



Analiza danych parlamentarnych

Warsztat pokonkursowy

Instytut Podstaw Informatyki PAN
Warszawa 2023

Analiza danych parlamentarnych

Warsztat pokonkursowy

pod redakcją Macieja Ogrodniczuka

Projekt okładki i stron tytułowych:
Bartosz Szymański

Ilustracja na okładce:
https://commons.wikimedia.org/wiki/File:Obchody_XX-lecia_odrodzonego_Senatu_01.JPG

Skład i łamanie w systemie \LaTeX :
Maciej Ogrodniczuk

ISBN 978-83-63159-32-0

Publikacja finansowana ze środków Ministerstwa Edukacji i Nauki w ramach wkładu własnego do europejskiej infrastruktury badawczej CLARIN ERIC — European Research Infrastructure Consortium: Common Language Resources and Technology Infrastructure—Wspólne Zasoby Językowe i Infrastruktura Technologiczna na lata 2022-2023” (CLARIN Q, nr umowy 2022/WK/09).

© Copyright by Instytut Podstaw Informatyki PAN, Warszawa 2023
Jana Kazimierza 5
01-248 Warszawa



Publikacja jest dostępna na licencji Creative Commons Uznanie autorstwa 4.0 (CC BY 4.0). Treść licencji dostępna jest na stronie <http://creativecommons.org/licenses/by-nc-sa/4.0>.

Warszawa 2023

Spis treści

Informacja o konkursie	5
Kapituła konkursu	7
Nagrodzone prace	9
Podziękowania	11
Program warsztatu	15
Zmiany w polszczyźnie ostatnich stu lat w zwierciadle debat parlamentarnych	
Marek Łaziński	17
Aplikacja do analizy sentymentu wypowiedzi parlamentarzystów RP	
Bartłomiej Klimowski	19
Analiza dynamiki emocji w debatach parlamentarnych o wojnie w Ukrainie	
Konrad Kiljan, Ewelina Gajewska	27
Język wrażliwy na płęć (<i>gender-fair language</i>) w debatach parlamentarnych w korpusie ParlaMint — badanie pilotażowe	
Aleksandra Tomaszewska, Anna Jamka	35
Zagraniczne orzeczenia sądów konstytucyjnych w debacie parlamentarnej parlamentów państw członkowskich Unii Europejskiej	
Mikołaj Wolanin	43
Analiza emocji występujących na posiedzeniach sejmowych	
Jakub Chojnacki, Bartłomiej Kruczek	49

Informacja o konkursie

Konkurs na analizę danych parlamentarnych trwał od 2 października do 15 listopada 2023 r. i został zorganizowany przez Instytut Podstaw Informatyki PAN w ramach konsorcjum CLARIN-PL. Jego celem było opracowanie oryginalnej analizy (jedno- lub wielojęzycznej, z obowiązkowym udziałem danych polskich) prezentującej ciekawy problem badawczy z dowolnej dyscypliny nauki. Źródłem danych do analizy był wielojęzyczny korpus ParlaMint w dostępnej w momencie ogłoszenia konkursu wersji 3.1 — zestaw wystąpień z parlamentów 26 krajów i regionów autonomicznych, zapisany we wspólnym formacie, opatrzone anotacją lingwistyczną i dostępny do pobrania i przeszukiwania. Aktualne wydanie korpusu zawierało także tłumaczenia wystąpień na język angielski.

W konkursie mogli wziąć udział studenci, doktoranci oraz młodzi naukowcy. Ogłoszenie prezentowało w ramach inspiracji kilka wcześniejszych badań powstałych dla korpusów ParlaMint. Zgłoszenia miały mieć postać zanonimizowanych krótkich artykułów gotowych do publikacji w tomie pokonkursowym.

Na konkurs przesłano 5 zgłoszeń, z których kapituła wybrała trzy, wyróżnione odpowiednio nagrodami w wysokości 5, 3 i 1 tys. zł.

W związku z niejednorodnym charakterem zgłoszeń oraz sugestiami niektórych uczestników, że przekazane teksty będą publikowane w czasopiśmie, podjęto decyzję o ograniczeniu publikacji do niniejszej broszury zawierającej wyłącznie abstrakty zgłoszonych tekstów i przekazano uczestnikom prośbę o przygotowanie wersji ograniczonych do 1200–1500 słów.

Warsztat podsumowujący konkurs, na który zostali zaproszeni wszyscy uczestnicy, członkowie kapituły konkursu (recenzenci prac) oraz pozostałe osoby zainteresowane przetwarzaniem danych parlamentarnych, został zaplanowany na 29 lutego 2024 r. w siedzibie Kancelarii Sejmu RP.

Kapituła konkursu

Zgłoszenia konkursowe oceniał zespół recenzentów w składzie:

- dr hab. Magdalena Derwojedowa, prof. Uniwersytetu Warszawskiego,
- dr hab. Agnieszka Hess, prof. Uniwersytetu Jagiellońskiego,
- dr hab. Krzysztof Kasianiuk, prof. Collegium Civitas,
- dr Marta Kołczyńska z Instytutu Studiów Politycznych PAN,
- dr Agnieszka Kwiatkowska z Uniwersytetu SWPS,
- dr hab. Maciej Ogrodniczuk, prof. Instytutu Podstaw Informatyki PAN,
- prof. dr hab. Adam Pawłowski z Uniwersytetu Wrocławskiego,
- dr hab. Piotr Pęzik, prof. Uniwersytetu Łódzkiego.

Nagrodzone prace

Pierwsze miejsce i nagrodę w wysokości 5 tys. zł otrzymuje autor tekstu *Aplikacja do analizy sentymentu wypowiedzi parlamentarzystów RP*, Bartłomiej Klimowski z Uniwersytetu Warszawskiego, za dobry przykład wykorzystania metody analizy składowych głównych w aplikacji prezentującej charakterystykę emocjonalną wypowiedzi parlamentarzystów. Recenzenci docenili dobre dostosowanie narzędzia do zbioru danych oraz możliwość jego dalszej rozbudowy.

Drugie miejsce i nagrodę w wysokości 3 tys. zł otrzymują autorzy tekstu *Analiza dynamiki emocji w debatach parlamentarnych o wojnie w Ukrainie*, Konrad Kiljan z Uniwersytetu Warszawskiego i Ewelina Gajewska z Politechniki Warszawskiej, za dopracowaną analizę porównującą emocje w wypowiedziach parlamentarzystów z Europy Zachodniej, Wschodniej i Ukrainy na przestrzeni czasu. Recenzenci docenili dobry dobór metody badawczej oraz jasne kryteria oceny wyników (z podaną dokładnością oraz informacją o walidacji).

Trzecie miejsce i nagrodę w wysokości 1 tys. zł otrzymują autorki tekstu *Język wrażliwy na płeć (gender-fair language) w polskim parlamencie: badanie korpusowe debat parlamentarnych w korpusie ParlaMint*, Aleksandra Tomaszewska z Instytutu Podstaw Informatyki Polskiej Akademii Nauk i Anna Jamka z Instytutu Lingwistyki Stosowanej Uniwersytetu Warszawskiego, za analizę tendencji dotyczących użycia języka inkluzywnego w debatach parlamentarnych. W zgłoszonym tekście recenzenci zwrócili uwagę na dobry opis tematu i czytelne pytania badawcze.

Podziękowania

Dziękujemy pracownikom Kancelarii Sejmu RP dzięki wsparciu których warsztat podsumowujący projekt mógł się odbyć na terenie Sejmu, w szczególności:

- p. Ministrowi Michałowi Deskurowi, Zastępcy Szefa Kancelarii Sejmu X kadencji,
- p. Ministrowi Christianowi Młynarkowi, Zastępcy Szefa Kancelarii Sejmu IX kadencji,
- p. Zbigniewowi Jabłońskiemu, Dyrektorowi Ośrodka Informatyki,

oraz pozostałym pracownikom Kancelarii wspierającym naszą inicjatywę na różnych etapach jej organizacji.

Dziękujemy członkom Kapituły Konkursu, przedstawicielom konsorcjum CLARIN-PL oraz współpracownikom z Zespołu Inżynierii Lingwistycznej i Działu Administracji Instytutu Podstaw Informatyki PAN wspomagających nas w przeprowadzeniu konkursu.

Informacja o finansowaniu prac

Prace nad Korpusem Dyskursu Parlamentarnego, jeszcze pod nazwą Polskiego Korpusu Sejowego, zawierającego wyłącznie treść stenogramów z posiedzeń Sejmu oraz interpelacji i zapytań, zostały rozpoczęte w projekcie badawczym CESAR (CEntral and South-east europeAn Resources — Elektroniczne Zasoby Językowe Europy Środkowej i Południowo-Wschodniej) finansowanym przez Komisję Europejską w ramach programu CIP – ICT PSP (Programu Szczegółowego „Konkurencyjność i Innowacje” na Rzecz Wspierania Polityki w Zakresie Technologii Informacyjnych i Komunikacyjnych; nr kontraktu: 271022, Call 4, Theme 6: Multilingual Web, Objective 6.1: Open linguistic infrastructure, CIP-ICT-PSP-2009-4) od 1 lutego 2011 do 31 stycznia 2013 r. i dodatkowo dofinansowanym przez MNiSW w ramach umowy o wykonanie projektu międzynarodowego współfinansowanego.

Od roku 2016 prace nad korpusem były finansowane przez MNiSW w ramach wkładu krajowego na rzecz CLARIN-PL — infrastruktury badawczej realizowanej w ramach wspólnego międzynarodowego przedsięwzięcia pn. CLARIN ERIC — European Research Infrastructure Consortium: Common Language Resources and Technology Infrastructure zgodnie z decyzjami nr DIR/WK/2016/02 (1 lipca 2016 — 30 czerwca 2018) oraz DIR/WK/2018/01 (1 lipca 2018 — 30 czerwca 2021).

Uzupełnienie danych dot. posłów VIII i IX kadencji oraz senatorów IX i X kadencji zawartych w Polskiej Ontologii Politycznej wykorzystywanej w indeksowaniu korpusu zrealizowano w latach 2020–23 w ramach dwóch edycji projektu ParlaMint finansowanego przez europejskie konsorcjum CLARIN ERIC.

W latach 2020–21 baza tekstowa korpusu została uzupełniona w ramach projektu MARCELL: Multilingual Resources for CEFAT in the legal domain (Zasoby wielojęzyczne dla CEFAT w domenie prawnej) finansowanego przez Agencję Wykonawczą Komisji Europejskiej ds. Innowacyjności i Sieci (grant nr INEA/CEF/ICT/A2017/1565710) wraz z wkładem MNiSW na realizację projektu międzynarodowego współfinansowanego (umowa nr 4082/CEF/2018/2).

W latach 2022–23 prace były finansowane w ramach Programu Operacyjnego Inteligentny Rozwój 2014–2020, Priorytet IV: Zwiększenie potencjału naukowo-badawczego, Działanie 4.2: Rozwój nowoczesnej infrastruktury badawczej sektora nauki, nr POIR.04.02.00-00C002/19 oraz w ramach wniesienia wkładu własnego do europejskiej infrastruktury badawczej CLARIN ERIC — European Research Infrastructure Consortium: Common Language Resources and Technology Infrastructure—Wspólne Zasoby Językowe i Infrastruktura Technologiczna na lata 2022–2023” (projekt CLARIN Q finansowany przez MEiN, nr umowy 2022/WK/09).

Program warsztatu

12:00–12:10 *Powitanie*

12:10–12:30 *Korpus Dyskursu Parlamentarnego i korpus ParlaMint*
Maciej Ogrodniczuk

Wykład na zaproszenie

12:30–13:30 *Zmiany w polszczyźnie ostatnich stu lat w zwierciadle debat parlamentarnych*
Marek Łaziński

Zgłoszenia konkursowe

13:30–13:50 *Aplikacja do analizy sentymentu wypowiedzi parlamentarzystów RP*
Bartłomiej Klimowski

13:50–14:10 *Analiza dynamiki emocji w debatach parlamentarnych o wojnie w Ukrainie*
Konrad Kiljan i Ewelina Gajewska

14:10–14:30 *Język wrażliwy na płęć (gender-fair language) w polskim parlamencie: badanie korpusowe debat parlamentarnych w korpusie ParlaMint*
Aleksandra Tomaszewska i Anna Jamka

14:30–14:50 *Zagraniczne orzeczenia sądów konstytucyjnych w debacie parlamentarnej parlamentów państw członkowskich UE*
Mikołaj Wolanin

Dyskusja panelowa

14:50–15:45 *Przetwarzanie danych parlamentarnych: co dalej?*
Członkowie Kapituły Konkursu

Zakończenie

15:45–15:50 *Wręczenie nagród*

15:50–16:00 *Podsumowanie warsztatu*

Zmiany w polszczyźnie ostatnich stu lat w zwierciadle debat parlamentarnych (streszczenie wykładu)

Marek Łaziński
(Uniwersytet Warszawski)

W wykładzie pokazano na przykładach debat parlamentarnych, jak zmieniała się polszczyzna w ostatnim stuleciu. Chodzi o zmiany gramatyczne: fleksyjne i składniowe, jak stopniowy wzrost udziału konstrukcji analitycznych, oraz zmiany leksykalne: pojawianie się nowych słów i zanikanie starych, w tym takich słów, o które toczyliśmy i do dziś toczymy spory. Analiza statystyczna Korpusu Dyskursu Parlamentarnego pozwala też prześledzić najważniejsze tematy debaty publicznej z ponad stu lat.

Debaty parlamentarne pozwalają dostrzec niekontrowersyjne w opisie i ocenie zmiany fleksyjne, jak choćby zamiana końcówek przymiotnika w N. lp. z *-em* na *-ym/im*, a także te, które były obiektem sporów nie tylko o język, ale i o relacje społeczne, które język w jakimś stopniu odzwierciedla. Do tej dziedziny należy rozwój feminatywów ze szczególnym uwzględnieniem formy *posłanka*, zmiany nazw etnicznych i innych przynależności grupowych.

W wykładzie zwrócono szczególną uwagę na te zjawiska, które występują w języku parlamentarzystów, jak feminatywy, konstrukcje analityczne czy szczególna skłonność do złożeń z członem *spec-*, oraz na te, które stały się tematem debat. 80 lat temu w Sejmie spierano się, czy nazywać sąsiedni naród Ukraińcami czy Rusinami; dziś spieramy się o inne nazwy etniczne, a nawet o przyimki.

Język się zmienia i nic dziwnego, że o tych zmianach dyskutujemy, że nie wszyscy od razu je akceptują. Korpus Dyskursu Parlamentarnego odzwierciedla proporcjonalnie i zgodnie z chronologią proces zmian w polszczyźnie ostatnich stu lat oraz dyskusje o tych zmianach.

Aplikacja do analizy sentymentu wypowiedzi parlamentarzystów RP

Bartłomiej Klimowski
(Uniwersytet Warszawski)

Streszczenie

Emocje są nieodłączną częścią życia politycznego. Niniejsza praca przedstawia koncept aplikacji internetowej pozwalającej na analizę sentymentu wypowiedzi parlamentarzystów RP z lat 2015–2022. Aplikacja umożliwia konstruowanie profili emocjonalnych wyrażających zrównoważenie występowania pięciu emocji: szczęścia, wstrętu, gniewu, smutku i strachu. Aplikacja bazuje na danych projektu PARLAMINT, zbiorze NENCKI AFFECTIVE WORD LIST (NAWL) i zrealizowana jest z użyciem biblioteki języka R SHINY.

Słowa kluczowe

przetwarzanie języka naturalnego, analiza sentymentu, analiza głównych składowych

1. Wstęp

*Marszałek Sejmu będzie obecny w mediach, będzie miał swój podcast [...] Drodzy Państwo, jeżeli tylko wyrazicie swoją wolę, będziecie gośćmi tego podcastu, bo jak dobrze wiemy, **emocje to jest coś, co współczesne media lubią najbardziej, a jestem przekonany, że ich tam nie zabraknie.***

Te słowa, wypowiedziane przez nowego marszałka sejmu X kadencji Szymona Hołownię, jasno określają istotność emocji w życiu politycznym. Politycy mają tego pełną świadomość i często w przemyślany sposób kreują swoje wypowiedzi, tak aby reprezentowały określony sentyment w określonym celu. Charakteryzowanie wydźwięku emocjonalnego wypowiedzi danego polityka stanowi zarówno ciekawe zagadnienie badawcze, jak i jest przydatną informacją z punktu widzenia obywatela.

W niniejszej pracy prezentuję narzędzie pozwalające na inspekcję sentymentu wypowiedzi poszczególnych parlamentarzystów — pod względem balansu różnych emocji w ich wypowiedziach, różnorodności emocjonalnego słownictwa i tego, jak się plasują w stosunku do pozostałych członków parlamentu. Z faktu, że charakterystyka posła pod tym względem jest wartościowa sama w sobie, projekt zrealizowano w formie webowego dashboardu, tak aby można było przystępnie przeglądać wyniki.

Aplikacja jest dostępna pod adresem <https://hrspxmc.shinyapps.io/parlamintvo15/>.

2. Źródła danych

Analiza bazuje na danych projektu PARLAMINT (Erjavec i in. 2023) dla polskiego parlamentu na pełnym dostępnym zakresie (od 2015-11-12 do 2022-06-30 i obu izb). Charakterystyka sentymentu oparta jest na metodzie słownikowej, wykorzystującej dane ze zbioru *Nencki Affective Word List (NAWL; Riegel i in. 2015)*. Zbiór opisuje 2902 słowa na pięciu wymiarach emocjonalnych — radości, gniewu, smutku, strachu i wstrętu, w skali od 1 do 5.

Dane ze zbioru PARLAMINT wydobyto z plików formatu CoNLLU. W analizie wykorzystano zlematyzowaną formę słów. Zakres danych z projektu *NAWL* zawężono do najbardziej emocjonalnie nacechowanych słów, określonych jako górne 5% pod względem wagi na jednej z emocji i górne 5% dla iloczynu wszystkich wag (aby uwzględnić też słowa o silnym nacechowaniu na wielu wymiarach). Ze zbioru prawie trzech tysięcy słów ostatecznie analizowałem podzbiór pięciuset czterech. **W tym miejscu podkreślam, że wszystkie wskaźniki opisane w tej pracy dotyczą wyżej wymienionego zakresu silnie nacechowanych słów.**

3. Metody

3.1. Profil emocjonalny wypowiedzi

Profil emocjonalny określono jako balans pomiędzy siłą poszczególnych emocji we wszystkich wypowiedziach parlamentarzysty. Aby otrzymać ten indeks, zliczyłem występowanie każdego słowa z wcześniej zdefiniowanego zbioru słów silnie nacechowanych emocjonalnie, następnie wyznaczona częstość była przemnożona przez wagę na danym wymiarze emocjonalnym. W kolejnym kroku iloczyny są sumowane, zgrupowane podług emocji, a następnie normalizowane. Wynikiem jest piątka liczb, sumujących się do jedności, z których każda określa siłę występowania konkretnej emocji, w ramach całości, w wypowiedziach parlamentarzysty.

W celu stworzenia prostej reprezentacji powyższych wartości w dwóch wymiarach, na powyższym zbiorze danych zaaplikowano Analizę Głównych Składowych (PCA; Wikipedia 2023c). Przed obliczeniem PCA wartości zostały podzielone przez średnią i zlogarytmowane.

3.2. Wskaźniki różnorodności

Poza samym balansem emocji scharakteryzowałem różnorodność słownictwa emocjonalnego, jak i samego sentymentu. Do tego celu użyłem wskaźnika zdefiniowanego w następujący sposób ($p(x_i)$ to częstość występowania podgrupy ze zbioru):

$$E(x) = \frac{1}{\sum_{i=1}^N p(x_i)^2}$$

Wskaźnik występuje w literaturze pod wieloma nazwami (p. Wikipedia 2023a), ale sprowadza się do prostej interpretacji — *efektywnej* liczby podgrup. Przeanalizujmy skrajny przykład — kiedy mamy do czynienia z równym występowaniem wszystkich podgrup, wtedy wskaźnik jest równy ich liczbie; gdy jedna z nich występuje skrajnie często, np. stanowi 99,999% całości, to *efektywnie* możemy mówić o istnieniu tylko jednej grupy. Wskaźnik liczbowo charakteryzuje wszystkie sytuacje pomiędzy — określając heterogeniczność badanego zbioru.

Ponadto do scharakteryzowania różnorodności wyznaczyłem liczbę unikatowych słów i słowa będące *hapax legomenon* (HL_1), czyli występujące tylko raz we wszystkich wypowiedziach parlamentarzysty (Wikipedia 2023b).

3.3. Wordcloud

Żadna analiza tekstowa nie może się obyć bez popularnej metody wizualizacji nazywanej *wordcloud*. W tej pracy słowa przedstawiane w ten sposób są selekcjonowane i ważone z użyciem podejścia TF-IDF (Manning i in. 2022). Taki sposób ważenia pozwala na wydobycie słów, które są szczególnie często używane w wypowiedziach wybranego parlamentarzysty, a nie są popularnie używane w wypowiedziach pozostałych.

4. Przykładowy profil

W ramach prezentacji metod przedstawię wyniki otrzymane dla jednego z parlamentarzystów — Macieja Koniecznego. Zachęcam do wybrania w aplikacji innego parlamentarzysty i przesłanie jego profilu.

Tabela 1: Profil emocjonalny

Emocja	Udział	Percentyl	Udział-średnia	Udział/średnia
Gniew	19.08%	80.4%	1.08%	1.06
Wstręt	14.94%	57.9%	0.16%	1.01
Strach	22.22%	96.5%	1.92%	1.09
Szczęście	23.57%	11.1%	-4.76%	0.83
Smutek	20.20%	92.1%	1.59%	1.09

Tabela 1 przedstawia profil emocjonalny wypowiedzi parlamentarzysty. Oprócz tego zawarłem w niej kolumny określające pozycję w rozkładzie jako percentyl (względem reszty parlamentarzystów), jak i różnice oraz proporcję udziału względem średniej.

W tym wypadku można zaobserwować, że w wypowiedziach posła słowa o silnym wydźwięku w wymiarze strachu stanowią znaczącą część. Jest on aż w 96 percentylu, czyli w ścisłej czołówce osób często powołujących się na tę emocję. Sytuacja ma się odwrotnie w przypadku uczucia radości; poseł jest na odwrotnej stronie rozkładu i proporcja tego sentymentu stanowi zaledwie 0,83 średniej.

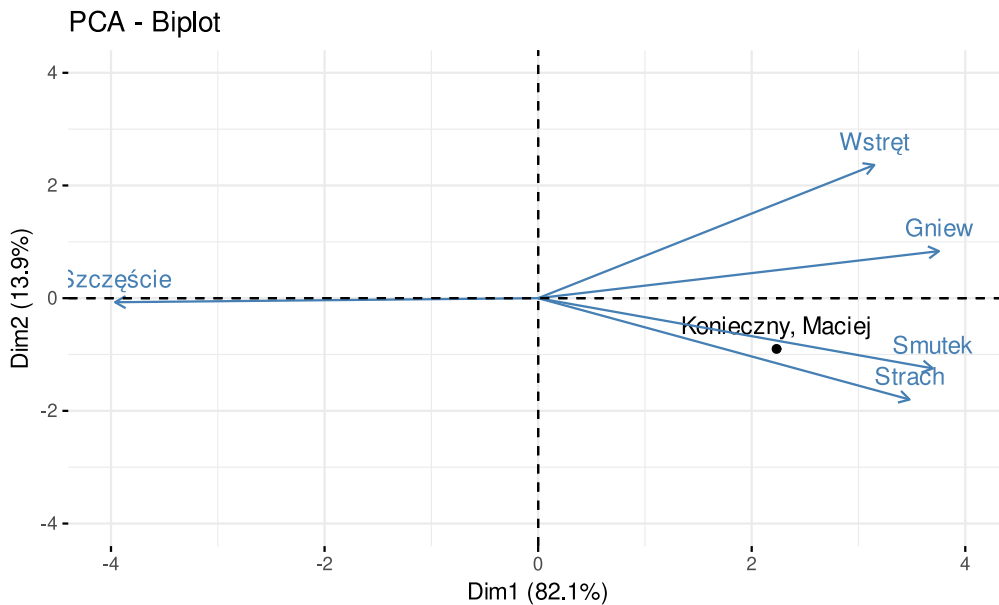
Różnorodność jest opisana w tabeli 2. Pod tym względem omawiany poseł wykazuje się dużym zasobem emocjonalnego słownictwa. Liczba efektywnych słów jest jedną z najwyższych w analizowanym zbiorze (99 percentyl), a pozostałe miary również plasują go wysoko.

Tabela 2: Wskaźniki różnorodności

Wskaźnik	Wartość	Percentyl
Efektywna liczba słów	41.0	99.4%
Efektywna liczba wyr. emocji	4.9	82.7%
Liczba unikatowych słów	150.0	76.9%
HL1	52.0	74.9%

Intuicyjnie możemy się spodziewać, że występowanie takich emocji jak radość i gniew jest ze sobą negatywnie skorelowane. Tę zależność można uwzględnić korzystając z analizy głównych składowych. Na podstawie PCA otrzymano dwuwymiarową reprezentację mogącą wyjaśnić ponad 90% wariacji badanego zbioru. W tym sensie PCA przedstawia dwie osie znacznie różnicujące parlamentarzystów — wymiar PC1 (Dim1) to biegun emocji pozytywnych i negatywnych, natomiast PC2 (Dim2) różnicuje wewnątrz emocji negatywnych na osi smutek/strach — wstręt/gniew (w zbiorze danych te dwie pary emocji często występują wspólnie).

Wykres 1 określa pozycję parlamentarzysty w tych wymiarach. Bliskość do kierunku wyznaczonego przez określone strzałki wyznacza charakterystykę konfiguracji emocjonalnej reprezentowanej przez posła.



Rysunek 1: Koordynaty PCA

Finalnie, chmura siedemdziesięciu słów z najwyższą wartością wagi TF-IDF (rys. 2). Przesłaniem tej wizualizacji jest to, że pan poseł Konieczny relatywnie często używał takich słów jak *kryzys*, *odwaga*, *odejście*, *raj* w porównaniu do reszty parlamentarzystów. Reprezentacja na tym wykresie nie wyraża bezpośredniej częstości występowania danych słów.

Tabela 3: Charakterystyka PCA

Wskaźnik	Wartość	Procent
PC1 (pozytywne – negatywne)	2.2	87.6%
PC2 (smutek/strach – gniew/wstręt)	-0.9	11.0%

5. Dyskusja i podsumowanie

W obecnym stanie projekt stanowi raczej *proof of concept* zarówno od strony technicznej, jak i merytorycznej. Niemniej stanowi dobry punkt wyjścia do dalszego rozwoju i już teraz może nakierować na serię ciekawych pytań badawczych. W podsumowaniu chcę zaadresować pewne niedociągnięcia i luki obecnego podejścia, jak i potencjalne kierunki rozwoju narzędzia wraz z pytaniami badawczymi, jakie mogą temu towarzyszyć.



Rysunek 2: Wordcloud

5.1. Uwagi i ograniczenia

- Funkcje pełnione przez parlamentarzystów nie są uwzględnione. Nawet proste rytualne „proszę o ciszę” w kontekście marszałków zakrzywia wyniki w stronę pozytywnego wydzwieku.
- Miary różnorodności nie są skorygowane o całkowitą liczbę wypowiedzianych słów.
- Zaproponowane podejście jest ślepe na kontekst używanych słów. Na przykład wypowiedź „ta propozycja to żart” będzie uwzględniona jako wkład w pozytywny sentyment.
- Metoda słownikowa ma swoje inherentne ograniczenie związane z ograniczonym zasobem scharakteryzowanych słów. Część mocno nacechowanych emocjonalnie słów może padać w parlamencie, ale nie ma ich po prostu w zakresie wykorzystywanego słownika.

5.2. Kierunki rozwoju i przykładowe pytania badawcze

- Jak skutecznie można wyjaśnić profile emocjonalne za pomocą innych cech — przynależności klubowej, wieku, płci itd.?
- Utworzenie szerszego raportu uwzględniającego wyżej wymienione cechy.
- Jaka jest charakterystyka wyżej wymienionych miar w czasie — czy można zaobserwować długotrwałe trendy lub różnice występujące między kadencjami?
- Jak wyglądają relacje powiązań między parlamentarzystami — czy istnieją tendencje do podobieństw w kwestii używanego języka i jego nacechowania emocjonalnego?

Literatura

Erjavec T., Ogrodniczuk M., Osenova P., Ljubešić N., Simov K., Pančur A., Rudolf M., Kopp M., Barkarson S., Steingrímsson S., Çöltekin Ç., de Does J., Depuydt K., Agnoloni T., Venturi G., Pérez M. C., de Macedo L. D., Navarr*etta C., Luxardo G., Coole M., Rayson P., Morkevičius V., Krilavičius T., Dargis R., Ring O., van Heusden R., Marx M. and Fišer D. (2023). *The ParlaMint Corpora of Parliamentary Proceedings*. „Language Resources and Evaluation”, 58, s. 415–448.

Manning C. D., Raghavan P. and Schütze H. (2022). *An Introduction to Information Retrieval*. Cambridge University Press.

Riegel M., Wierzba M., Wypych M., Żurawski Ł., Jednoróg K., Grabowska A. and Marchewka A. (2015). *Nencki Affective Word List (NAWL): The cultural adaptation of the Berlin Affective Word List–Reloaded (bawl-R) for Polish*. „Behavior Research Methods”, 47(4), s. 1222–1236.

Wikipedia (2023a). *Diversity index — Wikipedia, the free encyclopedia*. <http://en.wikipedia.org/w/index.php?title=Diversity%20index&oldid=1178233805>. (online; data dostępu: 15 listopada 2023).

Wikipedia (2023b). *Hapax legomenon — Wikipedia, the free encyclopedia*. <http://en.wikipedia.org/w/index.php?title=Hapax%20legomenon&oldid=1177336166>. (online; data dostępu: 15 listopada 2023).

Wikipedia (2023c). *Principal component analysis — Wikipedia, the free encyclopedia*. <http://en.wikipedia.org/w/index.php?title=Principal%20component%20analysis&oldid=1184414408>. (online; data dostępu: 15 listopada 2023).

Analiza dynamiki emocji w debatach parlamentarnych o wojnie w Ukrainie

Konrad Kiljan, Ewelina Gajewska
(Uniwersytet Warszawski, Politechnika Warszawska)

Streszczenie

Artykuł skupia się na zastosowaniu modeli językowych do analizy emocji wyrażanych w debatach parlamentarnych na temat wojny w Ukrainie. Autorzy analizują reakcje emocjonalne, w szczególności złości, smutku, strachu, obrzydzenia i zaskoczenia, z wykorzystaniem modelu RoBERTa. W badaniu uwzględniają także różnice między krajami NATO z historią okupacji rosyjskiej a tymi bez niej. Wyniki wskazują na zróżnicowane reakcje emocjonalne, co może mieć istotne konsekwencje dla jedności i współpracy w obrębie sojuszu NATO.

Słowa kluczowe

emocje, debaty, parlament, stosunki międzynarodowe, Rosja, Ukraina, NATO

1. Wstęp

Wiele teorii stosunków międzynarodowych wskazywało badanie emocji jako zagadnienie niesłusznie pomijane w dotychczasowych badaniach empirycznych, a kluczowe dla rozumienia interakcji państwowych oraz poszukiwaniu odpowiedzi na globalne wyzwania (Crawford 2000, Hutchison i Bleiker 2014, Mercer 2014). Emocje, szczególnie strach przed zagrożeniami, mają istotny wpływ na efekty odstraszenia i współpracę międzynarodową (Stein 1988, Rousseau i Garcia-Retamero 2007, Garcia-Retamero i in. 2012). Jednakże tradycyjne metody badawcze miały trudności z efektywnym uwzględnieniem emocji, ich natężenia i dynamiki.

Aby zapełnić tę lukę, nasze badania wykorzystują modele językowe do przetwarzania języka i analizy debat parlamentarnych z różnych krajów europejskich. Dyskusje te dostarczają

wglądu w dyskurs publiczny i sentymenty narodowe, szczególnie w kontekście wojen i konfliktów międzynarodowych (Finlayson 2014, 2017, Rai i Reinelt 2014, Proksch i Slapin 2015).

Dotychczasowe badania wskazywały, że postrzeganie Rosji przez poszczególne kraje europejskie ewoluowało, kształtowane przez dynamikę polityczną, kulturalną i geopolityczną. Była ona odbierana na Zachodzie jako barbarzyńskie zagrożenie (Cain Jr 2006, McNally 1958), ale też źródło odnowienia cywilizacji (Rubenstein 2009, Judt 2011). Zimnowojenna propaganda wykorzystywała do kształtowania negatywnego wizerunku Rosji kulturę popularną (Diesen 2022), a po 1989 roku strach przed jej działaniami był niejednokrotnie krytykowany jako paranoiczny i nierealistyczny (Mettan 2017, Wheeler 1984). Kraje wschodnie, z historią dominacji rosyjskiej w XX, a często także XIX wieku, w niektórych jednak pojawiał się także pozytywny sentyment związany z nostalgią za współpracą wojskową czy kulturalną (Vovchenko 2023, Stanchev 2023). Stosunek Polski do Rosji pozostawał dominująco negatywny, a jej elity polityczne autoidentyfikowały się w ostatnich dekadach jako „Kasandra Zachodu” (Taras 2014). Termin „rusofobia” został skonstruowany przez rosyjską dyplomację w XIX wieku by wspierać rosyjski dyskurs imperialny i tłumić krytykę (Darczewska i Żochowski 2015), do czego jest wykorzystywany także i w dzisiejszych realiach (Lucas i Pomeranzev 2016). Badania zróżnicowanych dyskursów wobec Rosji ograniczone jednak były skromnymi objętościami analizowanych materiałów, co nie pozwalało na oszacowanie skali ich oddziaływania i porównania między krajami.

Stopniowa eskalacja agresji Rosji względem Ukrainy od roku 2014 sprawiła, że analiza jej postrzegania stała się sprawą jeszcze bardziej skomplikowaną i pilną. Powstawanie sprzecznych narracji wewnątrz NATO wpłynęło na kwestionowanie solidarności i ograniczenie zaufania wewnątrz bloku. (Deriglazova i Mäkinen 2019, Nitou 2016). Nasze badanie ma na celu wyodrębnienie różnic emocjonalnych w dyskusjach na temat rozpoczętej w 2022 wojny w krajach NATO z historią okupacji rosyjskiej oraz bez niej. Podzieliliśmy kraje NATO na dwie grupy: te z historią okupacji rosyjskiej (wschodnie) oraz te bez takiej historii (zachodnie), aby zbadać, czy różne emocje dominowały w dyskusjach dotyczących wojny w ich parlamentach w 2022 roku. Wykorzystując korpusy ParlaMint, analizujemy wzorce emocjonalne między blokami krajów, aby zrozumieć, czy historyczne doświadczenia i geograficzna lokalizacja wpływają na dyskusje dotyczące wojny i stosunków międzynarodowych. Nasze podejście ma na celu dostarczenie wniosków dotyczących zachowań państw, strategii dyplomatycznych i współpracy międzynarodowej, przyczyniając się do rozwoju teorii komunikacji politycznej oraz relacji międzynarodowych.

2. Metody

Analiza emocji dokonana została na angielskiej wersji korpusu ParlaMint, dostępnego do pobrania z repozytorium CLARIN (Erjavec i in. 2023)¹. Czasowa analiza emocji obejmuje okres

¹<https://www.clarin.si/repository/xmlui/handle/11356/1810>

od stycznia do końca maja 2022 roku, aby uwzględnić sentyment dyskusji parlamentarnych przed rozpoczęciem wojny 21 lutego oraz zmiany emocji w pierwszych miesiącach od wybuchu konfliktu. Wykorzystano debaty parlamentarne z 17 państw, które spełniły powyższe kryterium czasowe. Dokonaliśmy analizy emocji zarówno dla indywidualnych korpusów, jak i bloków państw europejskich. Do bloku państw Europy Zachodniej zaliczono: Austrię, Belgię, Danię, Katalonię, Galicję, Francję, Włochy, Holandię i Norwegię. Blok państw Europy Wschodniej obejmuje: Bośnię, Bułgarię, Czechy, Węgry, Litwę i Polskę. Oddzielnie przeanalizowaliśmy także korpus debat z parlamentu ukraińskiego. Do analiz wykorzystaliśmy przemówienia parlamentarne odnoszące się do tematu wojny na Ukrainie, tj. posiadających co najmniej jedno słowo kluczowe².

Analiza emocji dokonana została w blokach czasowych. Dla każdego bloku zmierzaliśmy rozkład emocji podstawowych według klasyfikacji Paula Ekmana (Ekman 1999) z wykorzystaniem modelu RoBERTa³: złości, smutku, strachu, obrzydzenia i zaskoczenia. W analizie skupiliśmy się zatem na emocjach negatywnych (złość, smutek, strach, obrzydzenie) lub ambiwalentnych (zaskoczenia). Odrzuciliśmy również przypadki zdań neutralnych pod względem wyrażanych emocji. Z wykorzystaniem kalkulatora prawdopodobieństwa (ang. *log-likelihood calculation*)⁴ dokonaliśmy również analizy słów kluczowych dla dominujących emocji wyrażanych przez polityków w parlamentach państw Europy Wschodniej i Zachodniej.

Okres od początku marca do końca maja 2022 roku podzieliliśmy na bloki dwutygodniowe (oznaczenia według konwencji YYYY-MM-H1 dla pierwszych dwóch tygodni każdego miesiąca i YYYY-MM-H2 dla drugiej części miesiąca. Debaty parlamentarne ze stycznia i lutego przed wybuchem konfliktu zagregowane zostały do jednego okresu czasowego (oznaczony jako 2022-02-24 na rys. 1). Wyodrębniliśmy również oddzielny okres czasowy dla tygodnia następującego po wybuchu wojny (oznaczony jako 2022-03-01 na rys. 1, tj. okres od 24 lutego do 1 marca).

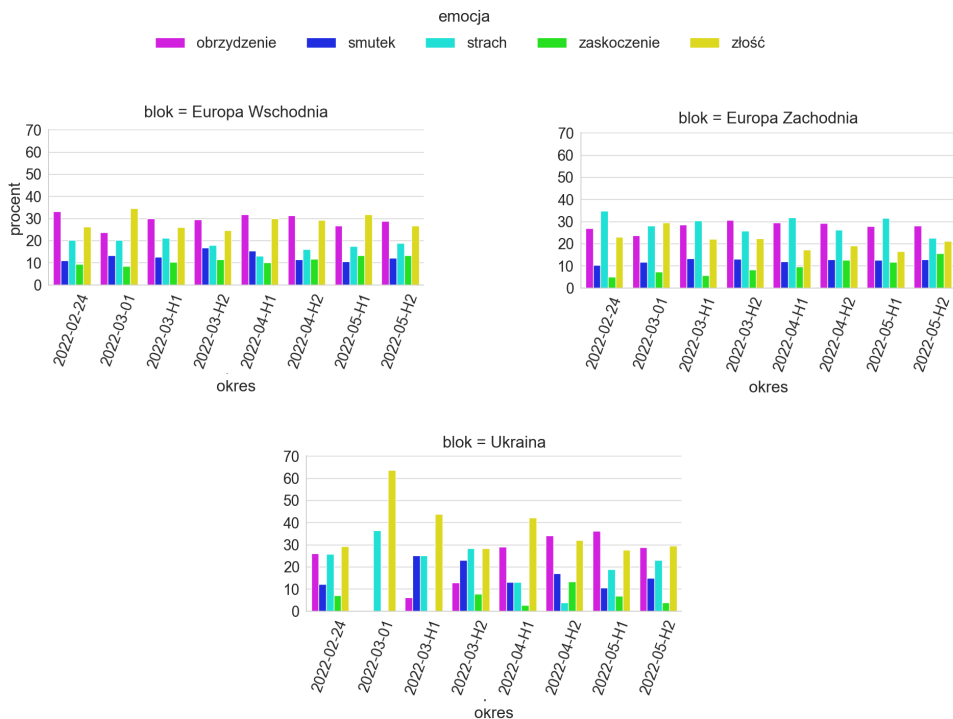
3. Wyniki

Wyniki analizy emocji wskazują na różny rozkład emocji dominujących debaty parlamentarne państw Europy Wschodniej i Zachodniej. Chociaż w obu blokach jednoznacznie dominują emocje negatywne, w pierwszym obserwujemy dominację emocji złości i obrzydzenia, których połączenie wskazuje na jest jako moralną pogardę wobec zachowań i czynów, pełniącą funkcję mobilizacyjną (Haidt i in. 2003, Miceli i Castelfranchi 2018). W drugim bloku główną rolę odgrywa strach i obrzydzenie, których wysoki poziom utrzymuje się przez cały okres czasowy analizy. Ponadto obserwujemy wzrost roli emocji złości w tygodniu następującym po rozpoczęciu wojny oraz jej spadek w kolejnych tygodniach.

²„Ukraine”, „Ukrainian”, „Russia”, „Russian”, „Putin”, „Russia's”, „Ukraine's”, „war”.

³<https://huggingface.co/j-hartmann/emotion-english-distilroberta-base>

⁴<https://ucrel.lancs.ac.uk/llwizard.html>



Rysunek 1: Rozkład emocji w przemówieniach parlamentarnych państw Europy Wschodniej, Zachodniej i Ukrainy na przestrzeni tygodni poprzedzających i następujących po wybuchu konfliktu zbrojnego na Ukrainie.

Rys. 2 przedstawia analizę słów kluczowe dla emocji dominujących w parlamentach Europy Wschodniej i Zachodniej. Oba bloki państw przedstawiają konflikt w Ukrainie jako zbrodnie, okrucieństwo, przemoc wobec ludzkości, złamanie zasad moralnych i humanitarnych. W krajach zachodnich jednak ujęcie to jest częściej abstrakcyjne, odnosi się do języka dyplomacji i sprawozdań, zaś w krajach wschodnich pojawia się więcej słów obrazujących cierpienie i wzywających do oporu.

4. Dyskusja

Niniejszy tekst poszerza ujęcia znane z tradycyjnej analizy dyskursu, wykorzystując nowoczesne metody językoznawstwa obliczeniowego, aby odkryć nowe spojrzenie na stosunki międzynarodowe i studia nad komunikacją. Poprzez analizę debat parlamentarnych, ujawnia różnice emocjonalne w reakcjach na wojnę Rosja-Ukraina między krajami NATO z i bez historii dominacji Rosji w XX wieku.



Rysunek 2: Słowa kluczowe dla emocji dominujących w dyskursie parlamentarnym państw Europy Zachodniej i Wschodniej.

Badanie udowadnia, że w parlamentach krajów z historią wpływu Rosji reakcje emocjonalne były bardziej zbliżone do tych obserwowanych w Ukrainie. Chociaż zarówno w kraje wschodnich, jak i zachodnich wyrażano silne negatywne emocje, pojawiają się istotne różnice. Na Zachodzie Europy częściej wyrażano obrzydzenie i strach, emocje charakterystyczne dla pragnienia ucieczki w stronę bezpieczeństwa. W krajach wschodnich dominowały obrzydzenie i złość, odzwierciedlając reakcję mobilizacyjną na nieuniknione zagrożenie. Różnice emocjonalne znajdują odzwierciedlenie lingwistyczne w abstrakcyjnym wyborze słów na Zachodzie oraz w ilustratywnym języku na Wschodzie.

Automatyczna analiza językowa pozwala na przetworzenie dużych zbiorów tekstów, jednak także niesie ze sobą ograniczenia dotyczące wiarygodności i doboru próby. Mimo że rozpoznawanie emocji jest rzetelne w stopniu przewyższającym większość analiz manualnych, może zawierać potencjalne błędy klasyfikacji. Ograniczona dostępność danych z różnych krajów europejskich wpływa na dokładność wniosków. Skupienie badania na debatach parlamentarnych nie w pełni uzupełnia szerszej analizy dyskursów medialnych i opinii publicznej. Ponadto koncentracja na segmentach dotyczących wojny może ograniczyć kompleksową analizę wszystkich związanych dyskusji w parlamentach, sugerując potencjalne ulepszenia poprzez zastosowanie metod takich jak klasyfikacja na bazie uczenia się bez nadzoru przez systemy oparte na BERT, w celu szerszego wykrywania tematów.

Zróznicowanie reakcji emocjonalnych wśród krajów NATO to czynnik istotny dla jedności i współpracy wewnątrz bloku. Zrozumienie tych niuansów ma kluczowe znaczenie dla zawie-

rania i przedstawiania obywatelom sojuszy w obliczu wyzwań geopolitycznych. Nasz artykuł wytacza nowe kierunki analizy dyskursu w relacjach międzynarodowych wspieranej przy wsparciu modeli sztucznej inteligencji, oferując cenne spojrzenie dla decydentów, zwłaszcza w kontekście kryzysów wymagających nowych sposobów budowania solidarności NATO.

Literatura

Cain Jr J. E. (2006). *Bram Stoker and Russophobia: Evidence of the British Fear of Russia in Dracula and The Lady of the Shroud*. McFarland.

Crawford N. C. (2000). *The passion of world politics: Propositions on emotion and emotional relationships*. „International Security”, 24(4), s. 116–156.

Darczewska J. i Żochowski P. (2015). *Russophobia in the Kremlin's Strategy. A Weapon of Mass Destruction*. Number 56 in Point of View. Ośrodek Studiów Wschodnich.

Deriglazova L. i Mäkinen S. (2019). *Still looking for a partnership? EU-Russia cooperation in the field of higher education*. „Journal of Contemporary European Studies”, 27(2), s. 184–195.

Diesen G. (2022). *The Foundational Stereotypes of Anti-Russian Propaganda*. In *Russophobia: Propaganda in International Politics*, s. 45–82. Springer.

Erjavec T., Ogrodniczuk M., Osenova P., Ljubešić N., Simov K., Pančur A., Rudolf M., Kopp M., Barkarson S., Steingrímsson S., Çöltekin Ç., de Does J., Depuydt K., Agnoloni T., Venturi G., Pérez M. C., de Macedo L. D., Navarretta C., Luxardo G., Coole M., Rayson P., Morkevičius V., Krilavičius T., Dargis R., Ring O., van Heusden R., Marx M. i Fišer D. (2023). *The ParlaMint Corpora of Parliamentary Proceedings*. „Language Resources and Evaluation”, 58, s. 415–448.

Finlayson L. (2014). *How to screw things with words*. „Hypatia”, 29(4), s. 774–789.

Finlayson L. (2017). *With radicals like these, who needs conservatives? Doom, gloom, and realism in political theory*. „European Journal of Political Theory”, 16(3), s. 264–282.

Garcia-Retamero R., Müller S. M. i Rousseau D. L. (2012). *The impact of value similarity and power on the perception of threat*. „Political Psychology”, 33(2), s. 179–193.

Haidt J. i in. (2003). *The moral emotions*. „Handbook of affective sciences”, 11(2003), s. 852–870.

Hutchison E. i Bleiker R. (2014). *Theorizing emotions in world politics*. „International theory”, 6(3), s. 491–514.

Judt T. (2011). *The Sacrifices of the Russian People: A Phenomenology of Intellectual Russophilia*. In *Past Imperfect*, s. 153–167. New York University Press.

Lucas E. i Pomeranzev P. (2016). *Winning the information war*. „Techniques and Counter-strategies to Russian Propaganda in Central and Eastern Europe”, s. 1–66.

- McNally R. T. (1958). *The origins of russophobia in France: 1812-1830*. „American Slavic and East European Review”, 17(2), s. 173–189.
- Mercer J. (2014). *Feeling like a state: Social emotion and identity*. „International Theory”, 6(3), s. 515–535.
- Mettan G. (2017). *Creating Russophobia: From the great religious schism to anti-Putin hysteria*. SCB Distributors.
- Miceli M. i Castelfranchi C. (2018). *Contempt and disgust: The emotions of disrespect*. „Journal for the Theory of Social Behaviour”, 48(2), s. 205–229.
- Nitoiu C. (2016). *Towards conflict or cooperation? The Ukraine crisis and EU-Russia relations*. „Southeast European and Black Sea Studies”, 16(3), s. 375–390.
- Proksch S.-O. i Slapin J. B. (2015). *The politics of parliamentary debate*. Cambridge University Press.
- Rai S. M. i Reinelt J. (2014). *The Grammar of Performance and Politics*. Routledge.
- Rousseau D. L. i Garcia-Retamero R. (2007). *Identity, power, and threat perception: A cross-national experimental study*. „Journal of Conflict Resolution”, 51(5), s. 744–771.
- Rubenstein R. (2009). *Virginia Woolf and the Russian point of view*. Springer.
- Stanchev E. (2023). *The growing Russophilia of post-communist Bulgarian nationalism: between entanglements and paradoxes: La creciente rusofilia del nacionalismo búlgaro poscomunista: entre enredos y paradojas*. „Araucaria”, 25(53).
- Stein J. G. (1988). *Building politics into psychology: The misperception of threat*. „Political psychology”, s. 245–271.
- Taras R. (2014). *Russia resurgent, Russophobia in Decline? Polish perceptions of relations with the Russian Federation 2004–2012*. „Europe-Asia Studies”, 66(5), s. 710–734.
- Vovchenko D. (2023). *From Ethnic Hatred to Religious Solidarity: The Rise and Fall of Russophobia in Greece, 1830–1924*. „European History Quarterly”, 53(3), s. 495–519.
- Wheeler G. (1984). *Russophobia in the western world: A brief case history*. „Asian Affairs”, 15(2), s. 138–143.

Język wrażliwy na płeć (*gender-fair language*) w debatach parlamentarnych w korpusie ParlaMint — badanie pilotażowe

Aleksandra Tomaszewska (Instytut Podstaw Informatyki PAN)

Anna Jamka (Instytut Lingwistyki Stosowanej, Uniwersytet Warszawski)

Streszczenie

W tekście przedstawiono wyniki badania pilotażowego, którego celem było zdobycie wiedzy na temat użycia języka wrażliwego na płeć (ang. *gender-fair language*) w polskim dyskursie parlamentarnym. Zbadano strategie językowe neutralizacji i feminizacji w korpusie *ParlaMint* 4.0, który składa się ze stenogramów debat z Sejmu i Senatu z lat 2015–2022. Badanie dostarcza wstępnych danych na temat rzadko badany korpusowo, a budzący duże zainteresowanie społeczne.

Słowa kluczowe

język wrażliwy na płeć, dyskurs parlamentarny, ParlaMint, badania korpusowe

1. Wstęp

Język wrażliwy na płeć (ang. *gender-fair language*; Chmura-Rutkowska i Szpyra-Kozłowska 2023) to podejście lingwistyczne, które ma na celu redukcję stereotypów płciowych i dyskryminacji (Lindqvist i in. 2018). Dążeniu do tego sprzyjają strategie feminizacji, czyli tworzenia i używania form żeńskich (tzw. feminatywów) oraz neutralizacji — wyboru słów, które mogą nazywać kobiety i mężczyzn (Szpyra-Kozłowska 2019: 23). Do tej pory nie powstały zalecenia instytucjonalne dotyczące użycia strategii języka wrażliwego na płeć w polskim parlamencie. W badaniu jako punkt odniesienia wykorzystujemy zatem rekomendacje opublikowane przez Parlament Europejski (2018). Mimo że dotyczą one unijnej komunikacji parlamen-

tarnej, są jedynymi zaleceniami uwzględniającymi specyfikę języka polskiego w kontekście parlamentarnym.

Parlament Europejski zaleca dostosowanie strategii języka wrażliwego na płeć do specyfiki danego języka (inne są rekomendowane strategie w przypadku języków, w których występuje rodzaj naturalny niż w przypadku tych, w których występuje rodzaj gramatyczny czy też języków, w których nie występuje rodzaj) (Parlament Europejski 2018). Najistotniejsze zalecenia dla języka polskiego to feminizacja w odniesieniu do konkretnej kobiety lub grupy kobiet oraz użycie *splittingu*, czyli formy podwójnej, w bezpośrednich zwrotach grzecznościowych — a także neutralizacja, czyli używanie określeń neutralnych w odniesieniu do grup mieszanych lub osoby, której płeć nie jest znana (Parlament Europejski 2018).

2. Metoda i materiały

Związany z językiem wrażliwym na płeć temat form żeńskich podejmowano niejednokrotnie (zob. m.in. Łaziński 2006, Karwatowska i Szpyra 2005, Szpyra-Kozłowska 2023). Mimo popularności tej kwestii w polskich pracach i różnorodności opisywanych zagadnień¹, a także jej obecności w debacie publicznej, użycie języka wrażliwego na płeć nie było do tej pory badane w polskim dyskursie parlamentarnym. Wyniki tego pilotażowego badania korpusowego dostarczają zatem informacji ilościowych i jakościowych na temat istotny społecznie, a rzadko podejmowany w literaturze. Stanowią one również wstęp do dalszych badań — polskojęzycznych, ale również wielojęzycznych — nad językiem wrażliwym na płeć w komunikacji parlamentarnej.

2.1. Cel badania

Celem badania jest dostarczenie nowej wiedzy na temat użycia języka wrażliwego na płeć w polskim dyskursie parlamentarnym. Wykorzystując korpus *ParlaMint* 4.0 z lat 2015–2022 oraz aplikację NoSketchEngine, sprawdzono, czy strategie językowe neutralizacji i feminizacji używane są w debatach w polskim parlamencie oraz jakie są tendencje związane z ich użyciem w badanym okresie. W tym celu przeanalizowano pilotażowo rzeczowniki osobowe (nazwy profesji, funkcji, tytułów zawodowych oraz rzeczowniki desygnujące przynależność do wspólnoty), a także częstość ich występowania i konteksty ich użycia w dyskursie parlamentarnym.

¹Opisywano na przykład historię form żeńskich i ich poprawność językową (Woźniak 2014, Grochowska i Wierzbicka 2015), uwarunkowania społeczne powstawania feminatywów i ich recepcję w Polsce (zob. Łaziński 2006, Małocha-Krupa 2018, Szpyra-Kozłowska 2021, Szpyra-Kozłowska 2019, 2023).

2.2. Korpus ParlaMint

ParlaMint to opracowany w ramach konsorcjum CLARIN zbiór porównywalnych korpusów obejmujących transkrypcje debat parlamentarnych z 29 krajów europejskich (w tym Polski) oraz regionów autonomicznych. W tym badaniu korzystamy z jego najnowszej dostępnej obecnie wersji (4.0), która obejmuje okres od roku 2015 aż do połowy 2022 r. (Erjavec i in. 2023). Wielkość poszczególnych podkorpusów w obrębie korpusu *ParlaMint* waha się od 9 do 126 milionów słów, a całość zbioru zawiera ponad 1,1 miliarda słów. Dane *ParlaMint* dla języka polskiego zaczerpnięto z Korpusu Dyskursu Parlamentarnego (Ogrodniczuk 2018). Zawierają one transkrypcje z obrad obu izb Parlamentu RP: Sejmu oraz Senatu. Polski podkorpus *ParlaMint* to jedyny taki zasób w Polsce — szczególnie ze względu na jego spójny i porównywalny charakter, a także bogactwo metadanych i anotacji.

Z danych frekwencyjnych z list słów w *NoSketchEngine* wynika, że 68,1% polskiego podkorpusu to wypowiedzi z Sejmu, a nieco ponad 31,9% z Senatu; ponad 89% to wypowiedzi posłanek i posłów czy senatorek i senatorów; reszta to natomiast wypowiedzi innych osób, które występowały podczas posiedzeń izb. W polskim podkorpusie *ParlaMint* 4.0 zaobserwowano istotne różnice w liczbie słów przypisanych do płci. Stosunek wypowiedzi mężczyzn do wypowiedzi kobiet wynosi 70% do 30%. Ta dysproporcja sugeruje, że głos mężczyzn jest bardziej reprezentowany w debatach parlamentarnych w Polsce. Obserwuje się również zróżnicowanie w zakresie frekwencji słów w zależności od wieku mówcy. Największą frekwencję obserwuje się w grupie osób urodzonych w latach 50. i 60. XX wieku. Wynik ten jest już mniejszy w przypadku osób urodzonych w latach 70. i 80. XX wieku. Natomiast osoby urodzone w latach 90. — najmłodsze wśród wypowiadających się w korpusie — są w nim najsłabiej reprezentowane. Osoby urodzone przed rokiem 1940 mają również ograniczoną reprezentację w korpusie. Dominującą grupą wiekową w polskim podkorpusie są zatem osoby urodzone w latach 50. i 60. XX wieku, czyli obecnie w przedziale wiekowym 54–73 lata. Wiek może odzwierciedlać pewne specyficzne dla tych pokoleń podejście do języka (zob. badania Szpyry-Kozłowskiej o preferencjach językowych młodego pokolenia (Szpyra-Kozłowska 2019). Aby jednak uzyskać pełniejszy obraz wpływu wieku mówców na specyfikę językową w parlamencie, konieczne byłyby bardziej pogłębione analizy.

2.3. Procedura badawcza

Badanie podzielono na kilka etapów:

1. ustalono kategorie i podkategorie do analizy;
2. w oparciu o rekomendacje Parlamentu Europejskiego wytypowano jednostki leksykalne do badania (uwzględniono również leksemy istotne z perspektywy kontekstu socjopolitycznego);
3. na potrzeby badania pilotażowego sprawdzono, jak często wybrane jednostki z grupy wytypowanych leksemów pojawiają się w korpusie;

4. przeprowadzono analizę testową dotyczącą spodziewanych tendencji w użyciu tych leksemów.

Wyodrębnione w kroku 2. pary leksemów podzielono na następujące kategorie:

A. Zwyczajowe użycie rodzaju męskiego:

- (a) nazwy zawodów istotne ze względu na występowanie w debacie publicznej oraz kontekst społeczno-polityczny (np. *lekarz–lekarka*),
- (b) rzeczowniki desygnujące przynależność do wspólnoty (narodowej/zawodowej/ideologicznej, np. *Polak–Polka*).

B. Nazwy związane z pełnieniem funkcji publicznych:

- (a) wymienione w rekomendacjach Parlamentu Europejskiego (np. *poseł–posłanka*),
- (b) związane ze specyfiką Sejmu i Senatu RP (np. *marszałek–marszałkini*).

Następnie za pomocą funkcji *Wordlist* sprawdzono frekwencję wytypowanych rzeczowników testowych. W dalszej kolejności przeanalizowano konkordancje i kolokacje, w jakich występują wybrane leksemy.

3. Wyniki

Wyniki badania pilotażowego wskazują na przewagę strategii neutralizacji nad feminizacją, co potwierdzają poniższe przykłady wyników dla każdej z badanych kategorii:

1. nazwy zawodów istotne ze względu na występowanie w debacie publicznej oraz kontekst społeczno-polityczny: *lekarz* (*lekarz* — 10 202, *lekarka* — 66);
2. rzeczowniki desygnujące przynależność do wspólnoty (narodowej/zawodowej/ideologicznej): *Polak* (*Polak* — 28 003, *Polka* — 1548);
3. nazwy związane z pełnieniem funkcji publicznych: leksemy wymienione w rekomendacjach Parlamentu Europejskiego: *poseł* (*poseł* — 200 955, *posłanka* — 3581); leksemy związane ze specyfiką Sejmu i Senatu RP: *marszałek* (*marszałek* — 117 729, *marszałkini* — 227).

Wskazano na możliwe tendencje związane z użyciem języka wrażliwego na płeć w polskim dyskursie parlamentarnym:

1. użycie formy męskiej nie jest uzależnione od stopnia maskulinizacji bądź feminizacji danego zawodu (np. zgodnie z danymi GUS w Polsce jest więcej lekarek od lekarzy (GUS 2023);

2. formy żeńskiej w przypadku nazw zawodów używa się najczęściej w odniesieniu do konkretnej kobiety (w korpusie jest to 47% wyników, czyli 31 wystąpień leksemu *lekarka*) oraz w *splittingu* (27%, czyli 18 wystąpień leksemu *lekarka* to przykłady *splittingu*);
3. forma męska w przeważającej większości (92% wyników z losowej próby, czyli 183 na 200 wystąpień leksemu *lekarz*) określa prototypowego przedstawiciela całej grupy zawodowej lub jej podgrup, np. *lekarz rodzinny*;
4. użycie formy męskiej, tj. generycznej, w strategii neutralizacji może być podyktowane zasadą ekonomii języka, czyli dążeniem do najbardziej zwartej formy wypowiedzi, na co wskazuje np. użycie formy męskiej *Polak* w odniesieniu do ogółu społeczeństwa (analiza konkordancji leksemu *Polak*, próba losowa, 200 konkordancji, bez *Polka*, *Polki*);
5. na wybór formy męskiej w strategii neutralizacji może mieć również wpływ postrzegany prestiż funkcji oraz forma rodzajowa stosowana w aktach prawnych (np. *pani poseł* — 41 509, *pani posłanka* — 2180; *pani marszałek* — 11 975, *pani marszałkini* — 24);
6. formy żeńskie nazw bardzo wysokich stanowisk i/lub urzędów mają bardzo niską częstość w korpusie (np. *marszałkini* — 227, *marszałek* — 117 729). Wskazuje to, że brak feminitywów w legislacji i na stronach oficjalnych parlamentu, czynniki pragmatyczno-semantyczne oraz indywidualne preferencje mówiących niejako hamują stosowanie form żeńskich w debatach parlamentarnych.

4. Wnioski

W raporcie zaprezentowano wyniki badania pilotażowego, którego celem było dostarczenie wiedzy na temat użycia języka wrażliwego na płeć w polskim dyskursie parlamentarnym. Wyniki analiz, do przeprowadzenia których wykorzystano korpus *ParlaMint*, wskazują, że w polskim dyskursie parlamentarnym występują strategie właściwe dla języka wrażliwego na płeć.

W korpusie przeważa neutralizacja rozumiana jako rezygnacja z rzeczowników żeńskich na rzecz określeń, które odnoszą się do obu płci. Strategia feminizacji natomiast, polegająca na tworzeniu derywatów żeńskich od rzeczowników męskich, występuje znacznie rzadziej. Wyniki wskazują na pewne tendencje związane z występowaniem tych strategii w dyskursie parlamentarnym. Przewaga formy męskiej w odniesieniu do grup różnopłciowych i prototypowych przedstawicieli danej grupy w polskim dyskursie parlamentarnym (neutralizacja) zdaje się być uwarunkowana kilkoma czynnikami. Po pierwsze funkcją generyczną rzeczowników męskich, która pozwala na tworzenie krótszych wypowiedzi w opozycji do *splittingu*, czyli równoległego stosowania form żeńskiej i męskiej, a po drugie formami rodzajowymi występującymi w aktach prawnych. W korpusie wskaźnik feminizacji zawodu nie ma natomiast większego wpływu na wybór formy rodzajowej. Formy żeńskie najczęściej stosowane są w odniesieniu do konkretnej kobiety lub grupy kobiet oraz w strategii *splittingu*.

Mimo istnienia strategii językowych i możliwości słowotwórczych polszczyzny, używanie języka podkreślającego obecność kobiet w polskim dyskursie parlamentarnym jest ograniczone. Wyniki wskazują na potrzebę dalszych badań dotyczących języka wrażliwego na płęć zarówno w sferze instytucjonalnej, jak i w szerszym kontekście społecznym. Należy podkreślić, że — ze względu na pilotażowy charakter badania — przeanalizowano jedynie jednostki testowe w ramach wybranych grup leksemów. W trakcie są dalsze analizy, uwzględniające większą liczbę jednostek, które pozwolą zdobyć więcej informacji na temat zidentyfikowanych tendencji oraz innych możliwych trendów w użyciu języka wrażliwego na płęć w polskim dyskursie parlamentarnym.

Literatura

Chmura-Rutkowska I. i Szpyra-Kozłowska J. (2023). *Nierówność płci w języku: Poradnik dla nauczycielek i nauczycieli*. Urząd m.st. Warszawy.

Erjavec T., Kopp M., Ogrodniczuk M., Osenova P., Agirrezabal M., Agnoloni T., Aires J., Albini M., Alkorta J., Antiba-Cartazo I., Arrieta E., Barcala M., Bardanca D., Barkarson S., Bartolini R., Battistoni R., Bel N., Bonet Ramos M. d. M., Calzada Pérez M., Cardoso A., Çöltekin Ç., Coole M., Dargis R., de Libano R., Depoorter G., Diwersy S., Dodé R., Fernandez K., Fernández Rei E., Frontini F., Garcia M., García Díaz N., García Louzao P., Gavriilidou M., Gkoumas D., Grigorov I., Grigorova V., Haltrup Hansen D., Iruskieta M., Jarlbrink J., Jelencsik-Mátyus K., Jongejan B., Kahusk N., Kirnbauer M., Kryvenko A., Ligeti-Nagy N., Ljubešić N., Luxardo G., Magariños C., Magnusson M., Marchetti C., Marx M., Meden K., Mendes A., Mochtak M., Mölder M., Montemagni S., Navarretta C., Nitoń B., Norén F. M., Nwadukwe A., Ojsteršek M., Pančur A., Papavassiliou V., Pereira R., Pérez Lago M., Piperidis S., Pirker H., Pisani M., Pol H. v. d., Prokopidis P., Quochi V., Rayson P., Regueira X. L., Rudolf M., Ruisi M., Rupnik P., Schopper D., Simov K., Sinikallio L., Skubic J., Tungland L. M., Tuominen J., van Heusden R., Varga Z., Vázquez Abuín M., Venturi G., Vidal Miguéns A., Vider K., Vivel Couso A., Vladu A. I., Wissik T., Yrjänäinen V., Zevallos R. i Fišer D. (2023). *Multilingual comparable corpora of parliamentary debates ParlaMint 4.0*. Slovenian language resource repository CLARIN.SI: <http://hdl.handle.net/11356/1859>.

Grochowska M. i Wierzbicka A. (2015). *Produktywne typy słowotwórcze nazw żeńskich we współczesnej polszczyźnie*. „Acta Universitatis Lodziensis. Folia Linguistica”, (049).

GUS (2023). *Zasoby kadrowe w wybranych zawodach medycznych na podstawie źródeł administracyjnych w latach 2019–2020*. Dostęp: 15.11.2023.

Karwatowska M. i Szpyra J. (2005). *Lingwistyka płci: Ona i on w języku polskim*. Lublin: UMCS.

Łaziński M. (2006). *O panach i paniach: Polskie rzeczowniki tytułowe i ich asymetria rodzajowo-płciowa*. Warszawa: PWN.

- Lindqvist A., Renström E. A. i Gustafsson Sendén M. (2018). *Reducing a male bias in language? Establishing the efficiency of three different gender-fair language strategies*. „Sex Roles”, 81(1–2), s. 109–117.
- Małocha-Krupa A. (2018). *Feminytywum w uwikłaniach językowo-kulturowych*. Oficyna Wydawnicza ATUT-Wrocławskie Wydawnictwo Oświatowe.
- Ogrodniczuk M. (2018). *Polish Parliamentary Corpus*. In *Proceedings of the LREC 2018 Workshop ParlaCLARIN: Creating and Using Parliamentary Corpora*, s. 15–19. European Language Resources Association.
- Parlament Europejski (2018). *Język neutralny płciowo w Parlamencie Europejskim*.
- Szpyra-Kozłowska J. (2019). „Premiera”, „premierka” czy „pani premier”? *Nowe feminytywy w badaniu ankietowym*. „Język Polski”, (2), s. 22–40.
- Szpyra-Kozłowska J. (2023). *Pielęgniarka ery cyfrowej — feminytywy w funkcji generycznej*. „Język Polski”, (1), s. 71–85.
- Szpyra-Kozłowska J. (2021). *Nianiek, ministra i japonki. Eseje o języku i płci*. Universitas, Kraków.
- Woźniak E. (2014). *Język a emancypacja, feminizm, gender*. „Rozprawy Komisji Językowej”, (60), s. 295–312.

Zagraniczne orzeczenia sądów konstytucyjnych w debacie parlamentarnej parlamentów państw członkowskich Unii Europejskiej

Mikołaj Wolanin
(Uniwersytet Warszawski)

Streszczenie

Artykuł omawia wpływ orzeczeń sądów konstytucyjnych państw członkowskich Unii Europejskiej na debatę parlamentarną. Autor analizuje, jak decyzje tych sądów oddziałują na krajowe porządki prawne, szczególnie w kontekście globalizacji i quasi-federalizacji w UE. Badanie koncentruje się na odwoływaniu się do orzeczeń konkretnych sądów konstytucyjnych w debatach parlamentarnych innych państw członkowskich. Przykładowo, analizuje, czy polscy parlamentarzyści odwoływali się do orzeczeń francuskiej Rady Konstytucyjnej czy niemieckiego Federalnego Trybunału Konstytucyjnego. Autor sugeruje, że taka wymiana odniesień do orzeczeń sądów konstytucyjnych może wpływać na kształtowanie się prawa w poszczególnych państwach. Wyniki badań wstępnych obejmują cytaty polityków, którzy wypowiadają się na temat dialogu między sądami konstytucyjnymi, kompetencji Unii Europejskiej, oraz kwestii związanych z prawem międzynarodowym i społecznymi.

Słowa kluczowe

sądy konstytucyjne, debata parlamentarna, wpływ orzeczeń międzynarodowych

1. Wstęp

Bezspornie ostatnie dziesięciolecie wyznaczają trend globalizacyjny w szerokim spektrum zagadnień życia społecznego. Nie inaczej jest w przypadku nauk prawnych. Systemy prawne czerpią z siebie i inspirują się sobą wzajemnie. Szczególnie widać to w zakresie prawa cywilnego. Przykładowo, polski Kodeks cywilny czerpie bardzo dużo koncepcji z niemieckiego

Bürgerliches Gesetzbuch uchwalonego w 1896 roku i nadal obowiązującego na terenie Republiki Federalnej Niemiec. Podobnie jednak, część instytucji prawnych obecnych także w Polsce wywodzi się z francuskiego Kodeksu Napoleona z 1804 roku. Przykłady można by zresztą mnożyć.

W przypadku państw członkowskich Unii Europejskiej zagadnienie to jest nawet mocniejsze. Państwa te współpracują bowiem w coraz większej liczbie spraw (zob. art. 3–6 Traktatu o funkcjonowaniu Unii Europejskiej¹). Mają ponadto ogólnounijne sądy — Trybunał Sprawiedliwości sensu stricto i Sąd tworzące Trybunał Sprawiedliwości Unii Europejskiej sensu largo. Wszystkie państwa członkowskie UE są ponadto członkami Rady Europy, w obrębie której działają Europejski Trybunał Praw Człowieka oraz Europejska Komisja na rzecz Demokracji przez Praw (tzw. Komisja Wenecka; ta jednak nie jest sądem w rozumieniu art. 6 Europejskiej Konwencji Praw Człowieka; to raczej organ doradczy, stanowiący o dobrych praktykach).

Oczywisty jest zatem wniosek, że orzeczenia wydawane przez te międzynarodowe sądy mają wpływ na krajowe porządki prawne, które są zbliżane do siebie. Przykładowo, orzeczenie Trybunału Sprawiedliwości Unii Europejskiej w sprawie Xero Flor przeciwko Polsce nie będzie dotyczyło przecież tylko sytuacji prawnej i powołań sędziów polskich, ale wypracowane tam standardy będą miały wpływ na to, jak sędziów powołuje się we wszystkich państwach Unii Europejskiej (szczególnie że są oni równocześnie sędziami samej Unii). Tak samo, wyrok Europejskiego Trybunału Praw Człowieka w sprawie Lautsi i in. przeciwko Włochom wyznacza standardy obecności symboli religijnych w przestrzeni publicznej nie tylko tam, ale we wszystkich państwach, które ratyfikowały Europejską Konwencję Praw Człowieka (co jest warunkiem *sine qua non* należenia do Rady Europy).

Znacznie ciekawsze badawczo jest jednak badanie nieco innego zagadnienia. Bo tak, jak zagadnienie przenikania do krajowych systemów prawnych orzeczeń sądów międzynarodowych jest oczywiste, tak interesujące i dotychczas mniej badane zdaje się analizowanie tego, jak orzeczenia poszczególnych sądów krajowych wpływają — w dobie globalizacji i nawet częściowej quasi-federalizacji w zakresie Unii Europejskiej — na systemy prawne pozostałych państw — w tym przykładzie członkowskich Unii Europejskiej. Czy jest bowiem tak, że państwa wzajemnie spoglądają na orzeczenia wydawane przez, przede wszystkim, najwyższe sądy pozostałych państw? Zdaje się, że tak.

Autor w ramach opracowania konkursowego przeprowadzi jednak badania węższe. Przedmiotem jego zainteresowania będą bowiem sądy konstytucyjne, czyli naczelne sądy krajowe badające zgodność prawa z aktami wyższego rzędu (w Polsce Trybunał Konstytucyjny). To orzeczenia polskiego Trybunału Konstytucyjnego i jego odpowiedników w państwach członkowskich Unii Europejskiej będą zatem analizowane. Analiza zostanie natomiast przeprowadzona w oparciu o korpus ParlaMint (Erjavec i in. 2023). Z tego względu dokładnym przedmiotem badania będzie debata parlamentarna w parlamentach państw członkowskich

¹Z uwagi na charakter dokumentu — zgłoszenie konkursowe o charakterze konceptualnym, autor celowo pomija m.in. wskazywanie adresów publikacyjnych we właściwych dziennikach urzędowych. W razie realizacji projektu adresy te będą umieszczone w artykule.

Unii Europejskiej (rzecz jasna, tych uwzględnionych w ParlaMint), w trakcie której takie odniesienia miały miejsce.

Celem przykładu należy wskazać, że badanie będzie polegało na sprawdzeniu, czy w ramach obrad polskiego Sejmu czy Senatu ktoś odwoływał się do orzeczenia np. francuskiej Rady Konstytucyjnej czy niemieckiego Federalnego Trybunału Konstytucyjnego, albo czy w trakcie debaty w litewskim Seimasie dany poseł czy dana posłanka odwoływał/a się do orzeczenia polskiego Trybunału Konstytucyjnego. Zanim badanie zostanie przeprowadzone, trudno jest wskazywać, na ile to zjawisko jest częste i jakie wnioski z takich debat płyną. Roboczą hipotezą autora jest jednak to, że przede wszystkim parlamenty innych państw członkowskich Unii Europejskiej będą w swoich debatach odnosiły się do orzeczeń prawnie lub społecznie doniosłych (tzw. *landmark judgments*), które potencjalnie mogą dotyczyć ich samych, albo takich, które są istotne z punktu widzenia międzynarodowego. W pierwszym przypadku mowa zatem m.in. o krajowych orzeczeniach dotyczących np. relacji prawa krajowego do unijnego czy — szczególnie wśród państw federacyjnych — zagadnień związanych z potencjalną secesją danego regionu. Drugi przypadek to zaś sprawy wzbudzające międzynarodowe zainteresowanie, dotyczące szeregu zagadnień, w tym np. relacji międzynarodowych, a dla jednostek aborcji czy eutanazji.

Jednocześnie trzeba wskazać, że autor badał będzie wyłącznie orzeczenia sądu konstytucyjnego. Jakiegokolwiek inne sądy krajowe, nawet najwyższe lub naczelne, w zakres tego badania nie będą wchodziły. Uzasadnieniem dla takiego podejścia jest specyfika sądu konstytucyjnego, który czasem nazywany jest tzw. negatywnym ustawodawcą. Nie orzeka on bowiem, co do zasady, w sferze faktów, ale litery prawa. Analiza będzie w sobie łączyła zagadnienia prawne i politologiczne. Tylko taka interdyscyplinarność pozwoli autorowi dojść do wniosków w zakresie celowości opisywanego działania oraz tego, jakie funkcje pełni taka debata — zarówno dla państwa, w którym debata się toczy, jak i dla pozostałych, w szczególności dla tego, w którym orzeczenie zostało wydane.

Opracowanie ma zatem odpowiedzieć na pytania nie tylko dotyczące częstotliwości oraz treści takich debat, ale także na zagadnienia bardziej generalne: dlaczego taka debata się odbywa? Dlaczego przykład zagraniczny jest podawany? Jaki wpływ taki zagraniczny przykład ma na krajowy porządek prawny? Czy krajowy prawodawca poszukuje innego rozwiązania, gdy zagraniczny przykład pokazuje niezgodność jakiegoś działania np. z prawem wspólnotowym? I odwrotnie, czy krajowy prawodawca inspirował się zagranicznymi rozwiązaniami zawartymi w orzeczeniach sądów konstytucyjnych? Celem naukowym artykułu będzie uzyskanie odpowiedzi na te pytania.

Sama analiza zostanie natomiast przeprowadzona w języku angielskim. Dzięki zawężeniu tematu tylko do sądów konstytucyjnych, możliwe będzie przebadanie wszystkich parlamentów państw członkowskich Unii Europejskiej, których stenogramy z obrad są dostępne w bazie ParlaMint.

2. Wyniki badań wstępnych

W ramach zgłoszenia konkursowego autor przeprowadził badania wstępne. Ich zakresem były dane objęte bazą ParlaMint-PL 4.0, czyli te dotyczące wypowiedzi na salach plenarnych obydwu izb polskiego parlamentu w ostatnich dwóch kadencjach. Słowami-kluczami w badaniu były „sąd konstytucyjny” i jego wariacje. Za takim wyborem stoi fakt, że jest to utarte w naukach prawnych pojęcie obejmujące wszelkie sądy i trybunały, które mają kompetencję do badania hierarchicznej zgodności norm prawnych.

Zanim jednak przedstawione zostaną wyniki, wskazać trzeba, że w ramach badań pomijane były cytaty z dokumentów, np. ze sprawozdania rocznego Trybunału Konstytucyjnego czy w ogóle sprawozdań prezesów Trybunału wypowiadanych na posiedzeniu parlamentu. Celowo nie odnoszono się też do samych debat na temat tego, że w danym państwie kontrola konstytucyjna odbywa się w taki, a nie inny sposób (kwestie organizacji i ustroju). Celem badania jest badanie odwoływania się do konkretnych orzeczeń, a nie do zagadnień systemowych.

I tak, wskazać należy następujące najważniejsze — zdaniem autora — wypowiedzi dotyczące tematyki planowanego badania:

1. „Nasze orzecznictwo było śledzone przez Sąd Konstytucyjny Niemiec i Sąd Konstytucyjny Litwy. Wyrazem tego było to, że w każdym z tych sądów byli prawnicy świetnie znający język polski, którzy śledzili nasze orzecznictwo. I to było zjawisko, które wtedy nazywane było dialogiem sądów. To był dialog wszystkich sądów konstytucyjnych, dość częsty, obfity, to była wymiana doświadczeń, refleksji w sprawie orzeczeń i przekonanie, że wspólnie służymy prawie 500 milionom obywateli Unii Europejskiej” — Andrzej Rzepliński, 29.06.2022, Senat, nieparlamentarzysta.
2. „Traktaty jasno określają kompetencje Unii Europejskiej, kompetencje dzielone oraz te pozostające w dyspozycji państw członkowskich. Rolą każdego rządu jest dbanie o interes własnego kraju. Jest to całkowicie zrozumiałe. Niemiecki sąd konstytucyjny przyznał, że narodowe sądy muszą interweniować w rzadkich sytuacjach nadzwyczajnych, gdy dochodzi do poważnego naruszenia kompetencji” — Jan Mosiński, 15.12.2021, Sejm, rządzący.
3. „Dochodziło także do istotnych rozstrzygnięć polemicznych między np. trybunałem w Karlsruhe a trybunałem w Luksemburgu. Rok temu trybunał Republiki Federalnej Niemiec zakwestionował program skupu obligacji na rzecz spłaty długu greckiego. Ale żadna z tych spraw nie dotyczyła niezawisłości i niezależności sądownictwa. Nigdy sądy konstytucyjne w Europie nie wydawały orzeczeń po to, by osłonić destrukcję niezależnego sądownictwa we własnym kraju” — Kazimierz Ujazdowski, 08.10.2021, Senat, opozycja.
4. „Z kolei w międzyczasie Federalny Trybunał Konstytucyjny, niemiecki, wypowiedział się o tym, że Trybunał Sprawiedliwości nie ma kompetencji, by orzekać o sprawach sądow-

nictwa w Niemczech. Stwierdził, że każda dokonana przez Trybunał Sprawiedliwości interpretacja kompetencji Unii Europejskiej w tym zakresie będzie następnie podlegała ocenie sądu konstytucyjnego niemieckiego. [...] Zatem można powiedzieć, że niemiecki trybunał antycypował możliwość prawotwórczej, ustrojotwórczej wręcz roli Trybunału Sprawiedliwości Unii Europejskiej” — Sebastian Kaleta, 15.01.2020, Senat, rządzący.

5. „To, że trybunał w Luksemburgu zawłaszcza sobie coraz dalej idące prerogatywy, to każdy widzi. Mówią już o tym trybunały, sądy konstytucyjne z innych państw, m.in. z Niemiec” — Jerzy Czerwiński, 26.04.2019, Senat, rządzący.
6. „Sąd konstytucyjny RFN uzależnił możliwość podsłuchiwanie i pobierania wrażliwych danych od niezależnej kontroli sądowej nad ich wykorzystaniem” — Sławomir Rybicki, 15.06.2016, Senat, opozycja.
7. „No, przecież Sąd Najwyższy w Stanach Zjednoczonych jest sądem konstytucyjnym, a to właśnie on, a nie Kongres Stanów Zjednoczonych, swoim orzeczeniem zalegalizował aborcję. [...] To właśnie Sąd Najwyższy – w ramach innej procedury, w ramach takiego bardzo konkretnego precedensu prawnego – podjął decyzję, że aborcja w Stanach Zjednoczonych ma być dopuszczalna. I też to nie była decyzja ustawodawcy. To nie była decyzja Kongresu, to nie było decyzja prezydenta Stanów Zjednoczonych. A więc bardzo często zdarza się tak, że sądy powszechne, a przede wszystkim sądy konstytucyjne, podejmują pewnego rodzaju decyzje, w wyniku których tworzy się nowa sytuacja prawna” — Krzysztof Mróz, 28.10.2020, Senat, rządzący.
8. „Na temat tożsamości konstytucyjnej i suwerenności wypowiadały się już sądy konstytucyjne wielu krajów, w tym również Trybunał Konstytucyjny. Trybunał Sprawiedliwości odwołuje się w swoim orzecznictwie do konieczności szanowania tożsamości konstytucyjnej i tożsamości narodowej poszczególnych krajów członkowskich” — Marcin Warchoł, 20.12.2019, Sejm, rządzący.

Dzięki analizie wskazanych wyżej cytatów potwierdzenie znajduje wskazana przez autora hipoteza dotycząca tego, że jeśli odwoływanie się do orzeczeń z innych państw następuje, to dotyczy ono najczęściej spraw wspólnotowych. Stąd właśnie tak wysoki odsetek wypowiedzi dotyczących prawa Unii Europejskiej oraz jego relacji do prawa krajowego. Oczywiście jednak, polityczki i politycy nie mówili tylko o relacji do prawa unijnego. W powyższym zbiorze mamy przecież ponadto wypowiedź dotyczącą konkretnego niemieckiego orzeczenia w sprawie podsłuchu jako metodzie kontroli operacyjnej, a także amerykańskiego wyroku poświęconemu dopuszczalności aborcji. W nich także potwierdza się ta hipoteza. Kwestie aborcyjne czy dotyczące kontroli operacyjnej dokonywanej przez służby są przecież uniwersalne i w konsekwencji, uwzględniając oczywiście kontekst krajowy, pewne tezy będą prawdziwe czy słuszne, niezależnie od granic państwowych.

Znakomita większość wypowiedzi faktycznie skupia się jednak na zagadnieniach związanych z relacją prawa unijnego do prawa krajowego (w tym do konstytucji). Nie jest to jednak, zdaniem autora, niespodziewane. Skoro bowiem mowa o porządku prawnym wspólnym dla

grupy 27 państw, to jasne jest, że sądy konstytucyjne tych państw będą się na ten temat wypowiadały, a pozostałe sądy konstytucyjne będą z tych tez czerpały. Do tego, oczywiście, dochodzi kontekst polityczny, związany z — jak mówią media — konfliktem władz polskich z organami Unii Europejskiej.

Podobna analiza zostanie przeprowadzona w zakresie pozostałych parlamentów krajowych państw członkowskich Unii Europejskiej, z których stenogramy są dostępne w bazie ParlaMint.

Literatura

Erjavec T., Ogrodniczuk M., Osenova P., Ljubešić N., Simov K., Pančur A., Rudolf M., Kopp M., Barkarson S., Steingrímsson S., Çöltekin Ç., de Does J., Depuydt K., Agnoloni T., Venturi G., Pérez M. C., de Macedo L. D., Navarretta C., Luxardo G., Coole M., Rayson P., Morkevičius V., Krilavičius T., Dargis R., Ring O., van Heusden R., Marx M. i Fišer D. (2023). *The ParlaMint Corpora of Parliamentary Proceedings*. „Language Resources and Evaluation”, 58, s. 415–448.

Analiza emocji występujących na posiedzeniach sejmowych

Jakub Chojnacki, Bartłomiej Kruczek
(Akademia Górniczo-Hutnicza)

Streszczenie

Artykuł przedstawia analizę emocji występujących na posiedzeniach Sejmu z wykorzystaniem modelu BERT dostrojonego na zbiorze GoEmotions. Dominującymi emocjami w zbiorze polskich danych parlamentarnych okazują się wdzięczność, neutralność i ciekawość. Analiza uwzględnia również wpływ wydarzeń geopolitycznych na emocje. Autorzy zwracają uwagę na możliwości rozbudowy systemu, umożliwiającego analizę słów kluczowych, przypisanie emocji do płci czy skorelowanie wyników z partiami politycznymi.

Słowa kluczowe

analiza emocji, model BERT, wpływ wydarzeń geopolitycznych na wyrażane emocje

1. Wstęp

Ogólny i łatwy dostęp do informacji pozwala dzisiaj na analizę danych parlamentarnych nie tylko swojego kraju, ale całej Europy. Pozwala to badaczom na prowadzenie badań na szeroką skalę. Podczas każdego zebrania głów państw, omawianych jest wiele ważnych tematów. Jednocześnie podczas każdego takiego posiedzenia generowane są rozmaite dane; począwszy od liczby zebranych osób na sali, przynależności do partii politycznych, kończąc na liczbie wypowiedzianych słów, zdań, itp. Stenograf, osoba za to odpowiedzialna, ma za zadanie notować usłyszane zdania, słowa, reakcje publiczności. Baza danych, zawierająca lata dokumentacji stenograficznych, może posiadać nawet setki tysięcy jak nie miliony słów. Warto zwrócić także uwagę na dostępne tłumaczenia posiedzeń, najczęściej w języku angielskim.

Każde wypowiedziane przez nas sformułowanie niesie ze sobą jakąś emocję, silniejszą bądź słabszą. Żywsza dyskusja może w skrajnych przypadkach doprowadzić nas do granic kontroli emocjonalnej, co może mieć realny wpływ na rozumienie i skuteczną ocenę otaczającej nas rzeczywistości. Może mieć to poważne przełożenie na zrozumienie sytuacji z perspektywy drugiego rozmówcy.

Na szczęście żyjemy w czasach dobrze rozwiniętych algorytmów uczenia maszynowego, które są w stanie scharakteryzować oraz wyekstrahować pożądane przez nas dane z ogromnych baz danych skuteczniej od człowieka. W niektórych przypadkach możemy być świadkami korelacji na pierwszy rzut oka niepowiązanych ze sobą czynników zewnętrznych, rzutujących na to, co oraz w jaki sposób mówimy. W poniższym krótkim raporcie przedstawiono wybrane zależności statystyczne w oparciu o polską część korpusu danych parlamentarnych ParlaMint (Erjavec i in. 2023).

2. Cele oraz opis metodyki

Pierwszym ogólnym celem była próba pokazania liczebności występowania poszczególnych emocji na przestrzeni wybranych lat ze zbioru danych. Z naukowego punktu widzenia klasyfikacja emocji na podstawie tekstu pisanego jest dość skomplikowanym problemem. Do jego rozwiązania został wykorzystany pre-trenowany model oparty na algorytmach uczenia maszynowego. Przyjmuje on słowa, zdania lub całe wypowiedzi w formie pisanej jako dane wejściowe i z pewną dozą prawdopodobieństwa zwraca sklasyfikowaną emocje. Zebrane w ten sposób dane można wykorzystać na wiele interesujących sposobów. Jednym z nich jest próba zestawienia częstotliwości występowania jednej z możliwych do zidentyfikowania emocji z podziałem na miesiące. Możliwe jest nawet stworzenie zestawienia pokazującego jak (oraz czy) zmieniały się one podczas poważniejszych obrad, mając na uwadze tło geopolityczne. To podejście może być jak najbardziej wykorzystane do analizy dowolnego zbioru danych, aczkolwiek my skupiliśmy się tylko na polskiej bazie. Ze względu na jej obszerność (nawet 125 mln słów), czas potrzebny na wykonanie obliczeń był liczony w kilkunastu godzinach, a nawet dniach. Sam proces przetwarzania danych był realizowany na kartach graficznych NVIDIA A100.

3. Zbudowany system

W celu zrealizowania zadania powstał system w języku Python, który łączy w sobie zautomatyzowaną analizę tysięcy plików tekstowych z wykorzystaniem pre-trenowanego modelu oraz nakładkę statystyczną. Pozwala ona m.in. na generowanie pożądanych wykresów czy innych danych. Modelem wykorzystanym do otrzymania emocji był bazowy model BERT dotrenowany zbiorem danych GoEmotions (Demszky i in. 2020). Zbiór ten składa się z 58 tys. komentarzy

pozyskanych z forum internetowego Reddit. Tak dostosowana sieć przyjmuje jako dane wejściowe ciągi znaków, by następnie zwrócić jedną z 28 emocji klasyfikowanych przez model. Problemem użytkownika była czytelna bariera językowa. Wykorzystany model trenowany był oraz operuje na języku angielskim. W celu rozwiązania tego problemu zastosowano automatyczne tłumaczenie danych z wykorzystaniem biblioteki Deep Translate oraz API Google. Wyniki modelu zapisywane były do łatwych w odczycie oraz analizie plikach .xlsx. Proces analizy i klasyfikacja zostały rozdzielone ze względu na ograniczenia obsługi oraz dostępu architektury obliczeniowej. Dlatego pliki Excel wczytywane były do stworzonego w celu analizy oraz prezentacji graficznej notebooka Jupyter. Skrypty w nim zawarte bazowały na bibliotece *pandas* pozwalającej manipulować oraz czytać dane skoroszytowe oraz bibliotece *plotly* — narzędziu, które w przystępny sposób daje możliwość prezentacji danych. Napisane skrypty w znacznym stopniu pozwalają na wyciąganie przydatnych informacji z każdego pliku tekstowego oraz porównywanie ich z plikami referencyjnymi. W zależności od potrzeb użytkownika, istnieje możliwość filtrowania przekształconej bazy danych względem daty, płci, identyfikatora obrad czy nawet sytuacji geopolitycznej na świecie. Powstały notebook Python może być w przyszłości rozbudowywany w celu głębszej analizy całego zbioru językowego.

4. Wnioski

Zależnością widoczną na wykresach (p. rys. 1) jest dominacja niektórych emocji. W naszym przypadku można wyłonić stabilną trójkę najczęściej wskazywanych przez model ekspresji: *gratitude* (wdzięczność), *neutral* (neutralność) oraz *curiosity* (ciekawość). Ze względu na profesjonalny charakter miejsca jakim jest Sejm, należało przyjąć naturalne założenie, że wypowiedzi będą jak najbardziej neutralne. Tymczasem frekwencja neutralności została znacznie i w zaskakujący sposób przewyższona przez *wdzięczność*. Zauważyć również można, że żadna z trzech wymienionych emocji nie jest postrzegana jako negatywna. Miesięczna analiza występowania emocji jak najbardziej potwierdza powyższy wniosek. *Wdzięczność* występuje znacznie częściej w porównaniu do innych emocji, podobnie jak trójka najpopularniejszych emocji. Jedyną wartą wskazania anomalią jest spadek ilości danych w okolicy początku roku 2020. Oczywistym powodem jest nagły wybuch pandemii COVID-19 oraz związane z tym utrudnienia oraz spowolnienia odbywania obrad. Dane poprawiają się po około dwóch miesiącach, choć do końca roku występują pewne spadki informacji. Przeprowadzono również analizę rozkładu emocji uwzględniając rok oraz występowanie słów kluczowych. Słowa kluczowe zostały wybrane na podstawie występujących jednocześnie wydarzeń geopolitycznych. W tym przypadku zamiast czystego zliczania wystąpień, zdecydowano się na procentowy udział we wszystkich wyznaczonych emocjach. Dla przykładu, w przypadku lat 2015 oraz 2016 (wzmrożona aktywność terrorystyczna na terenach Europy) zauważamy znaczny spadek udziału *gratitude* w całości zbioru (choć dalej utrzymuje znaczącą przewagę) na rzecz występowania innych emocji, w tym bardziej negatywnych. Podobny efekt widoczny jest w przypadku słowa „COVID” dla roku 2020 (początek pandemii) oraz „wojna” dla 2022 (początek inwazji wojsk rosyjskich na Ukrainę), choć w nieco mniejszej skali. Powyższe wyniki

należy przyjmować z pewną dozą niepewności. Ze względu na ograniczone zasoby czasowe oraz obliczeniowe wprowadzono pewne kompromisy. Dla części wypowiedzi model nie był w stanie sklasyfikować poprawnie emocji, dlatego też nie są one wzięte pod uwagę w finalnym zestawieniu.

5. Dalsza analiza wyników

W przyszłości stworzony system może być wykorzystany do wielkoskalowego zestawienia występujących słów kluczowych, sprawdzenia, której płci przypisywana jest najczęściej występująca emocja czy skorelowanie odpowiedzi emocjonalnej z partiami politycznymi. Dodatkowo przy odpowiednio dużej mocy obliczeniowej możliwe jest pokazanie zestawienia, jak dany mówca reagował na daną wypowiedź. Przy odpowiednio długim okresie czasu taki system mógłby posłużyć jako podstawa do stworzenia interaktywnej strony internetowej, bądź aplikacji, która analizowałaby setki milionów słów z różnych krajów. Niestety, jest jeszcze za wcześnie na podobny system działający w czasie rzeczywistym.

Podziękowania

Jako studenci AGH pragniemy serdecznie podziękować doktorowi Stanisławowi Kacprzakowi, który sprawował nad nami pieczę opiekuna naukowego. Praca została wykonana z wykorzystaniem infrastruktury PL-Grid.

Literatura

Demszky D., Movshovitz-Attias D., Ko J., Cowen A., Nemade G. i Ravi S. (2020). *GoEmotions: A Dataset of Fine-Grained Emotions*. In Jurafsky D., Chai J., Schluter N. i Tetreault J. (eds.), *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, s. 4040–4054, Online. Association for Computational Linguistics.

Erjavec T., Ogrodniczuk M., Osenova P., Ljubešić N., Simov K., Pančur A., Rudolf M., Kopp M., Barkarson S., Steingrímsson S., Çöltekin Ç., de Does J., Depuydt K., Agnoloni T., Venturi G., Pérez M. C., de Macedo L. D., Navarretta C., Luxardo G., Coole M., Rayson P., Morkevičius V., Krilavičius T., Dargis R., Ring O., van Heusden R., Marx M. i Fišer D. (2023). *The ParlaMint Corpora of Parliamentary Proceedings*. „Language Resources and Evaluation”, 58, s. 415–448.



Rysunek 1: Przykładowy zbiór wykresów występowania emocji dla miesięcy — rok 2020

