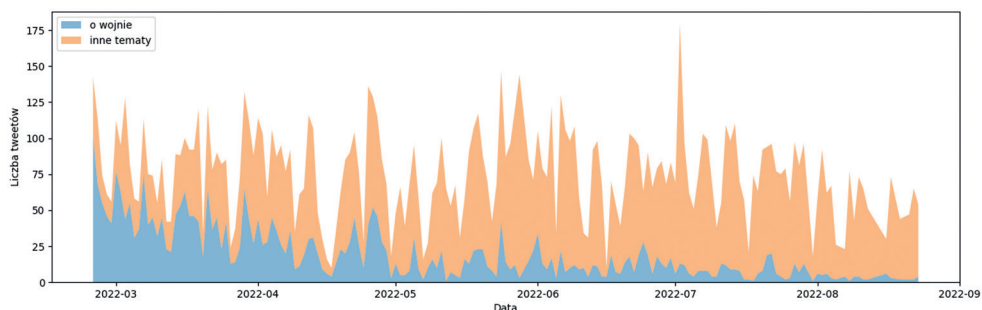
Należy zauważyć, że korpus utworzony na potrzeby badania jest wewnętrznie zróżnicowany – zawiera on komunikaty pięciu partii politycznych, które mogą mieć odmienne (przynajmniej w niektórych aspektach) poglądy na kwestie związane z wojną rosyjsko-ukraińską. Co więcej, każda z tych partii reprezentowana jest w korpusie przez różną liczbę tekstów. Jest to związane przede wszystkim z różnicami w aktywności poszczególnych ugrupowań w mediach społecznościowych. Mimo to traktuję zgromadzony korpus jako jeden zbiór danych i skupiam się raczej na przedstawieniu ogólnych tendencji w wypowiedziach polskich polityków niż na wykazaniu różnic między nimi. Najwięcej tweetów ogółem w badanym okresie opublikowało Prawo i Sprawiedliwość (4274), a najmniej Lewica (1201). Ponadto różnicę można zauważyć też w ilości uwagi, którą partie poświęcały Ukrainie. Największy udział wypowiedzi związanych z konfliktem toczącym się na terytorium tego państwa w zbiorze wszystkich tweetów danego konta w badanym okresie mają wypowiedzi zamieszczone przez PiS (38%), a najmniejsze te opublikowane na koncie Lewicy (11,57%).

Warto zaznaczyć, że temat wojny dominował w wypowiedziach polityków przede wszystkim w pierwszych tygodniach po rosyjskiej inwazji – od 24 do 28 lutego

wyświetlania treści należy założyć, że choć zebrany materiał stanowi reprezentatywną próbę, niekoniecznie jednak zawiera wszystkie opublikowane w danym czasie wiadomości. Niestety skorzystanie z tej metody nie jest obecnie możliwe ze względu na zmiany w kodzie Twittera – materiał został zgromadzony przed wprowadzeniem tych zmian, daje zatem unikalną możliwość zbadania dyskursu toczącego się wokół wojny w Ukrainie.

- 3 Celowo skorzystałem z wyszukiwania po ciągach znaków, a nie lematach (formach podstawowych), żeby przy decyzji o pobraniu wiadomości pod uwagę brane były też wyrazy pochodne.
- 4 Uwzględnienie tego ciągu znaków było potrzebne, aby w materiale znalazły się również formy, w których zastosowano transliterację nazwiska prezydenta Ukrainy w postaci *Zelenskyy*.
- 5 Wszelkie dane dotyczące frekwencji leksemów lub fraz odnoszą się do tego zawężonego tematycznie korpusu.

2022 r. poświęcono jej aż 71,08% wszystkich zebranych tweetów. W marcu tego roku wojny dotyczyło 53,92% komunikatów – a więc wciąż ponad połowa. W kolejnych miesiącach zagadnienia związane z Ukrainą stały się jednak coraz mniej obecne w tweetach polskich polityków (por. wykres 1). Można założyć, że jest to związane przede wszystkim z pojawieniem się nowych tematów w ramach agendy publicznej.



Wykres 1. Zmiana tematyki tweetów polskich partii politycznych w czasie (źródło: oprac. własne)

Analiza frekwencyjna

Analizę kreowanego przez polskich polityków obrazu wojny rosyjsko-ukraińskiej warto zacząć od sporządzenia listy frekwencyjnej przedstawiającej zasób najczęstszego słownictwa zawieranego przez badany korpus. W tym celu, skorzystawszy z biblioteki spaCy, wygenerowałem listę najczęstszych kolokacji zawierających rzeczowniki i ich określenia (rzeczownikowe lub przymiotnikowe⁶), co obrazuje tab. 1. Jak zauważa Laura Polkowska, na podstawie informacji o frekwencji występowania poszczególnych leksemów lub ich połączeń można wysnuć istotne wnioski na temat tekstów – odtworzyć najważniejsze obszary tematyczne, wskazać najczęściej wykorzystywane przez nadawców strategie komunikacyjne, a także zrekonstruować kreowaną przez nich wizję otaczającej rzeczywistości (Polkowska 2018: 32). Przy tworzeniu list frekwencyjnych pomijam wyrazy i połączenia wyrazowe o bardzo ogólnym znaczeniu, a także takie, których nadreprezentacja w korpusie wynika ze specyfiki badanego medium (są to przede wszystkim nazwy użytkowników i nazwiska polityków, których wypowiedzi są udostępniane za pośrednictwem partyjnych kont).

6 W modelu językowym spaCy liczebniki porządkowe są traktowane jako przymiotniki (przydzielane do klasy ADJ), co tłumaczy obecność w tab. 1 takich połączeń jak *pierwszy dzień*. Wyszukiwanie zawęziłem dodatkowo do zależności compound, amod, acl, nmod appos. Założyłem, że skorzystanie z zależności składniowych (wyszukiwanie par wyrazów nadrzędnych i podrzędnych) pozwoli na osiągnięcie bardziej precyzyjnych wyników od klasycznej listy bigramów.

Co ciekawe, wśród kolokacji, które znalazły się na liście rankingowej, nie dominują (ani nie zajmują najwyższych pozycji) połączenia odnoszące się do militarnego aspektu konfliktu toczącego się na terytorium Ukrainy. Znaczenia związane z działaniami zbrojnymi konotują przede wszystkim takie połączenia jak: *rosyjska agresja*, *wybuch wojny*, *początek wojny*, *wschodnia flank*a. Wydaje się, że polscy politycy więcej uwagi w swoich wypowiedziach poświęcają gospodarczym, społecznym i humanitarnym konsekwencjom wojny. Na podstawie listy można założyć, że częstym tematem badanych tweetów jest pomoc osobom uciekającym przed wojną, na co wskazują takie kolokacje jak m.in.: *pomoc uchodźcom* / *pomoc humanitarna* / *uchodźcy wojenni* / *program socjalny* / *ukraińskie dzieci*. Częste w zebranych materiale są również połączenia, które sugerują zaniepokojenie polityków wpływem konfliktu na gospodarkę, a przede wszystkim na europejski sektor energetyczny, który do czasu wojny w znacznym stopniu uzależniony był od surowców pochodzących z Rosji i Ukrainy: *rosyjski węgiel* / *ukraińskie zboże* / *rosyjski gaz* / *rosyjska ropa*. Ponadto kolokacje takie jak: *Unia Europejska* / *cała Europa* / *cały świat* / *polski rząd* / *misja pokojowa* świadczą o tym, że w badanych wypowiedziach omawiana jest też rola, jaką powinni w ramach konfliktu odgrywać partnerzy zagraniczni Ukrainy.

Lp.	Kolokacja	Frekwencja
1	europejski unia	109
2	polski granica	66
3	rząd rzecznik	62
4	uchodźca pomoc	55
5	cały świat	54
6	rosyjski węgiel	46
7	cały Europa	44
8	polski rząd	42
9	inny państwo	39
10	humanitarny pomoc	39
11	wojenny uchodźca	39
12	rosyjski agresja	38
13	ukraiński zboże	36
14	wojna wybuch	34
15	rosyjski gaz	33
16	sejm premier	33
17	sankcja pakiet	32
18	rosyjski ropa	32
19	wschodni granica	31

Lp.	Kolokacja	Frekwencja
20	minister wicepremier	31
21	energia źródło	30
22	pozarządowy organizacja	30
23	socjalny program	29
24	wojna początek	27
25	pomoc ustawa	27
26	ukraiński dziecko	27
27	ukraiński naród	27
28	pokojowy misja	27
29	polski rodzina	26
30	rosyjski surowiec	26
31	cena wzrost	25
32	praca rynek	25
33	trudny czas	25
34	recepcyjny punkt	25
35	inny kraj	24
36	węgiel import	24
37	cały czas	23
38	uchodźca program	22
39	rosyjski oligarcha	22
40	obecny sytuacja	21
41	wschodni flank	21
42	odnawialny źródło	20
43	ostatni rok	20
44	oligarcha majątek	20
45	rosyjski wojsko	19
46	paliwo cena	18
47	stan racja	18
48	cały Polska	17
49	pierwszy dzień	17
50	rząd polityka	16
51	zboże eksport	16
52	uchodźca wsparcie	16
53	polski gospodarka	16
54	energetyczny bezpieczeństwo	16

Lp.	Kolokacja	Frekwencja
55	uchodźca fala	16
56	światowy wojna	16
57	wojenny działanie	16
58	wolny Ukraina	16
59	gospodarczy podmiot	16
60	sejm wiceminister	16
61	wysoki cena	15
62	uchodźca liczba	15
63	rosyjski propaganda	15
64	migracyjny kryzys	15
65	europejski solidarność	15
66	wojna dzień	15
67	zbrojny siła	15
68	eksportowy korytarz	15
69	kolejny raz	14
70	polski państwo	14
71	premier spotkanie	14
72	wielki Brytania	14
73	polski język	14
74	sejm marszałek	14
75	sejm minister	14
76	polski rolnik	13
77	polski firma	13
78	międzynarodowy arena	13
79	polski rynek	13

Tabela 1. Najczęstsze połączenia rzeczowników i ich określeń

Analiza jednostek nazewniczych (NER)

Biblioteka spaCy pozwala również na rozpoznawanie, ekstrakcję i klasyfikację jednostek nazewniczych występujących w dużych korpusach tekstów (*Named Entity Recognition*) (por. Schmitt, Kubler, Robert, Papadakis, LeTraon 2019). Korzystając z tej funkcjonalności⁷, sporządziłem zestawienia pojawiających się w analizowanych tweetach nazw osobowych (por. tab. 2), a także nazw geograficznych i miejscowych (por. tab. 3).

Lp.	Nazwy osobowe	Frekwencja
1	@MorawieckiM	697
2	Putin	390
3	@szefernaker	207
4	Kaczyński	205
5	@KosiniakKamysz	199
6	@donaldtusk	174
7	@WTBartoszewski	107
8	@PZgorzelskiP	103
9	@mblaszczak	81
10	@michaldworczyk	75
11	@SawickiMarek	67
12	@PiotrMuller	66
13	@RobertWinnicki	59
14	@pawlakwaldemar	48
15	@Kaminski_M	42
16	@MarlenaMalag	40
17	Orban	36
18	@pawelbejda	31
19	@Paslowski	30
20	@CzarnekP	29
21	Łukaszenka	28
22	Morawiecki	26
23	@Kowalczyk_H	26

7 Ponieważ nazwy własne ze względu na swoją specyfikę są często błędnie rozpoznawane w procesie lematyzacji, wyniki analizy zostały przeze mnie dodatkowo zweryfikowane przez przeszukanie korpusu po poszczególnych formach fleksyjnych leksemów znalezionych w trakcie automatycznego rozpoznawania jednostek nazewniczych.

Lp.	Nazwy osobowe	Frekwencja
24	@WasikMaciej	25
25	@DariuszKlimczak	23
26	@KrzysztofPaszyk	21
27	@ZelenskyyUa	20
28	@PiotrGlinski	20
29	@Bozenazelazowsk	20
30	@AMAdamczyk	19
31	@RzepaJarek	19
32	@jbrudzinski	19
33	@a_niedzielski	18
34	@Hetman_K	18
35	@SasinJacek	18
36	Tusk	17
37	@JarubasAdać	17
38	@jflibicki	17
39	@paweljabIonski	16
40	Grodzki	16
41	Bandera	16
42	@JanekLopat	16
43	@KonradBerkowicz	15
44	@BeataSzydlo	15
45	@StefanKrajewski	14
46	@jciesz	14

Tabela 2. Najczęstsze nazwy osobowe

Zestawienie nazw osobowych dostarcza informacji o głównych aktorach badanego dyskursu – osobach zabierających głos w sprawie rosyjsko-ukraińskiego konfliktu, a także tych, którzy stają się tematem takich wypowiedzi. Nie dziwi obecność na tej liście najważniejszych polskich polityków, których fragmenty przemówień często są rozpowszechniane za pomocą partyjnych kont, m.in.: premiera Morawieckiego, ministrów⁸ (przede wszystkim: Mariusza Błaszczaka⁹, Michała

8 Warto zwrócić uwagę na to, że choć w tej grupie najczęściej przywoływane było nazwisko (a właściwie nazwa konta) ministra obrony narodowej, to na stosunkowo wysokich pozycjach na liście pojawiają się ministrowie niezwiązani bezpośrednio z obronnością. Wskazuje to na dużą wagę, jaką w badanych wypowiedziach miał kontekst społeczny.

9 Minister obrony narodowej od 9 stycznia 2018 r. do 13 grudnia 2023 r.

Dworczyka¹⁰, Mariusza Kamińskiego¹¹, Marleny Małąg¹², Przemysława Czarnka¹³, Piotra Glińskiego¹⁴, Adama Niedzielskiego¹⁵, Jacka Sasina¹⁶) czy liderów partii politycznych – Jarosława Kaczyńskiego (PiS), Donalda Tuska (PO), Władysława Kosiniaka-Kamysza (PSL). Nie zaskakuje także duża frekwencja nazwiska prezydenta Rosji – Władimira Putina. Co ciekawe, z danych ilościowych jasno wynika, że znacznie rzadziej przywoływana była postać prezydenta Ukrainy – Wołodymyra Zełenskiego. Można przypuścić, że w dyskursie o wojnie eksponowane są przede wszystkim potencjalne zagrożenia dla Polski, których symbolem jest przywódca Rosji. Jego zachowania, postawy i wypowiedzi są zatem istotnym punktem odniesienia dla polskich polityków. Takie założenie może tłumaczyć także, dlaczego w badanych wypowiedziach stosunkowo często wymieniane są nazwiska sprzyjających Rosji prezydenta Białorusi (Aleksandra Łukaszenki) i premiera Węgier (Viktora Orbána), rzadko zaś mówi się o politykach zachodnich, mimo ważnej roli, jaką w konflikcie odgrywają kraje NATO. Warte uwagi jest również pojawienie się na liście frekwencyjnej nazwiska Stepana Bandery – postaci, która w polskim dyskursie stanowi symbol dokonanego przez Ukraińców ludobójstwa polskiej ludności cywilnej na Wołyniu. Obecność tego onimu w korpusie wskazuje na to, że tematem badanych tweetów były też trudne karty wspólnej polsko-ukraińskiej historii.

Lp.	Nazwy geograficzne i miejscowe	Frekwencja
1	Ukraina	1993
2	Polska	992
3	Rosja	878
4	Europa	313
5	Warszawa	290
6	Kijów	107
7	Bruksela	93
8	Niemcy	91
9	USA	72

10 Minister-członek Rady Ministrów od 4 czerwca 2019 r. do 27 listopada 2023 r. i szef Kancelarii Prezesa Rady Ministrów od 19 grudnia 2017 r. do 13 października 2022 r.

11 Minister spraw wewnętrznych i administracji od 14 sierpnia 2019 r. do 27 listopada 2023 r. i koordynator służb specjalnych od 16 listopada 2015 r. do 27 listopada 2023 r.

12 Minister rodziny i polityki społecznej od 6 października 2020 r. do 27 listopada 2023 r.

13 Minister edukacji i nauki od 19 października 2020 r. do 27 listopada 2023 r.

14 Wiceprezes Rady Ministrów od 16 listopada 2015 r. do 21 czerwca 2023 r. i minister kultury i dziedzictwa narodowego od 26 października 2021 r. do 27 listopada 2023 r.

15 Minister zdrowia od 26 sierpnia 2020 r. do 10 sierpnia 2023 r.

16 Wiceprezes Rady Ministrów od 4 czerwca 2019 r. do 21 czerwca 2023 r. i minister aktywów państwowych od 15 listopada 2019 r. do 27 listopada 2023 r.

Lp.	Nazwy geograficzne i miejscowe	Frekwencja
10	Francja	40
11	Berlin	34
12	Moskwa	34
13	Afryka	33
14	Białoruś	30
15	Gruzja	27
16	Lwów	26
17	Czechy	25
18	Węgry	24
19	Słowacja	22
20	Wielka Brytania	22
21	Bucza	21
22	Przemyśl	20
23	Kraków	19
24	Wołyń	19
25	Katowice	16
26	Słowenia	15
27	Litwa	14
28	Madryt	13
29	Paryż	13
30	Płock	13
31	Strasburg	12
32	Finlandia	12
33	Hiszpania	11
34	Rumunia	11
35	Budapeszt	11
36	Toruń	11
37	Borodzianka	10
38	Mariupol	10
39	Korczowa	10

Tabela 3. Najczęstsze nazwy geograficzne i miejscowe

Analiza najczęściej pojawiających się nazw geograficznych i miejscowych pozwala wskazać regiony, które w kontekście wojny zyskały szczególną uwagę polityków. Są to nie tylko kraje bezpośrednio zaangażowane w konflikt – Rosja oraz

Ukraina – lecz także uzależniona od Rosji i pośrednio biorąca udział w prowadzeniu wojennych działań Białoruś, a także państwa wspierające Ukrainę militarnie i gospodarczo, m.in.: Polska, Niemcy, Francja, Czechy, Słowacja, Słowenia, Rumunia, Litwa, Finlandia, Wielka Brytania i USA.

Wydaje się ponadto, że pewna grupa nazw może być przywoływana przez polityków przede wszystkim w znaczeniu symbolicznym (por. Wolnicz-Pawłowska 2007: 152–154). Do tej kategorii można zaliczyć m.in. nazwę Gruzji, państwa zaatakowanego przez Rosję w 2008 r. – w ramach współczesnego polskiego dyskursu politycznego jest ona czytelnym symbolem rosyjskich ambicji imperialnych¹⁷. W funkcji symbolicznej najprawdopodobniej wykorzystywane są także takie nazwy jak Bucza¹⁸ i Borodzianka¹⁹ – symbole rosyjskiego okrucieństwa, Mariupol²⁰ – symbol ukraińskiej odwagi, a także przywoływany już Wołyń²¹. Jak zauważa Andrzej Czerny, funkcja tego rodzaju nazw nie polega wyłącznie na wskazywaniu konkretnego miejsca, ale też na wywoływaniu emocji przez wzbudzanie określonych skojarzeń i sygnalizowaniu wartości (Cieślíkowa 2001: 99–100; Czerny 2011: 84–85; por. Rejter 2016: 38). Kontekst geopolityczny zawarty w analizowanych wypowiedziach rysują też nazwy kontynentów. Na liście frekwencyjnej zanotowano dwa takie onimy: *Europa* i *Afryka*. Duża frekwencja drugiego z nich zapewne wynika z gospodarczych reperkusji wojny, której wybuch uniemożliwił transport ukraińskiego zboża do krajów afrykańskich drogą morską.

17 Jak się wydaje – szczególnie dla dyskursu prawnicowego. Symboliczne znaczenie Gruzji ugruntowała wizyta prezydenta Lecha Kaczyńskiego w Tbilisi w 2008 r., podczas której padły słowa często cytowane w nawiązaniu do konfliktu rosyjsko-ukraińskiego: „I my też świetnie wiemy, że dziś Gruzja, jutro Ukraina, pojutrze Państwa Bałtyckie, a później może i czas na mój kraj, na Polskę!” (Prezydent.pl 2008). Jak podkreśla Andrzej Czerny, „nazwy funkcjonują jako symbole w szerszym lub węższym kręgu użytkowników należących do pewnej wspólnoty kulturowej, na przykład etnicznej albo religijnej” (2011: 85–86).

18 Miejscowość, w której wojska Federacji Rosyjskiej dokonały zbrodni wojennej na ludności cywilnej po inwazji na Ukrainę w 2022 r.

19 Miejsce intensywnych bombardowań w marcu 2022 r. podczas walk o Kijów.

20 Miasto w obwodzie donieckim oblegane przez wojska rosyjskie od 24 lutego do 20 maja 2022 r.

21 Por. na przykład wypowiedzi premiera Mateusza Morawieckiego: „Po 40 dniach wojny nie mamy wątpliwości, że to, co dzieje się na Ukrainie, to czyste zło. Bucza, Irpień, Hostomel, Motyzin to nazwy miejscowości, które każdy z nas zapamięta do końca życia” (PiSorg 2022a); „Rok temu Mariupol nosił tytuł ukraińskiego miasta kultury. To miasto pozostanie symbolem ukraińskiego heroizmu, walki o bezpieczeństwo i pokój. Razem z @antonioconstapm zanosimy Ukrainie nadzieję na jej odbudowę” (PiSorg 2022b). Symboliczne znaczenie omawianych onimów można zauważyć także w tweecie opublikowanym na koncie nowePSL: „Mariupol wygląda dziś jak Warszawa w X/XI 1944 r. N. Davies porównał działania wojsk rosyjskich do SS. Działania Rosji sprzed 8 lat w Ukrainie nie spowodowały reakcji. Apeluję do polityków Europy Zachodniej, musicie się wreszcie obudzić!” (nowePSL).

Modelowanie tematyczne (LDA)

Wstępne obserwacje na temat zawartości zebranego materiału zostały zweryfikowane za pomocą analizy z zastosowaniem modelowania tematycznego przy pomocy algorytmu Latent Dirichlet Allocation (LDA)²². Metoda ta wykorzystuje model probabilistyczny do zbadania frekwencji słów pojawiających się w dużych zbiorach dokumentów, co pozwala na identyfikację i wizualizację najważniejszych tematów²³ występujących w tych dokumentach. Algorytm działa, opierając się na założeniach, że słowa, które często współwystępują w wypowiedziach, są prawdopodobnie związane z tym samym tematem, a także na tym, że wypowiedzi zawierające te same słowa również poświęcone są tej samej tematyce (por. AlAgha 2021: 40–41). Teksty przed przeprowadzeniem analizy zostały odpowiednio przygotowane. Zebrane tweety najpierw poddałem procesowi tokenizacji i lematyzacji z wykorzystaniem biblioteki spaCy, następnie najczęstsze bigramy i trigramy połączyłem znakiem podkreślenia (tak, by przez algorytm były traktowane jako całość), w dalszej kolejności ze wszystkich tekstów usunąłem słowa najczęstsze dla języka polskiego (tzw. stoplista²⁴), a także znaki i symbole, które mogłyby zaburzyć przebieg badania (nazwy użytkowników, znaki interpunkcyjne, symbole, łączy URL itp.).

Każdy z tematów reprezentowany jest przez grupę 30 słów o największej istotności (ang. *relevance*) – sposób sortowania tych leksemów można kontrolować, zmieniając wartość współczynnika λ ²⁵. Przeprowadzona przez badacza analiza i interpretacja znaczeń i konotacji tych słów kluczowych pozwala na wyciągnięcie wniosków na temat zawartości danego tematu i nadanie każdemu z nich nazwy²⁶.

Temat 1 – Reakcja państw NATO i UE na zagrożenie wywołane wybuchem wojny: przewodniczący, unia_europejski, Ukraina, sytuacja, państwo, Rosja, Polska, UE, chcieć, rosyjski, bezpieczeństwo, wielki, europejski, sankcja, pierwszy,

22 Skorzystałem w tym celu z biblioteki pyLDAvis.

23 Tematy generowane za pomocą tej metody mają charakter abstrakcyjnych zbiorów słów kluczowych – ich interpretacja leży w gestii badacza.

24 Posłużyłem się stoplistą stworzoną przez Marcina Bielaka (2013).

25 W analizie LDA współczynnik λ dla parametru *relevance* może przyjmować wartości od 0 do 1. Ustawienie niższej wartości tego współczynnika (bliżej 0) pozwala na posortowanie słów kluczy przydzielonych do danego tematu, tak żeby preferowane były słowa, których prawdopodobieństwo wystąpienia w danym temacie jest większe niż w reszcie korpusu (zgodnie z parametrem *lift*). Przy ustawieniu wyższej wartości słowa są sortowane zgodnie z prawdopodobieństwem ich wystąpienia w danym temacie, niezależnie od częstości w całym korpusie (Sievert, Shirley 2014: 66).

26 Na potrzeby artykułu przedstawiam wyniki analizy dla parametru $\lambda = 0,4$ (przy takim ustawieniu udało się osiągnąć najbardziej czytelne wyniki). Trzeba jednak podkreślić, że wyniki analizy LDA należy interpretować wielowymiarowo, co jest możliwe dzięki interaktywnej wizualizacji, która generowana jest przez bibliotekę pyLDAvis.

wojna, mieć, nowy, wojenny, widzieć, należeć, musieć, stan, Rosjanin, wspierać, NATO, zagrożenie, czas, plan, miejsce.

Temat 2 – Reakcja Polski i Polaków na zagrożenie wywołane wybuchem wojny: na_ŻYWO, solidarność, Polak, musieć, premier, pomoc_uchodźca, działać, agresja, Ukrainiec, jedność, Ukraina, kwestia, pomoc, swój, obywatel, władza, obrona, przed_wojna, Ukrainka, taki_sam, słowo, jak_szybko, na_Ukraina, Warszawa, walczyć, moment, Putin, przeciwko, rozmawiać, demokratyczny.

Temat 3 – Wpływ wojny na gospodarkę i ceny surowców: dzień, spotkanie, możliwość, gaz, przygotować, Orban, premier, rosyjski_agresja, polityczny, rząd, zrobić, otrzymać, węgiel, agresja_Rosja_na, siła, zmienić, pieniądz, trwać, Putin, osoba_który, obecnie, proces, finansować, dzięki, dołączyć_do, wprowadzić, robić, nadzieja, daleki, bliski.

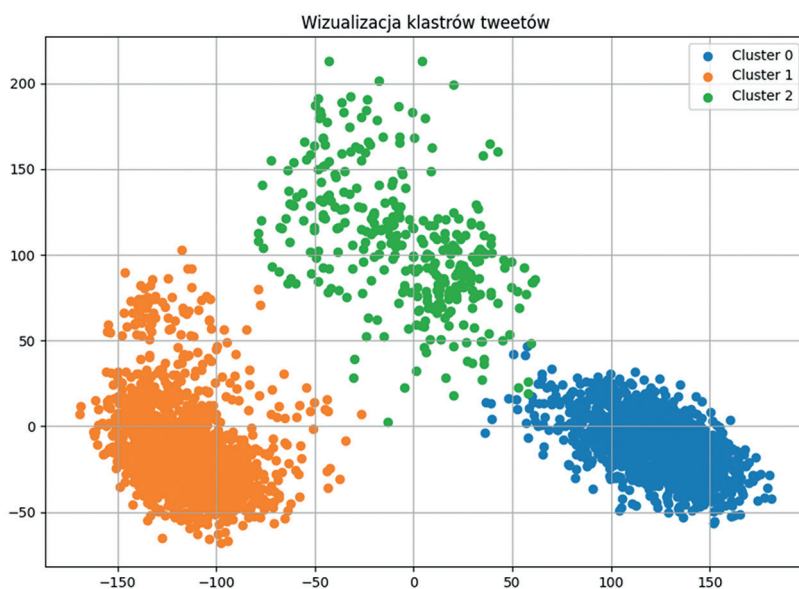
Temat 4 – Przyjęcie uchodźców z Ukrainy: minister, uchodźca, wiceminister_w, dziecko, osoba, obywatel_Ukraina, ws, kobieta, samorząd, praca, cel, pomagać, siebie, Ukraina, Ukrainiec, granica, pomoc_humanitarny, zostać, serce, transport, organizacja_pozarządowy, punkt, nasz_kraj, numer_PESEL, pełny, na_teren, uchodźcy, ręka, punkt_recepcyjny, który_uciekać.

Temat 5 – Organizacja pomocy dla uchodźców: Europa, kraj, system, wolność, nie_tylko, dla_uchodźca, życie, przyjąć, specjalny, suwerenność, miasto, specustawa, cały_Europa, poważny, zapewnić, zmiana, szpital, obowiązek, edukacja, wsparcie, mocny, wartość, Budapeszt, ale_także, inwazja, szczególnie, rada, uciekać_przed_wojna, zobaczyć, rządzący.

Wyniki analizy LDA wydają się w dużym stopniu potwierdzać wnioski, których dostarczyła interpretacja list frekwencyjnych. Politycy stosunkowo rzadko odnoszą się bezpośrednio do działań wojennych prowadzonych na terytorium Ukrainy. Skupiają się zaś przede wszystkim na komentowaniu politycznych (tematy 1 i 2), ekonomicznych (temat 3) oraz społecznych (tematy 4 i 5) skutków wybuchu wojny. Analiza LDA uwidoczniła jednak ponadto ważną rolę, jaką w dyskursie o wojnie odgrywają nazwy wartości oraz określenia zawierające sem pozytywnego wartościowania w obrębie swojej podstawowej treści, np.: *jedność, nadzieja, solidarność, wolność, bezpieczeństwo, suwerenność, wsparcie, wspierać, obrona, pomoc, pomagać, życie* (por. Puzyrnik 1992: 119). Nazwy antywartości i określenia wartościujące negatywnie nie są natomiast tak licznie obecne wśród słów, które zostały przez algorytm uznane za reprezentatywne dla wygenerowanych tematów. Należą do nich m.in. wyrazy *agresja* czy *zagrożenie*. Niekoniecznie musi oznaczać to jednak, że politycy w swoich wypowiedziach wykorzystują retorykę odwołującą się wyłącznie do pozytywnych emocji odbiorców. Wręcz przeciwnie – używanie określeń takich jak *bezpieczeństwo, suwerenność* czy *wolność* może służyć budowaniu wizerunku wroga, który dokonuje zamachu na te ważne dla grupy wartości.

Analiza semantyczna z wykorzystaniem dużego modelu językowego (LLM)

Warto zauważyć, że model LDA opiera się na założeniach probabilistycznych i nie uwzględnia semantycznych relacji między słowami – skupia się wyłącznie na częstości współwystępowania wyrazów w dokumentach. Jego skuteczność może być ograniczona w przypadku krótkich tekstów, takich jak tweety, ze względu na niewielką liczbę informacji zawartych w pojedynczym wpisie. Z tego względu wyniki zostały zweryfikowane przy użyciu analizy treści z wykorzystaniem dużego modelu językowego. Każdy tweet został przekształcony w wektor osadzenia (tzw. *embedding*) za pomocą modelu językowego Bielik 11B (por. Ociepa, Flis, Gwoździej, Wróbel, Kinas 2024). Następnie zastosowałem technikę KMeans, która pozwala na podział danych na klastry (grupy) na podstawie odległości pomiędzy punktami wektorowymi w tej wielowymiarowej przestrzeni. Wyniki zostały zwizualizowane za pomocą metody PCA (Principal Component Analysis), która redukuje liczbę wymiarów do dwóch, zachowując przy tym jak najwięcej informacji (por. wykres 2). W celu interpretacji klastrów zebrałem zlematyzowane tweety przypisane do każdego z nich i obliczyłem częstość występowania słów. Następnie przeprowadziłem porównanie częstości między klastrami oraz normalizację wyników w celu uwzględnienia różnic w ich liczności. Za słowa kluczowe uznałem te, które występowały istotnie częściej w danym klastrze niż w pozostałych. Dla każdego klastra wybrałem 150 reprezentujących go słów kluczowych.



Wykres 2. Wizualizacja klastrów (źródło: oprac. własne)

Analiza tematyczna tweetów przeprowadzona z wykorzystaniem wektorowych reprezentacji znaczeniowych i klasteryzacji ujawniła trzy wyraźnie zróżnicowane obszary narracyjne pojawiające się w badanych wypowiedziach. Klaster 0 obejmuje tweety, w których dominują pojęcia związane z geopolityką, stosunkami międzynarodowymi oraz polityką gospodarczą. Często pojawiały się takie słowa jak: *Rosja, rząd, Putin, państwo, polski, Polak, UE, sankcja, NATO, granica, polityka, węgiel, Niemiec, PiS, Zachód, inflacja*. Klaster 1 charakteryzuje się obecnością terminów odnoszących się do pomocy humanitarnej oraz wsparcia obywateli Ukrainy, np.: *Ukraina, premier, Polska, minister, wiceminister, uchodźca, pomoc, dziecko, samorząd, suwerenność, kobieta, obywatel_Ukraina, działania, wojenny, numer PESEL*. Może to sugerować, że politycy w tych wypowiedziach skupiali się na reakcjach rządu i samorządu na kryzys wywołany wojną, zwracali uwagę na potrzebę organizacji pomocy humanitarnej dla uchodźców. W klastrze 2 dominowały natomiast słowa związane z aspektami militarnymi, bezpieczeństwem i konsekwencjami gospodarczymi wojny, takie jak: *wojna, Rosja, Putin, polski, rząd, inflacja, sprawa, sytuacja, bezpieczeństwo, obrona, armia, NATO, problem, kryzys, propaganda*. Wskazuje to na koncentrację wypowiedzi na działaniach zbrojnych, zagrożeniach militarnych oraz dyskusji o skutkach wojny dla Polski – zarówno w wymiarze bezpieczeństwa, jak i ekonomii oraz polityki informacyjnej.

Analiza zawartości na podstawie frekwencji słów nie pozwala na uchwycenie kontekstu, co może prowadzić do nadmiernego uproszczenia wyciągniętych na tej podstawie wniosków, dlatego do interpretacji treści klastrow ponownie wykorzystałem możliwości oferowane przez duże modele językowe. Dla każdego klastra zebrałem wszystkie tworzące go tweety. Następnie podzieliłem powstały na ich podstawie tekst na zdania i wygenerowałem dla nich reprezentacje wektorowe. Utworzyłem macierz podobieństw kosinusowych między wektorami, co pozwoliło ocenić, jak bardzo treść poszczególnych zdań jest do siebie zbliżona. Z wykorzystaniem algorytmu PageRank wybrałem 50 najważniejszych zdań w każdym klastrze, na podstawie których wygenerowałem spójne streszczenia przy użyciu modelu Bielik 11B²⁷.

Klaster 0: W świetle tweetów polskich polityków widać wyraźną troskę o bezpieczeństwo narodowe i wzmocnienie obronności Polski. Priorytetem jest też reagowanie na kryzys uchodźczy z Ukrainy, wspierając ofiary wojny przy jednoczesnym uniknięciu bezpośredniego zaangażowania Polski w konflikt. Politycy postulują izolację Rosji poprzez ograniczenie handlu i nałożenie sankcji, a także zwiększenie niezależności energetycznej kraju. Ważnym elementem jest utrzymanie i wzmocnienie sojuszniczych relacji międzynarodowych, zwłaszcza z państwami UE i NATO. Dyskusje obejmują również

27 Teksty przywołuję z zachowaniem oryginalnej pisowni, po wygenerowaniu nie poddawałem ich dodatkowej redakcji.

kwestie polityki migracyjnej, potrzebę zmian prawnych oraz dostosowanie przepisów do wyzwań wynikających z wojny rosyjsko-ukraińskiej. Część polityków krytykuje opieszałość i brak zdecydowanych działań rządu w obliczu narastającego kryzysu, podkreślając przy tym solidarność społeczną Polaków w obliczu trudności.

Klaster 1: Podsumowując analizę tweetów polskich polityków komentujących wojnę rosyjsko-ukraińską, można stwierdzić, że najczęściej skupiają się oni na aspektach bezpieczeństwa narodowego Polski oraz humanitarnego wymiaru konfliktu. W ich wypowiedziach często podkreślana jest kwestia odstraszenia Rosji, wsparcia dla Ukrainy w obronie jej terytorium oraz przyjmowania uchodźców wojennych. Politycy zwracają uwagę na konieczność koordynacji działań między państwami i organizacjami pozarządowymi w celu zapewnienia efektywnej pomocy. Ważnym aspektem jest również podkreślanie korzyści gospodarczych i geopolitycznych, jakie może przynieść stabilna i silna Ukraina dla Polski oraz UE. Nie zapominają także o wyzwaniach logistycznych związanych ze wzrostem liczby uchodźców oraz konieczności utrzymania stabilności kraju w obliczu zwiększonego zagrożenia.

Klaster 2: Polscy politycy skupiają się głównie na aspektach wojny rosyjsko-ukraińskiej związanych z bezpieczeństwem Polski, pomocą uchodźcom oraz sankcjami wobec Rosji. Wiele wypowiedzi dotyczy konieczności zapewnienia bezpieczeństwa kraju, wsparcia Ukrainy i obrony przed rosyjską propagandą. Politycy podkreślają potrzebę jedności NATO i UE oraz mówią o przygotowaniach do długotrwałego kryzysu. Wiele głosów wskazuje na konieczność wspierania Ukrainy, ale także ostrożnego zarządzania krajowymi zasobami i kontrolowania granic. Pojawiają się też apele o odpowiedzialne działanie i unikanie prowokacji wobec Rosji. Tweety poruszają również temat wsparcia uchodźców, ale zwracają uwagę na potrzebę dbania także o polskich obywateli i krajowe interesy gospodarcze.

Treść streszczeń wydaje się potwierdzać wcześniejsze wnioski – w badanych wypowiedziach najważniejszy jest wpływ konfliktu na sytuację wewnętrzną i zagraniczną Polski. Jednocześnie jednak można zauważyć, że pojawiają się różnice zdań na temat właściwej reakcji na kryzys – krytyka rządu za opieszałość, obawa o zbytne zaangażowanie w konflikt. Pokazuje to, że analizowany korpus cechuje się wewnętrznym zróżnicowaniem. Pogłębiona analiza sposobów wartościowania poszczególnych zagadnień mogłaby stać się przedmiotem dalszych badań jakościowych.

Wnioski

Przeprowadzone badanie dowodzi, że istotne dla pogłębionej analizy treści może być łączenie różnych metod ilościowych. Wstępna analiza frekwencyjna ujawniła, że polscy politycy koncentrują się przede wszystkim na gospodarczych, społecznych i humanitarnych aspektach konfliktu. W badanych wypowiedziach często pojawiały się odniesienia do pomocy uchodźcom wojennym, skutków gospodarczych wojny dla Europy oraz roli Polski i innych państw europejskich w zarządzaniu kryzysem migracyjnym i politycznym.

Analiza jednostek nazewniczych pozwoliła na identyfikację kluczowych aktorów dyskursu. Zaskakiwać może wysoka frekwencja nazwisk polityków prorosyjskich, w tym prezydenta Rosji, przy jednocześnie relatywnie niskim natężeniu odniesień do prezydenta Ukrainy oraz liderów państw wspierających Ukrainę. Może to sugerować, że analizowane wypowiedzi bardziej koncentrowały się na potencjalnych zagrożeniach wynikających z agresji Rosji niż na analizowaniu sytuacji Ukrainy. Taki rozkład akcentów może odzwierciedlać dominującą narrację, w ramach której priorytetem jest bezpieczeństwo Polski. Warto zwrócić uwagę również na to, że w badanych wypowiedziach pojawiały się onimy, które niosą ze sobą znaczenia symboliczne. Niektóre z nich, jak *Wołyń* czy *Stepan Bandera*, mają znaczenia ugruntowane historycznie i kulturowo – odnoszą się do wydarzeń i postaci głęboko osadzonych w pamięci zbiorowej Polaków. W obliczu wojny ich sens ulega aktualizacji, stają się one narzędziem interpretacji współczesnych wydarzeń poprzez pryzmat przeszłości, często nacechowanej emocjonalnie i ideologicznie. Inne onimy, takie jak *Bucza*, *Mariupol* czy *Borodzianka*, nabierają znaczenia w trakcie wojny – są to nazwy miejsc przywołujące konkretne, często traumatyczne wydarzenia. Ich pojawienie się w dyskursie buduje nową przestrzeń semantyczną wokół doświadczenia wojny.

Użycie metody LDA (Latent Dirichlet Allocation) umożliwiło identyfikację głównych tematów dyskursu i potwierdziło dominację aspektów gospodarczych, społecznych i humanitarnych nad militarnymi. Ale jednocześnie uwidoczniło rolę, jaką w badanych wypowiedziach odgrywają nazwy wartości.

Analiza treści z wykorzystaniem modelu Bielik pomogła w uchwyceniu głębszych zależności semantycznych między tekstami, co pozwoliło na bardziej trafną segmentację wypowiedzi polityków. Wykorzystanie tego podejścia ujawniło, że w badanych wypowiedziach pojawiają się różnice zdań na temat tego, jak powinna wyglądać odpowiednia reakcja na kryzys wywołany wojną.

Należy jednak zwrócić uwagę także na pewne ograniczenia związane z zastosowaną metodologią. Przede wszystkim z wykorzystaniem wyłącznie danych ilościowych niezwykle trudne jest zbadanie różnic w sposobie budowania narracji na temat wojny przez poszczególne siły polityczne. Wiąże się to ze strukturą korpusu, która zaś wynika z tego, że konta partyjne publikują tweety z różną frekwencją i częstotliwością. Porównywanie danych liczbowych pomiędzy poszczególnymi partiami

mogłoby prowadzić do błędnych wniosków. Wydaje się jednak, że badanie ilościowe może stać się dobrym punktem wyjścia do przeprowadzenia w przyszłości analiz jakościowych, które mogłyby skupić się na analizie wartości i wartościowania zawartego w zebranych wypowiedziach. Interesujących wniosków mogłaby dostarczyć także analiza sposobów ramowania²⁸ wojny przez poszczególnych aktorów politycznych w badanym dyskursie. Na tej podstawie można spróbować odtworzyć zawarty w zebranych tekstach dyskursywny obraz świata (por. Czachur 2011, 2020). Uzupełnienie badań ilościowych o analizę jakościową mogłoby służyć także weryfikacji wstępnych wniosków na temat analizowanego dyskursu. Trzeba podkreślić, że analiza ilościowa pozwala spojrzeć na materiał z przyjęciem szerokiej perspektywy, zapewnia kontekst niezbędny do zrozumienia zawartości korpusu, a tym samym treści poruszanych w ramach badanego dyskursu, jednak ograniczenie się wyłącznie do danych liczbowych może niekiedy prowadzić do zbyt uproszczeń i uogólnień. Wreszcie należy zaznaczyć, że niektóre prezentowane w artykule metody badawcze mają charakter eksperymentalny. Z dystansem należy podejść przede wszystkim do danych uzyskanych za pomocą modelu językowego Bielek. Duże modele językowe oparte na architekturze transformerowej mają skłonność do tzw. halucynacji, czyli przedstawiania informacji nieprawdziwych lub niezwiązanych z treściami dostarczonymi przez użytkownika. Prezentacja tego rodzaju metodologii służy głównie przedstawieniu nowych propozycji analizy dyskursu politycznego.

Podsumowując: przeprowadzone badanie wykazało, że kompleksowe wykorzystanie różnych metod ilościowych – łączących analizę frekwencyjną, badanie jednostek nazewniczych, klasteryzację tematyczną oraz wykorzystanie dużych modeli językowych – pozwala na pełniejsze uchwycenie niuansów dyskursu politycznego. Takie wieloaspektowe podejście analityczne nie tylko potwierdza złożoność omawianego dyskursu, lecz także otwiera perspektywy na dalsze badania nad dynamiką komunikacji politycznej w kontekście wojny rosyjsko-ukraińskiej.

28 Ramowanie (ang. *framing*) definiuję za Robertem Entmanem jako proces selekcji i eksponowania pewnych aspektów rzeczywistości w przekazie medialnym, którego celem jest reinterpretacja danej kwestii, nadanie jej określonej oceny moralnej lub sugestia sposobu jej rozwiązania (Entman 1993: 53). Trzeba zaznaczyć, że ramowanie jest procesem świadomym i manipulacyjnym – aktorzy dyskursu, politycy i dziennikarze używają ram do narzucenia społeczeństwu określonej interpretacji rzeczywistości. Kontrolowanie ram umożliwia sterowanie społeczną uwagą, a tym samym wpływanie na światopogląd. Przeramowanie (ang. *reframing*) nie tylko nadaje przedstawianym wydarzeniom nowe znaczenie lub kontekst, lecz także może się wiązać z kreowaniem ważnych zmian społecznych (Lakoff 2004; Pluwak 2009: 62; Franczak 2014: 97).

Literatura

- ALAGHA I., 2021, *Topic Modeling and Sentiment Analysis of Twitter Discussions on COVID-19 from Spatial and Temporal Perspectives*, „Journal of Information Science Theory and Practice” 9 (1), s. 35–53, <https://doi.org/10.1633/JISTaP.2021.9.1.3>.
- BIELAK M., 2013, *Stopwords*, [on-line:] <https://github.com/bieli/stopwords/blob/master/polish.stopwords.txt> (dostęp: 19 IX 2023).
- CIEŚLIKOWA A., 2001, *Nazwa w tekście a tekst w nazwie*, [w:] A. Pajdzińska, R. Tokarski (red.), *Semantyka tekstu artystycznego*, „Czerwona Seria» Instytutu Filologii Polskiej UMCS”, 15, Lublin, s. 99–108.
- CZACHUR W., 2011, *Dyskursywny obraz świata. Kilka refleksji*, „Tekst i Dyskurs” vol. 4, s. 79–97.
- CZACHUR W., 2020, *Lingwistyka dyskursu jako integrujący program badawczy*, Wrocław.
- CZERNY A., 2011, *Teoria nazw geograficznych*, „Prace Geograficzne – Polska Akademia Nauk”, nr 226, Warszawa.
- ENTMAN R.M., 1993, *Framing. Toward Clarification of a Fractured Paradigm*, „Journal of Communication” vol. 43 (4), s. 51–58, <https://doi.org/10.1111/j.1460-2466.1993.tb01304.x>.
- FRANCZAK K., 2014, *Medialne objaśnienia kryzysu gospodarczego w perspektywie framing analysis*, [w:] M. Czyżewski, K. Franczak, M. Nowicka, J. Stachowiak (red.), *Dyskurs elit symbolicznych. Próba diagnozy*, „Biblioteka Dyskursu Publicznego”, Warszawa, s. 91–122.
- KIERAŚ W., KOBYLIŃSKI Ł., 2021, *Korpusomat – stan obecny i przyszłość projektu*, „Język Polski” CI, z. 2, s. 49–58.
- LAKOFF G., 2004, *Don't Think of an Elephant! Know Your Values and Frame the Debate. The Essential Guide for Progressives*, Vermont.
- OCIEPA K., FLIS Ł., WRÓBEL K., GWOŹDZIEJ A., KINAS R., 2024, *Bielik 7B vo.1. A Polish Language Model – Development, Insights, and Evaluation*, arXiv, <https://doi.org/10.48550/arXiv.2410.18565>.
- PLUWAK A., 2009, *Geneza i ewolucja pojęcia framing w naukach społecznych*, „Global Media Journal – Polish Edition” nr 1 (5), s. 49–79.
- POLKOWSKA L., 2018, *Wartości i antywartości w tygodnikach opinii. Komparatywna analiza dyskursu z elementami lingwistyki kwantytatywnej*, Warszawa.
- ORĘDZIE: Orędzie Prezydenta RP, [on-line:] https://www.youtube.com/watch?v=GLuGTYYgrpM&ab_channel=PrezydentRPAAndrzejDuda (dostęp: 22 II 2024).
- PISORG 2022A: @pisorgpl, *Po 40 dniach wojny...*, [on-line:] <https://twitter.com/pisorgpl/status/1510881222590926848> (dostęp: 19 IX 2023).
- PISORG 2022B: @pisorgpl, *Rok temu Mariupol nosił...*, [on-line:] <https://twitter.com/pisorgpl/status/1527575607231619074> (dostęp: 19 IX 2023).
- PREZYDENT.PL 2008, *Wizyta Prezydenta RP w Gruzji*, <http://prezydent.pl/kancelaria/archiwum/archiwum-lecha-kaczynskiego/aktualnosci/rok-2008/wizyta-prezydenta-rp-w-gruzji,26753,archive> (dostęp: 19 IX 2023).
- NOWEPSL: @nowepsl, 2022, *Mariupol wygląda dziś jak...*, [on-line:] <https://twitter.com/nowepsl/status/1554111445054504960> (dostęp: 19 IX 2023).
- PUZYNINA J., 1992, *Język wartości*, Warszawa.
- REJTER A., 2016, *Nazwa własna wobec gatunku i dyskursu*, „Prace Naukowe Uniwersytetu Śląskiego w Katowicach”, nr 3408, Katowice.

- SCHMITT X., KUBLER S., ROBERT J., PAPADAKIS M., LETRAON Y., 2019, *A Replicable Comparison Study of NER Software: StanfordNLP, NLTK, OpenNLP, SpaCy, Gate*, [w:] *2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, s. 338–343, <https://doi.org/10.1109/SNAMS.2019.8931850>.
- SIEVERT C., SHIRLEY K., 2014, *LDavis: A Method for Visualizing and Interpreting Topics*, [w:] J. Chuang, S. Green, M. Hearst, J. Heer, Ph. Koehn (red.), *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*, Baltimore, s. 63–70.
- WOLNICZ-PAWŁOWSKA E., 2007, *Nazwy geograficzne w kulturze narodu*, „Prace Językoznawcze” nr 9, s. 143–156.

**Russo-Ukrainian War in Polish Political Discourse on Twitter –
Methods of Quantitative Content Analysis
Abstract**

This article explores the use of quantitative methods in analysing political discourse on social media. The study focuses on tweets published by official political party accounts on Twitter during the first six months following Russia's full-scale invasion of Ukraine. The primary objective was to identify the main topics, discourse actors, and values present in political messages related to the armed conflict, as well as to reconstruct the structure and dynamics of the discourse. The analysis included a combination of quantitative techniques, including frequency analysis, named entity recognition, topic modelling with the LDA algorithm, and semantic analysis based on a large language model (LLM). This methodological approach allowed for the identification of dominant themes and their evolution over time, offering deeper insights into the complexity of the discourse. The study serves as an example of how natural language processing (NLP) tools can be applied to the analysis of contemporary political communication.